ORAL ARGUMENT NOT YET SCHEDULED

No. 19-5031

IN THE UNITED STATES COURT OF APPEALS
FOR THE DISTRICT OF COLUMBIA CIRCUIT

ELECTRONIC PRIVACY INFORMATION CENTER,
Plaintiff-Appellant,

v.

UNITED STATES DEPARTMENT OF COMMERCE, et al.,
Defendants-Appellees.

On Appeal from an Order of the
U.S. District Court for the District of Columbia
Case No. 18-cv-2711-DLF

JOINT APPENDIX

MARC ROTENBERG
ALAN BUTLER
JOHN DAVISSON
Electronic Privacy Information Center
1718 Connecticut Ave. NW, Suite 200
Washington, DC 20009
(202) 483-1140
Counsel for Plaintiff-Appellant

INDEX TO JOINT APPENDIX

U.S. District Court
District of Columbia (Washington, DC)
CIVIL DOCKET FOR CASE #: 1:18-cv-02711-DLF

ELECTRONIC PRIVACY INFORMATION CENTER v. UNITED STATES DEPARTMENT OF COMMERCE et al
Assigned to: Judge Dabney L. Friedrich
Case in other court: 19-05031
Cause: 05:702 Administrative Procedure Act

Date Filed: 11/20/2018
Jury Demand: None
Nature of Suit: 899 Administrative Procedure Act/Review or Appeal of Agency Decision
Jurisdiction: U.S. Government Defendant

Date Filed	#	Docket Text
11/20/2018	1 R	COMPLAINT against All Defendants (Filing fee $ 400 receipt number 0090-5802923) filed by ELECTRONIC PRIVACY INFORMATION CENTER. (Attachments: # 1 Civil Cover Sheet, # 2 Summons United States Attorney for the District of Columbia, # 3 Summons Attorney General of the United States, # 4 Summons United States Department of Commerce, # 5 Summons Bureau of the Census)(Davisson, John) (Entered: 11/20/2018)
11/21/2018	2	LCvR 26.1 CERTIFICATE OF DISCLOSURE of Corporate Affiliations and Financial Interests by ELECTRONIC PRIVACY INFORMATION CENTER (Davisson, John) (Entered: 11/21/2018)
11/21/2018		Case Assigned to Judge Dabney L. Friedrich. (zrdj) (Entered: 11/21/2018)
11/21/2018	3	SUMMONS (4) Issued Electronically as to BUREAU OF THE CENSUS, UNITED STATES DEPARTMENT OF COMMERCE, U.S. Attorney and U.S. Attorney General (Attachments: # 1 Notice and Consent)(zrdj) (Entered: 11/21/2018)
12/03/2018	4	RETURN OF SERVICE/AFFIDAVIT of Summons and Complaint Executed as to the United States Attorney. Date of Service Upon United States Attorney on 11/26/2018. Answer due for ALL FEDERAL DEFENDANTS by 1/25/2019. (Davisson, John) (Entered: 12/03/2018)
12/03/2018	5	RETURN OF SERVICE/AFFIDAVIT of Summons and Complaint Executed. UNITED STATES DEPARTMENT OF COMMERCE served on 11/29/2018 (Davisson, John) (Entered: 12/03/2018)
12/03/2018	6	RETURN OF SERVICE/AFFIDAVIT of Summons and Complaint Executed. BUREAU OF THE CENSUS served on 12/3/2018 (Davisson, John) (Entered: 12/03/2018)
12/03/2018	7	RETURN OF SERVICE/AFFIDAVIT of Summons and Complaint Executed on United States Attorney General. Date of Service Upon United States Attorney General 11/30/18. (Davisson, John) (Entered: 12/03/2018)
01/18/2019	8	MOTION for Preliminary Injunction by ELECTRONIC PRIVACY INFORMATION CENTER (Attachments: # 1 Memorandum in Support, # 2 Text of Proposed Order, # 3 Exhibit List, # 4 Exhibit Nos. 1-4, # 5 Exhibit Nos. 5-9, # 6 Exhibit Nos. 10-14, # 7 Exhibit Nos. 15-19)(Davisson, John) (Entered: 01/18/2019)
01/18/2019		MINUTE ORDER. Having reviewed the plaintiff's 8 Motion for Preliminary

		Injunction, it is ORDERED that the parties shall appear for a preliminary injunction hearing on February 8, 2019 at 11:00 A.M. in Courtroom 12. See LCvR 65.1(d). In accordance with Local Civil Rule 65.1(c), the defendants' opposition, if any, shall be filed within 7 days after service of the motion. Signed by Judge Dabney L. Friedrich on January 18, 2019. (lcdlf1) Modified on 1/22/2019 to reflect updated hearing start time (jl). (Entered: 01/18/2019)
01/22/2019		Set/Reset Deadlines/Hearings: Preliminary Injunction Hearing set for 2/8/2019 at 11:00 AM in Courtroom 12 before Judge Dabney L. Friedrich. (jl). (Entered: 01/22/2019)
01/22/2019	9	NOTICE of Appearance by Kristina Ann Wolfe on behalf of BUREAU OF THE CENSUS, UNITED STATES DEPARTMENT OF COMMERCE (Wolfe, Kristina) (Entered: 01/22/2019)
01/22/2019	10	MOTION to Stay in Light of the Lapse of Appropriations or, in the Alternative,, MOTION for Extension of Time to File a Response to Plaintiff's Motion for a Preliminary Injunction and to Respond to the Complaint by BUREAU OF THE CENSUS, UNITED STATES DEPARTMENT OF COMMERCE (Attachments: # 1 Text of Proposed Order #1, # 2 Text of Proposed Order #2)(Wolfe, Kristina) (Entered: 01/22/2019)
01/23/2019	11	NOTICE of Appearance by Joseph Evan Borson on behalf of BUREAU OF THE CENSUS, UNITED STATES DEPARTMENT OF COMMERCE (Borson, Joseph) (Entered: 01/23/2019)
01/23/2019		MINUTE ORDER granting in part and denying in part the defendant's 10 Motion for a Stay in Light of the Lapse of Appropriations or, in the Alternative, an Extension of Time to File a Response to Plaintiff's Motion for a Preliminary Injunction and to Respond to the Complaint. Accordingly, it is ORDERED that the defendant's request for a stay is DENIED. It is further ORDERED that (1) the defendant shall file its response to the plaintiff's motion on or before January 30, 2019, (2) the plaintiff shall file its reply on or before February 5, 2019, and (3) the defendant shall file its answer or other response to the plaintiff's complaint within 30 days after the Court's resolution of the plaintiff's motion for a preliminary injunction. The preliminary injunction hearing will take place as scheduled on February 8, 2019 at 11:00 A.M. Signed by Judge Dabney L. Friedrich on January 23, 2019. (lcdlf1) (Entered: 01/23/2019)
01/23/2019		Set/Reset Deadlines/Hearings: Defendant's response to plaintiff's motion due by 1/30/2019. Plaintiff's reply due by 2/5/2019. (jl) (Entered: 01/23/2019)
01/30/2019	12 R	Memorandum in opposition to re 8 MOTION for Preliminary Injunction filed by BUREAU OF THE CENSUS, UNITED STATES DEPARTMENT OF COMMERCE. (Attachments: # 1 Declaration of Robin J. Bachman, # 2 Exhibit A to Bachman Decl., # 3 Exhibit B to Bachman Decl., # 4 Text of Proposed Order)(Borson, Joseph) (Entered: 01/30/2019)
02/05/2019	13	REPLY to opposition to motion re 8 MOTION for Preliminary Injunction filed by ELECTRONIC PRIVACY INFORMATION CENTER. (Davisson, John) (Entered: 02/05/2019)
02/06/2019		MINUTE ORDER. Local Civil Rule 65.1(d) requires the Court to hold a hearing within 21 days of an application for a preliminary injunction upon the request of the moving party "unless the Court earlier decides the motion on the papers." LCvR 65.1(d). Having reviewed the parties' briefs, the Court intends to decide the plaintiff's 8 Motion for Preliminary Injunction on the papers and issue a decision on or before February 8, 2019. Accordingly, it is ORDERED that the preliminary injunction hearing currently set for February 8, 2019 is VACATED. Signed by Judge Dabney L. Friedrich on February 6, 2019. (lcdlf1) (Entered: 02/06/2019)

02/06/2019	14	NOTICE of Appearance by Marc Rotenberg on behalf of ELECTRONIC PRIVACY INFORMATION CENTER (Rotenberg, Marc) (Entered: 02/06/2019)
02/06/2019	15	NOTICE of Appearance by Alan Jay Butler on behalf of ELECTRONIC PRIVACY INFORMATION CENTER (Butler, Alan) (Entered: 02/06/2019)
02/08/2019	16	ORDER denying the plaintiff's 8 Motion for a Preliminary Injunction. See text for details. Signed by Judge Dabney L. Friedrich on February 8, 2019. (lcdlf1) (Entered: 02/08/2019)
02/08/2019	17	MEMORANDUM OPINION regarding the plaintiff's 8 Motion for a Preliminary Injunction. See text for details. Signed by Judge Dabney L. Friedrich on February 8, 2019. (lcdlf1) (Entered: 02/08/2019)
02/12/2019	18	NOTICE OF APPEAL TO DC CIRCUIT COURT as to 16 Order on Motion for Preliminary Injunction by ELECTRONIC PRIVACY INFORMATION CENTER. Filing fee $ 505, receipt number 0090-5942771. Fee Status: Fee Paid. Parties have been notified. (Davisson, John) (Entered: 02/12/2019)
02/13/2019	19	Transmission of the Notice of Appeal, Order Appealed (Memorandum Opinion), and Docket Sheet to US Court of Appeals. The Court of Appeals fee was paid this date 2/12/19 re 18 Notice of Appeal to DC Circuit Court,. (ztd) (Entered: 02/13/2019)
02/21/2019		USCA Case Number 19-5031 for 18 Notice of Appeal to DC Circuit Court, filed by ELECTRONIC PRIVACY INFORMATION CENTER. (zvt) (Entered: 02/25/2019)

PACER Service Center			
Transaction Receipt			
02/26/2019 14:15:26			
PACER Login:	ep0116:2545265:0	**Client Code:**	
Description:	Docket Report	**Search Criteria:**	1:18-cv-02711-DLF
Billable Pages:	3	**Cost:**	0.30

UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

ELECTRONIC PRIVACY INFORMATION CENTER,

Plaintiff,

v.

U.S. DEPARTMENT OF COMMERCE, *et al.*,

Defendants.

No. 18-cv-2711 (DLF)

MEMORANDUM OPINION

Plaintiff Electronic Privacy Information Center (EPIC), a non-profit organization dedicated to privacy and civil liberties issues, brings this action against the U.S. Department of Commerce and the U.S. Census Bureau under the Administrative Procedure Act (APA) and the Declaratory Judgment Act. The plaintiff claims that the E-Government Act requires the defendants to conduct and release "privacy impact assessments" addressing Secretary of Commerce Wilbur Ross's March 26, 2018 decision to include a citizenship question in the 2020 Census. The defendants agree, but insist they still have plenty of time to do so "before" actually "initiating a new collection of information" within the meaning of the E-Government Act.[1] Before the Court is the plaintiff's Motion for a Preliminary Injunction, Dkt. 8, seeking to enjoin Commerce and the Bureau from implementing Secretary Ross's decision to add a citizenship question to the Census, *see* Dkt. 8-2. For the following reasons, the Court will deny the motion.

[1] E-Government Act of 2002, § 208(b)(1)(A), Pub. L. 107-347, 116 Stat. 2899 (2002), *codified at* 44 U.S.C.A. § 3501 note (hereinafter "E-Government Act").

I. BACKGROUND

A. Statutory Background

The E-Government Act requires federal agencies to "conduct a privacy impact assessment," "ensure the review of the privacy impact assessment," and, "if practicable, . . . make the privacy impact assessment publicly available" "before" "initiating a new collection of information" that "will be collected, maintained or disseminated using information technology" and that "includes any information in an identifiable form permitting the physical or online contacting of a specific individual, if identical questions have been posed to[] . . . 10 or more persons." E-Government Act § 208(b)(1)(A)–(B).

The term "collection of information" is defined by statute as "the obtaining, causing to be obtained, soliciting, or requiring the disclosure to third parties or the public, of facts or opinions . . . regardless of form or format, calling for" "answers to identical questions posed to . . . ten or more persons[.]" 44 U.S.C. § 3502(3)(A); *see also* E-Government Act § 201 (incorporating § 3502 definitions by reference). The same term is also used in OMB regulations to "refer[] to the act of collecting or disclosing information, to the information to be collected or disclosed, to a plan and/or an instrument calling for the collection or disclosure of information, or any of these, as appropriate." 5 C.F.R. § 1320.3(c). The term "initiating" has no statutory or regulatory definition.

A privacy impact assessment—or "PIA"—must "address" "what information is to be collected;" "why the information is being collected;" "the intended use of the agency of the information;" "with whom the information will be shared;" "what notice or opportunities for consent would be provided to individuals regarding what information is collected and how that information is shared;" "how the information will be secured;" and "whether a system of records is being created under [the Privacy Act]." E-Government Act § 208(b)(2)(B)(ii).

B. Factual Background

On March 26, 2018, Secretary of Commerce Wilbur Ross announced his decision to include a citizenship question on the 2020 Decennial Census questionnaire. *See* Bachman Decl. ¶ 12, Dkt. 12-1. That decision has been challenged elsewhere on a number of grounds.[2] For present purposes, all that matters is whether—and, more importantly, *when*—the decision to collect citizenship information had to be addressed in one or more PIAs.

The Bureau is no stranger to PIAs. When Secretary Ross announced the inclusion of the citizenship question in March 2018, the Bureau was already planning to conduct an annual PIA for the primary information technology system used for the decennial census. Bachman Decl. ¶¶ 3, 9. That system—called "CEN08"—shares Census-related information with four other systems: "CEN21," "CEN05," "CEN11," and "CEN13." *Id.* ¶ 14. And a sixth information technology system—called "CEN18"—enables the flow of information between CEN08 and the other four systems. *Id.*

The Bureau maintains and regularly updates PIAs for each of these systems. *See id.* ¶¶ 9, 15. The PIA for CEN08 was updated in June and September of 2018, and another update is in progress and scheduled for release in February or March of 2019. *Id.* ¶ 9. The PIAs for the remaining systems were all updated in June 2018 and will be reviewed and updated again "within the next two months" as part of the Bureau's annual PIA process. *Id.* ¶ 15. In the meantime, the current PIAs for these systems are available to the public online.[3]

---

[2] *See New York v. U.S. Dep't of Commerce*, No. 18-CV-2921 (JMF), 2019 WL 190285 (S.D.N.Y. Jan. 15, 2019).

[3] http://www.osec.doc.gov/opog/privacy/Census-pias.html?#.

The existing PIAs say little about the collection of citizenship information in particular. The PIAs for CEN05,[4] CEN13,[5] and CEN18[6] do not mention citizenship at all. And the PIAs for CEN08[7] and CEN11[8] mention citizenship only once, in a field labeled "Other general personal data (specify)," without any analysis or further context.[9]

Unsatisfied with this level of treatment, EPIC filed this action on November 20, 2018. The complaint asserts two counts under the APA and one count under the Declaratory Judgment Act. Count I alleges that the defendants acted unlawfully by adding the citizenship question to the Census without first conducting, reviewing, and releasing PIAs to address that decision. Compl. ¶¶ 64–70 (citing 5 U.S.C. § 706(2)(a), (c)). Count II alleges that the defendants unlawfully withheld agency action by failing to conduct, review, or release PIAs as required. *Id.* ¶¶ 71–76 (citing 5 U.S.C. § 706(1)). And Count III seeks a declaration of rights and relations consistent with counts I and II. *Id.* ¶¶ 77–78 (citing 28 U.S.C. § 2201(a)).

On January 15, 2019, a federal district court in New York permanently enjoined Commerce and the Bureau from including the citizenship question on the Census. *See New York v. U.S. Dep't of Commerce*, 2019 WL 190285, at *125. Three days later, EPIC filed this motion for a preliminary injunction, which the Court now resolves.

---

[4] http://www.osec.doc.gov/opog/privacy/Census%20PIAs/CEN05_PIA_SAOP_Approved.pdf.

[5] http://www.osec.doc.gov/opog/privacy/Census%20PIAs/CEN13_PIA_SAOP_Approved.pdf.

[6] http://www.osec.doc.gov/opog/privacy/Census%20PIAs/CEN18_PIA_SAOP_Approved.pdf.

[7] http://www.osec.doc.gov/opog/privacy/Census%20PIAs/CEN08_PIA_SAOP_Approved.pdf.

[8] http://www.osec.doc.gov/opog/privacy/Census%20PIAs/CEN11_PIA_SAOP_Approved.pdf.

[9] The plaintiffs do not challenge the PIA for CEN21. *See* Compl. ¶¶ 49, 51–62, Dkt.1.

II. LEGAL STANDARD

A preliminary injunction is "an extraordinary remedy that may only be awarded upon a clear showing that the plaintiff is entitled to such relief." *Winter v. Nat. Res. Def. Council, Inc.*, 555 U.S. 7, 22 (2008). To warrant a preliminary injunction, a plaintiff "must make a clear showing" that (1) he "is likely to succeed on the merits"; (2) he "is likely to suffer irreparable harm in the absence of preliminary relief"; (3) the "balance of equities" tips in his favor; and (4) "an injunction is in the public interest." *Id.* at 20; *League of Women Voters of United States v. Newby*, 838 F.3d 1, 6 (D.C. Cir. 2016). The last two factors "merge when the Government is the opposing party." *Nken v. Holder*, 556 U.S. 418, 435 (2009). The plaintiff "bear[s] the burdens of production and persuasion" when moving for a preliminary injunction. *Qualls v. Rumsfeld*, 357 F. Supp. 2d 274, 281 (D.D.C. 2005) (citing *Cobell v. Norton*, 391 F.3d 251, 258 (D.C. Cir. 2004)).

"Before the Supreme Court's decision in *Winter*, courts weighed the preliminary injunction factors on a sliding scale, allowing a weak showing on one factor to be overcome by a strong showing on another factor." *Standing Rock Sioux Tribe v. U.S. Army Corps of Eng'rs*, 205 F. Supp. 3d 4, 26 (D.D.C. 2016). The D.C. Circuit, however, has "suggested, without deciding, that *Winter* should be read to abandon the sliding-scale analysis in favor of a 'more demanding burden' requiring a plaintiff to independently demonstrate both a likelihood of success on the merits and irreparable harm." *Id.* (quoting *Sherley v. Sebelius*, 644 F.3d 388, 392–93 (D.C. Cir. 2011)); *see also Davis v. Pension Benefit Guar. Corp.*, 571 F.3d 1288, 1292 (D.C. Cir. 2009).

"Both before and after *Winter*, however, one thing is clear: a failure to show a likelihood of success on the merits alone is sufficient to defeat the motion." *Hudson v. Am.*

Fed'n of Gov't Employees, 308 F. Supp. 3d 121, 127 (D.D.C. 2018) (citing *Ark. Dairy Co-op Ass'n, Inc. v. USDA*, 573 F.3d 815, 832 (D.C. Cir. 2009)). "[A]bsent a substantial indication of likely success on the merits, there would be no justification for the Court's intrusion into the ordinary processes of administration and judicial review.'" *Archdiocese of Washington v. Washing Metro. Area Transit Auth.*, 281 F. Supp. 3d 88, 99 (D.D.C. 2017) (internal quotation marks omitted), *aff'd*, 897 F.3d 314 (D.C. Cir. 2018). Accordingly, "[u]pon finding that a plaintiff has failed to show a likelihood of success on the merits, the Court may deny a motion for preliminary injunction without analyzing the remaining factors." *In re Akers*, 487 B. R. 326, 331 (D.D.C. 2012); *see also Hudson*, 308 F. Supp. 3d at 131–32 (same).

Likewise, "it is clear" before and after *Winter* "that failure to show a likelihood of irreparable harm remains, standing alone, sufficient to defeat the motion." *Navajo Nation v. Azar*, 292 F. Supp. 3d 508, 512 (D.D.C. 2018).

III. ANALYSIS

A. Likelihood of Success on the Merits

The defendants concede that they must eventually prepare PIAs that adequately address the collection of citizenship data in the 2020 Census. *See, e.g.*, Defs.' Opp'n at 1, 12, Dkt. 12. But they disagree with the plaintiff that they were required to do so before Secretary Ross announced his decision to add the citizenship question on March 26, 2018. As the defendants point out, the E-Government Act requires agencies to conduct (and, if practicable, release) a PIA only before "*initiating* a new collection of information." E-Government Act § 208(b)(1)(A)(ii) (emphasis added). And "initiating" the collection of information, the defendants argue, means more than just announcing a decision to collect information at some point in the future. It requires at least one instance of obtaining, soliciting, or requiring the disclosure of information,

which in the defendants' view will not occur until the Bureau mails its first batch of Census questionnaires to the public. *See* Defs.' Opp'n at 11–14. The Court agrees.

"A fundamental canon of statutory construction is that, unless otherwise defined, words will be interpreted as taking their ordinary, contemporary, common meaning." *Perrin v. United States*, 444 U.S. 37, 42 (1979); *see also New Prime Inc. v. Oliveira*, 139 S. Ct. 532, 539 (2019) (same). Contemporary dictionaries define "initiate" as "[t]o begin, commence, enter upon; to introduce, set going, give rise to, originate, 'start' (a course of action, practice, etc.)." Oxford English Dictionary, http://www.oed.com/view/Entry/96066?rskey=wxG1jD&result=2&isAdvanced=false#eid; *see also* Merriam-Webster, https://www.merriam-webster.com/dictionary/initiate ("to cause or facilitate the beginning of : set going"). Black's Law Dictionary similarly defines "initiate" as to "[c]ommence, start; originate; introduce[.]" Black's Law Dictionary 784 (6th ed. 1990). These definitions share a focus on the *beginning*, *starting*, or *commencing* of a course of conduct. In the words of *Webster's Third*, they contemplate "the first actions, steps, or stages of" the activity initiated. Webster's Third New International Dictionary 1164 (3d ed. 1976)).

Combining this ordinary meaning with the statutory definition of "collection of information," an agency must conduct (and, if practicable, release) a PIA before it *begins* "obtaining, causing to be obtained, soliciting, or requiring the disclosure to third parties or the public, of facts or opinions[.]" 44 U.S.C. § 3502(3)(A). Commerce and the Bureau have not yet gone so far. While Secretary Ross decided to collect citizenship information—and announced that decision in a letter that the parties agree constitutes final agency action, *see* Pl.'s Mot. at 24–25, Dkt. 8-1; Defs.' Opp'n at 18—the defendants have yet to actually begin obtaining, soliciting, or requiring the disclosure of any citizenship data. Those actions will not occur until the Bureau

mails its first set of questionnaires to the public in January 2020. *See* Pl.'s Reply at 2, 13, Dkt. 13 (acknowledging that the questionnaires will be sent to the public in January 2020); U.S. Census Bureau, 2020 Census Operational Plan: A New Design for the 21st Century 97 (Dec. 2018), https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-oper-plan4.pdf (stating that the "printing, addressing, and mailing of Internet invitations, reminder cards or letters, and paper questionnaire packages" will occur between June 2019 and April 2020).

A simple hypothetical offered by the defendants illustrates why this interpretation tracks the plain meaning of the statute. Imagine a happy couple is planning a wedding, and a friend asks if they have "initiated the collection of RSVPs." Ordinarily, they would not say yes if they had merely finalized the guest list, chosen a font for the invitations, or decided to include a dinner selection on the RSVP cards. At that point, they have not "initiated the collection" of any RSVPs. They have merely made antecedent decisions about what information to collect—and from whom—in the future. Likewise, when Secretary Ross decided to add a citizenship question to the yet-to-be-mailed Census questionnaires—the equivalent of adding a dinner selection to an un-mailed RSVP card—he did not "initiate a new collection of information" but merely decided what new information the Bureau would collect later.

The plaintiff resists this analogy because Secretary Ross's decision was final and made the collection of information all but inevitable. *See* Reply at 5. For the analogy to hold, the plaintiff argues, the couple would have had to place an order with a full-service printer who will mail the invitations on a fixed date in the future unless the couple cancels the order. *Id.* But this change would not alter the couple's response because the fact that an event is certain to occur in the future does not mean it has already begun. To build on the wedding analogy, a couple does

not "initiate" their marriage by getting engaged or choosing a wedding date, even if those actions ordinarily serve as a final—and binding—decision to tie the knot. As each subsequent anniversary celebration makes clear, they will not have "initiated" their marriage until the wedding day.

A similar usage applies in the legal context. Courts routinely use the phrase "initiating an action" to refer the filing of the complaint. *See, e.g.*, *Horne v. Dep't of Agric.*, 569 U.S. 513, 520 (2013) (an agency "initiated an enforcement action" on the date the complaint was filed); *Arnold v. U.S. Secret Serv.*, 524 F. Supp. 2d 65, 66 (D.D.C. 2007) (the plaintiff "initiated this action" on the date the complaint was filed). And it would be unusual—if not downright misleading—to claim to have "initiated" a lawsuit when in fact one had merely decided which claims to allege in the complaint. That is because "initiating" normally means "beginning"—in the law as everywhere else. And there is a meaningful difference between deciding or preparing to bring a lawsuit and actually *initiating* it.

Congress must have been aware of this distinction. After all, it had a range of terms at its disposal if it wanted agencies' assessment and reporting obligations to arise earlier in the data-collection process. For instance, Congress could have required a PIA before "planning" or "providing for" a new collection of information. *See* E-Government Act (132 references to variations of the words "plan" or "provide"). Alternatively, Congress could have required a PIA whenever an agency makes a "determination" or "decision" to initiate a new collection of information. *See id.* (40 references). "The fact that [Congress] did not adopt th[ese] readily available and apparent alternative[s] strongly supports rejecting" an interpretation that would substitute them for the word Congress did choose. *Knight v. Comm'r*, 552 U.S. 181, 188 (2008).

Indeed, the only other use of "initiate" in the E-Government Act confirms that Congress uses that word deliberately to refer to actions beyond mere decisionmaking or planning. *See* Antonin Scalia & Bryan A. Garner, Reading Law: The Interpretation of Legal Texts 170 (2012) ("A word or phrase is presumed to bear the same meaning throughout a text[.]"). Section 214(c) requires the Administrator of the Office of Electronic Government to "initiate pilot projects or report to Congress on other activities that further the goal of maximizing the utility of information technology in disaster management." E-Government Act § 214(c). Plainly, this obligation would not be satisfied if the Administrator merely announced a decision to initiate a pilot project at some point in the future. The natural interpretation of § 214(c) is that the Administrator must either actually commence a pilot project or else perform "other activities" that serve the same goals.

Although the plaintiff does not address § 214(c), it notes that elsewhere in Title 44 Congress apparently drew a distinction between "initiating," "carrying out," and "completing." *See* Pl.'s Mot. at 19 (quoting 44 U.S.C. § 3902(a)). The relevant provision states that the "Director of the Government Publishing Office shall have no authority to prevent or prohibit the Inspector General from *initiating*, *carrying out*, or *completing* any audit or investigation[.]" 44 U.S.C. § 3902(a) (emphasis added). In the plaintiff's view, this sentence proves that Congress uses "initiating" to mean something different and less than "carrying out"; thus, it must include the decision to carry out an activity in the future. The Court is unconvinced. To be sure, "[i]t is a cardinal principle of statutory construction that [a court] must give effect, if possible, to every clause and word of a statute." *NLRB. v. SW Gen., Inc.*, 137 S. Ct. 929 (2017) (alteration adopted and internal quotation marks omitted). But it would not produce any redundancy to interpret "initiating" in § 3902(a) to refer to the actual commencement of an audit or investigation.

Section 3902(a) describes the beginning, middle, and end of an audit or investigation, and it makes clear that the Director cannot prevent the Inspector General from proceeding at any point in that process. If the Inspector General has not yet begun an audit or investigation, the Director cannot prevent him from "initiating" one; if he has already begun an audit or investigation, the Director cannot prevent him from "carrying [it] out"; and if he is nearing the end of an audit or investigation, the Director cannot prevent him from "completing" it. While the words "carrying out" might technically be used to describe the first or last step of an audit or investigation—just as it describes every step in between—it is more natural to refer to those steps as "initiating" and "completing" the audit or investigation. And there is nothing surprising about using the three terms together to emphasize the Inspector General's freedom from interference from beginning to end.

The plaintiff raises a number of additional arguments to support its interpretation, but none are persuasive. *First*, the plaintiff attempts to show that the text itself encompasses a decision to collect information at some point in the future. The plaintiff highlights the use of gerunds in the definition of "collection of information," *see* 44 U.S.C. § 3502(3)(A) ("obtaining," "causing," "soliciting," or "requiring"), and argues that this grammatical choice connotes "a process, not a one-off action," Pl.'s Reply at 4. But even so, the statute makes clear what that process consists of: the "obtaining" of information, the "causing" of information to be obtained, the "soliciting" of information, and the "requiring" of the disclosure of information. 44 U.S.C. § 3502(3)(A). Consequently, "initiating" a "collection of information"—even if viewed as a process—still requires the beginning of at least one of these actions.

The plaintiff also argues that Congress would not have used the six-word phrase "initiating a new collection of information" if it meant "collecting new information" and could

have said so directly in three fewer words. *See* Reply at 4. But this observation ignores that the noun form "collection of information" has a statutory definition that Congress may have used for clarity or consistency. Moreover, the defendants have never argued that the agency must actually "collect"—that is, obtain or receive—information to have initiated a new collection of information under § 208. They acknowledge that performing any one of the gerunds listed in 44 U.S.C. § 3502(3)(A) would qualify as "initiating" the collection of information. Thus, "soliciting" or "requiring the disclosure" of citizenship data—here, by mailing Census questionnaires—would require a PIA even if no information has been obtained in response. *See* Defs.' Opp'n at 12.

Next, the plaintiff argues that Secretary Ross literally "requir[ed] the disclosure of facts or opinions to third parties" when he issued the March 26, 2018 decision adding a citizenship question to the Census. *See* Pl.'s Reply at 7. That is simply not true. By the plaintiff's own admission, the public will not be obligated to disclose information to third-parties until the Bureau actually implements the 2020 Census. *See id.* ("[M]embers of the public will inevitably come under an obligation to disclose their citizenship status via the 2020 Census"); *id.* at 14 ("[O]nce [the Bureau] sends out the questionnaires, individuals will be legally obligated to respond.").

Second, EPIC attempts to draw various inferences from statutory structure. For instance, the plaintiff points to other provisions in Title 44 that describe the "collection of information" in contexts where an agency clearly has not begun obtaining or soliciting information. *See* Pl.'s Reply at 5–6 (citing, *e.g.*, 44 U.S.C. § 3505). But these provisions are both unsurprising and irrelevant because none use the critical word "initiate." Of course, an agency can "propose," "review," "approve," or "reject" a collection of information without "initiating" it, just as one

can propose or reject a marriage without initiating one. But that possibility says nothing about what it means to initiate a collection of information.

The plaintiff also highlights the provision directly adjacent to § 208(b)(1)(A)(ii), which requires a PIA before "developing or procuring information technology that collects, maintains, or disseminates information[.]" E-Government Act § 208(b)(1)(A)(i). In the plaintiff's view, the choice to require a PIA before "developing" or "procuring" technology—and not merely before "activating" or "deploying" it—shows that Congress intended PIAs to be completed early on in an agency's decisionmaking process. *See* Pl.'s Reply at 6. But one could just as easily draw the opposite inference and conclude that when Congress wants to require a PIA at a preliminary stage, like development or procurement, it does so explicitly.

Third, the plaintiff invokes OMB regulations that implement a related statute, the Paperwork Reduction Act, whose definitions are incorporated into the E-Government Act. *See* Pl.'s Mot. at 20; *see also* 5 C.F.R. § 1320.3 (implementing the Paperwork Reduction Act); 44 U.S.C. § 3502 (defining terms in Paperwork Reduction Act); E-Government Act § 201 (incorporating definitions in § 3502 by reference). Those regulations explain that OMB uses the term "collection of information" to refer not only to the "act of collecting or disclosing information" but also "to the information to be collected or disclosed" or to a "*plan* and/or an instrument calling for the collection or disclosure of information." 5 C.F.R. § 1320.3(c) (emphasis added). Applying this expansive regulatory definition, the plaintiff argues that Secretary Ross "introduced a definite plan . . . calling for the collection or disclosure of information" and thereby initiated a collection of information under § 208. Pl.'s Mot. at 21 (internal quotation marks omitted). Again, the Court is unpersuaded.

The OMB regulations define "collection of information" only "[a]s used in this Part"—that is, in the Paperwork Reduction Act regulations themselves. 5 C.F.R. § 1320.3(c). They do not purport to define the terms of the E-Government Act. This limitation is not just a technicality. Unlike § 208, the regulations implementing the Paperwork Reduction Act use the phrase "collection of information" to refer both to the act of collecting information and as a noun to describe materials submitted by an agency to OMB for approval. *See, e.g.*, *id.* § 1320.10. Given these multiple meanings, it makes sense for OMB to provide separate definitions for each. But it would be nonsensical to import these specialized, regulation-specific uses to § 208, which plainly uses "the collection of new information" to describe an event. To illustrate, it would be incoherent to speak of "initiating" "information" or "initiating" an "instrument." Yet that is the result of inserting the OMB definitions into § 208, where they were not meant to apply. And while one can "initiate" a "plan," it would be unwise to cherry-pick one component of a definition that, as a whole, was clearly designed for another purpose. Indeed, even OMB does not ordinarily invoke all three regulatory meanings of "collection of information" at once; rather, it uses the phrase to refer to "any" one of them, "as appropriate." *Id.* § 1320.3(c). Since in context, § 208 clearly refers to "the act of collecting or disclosing information," it is irrelevant that OMB sometimes uses the same phrase to refer to something else, like a "plan."

In any event, even if the OMB regulations did apply, they would not change the outcome here. To "initiate" a "plan" would still mean to commence it or put it into action, not merely to announce it, as EPIC suggests, *see* Pl.'s Mot. at 20–21. Thus, a "plan . . . calling for the collection or disclosure of information" would not be "initiated" until the "collection or disclosure" "call[ed] for" actually begins—in this case, with the mailing of questionnaires to the public.

Fourth, the plaintiff invokes precedent, pointing to a D.C. Circuit decision that mentioned in passing that an agency "need not prepare a privacy impact assessment unless it *plans* to collect information." *EPIC v. Presidential Advisory Comm'n on Election Integrity*, 878 F.3d 371, 380 (D.C. Cir. 2017) (emphasis added). Setting aside that this quote addresses *whether* an agency must prepare a PIA—not *when*—EPIC overlooks that the same decision elsewhere describes the E-Government Act as requiring an agency to "conduct, review and, if practicable, publish a privacy impact assessment before it *collects* information." *Id.* at 375 (emphasis added and internal quotation marks omitted); *see also id.* (describing the Act as "requiring an agency to fully consider [individuals'] privacy before *collecting* their personal information" (emphasis added)). If anything, *EPIC* supports the defendants' interpretation, although the Court declines to attach significance either way to a decision that had no occasion to interpret the statutory language.

Fifth, the plaintiff argues that allowing agencies to wait until after deciding to collect information to conduct and publish a PIA would frustrate the purpose of the E-Government Act's privacy provisions. *See* Pl.'s Reply at 9. But "[e]ven the most formidable argument concerning the statute's purposes could not overcome the clarity" of "the statute's text." *Kloeckner v. Solis*, 568 U.S. 41, 55 n.4 (2012). At any rate, here the statutory purpose and plain text are perfectly compatible. The E-Government Act has many purposes—eleven to be exact—and nearly all focus on improving Government efficiency, transparency, and performance through the use of the Internet and emerging technologies. *See* E-Government Act § 2(b)(1)–(11). Congress recognized, however, that this shift to "electronic Government" could create privacy concerns, and it addressed those concerns through the "Privacy Provisions" embodied in § 208. *Id.* § 208(a). Importantly, § 208 is not a general privacy law; nor is it meant to minimize the

collection of personal information. Rather, its express purpose is "to ensure sufficient protections for the privacy of personal information as agencies implement citizen-centered electronic Government." *Id*. Congress's focus on ensuring "protections" when agencies "implement" electronic Government shows that § 208's provisions—including the requirement to prepare PIAs—were not meant to discourage agencies from collecting personal information but rather to ensure that they have sufficient protections in place before they do. It is no surprise, then, that Congress would require agencies to prepare PIAs only before they actually begin to gather, store, and potentially share personal information.

The plaintiff advocates a much broader conception of § 208's purpose aimed at influencing agency decisionmaking. To support that vision, it cites cases discussing the National Environmental Policy Act (NEPA), 42 U.S.C. § 4321 *et seq.*, which requires agencies to prepare "environmental impact statements." *See* Pl.'s Reply at 9 (citing *Jones v. D.C. Redevelopment Land Agency*, 499 F.2d 502, 512 (D.C. Cir. 1974) and *Lathan v. Volpe*, 455 F.2d 1111, 1121 (9th Cir. 1971)). But the E-Government Act and NEPA are hardly analogous. Although they both require a form of "impact" assessment, the role and timing of those assessments differ sharply. Unlike the E-Government Act, NEPA explicitly requires an impact statement to be included "in every *recommendation* or *report* on *proposals* for legislation and other Federal actions" that meet certain criteria. 42 U.S.C. § 4332(C) (emphasis added). EPA regulations further specify that "[a]n agency shall commence preparation of an environmental impact statement *as close as possible* to the time the agency is *developing or is presented with a proposal*," and the statement must "be prepared *early enough* so that it can serve practically as an *important contribution to the decisionmaking process* and will not be used to rationalize or justify decisions already made." 40 C.F.R. 1502.5 (emphasis added). The regulations go on to provide specific deadlines

for preparing environmental statements depending on the type of agency action proposed. *Id.* (a)–(d). This language—explicitly tying impact statements to agency decisionmaking and imposing clear and specific deadlines as early as possible in the decisionmaking process—is notably absent from the E-Government Act, which requires only that agencies conduct and, if practicable, release a privacy impact assessment before "initiating the new collection of information" and only then for the purpose of "ensuring sufficient protections" for the information collected.

That is not to say that negative policy consequences cannot ever result if an agency drags its feet in performing its PIA obligations. *See* Pl.'s Reply at 3. But publishing a PIA shortly before commencing a new collection of information does not make the PIA "useless," as EPIC claims. *See id.* Indeed, publishing a PIA even belatedly would support one of the purposes of the E-Government Act to "make the Federal Government more transparent and accountable," E-Government Act § 2(b)(9), and would inform citizens why their data is being collected, how it is secured, and with whom it will be shared. *See id.* § 208(b)(2)(B)(ii).

For all of these reasons, the Court interprets "initiating a new collection of information," E-Government Act § 208(b)(1)(A)(ii), to require at least one instance of "obtaining, causing to be obtained, soliciting, or requiring the disclosure . . . of facts or opinions," 44 U.S.C. § 3502(3)(A). This interpretation is fatal to the plaintiff's APA claims. The Bureau did not act contrary to the E-Government Act by deciding to collect citizenship data before conducting, reviewing, or releasing a PIA addressing that decision. *See* 5 U.S.C. § 706(2). Nor have the defendants "unlawfully withheld" agency action by declining to conduct or release a PIA earlier

than they were required to under the statute. *See id.* § 706(1). EPIC is therefore unlikely to succeed on the merits.[10]

B. Likelihood of Irreparable Harm

"Having concluded that plaintiff has no likelihood of success on the merits, the Court finds it unnecessary to weight the remaining preliminary injunction factors." *Doe v. Hammond*, 502 F. Supp. 2d 94, 102 (D.D.C. 2007). Nonetheless, the Court will briefly address the plaintiff's three theories of irreparable harm—none of which are persuasive.

First, the plaintiff argues that the Bureau's ongoing failure to publish adequate PIAs irreparably harms its members by denying them information vital to a national debate. Pl.'s Mot. at 27. But even assuming this harm is irreparable, it will not be redressed by the relief requested. The plaintiff does not seek an affirmative injunction directing the defendants to perform or publish a PIA. It seeks only negative injunctions preventing the Bureau from "implementing" Secretary Ross's "decision to add a citizenship question to the 2020 Census" and from "initiating any collection of citizenship status information that would be obtained through the 2020 Census." Pl.'s Proposed Order, Dkt. 8-2. As the D.C. Circuit has explained, "halting" the "collection of . . . data" cannot redress an informational injury under the E-Government Act because "ordering the defendants *not* to collect . . . data only *negates* the need (if any) to prepare an impact assessment, making it *less* likely that EPIC will obtain the information it says is

[10] The defendants argue that this interpretation of § 208 also leads to certain prudential and jurisdictional consequences—namely, a lack of ripeness or final agency action. *See* Defs.' Opp'n at 16–21. But these arguments would only be relevant if EPIC sought to challenge, prospectively, the agencies' failure to conduct or release adequate PIAs in the future. It does not. *See* Pl.'s Reply at 13. EPIC challenges only the defendants' past failure to conduct or release adequate PIAs before Secretary Ross issued his decision on March 26, 2018. *See, e.g.*, Pl.'s Reply at 10–13; Compl. ¶¶ 64–76. The Court therefore need not consider whether a different claim premised on future acts or omissions could proceed.

essential." *EPIC*, 878 F.3d at 380 (emphasis in original). Because the purported deprivation of information is not redressable through the relief requested, the Court cannot rely on it to establish irreparable harm.

Second, the plaintiff argues that its members suffered irreparable harm from Secretary Ross's failure to conduct a PIA and take privacy considerations into account before deciding to collect citizenship data. *See* Pl.'s Mot. at 29–31. The plaintiff acknowledges that this harm has already "mature[d]", *id.* at 30 (internal quotation marks omitted), and that the defendants will not change course absent judicial intervention, *see* Pl.'s Reply at 5, 7, but it nonetheless argues that "equitable intervention is necessary" before an "irretrievable commitment of resources" occurs that might render any future PIA a rubber stamp, *id.* at 15 (internal quotation marks omitted). The problem, however, is that the earliest "irretrievable commitment" the plaintiff identifies is the "printing, addressing, and mailing" of Census materials in June 2019. Pl.'s Mot. at 30 (internal quotation marks omitted). That event, still four months away, is not "of such *imminence* that there is a clear and present need for equitable relief to prevent irreparable harm," *Wisconsin Gas Co. v. FERC*, 758 F.2d 669, 674 (D.C. Cir. 1985) (internal quotation marks omitted), particularly in an APA suit where summary judgment typically "serves as the mechanism for deciding, as a matter of law, whether the agency action is . . . consistent with the APA standard of review," *Sierra Club v. Mainella*, 459 F. Supp. 2d 76, 90 (D.D.C. 2006). Given the possibility of resolving this suit on the merits through expedited summary judgment briefing, the plaintiff has not shown a present need for equitable relief to maintain the status quo. Further, another court has already permanently enjoined the Bureau from implementing the Census with a citizenship question. *See New York v. U.S. Dep't of Commerce*, 2019 WL 190285, at *125. Thus, the prospect of printing and mailing questionnaires that include the citizenship question is

far from "certain," *Wisconsin Gas Co.*, 758 F.2d at 674, and will only occur if the Bureau successfully challenges the injunction on appeal.

Finally, the plaintiff argues that its members will be irreparably harmed if and when their own citizenship data is collected. But this harm, too, is neither imminent nor certain. The parties agree that the Bureau will not mail any questionnaires until January 2020 at the earliest. Pl.'s Reply at 2, 14; Defs.' Opp'n at 26–27. And, again, even that will only happen if the permanent injunction already in effect is vacated or reversed on appeal.

In short, the plaintiff has not demonstrated a "certain" injury "of such imminence that there is a clear and present need for equitable relief to prevent irreparable harm." *Wisconsin Gas Co.*, 758 F.2d at 674 (emphasis and internal quotation marks omitted). That failure alone, like the failure to show a likelihood of success on the merits, provides an independent ground for denying its motion. *Navajo Nation*, 292 F. Supp. 3d at 512.

CONCLUSION

For the foregoing reasons, the Court will deny the plaintiff's motion for a preliminary injunction. A separate order accompanies this memorandum opinion.

Dabney L. Friedrich

DABNEY L. FRIEDRICH
United States District Judge

February 8, 2019

UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

ELECTRONIC PRIVACY INFORMATION CENTER,

Plaintiff,

v.

U.S. DEPARTMENT OF COMMERCE, *et al.*,

Defendants.

No. 18-cv-2711 (DLF)

ORDER

For the reasons stated in the accompanying memorandum opinion, it is

ORDERED that the plaintiff's Motion for a Preliminary Injunction, Dkt. 8, is **DENIED**.

DABNEY L. FRIEDRICH
United States District Judge

February 8, 2019

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

ELECTRONIC PRIVACY INFORMATION CENTER
1718 Connecticut Avenue, N.W., Suite 200
Washington, D.C. 20009

Plaintiff,

v.

UNITED STATES DEPARTMENT OF COMMERCE,
1401 Constitution Avenue, N.W.
Washington, D.C. 20230

BUREAU OF THE CENSUS,
4600 Silver Hill Road
Suitland, Md. 20746

Defendants.

Civ. Action No. 18-2711

COMPLAINT FOR INJUNCTIVE RELIEF

1. This is an action under the Administrative Procedure Act ("APA"), 5 U.S.C. §§ 551–706; the E-Government Act of 2002, Pub. L. 107-347, 116 Stat. 2899 (codified as amended at 44 U.S.C. § 3501 note); and the Declaratory Judgment Act, 28 U.S.C. § 2201(a), to secure the creation and publication of multiple Privacy Impact Assessments ("PIAs") from the United States Department of Commerce and the Bureau of the Census.

2. The Electronic Privacy Information Center ("EPIC") seeks the release of certain Privacy Impact Assessments pertaining to the Defendants' attempted collection of personal data concerning citizenship status by means of the 2020 Census.

Jurisdiction and Venue

3. This Court has subject matter jurisdiction over this action under 28 U.S.C. § 1331, 5 U.S.C. § 702, and 5 U.S.C. § 704. This Court has personal jurisdiction over the Department of Commerce and the Census Bureau.

4. Venue is proper in this district under 5 U.S.C. § 703 and 28 U.S.C. § 1391.

Parties

5. Plaintiff Electronic Privacy Information Center is a non-profit, public interest research center established in 1994 to focus public attention on emerging privacy and civil liberties issues. Central to EPIC's mission is oversight of government activities that impact individual privacy, free expression, and democratic values.[1] EPIC has a particular interest in preserving legal privacy protections for personal data, including those established in the E-Government Act.

6. EPIC maintains one of the most popular privacy websites in the world, epic.org, which provides the public with information about emerging privacy and civil liberties issues. EPIC also provides a wide range of publications through the EPIC Bookstore.[2]

7. EPIC has a robust open government practice and routinely disseminates to the public information obtained under the Freedom of Information Act ("FOIA"), the E-Government Act, and other federal and state transparency statutes. EPIC makes this information available through the EPIC website, the biweekly EPIC Alert newsletter, and various news organizations.

8. EPIC has brought numerous successful cases seeking the release of Privacy Impact Assessments. In *EPIC v. DHS*, 926 F. Supp. 2d 311 (D.D.C. 2013), EPIC obtained a PIA and related records concerning an effort by the Department of Homeland Security to track social

[1] *See* EPIC, *About EPIC* (2018), https://epic.org/epic/about.html.
[2] EPIC, *EPIC Bookstore* (2018), https://epic.org/bookstore/.

media users and journalists.[3] EPIC made the previously undisclosed documents available to the public through the EPIC website. In *EPIC v. FBI*, 235 F. Supp. 3d 207 (D.D.C. 2017), EPIC obtained unpublished PIAs from the Federal Bureau of Investigation concerning facial recognition technology, which EPIC also made available to the public through the EPIC website.[4] In *EPIC v. DEA*, 208 F. Supp. 3d 108 (D.D.C. 2016), EPIC learned that the Drug Enforcement Administration had failed to produce PIAs for the agency's license plate reader program, a telecommunications records database, and other systems of public surveillance.[5] EPIC reported the agency's failure to produce a PIA through the EPIC website.[6] In *EPIC v. Presidential Advisory Commission on Election Integrity*, 266 F. Supp. 3d 297 (D.D.C.), *aff'd on other grounds*, 878 F.3d 371 (D.C. Cir. 2017), EPIC challenged the failure of the Presidential Advisory Commission on Election Integrity to undertake and publish a PIA prior to the collection of state voter data.[7] EPIC's suit led the Commission to suspend data collection, discontinue the use of an unsafe computer server, and delete voter information that had been illegally obtained.[8] And in *EPIC v. DHS*, No. 18-1268 (D.D.C. filed May 30, 2018), EPIC is currently challenging the failure of the Department of Homeland Security to publish a PIA for a system designed to track journalists, bloggers, and social media users.[9] EPIC's suit revealed that

---

[3] *See* EPIC, *EPIC v. Department of Homeland Security: Media Monitoring* (2015), https://www.epic.org/foia/epic-v-dhs-media-monitoring/.
[4] *See* EPIC, *EPIC v. FBI – Privacy Assessments* (2016), https://epic.org/foia/fbi/pia/.
[5] *See* EPIC, *EPIC v. DEA – Privacy Impact Assessments* (2016), https://epic.org/foia/dea/pia/.
[6] *See id.*
[7] *See* EPIC, *EPIC v. Presidential Election Commission* (2018), https://epic.org/privacy/litigation/voter/epic-v-commission/.
[8] *Id.*
[9] *See* EPIC, *EPIC v. DHS (Media Monitoring Services)* (2018), https://epic.org/foia/dhs/media-monitoring-services/.

the DHS had unlawfully failed to conduct a PIA prior to developing the "Media Monitoring" system.[10]

9. EPIC has long advocated for robust privacy protections for census respondents. EPIC was directly involved in the 2004 effort to revise the Census Bureau "sensitive data" policy after an EPIC FOIA lawsuit revealed that the DHS had acquired data on Arab Americans from the Census Bureau after 9/11.[11] In formal comments to the Census Bureau this year, EPIC opposed the decision to add a citizenship question to the 2020 census.[12] Through a recent FOIA request, EPIC also uncovered emails from Kansas Secretary of State Kris Kobach urging Commerce Secretary Wilbur Ross, on the direction of Chief White House Strategist Steve Bannon, to add a citizenship question to the 2020 Census.[13]

10. EPIC is a dues-paying membership organization, as set forth in the EPIC Articles of Incorporation[14] and the EPIC Bylaws.[15] EPIC is governed by a Board of Directors, all of whom "must be Members of" EPIC.[16]

---

[10] *Id.*
[11] EPIC, *Department of Homeland Security Obtained Data on Arab Americans From Census Bureau* (2004), https://epic.org/privacy/census/foia/.
[12] See EPIC, *Comments of the Electronic Privacy Information Center to the U.S. Census Bureau: 2020 Census* (Aug. 7, 2018), https://epic.org/apa/comments/EPIC-Census-2020-August2018.pdf.
[13] EPIC, *EPIC FOIA: EPIC Obtains Documents About Decision to Add Census Citizenship Question* (2018), https://epic.org/2018/06/epic-foia-epic-obtains-documen.html.
[14] EPIC, *Articles of Incorporation* (2018) ("The Corporation [EPIC] . . . may refer to people as 'members' pursuant to D.C. Code § 29-404.01, and the qualifications, rights, and privileges of such people shall be as set forth in the bylaws.").
[15] EPIC, *Bylaws of the Electronic Privacy Information Center* § 2.02 (as amended Jan. 26, 2018), https://epic.org/epic/bylaws.pdf [hereinafter "*EPIC Bylaws*"].
[16] *EPIC Bylaws* § 2.02.

11. EPIC's Members are "distinguished experts in law, technology, and public policy."[17] New Members are designated by EPIC following "nomination by the current Members and a vote of the Board [of Directors.]"[18]

12. Defendant United States Department of Commerce is a federal agency under 5 U.S.C. § 551(1) and 44 U.S.C. § 3502(1).[19] The Department of Commerce is headquartered in Washington, D.C.

13. The Bureau of the Census is a federal agency under 5 U.S.C. § 551(1) and 44 U.S.C. § 3502(1).[20] The Census Bureau is an agency within the Department of Commerce and is headquartered in Suitland, Md.

Facts

The Defendants' Obligation to Conduct and Publish Privacy Impact Assessments

14. Under section 208 of the E-Government Act, federal agencies—including the Department of Commerce and the Census Bureau—are required to "conduct," "ensure the review of," and "make . . . publicly available" a Privacy Impact Assessment before "initiating a new collection of information" that will be digitally stored or transmitted "in an identifiable form."[21] Information is in an identifiable form if it "permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means."[22] This includes "combination[s] of gender, race, birth date, geographic indicator, and other descriptors" that would permit

[17] *EPIC Bylaws* § 5.01.
[18] *Id.*
[19] *See* 15 U.S.C. § 1501 ("There shall be at the seat of government an executive department to be known as the Department of Commerce[.]").
[20] *See* 13 U.S.C. § 2 ("The Bureau is continued as an agency within, and under the jurisdiction of, the Department of Commerce.").
[21] E-Government Act § 208(b)(1).
[22] *Id.* § 208(d).

individuals to be uniquely identified, even in the absence of names, social security numbers, or other direct identifiers.[23]

15. The Office of Budget and Management ("OMB"), which oversees enforcement of the E-Government Act government-wide, defines a PIA as:

> [A]n analysis of how information is handled: (i) to ensure handling conforms to applicable legal, regulatory, and policy requirements regarding privacy, (ii) to determine the risks and effects of collecting, maintaining and disseminating information in identifiable form in an electronic information system, and (iii) to examine and evaluate protections and alternative processes for handling information to mitigate potential privacy risks.[24]

16. Section 208, in mandating that a PIA be conducted and published before an agency collects personally identifiable information, serves Congress's dual objectives of "mak[ing] the Federal Government more transparent and accountable," and "ensur[ing] sufficient protections for the privacy of personal information as agencies implement citizen-centered electronic Government."[25]

17. As the Census Bureau emphasizes in the Bureau's *Policy on Conducting Privacy Impact Assessments*, "PIAs are an important tool for assuring Census Bureau census and survey respondents, other agencies from whom we receive data, and the taxpayer, that the Census Bureau is minimizing privacy impacts and ensuring data confidentiality and security."[26]

18. To satisfy section 208, a PIA must specify, *inter alia*, "what information is to be collected"; "why the information is being collected"; "the intended use [by] the agency of the

---

[23] Joshua B. Bolten, Dir., Office of Mgmt. & Budget, Executive Office of the President, M03-22, Memorandum for Heads of Executive Departments and Agencies, Attachment A § II.A.2 (Sept. 26, 2003), https://georgewbush-whitehouse.archives.gov/omb/memoranda/m03-22.html [hereinafter *OMB Guidance*].
[24] *OMB Guidance* § II.A.6.
[25] E-Government Act §§ 2(b)(9), 208(a).
[26] U.S. Census Bureau, *Policy on Conducting Privacy Impact Assessments* 2 (Nov. 16, 2005), https://www2.census.gov/foia/ds_policies/ds019.pdf [hereinafter *Census Bureau PIA Policy*].

information"; "with whom the information will be shared"; "what notice or opportunities for consent would be provided"; and "how the information will be secured."[27] Additionally, a PIA "must identify what choices the agency made regarding an IT system or collection of information as a result of performing the PIA."[28]

19. A PIA must be "commensurate with the size of the information system being assessed, the sensitivity of information that is in an identifiable form in that system, and the risk of harm from unauthorized release of that information[.]"[29] As the OMB instructs, "The depth and content of the PIA should be appropriate for the nature of the information to be collected and the size and complexity of the IT system."[30]

20. The OMB also underscores that "[a]gencies must consider the information 'life cycle' (i.e., collection, use, retention, processing, disclosure and destruction) in evaluating how information handling practices at each stage may affect individuals' privacy. To be comprehensive and meaningful, privacy impact assessments require collaboration by program experts as well as experts in the areas of information technology, IT security, records management and privacy."

21. Where a PIA is required for a "major information system," the PIA should "reflect more extensive analyses of: 1. the consequences of collection and flow of information, 2. the alternatives to collection and handling as designed, 3. the appropriate measures to mitigate risks identified for each alternative and, 4. the rationale for the final design choice or business process."[31] A "major information system" includes any "system or project that requires special

[27] E-Government Act § 208(b)(2)(B)(ii).
[28] *OMB Guidance* § II.C.1.b.
[29] E-Government Act § 208(b)(2)(B)(i).
[30] *OMB Guidance* § II.C.2.a.
[31] *OMB Guidance* § II.C.2.a.ii.

management attention because of its: (i) importance to the agency mission, (ii) high development, operating and maintenance costs, (iii) high risk, (iv) high return, (v) significant role in the administration of an agency's programs, finances, property or other resources."[32] In these circumstances, a mere "checklist or template" will not satisfy an agency's PIA obligation.[33]

22. The Census Bureau further mandates that a PIA "cover the risks and effects of collecting, maintaining, and disseminating information in identifiable form in an electronic information system. A PIA must evaluate the protections and alternative processes for handling information to mitigate potential privacy risks."[34]

23. According to the OMB, "Agencies should commence a PIA when they begin to develop a new or significantly modified [information technology] system or information collection[.]"[35] Thus, when an agency intends to "develop[] . . . projects that collect, maintain or disseminate information in identifiable form from or about members of the public," the agency must first conduct a PIA.[36]

24. But an agency's privacy obligations under the E-Government Act do not end with the initial publication of a Privacy Impact Assessment. Rather, a PIA must be revised continually "to reflect changed information collection authorities, business processes or other factors affecting the collection and handling of information in identifiable form."[37] Specifically, a PIA must be "updated as necessary where a system change creates new privacy risks," including "when new

---

[32] *OMB Guidance* § II.A.4.
[33] *OMB Guidance* § II.C.2.a.iii.
[34] *Census Bureau PIA Policy* 2.
[35] *OMB Guidance* § II.C.2 (emphasis added).
[36] *OMB Guidance* § II.B.1.a.
[37] *OMB Guidance* § II.B.4; *accord* U.S. Dep't of Commerce, Office of Privacy & Open Gov't, *Privacy Compliance* (July 9, 2018), http://www.osec.doc.gov/opog/privacy/compliance.html [hereinafter *Commerce Dep't Privacy Compliance*]; *Census Bureau PIA Policy* 2.

information in identifiable form added to a collection raises the risks to personal privacy (for example, the addition of health or financial information)"; "when agencies work together on shared functions involving significant new uses or exchanges of information in identifiable form"; and "when agencies adopt or alter business processes so that government databases holding information in identifiable form are merged, centralized, matched with other databases or otherwise significantly manipulated."[38]

25. Compliance with section 208 is an essential step to fulfilling the *U.S. Census Bureau Privacy Principles*.[39] Under those principles, the Census Bureau has committed to "only collect information that is necessary for meeting the Census Bureau's mission and legal requirements"; to "be open about its programs, policies and practices to collect and protect identifiable data used to produce statistical information"; to "make it easy to access information about what [the Bureau] collect[s] and why"; and to "be considerate of respondents' time and desire for privacy."[40]

The Defendants' Collection of Personal Data Concerning Citizenship Status

26. In order to determine the apportionment of representatives "among the several States," the Census Clause of the U.S. Constitution, as amended, requires that an "actual Enumeration" of persons be undertaken every ten years "in such Manner as [Congress] shall by Law direct."[41]

[38] *OMB Guidance* §§ II.B.2.d, g, i; *accord Commerce Dep't Privacy Compliance*.
[39] U.S. Census Bureau, *U.S. Census Bureau Privacy Principles* (2006), https://www2.census.gov/foia/ds_policies/ds0pp.pdf.
[40] *Id.* at 1–2.
[41] U.S. Const. art. 1, § 2, cl. 3; *see also* U.S. Const. amend. XIV, § 2 ("Representatives shall be apportioned among the several States according to their respective numbers, counting the whole number of persons in each State, excluding Indians not taxed.").

27. In furtherance of the Census Clause, Congress has directed the Secretary of Commerce to "take a decennial census of population"[42] and to "determine the inquiries, and the number, form, and subdivisions" of the questionnaires to be used in the Census.

28. Congress has also established the Census Bureau as an agency under the Department of Commerce.[43] The Census Bureau will administer the next Census in 2020.[44]

29. By law, any person who refuses to answer "any of the questions . . . submitted to him in connection with any census"—or who willfully gives a false answer to a census question—is subject to criminal penalties.[45]

30. On March 26, 2018, Secretary of Commerce Wilbur Ross disclosed in an intra-agency letter that he "ha[d] determined that reinstatement of a citizenship question on the 2020 decennial census [wa]s necessary" and that he was "directing the Census Bureau to place the citizenship question last on the decennial census form."[46] No such question appeared on the 2010 Census,[47] nor has the Census Bureau posed a citizenship question to all census respondents since the 1950 Census.[48]

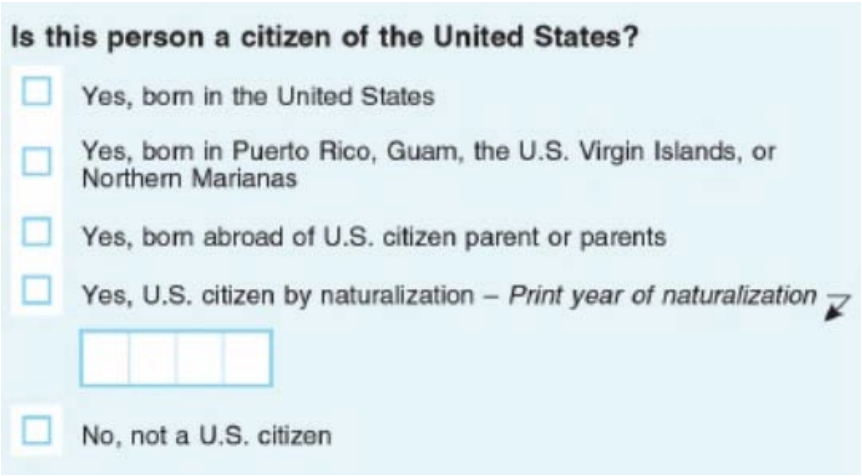
---

[42] 13 U.S.C. § 141(a).
[43] 15 U.S.C. § 1501.
[44] *See* U.S. Census Bureau, *2020 Census* (Oct. 19, 2018), https://www.census.gov/programs-surveys/decennial-census/2020-census.html.
[45] 13 U.S.C. § 221(a)–(b).
[46] Letter from Wilbur Ross, Secretary of Commerce, to Karen Dunn Kelley, Under Secretary for Economic Affairs, at 8 (March 26, 2018), *available at* https://www.scribd.com/document/374971353/Reinstatement-of-a-Citizenship-Question-on-the-2020-Decennial-Census-Questionnaire [hereinafter Ross Letter].
[47] U.S. Census Bureau, *History: 2010* (July 18, 2017), https://www.census.gov/history/www/through_the_decades/index_of_questions/2010.html.
[48] Tamara Keith, *FACT CHECK: Has Citizenship Been A Standard Census Question?*, NPR (Mar. 27, 2018), https://www.npr.org/2018/03/27/597436512/fact-check-has-citizenship-been-a-standard-census-question.

31. Secretary Ross stated that the decision to add the citizenship question was in response to a December 2017 request by the Department of Justice ("DOJ"),[49] which purportedly sought citizenship data to enable "more effective enforcement" of the Voting Rights Act.[50] The DOJ's request raised alarm and opposition from members of the U.S. Senate,[51] the attorneys general of at least twenty states,[52] and numerous mayors from across the country.[53] Moreover, Secretary Ross's explanation for his decision is at odds with his subsequent statement that he communicated with Chief White House strategist Steve Bannon and Kansas Secretary of State Kris Kobach about the citizenship question months before the DOJ made a request.[54]

[49] Letter from Arthur E. Gary, Gen. Counsel, Justice Mgmt. Div., Dep't of Justice, to Ron Jamin, U.S. Census Bureau (Dec. 12, 2017), *available at* https://www.documentcloud.org/documents/4340651-Text-of-Dec-2017-DOJ-letter-to-Census.html [hereinafter DOJ Letter].
[50] Ross Letter at 1.
[51] Letter from Sen. Dianne Feinstein et al. to Wilbur Ross, Secretary of Commerce (Jan. 5, 2018), https://www.feinstein.senate.gov/public/_cache/files/3/7/376f8dcd-7f35-4913-9e80-cd1e48e3b312/7E4C59B2988E2CC14866543EDD7E01A6.2018.01.05-census-citizeship-letter.pdf.
[52] Letter from Attorneys General of Twenty U.S. States to Wilbur Ross, Secretary of Commerce (Feb. 12, 2018), *available at* https://www.brennancenter.org/sites/default/files/legal-work/Multi-State-Attorney-General-Letter-re-2020-Census.pdf.
[53] U.S. Conference of Mayors, *Nation's Mayors to Secretary Ross: Don't Politicize Census. Remove the Citizenship Question* (Mar. 27, 2018), https://www.usmayors.org/2018/03/27/nations-mayors-to-secretary-ross-dont-politicize-census-remove-the-citizenship-question/.
[54] Defs.' Second Suppl. Resps. to Pls.' First Set of Interrogatories 2–3, *N.Y. Immigration Coal. v. U.S. Dep't of Commerce,* 18-5025 (Oct. 11, 2018), *available at* https://ag.ny.gov/sites/default/files/second_supp_res_to_rog_1_final_2018.10.11.pdf; *see also* Email from Kris Kobach, Sec'y, Kan. Dep't of State, to Wilbur Ross, Sec'y, Dep't of Commerce (Jul. 21, 2017), https://epic.org/foia/censusbureau/EPIC-18-03-22-Census-Bureau-FOIA-20180611-Production-Kobach-Emails.pdf.

32. On March 28, 2018, the Census Bureau reported to Congress the Bureau's intention to add a citizenship question to the 2020 Census.[55] The question was drafted as follows:[56]



33. The Census Bureau reported to Congress that the "question about a person's citizenship [would be] used to create statistics about citizen and noncitizen populations" and stated that "[t]hese statistics are essential for enforcing the Voting Rights Act and its protections against voting discrimination."[57]

34. The Bureau falsely implied that a citizenship question has been continuously asked on the census since 1890,[58] when in fact it has been nearly 70 years since the Bureau asked for the citizenship status of all census respondents.[59] As the Pew Research Center states plainly, "[f]or

[55] U.S. Census Bureau, *Questions Planned for the 2020 Census and American Community Survey* 7 (March 2018), https://www2.census.gov/library/publications/decennial/2020/operations/planned-questions-2020-acs.pdf.
[56] *Id.*
[57] *Id.* (capitalization altered).
[58] *Id.* at 7 n.1 ("Citizenship asked 1820, 1830, 1870, and 1890 to present.").
[59] *See* Catherine E. Shoichet, *Why putting a citizenship question on the census is a big deal*, CNN (Mar. 28, 2018), https://www.cnn.com/2018/03/27/politics/census-citizenship-question-explainer/index.html.

the first time since 1950, the U.S. Census Bureau is planning to ask everyone living in the United States whether they are citizens when it conducts its next decennial census in 2020."[60]

The Privacy Implications of Collecting Personal Data Concerning Citizenship Status

35. As the Supreme Court recognized in *Baldrige v. Shapiro*: "Although Congress has broad power to require individuals to submit responses, an accurate census depends in large part on public cooperation. To stimulate that cooperation Congress has provided assurances that information furnished . . . by individuals is to be treated as confidential."[61]

36. The Census Bureau's collection of personally identifiable information carries inherent privacy risks, but the addition of a citizenship question on the 2020 Census presents unique threats to privacy, personal security, and the accuracy of the United States Census.

37. The citizenship question would compel the release of respondents' citizenship and immigration status, potentially exposing individuals and their family members to investigation, sanction, and deportation.[62]

38. Indeed, Secretary Ross's stated basis for adding the citizenship question was to provide the DOJ with "census block level citizen voting age population ("CVAP") data."[63] The DOJ has also called on the Bureau to publicly "release this new data regarding citizenship at the same time it releases the other redistricting data[.]"[64]

39. Moreover, the most recent Privacy Impact Assessment for CEN08—a key Census Bureau division which "process[es] response data from census tests and 2020 Census operations"—

---

[60] D'Vera Cohen, *What to Know About the Citizenship Question the Census Bureau is Planning to Ask in 2020*, Pew Research Center (Mar. 30, 2018).
[61] *Baldrige v. Shapiro*, 455 U.S. 345, 354 (1982).
[62] U.S. Dep't of Commerce, U.S. Census Bureau, *QuickFacts* (2017), https://www.census.gov/quickfacts/fact/table/US/AGE775217.
[63] Ross Letter at 1; *see also* DOJ Letter at 1.
[64] DOJ Letter at 3.

states, for the first time, that the Bureau is "collecting, maintaining, or disseminating" personally identifiable information through CEN08 "[f]or criminal law enforcement activities":[65]

4.1 Indicate why the PII/BII in the IT system is being collected, maintained, or disseminated. *(Check all that apply.)*

Purpose			
For a Computer Matching Program		For administering human resources programs	x
For administrative matters		To promote information sharing initiatives	
For litigation		For criminal law enforcement activities	x
For civil enforcement activities		For intelligence activities	
To improve Federal services online		For employee or customer satisfaction	
For web measurement and customization technologies (single-session)		For web measurement and customization technologies (multi-session)	
Other (specify): For Statistical Purposes (i.e. Censuses/Surveys)			

The CEN08 PIA specifically notes that "[c]itizenship" status is among the "personally identifiable information . . . collected, maintained, or disseminated" by CEN08 (though the PIA fails completely to assess the privacy implications of handling that data):[66]

General Personal Data (GPD)					
a. Name	x	g. Date of Birth	x	m. Religion	
b. Maiden Name		h. Place of Birth		n. Financial Information	
c. Alias		i. Home Address	x	o. Medical Information	
d. Gender	x	j. Telephone Number	x	p. Military Service	
e. Age	x	k. Email Address	x	q. Physical Characteristics	
f. Race/Ethnicity	x	l. Education	x	r. Mother's Maiden Name	
s. Other general personal data (specify): Citizenship					

40. This admission in the CEN08 PIA is consistent with a June 12, 2018 email exchange between Department of Justice officials, disclosed in the course of litigation against Secretary Ross,[67] in which the officials "privately discussed the possibility that in the future census information could be shared with law enforcement."[68]

---

[65] U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN08 Decennial Information Technology Division (DITD)* 1, 7 (approved Sep. 28, 2018), http://www.osec.doc.gov/opog/privacy/Census%20PIAs/CEN08_PIA_SAOP_Approved.pdf.
[66] *Id.* at 5.
[67] Decl. of Andrew Case in Supp. of Pls.' Opp'n to Defs.' Mot. Summ. J. at Ex. B, *San Jose v. Ross*, 18-2279 (N.D. Cal. Filed Nov. 16, 2018), *available at* https://assets.documentcloud.org/documents/5193403/Nov-16-2018-Declaration-of-Andrew-Case-in.pdf.
[68] Tara Bahrampour, *Trump administration officials suggested sharing census responses with law enforcement, court documents show*, Wash. Post (Nov. 19, 2018), https://www.washingtonpost.com/local/social-issues/trump-administration-officials-suggested-

41. Transmitting personal data concerning citizenship status to the Department of Justice or other law enforcement agencies would violate the Bureau's statutory obligation not to disclose "personally identifiable information about an individual to any other individual or agency until 72 years after it was collected for the decennial census."[69]

42. Even if citizenship data were "deidentified" before dissemination, there is a material risk of reidentification. As Dr. Latanya Sweeney has demonstrated, the "practice of de-identifying data and of ad hoc generalization" used by the Census Bureau is "not sufficient to render data anonymous because combinations of attributes often combine uniquely to re-identify individuals."[70] Using Census summary data and information from other readily available sources at the time, Dr. Sweeney "found that 87% . . . of the population in the United States had reported characteristics that likely made them unique based only on {5-digit ZIP, gender, date of birth}."[71] Recent work by the National Academies of Sciences suggests that privacy-preserving techniques and privacy enhancing techniques could provide more robust approaches for deidentification, but the Census Bureau has given no indication that it will use such techniques with respect to personal data concerning citizenship status.[72]

---

sharing-census-responses-with-law-enforcement-court-documents-show/2018/11/19/41679018-ec46-11e8-8679-934a2b33be52_story.html.

[69] Pub. L. 94-416, 92 Stat. 915 (Oct. 5, 1978) (codified in relevant part at 44 U.S.C. § 2108); *see also* U.S. Dep't of Commerce, U.S. Census Bureau, *The "72-Year Rule"* (2018), https://www.census.gov/history/www/genealogy/decennial_census_records/the_72_year_rule_1.html.

[70] Latanya Sweeney, *Simple Demographics Often Identify People Uniquely* 2 (Carnegie Mellon Univ., Data Privacy Working Paper No. 3, 2000), https://dataprivacylab.org/projects/identifiability/paper1.pdf.

[71] *Id.*

[72] See Nat'l Academies of Sciences, Engineering, & Medicine, *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps* (2017), https://www.nap.edu/catalog/24893/federal-statistics-multiple-data-sources-and-privacy-protection-next-steps.

43. Historically, the misuse of census data has caused grave harm to certain populations.[73] For example, the 1910 census law prohibited the use of information supplied by businesses for non-statistical, non-census purposes, but there was no such prohibition regarding individual citizen data.[74] As a result, during World War I, the Census Bureau did in fact disclose census records to the Department of Justice and local draft boards to help enforce the draft.[75] Similarly, in 1920, the Department of Justice requested census data about individuals' citizenship for use in deportation cases.[76]

44. In 1930, Congress passed a census law that would become known as Title 13, which prohibited the Census Bureau from publishing any data identifying individuals.[77] However, the Second War Powers Act weakened this restriction and permitted the Census Bureau in 1943 to provide the U.S. Secret Service with the names, addresses, occupations, and citizenship status of every Japanese American residing in the Washington, D.C. area.[78] The Census Bureau also provided the War Department with census-block level data on Japanese Americans residing in western states to facilitate their internment.[79]

45. In 2004, an EPIC FOIA lawsuit revealed that the Census Bureau had provided the Department of Homeland Security with a list of cities containing more than 1,000 Arab-

---

[73] *See, e.g.,* Lutz. K. Berkner, *Review: The Use and Misuse of Census Data for the Historical Analysis of Family Structure*, 5 J. Interdisciplinary Hist. 721 (1975).
[74] Act of Jul. 2, 1909 (to provide for the expenses of the Thirteenth December Census, and for other purposes), ch. 2, § 25, 36 Stat. 1, 9.
[75] Margo Anderson & William Seltzer, *Challenges to the Confidentiality of U.S. Federal Statistics*, 1910-1965, 23 J Official Stat. 1, 6–7 (2007).
[76] *Id.* at 8–9.
[77] Act of Jun. 18, 1929 (to provide for the fifteenth and subsequent decennial censuses and to provide for apportionment of Representatives in Congress), ch. 28, § 11, 46 Stat. 21, 25.
[78] Margo Anderson & William Seltzer, *Census Confidentiality Under the Second War Powers Act (1942- 1947)* at 16 (Mar. 29-31, 2007) (unpublished manuscript).
[79] Comm'n on Wartime Relocation and Internment of Civilians, *Personal Justice Denied* 104-05 (1982).

American residents and a zip-code level breakdown of Arab-American populations throughout the United States, sorted by country of origin.[80] While the Census Bureau and Customs and Border Protection revised their data request policies following EPIC's FOIA case,[81] many Americans are justifiably fearful that their census responses will be used against them by other federal agencies, which can lead individuals to provide false or incomplete information.

46. The disclosure of personal data collected for census tabulation to other agencies for other purposes also threatens to undermine the integrity and accuracy of the census. In a 2018 report, the Census Bureau concluded that adding a citizenship question is "very costly, harms the quality of the census count, and would use substantially less accurate citizenship status data than are available" from other government sources.[82]

The Defendants' Failure to Assess the Privacy Impact of Collecting Personal Data Concerning Citizenship Status

47. The Department of Commerce and the Census Bureau have failed to conduct a legally sufficient—or in some cases, *any*—Privacy Impact Assessment for the collection, processing,

---

[80] EPIC, *Department of Homeland Security Obtained Data on Arab Americans From Census Bureau* (2004), https://epic.org/privacy/census/foia/; *see also* Lynette Clemetson, *Homeland Security Given Data on Arab-Americans*, N.Y. Times (Jul. 30, 2004).

[81] U.S. Customs and Border Protection, *Policy for Requesting Information of a Sensitive Nature from the Census Bureau* (Aug. 9, 2004), https://epic.org/privacy/census/foia/policy.pdf; Census Bureau News, *U.S. Census Bureau Announces Policy Regarding Sensitive Data*, press release CB04-145, August 30, 2004; Lynette Clemetson, *Census Policy On Providing Sensitive Data Is Revised*, N.Y. Times, (Aug. 31, 2004), http://www.nytimes.com/2004/08/31/us/census-policy-on-providing-sensitive-data-is-revised.html; Mikelyn Meyers, Center for Survey Management, U.S. Census Bureau, *Presentation on Respondent Confidentiality Concerns and Possible Effects on Response Rates and Data Quality for the 2020 Census*, presented at National Advisory Committee on Racial, Ethnic, and Other Populations Fall Meeting (Nov. 2, 2017), https://www2.census.gov/cac/nac/meetings/2017-11/Meyers-NAC-Confidentiality-Presentation.pdf.

[82] U.S. Dep't of Commerce, U.S. Census Bureau, *Technical Review of the Dep't of Justice Request to Add Citizenship Question to the 2020 Census* (Jan. 19, 2018), *available at* https://epic.org/foia/censusbureau/EPIC-18-03-22-Census-Bureau-FOIA-20180611-Production-Technical-Review-Memo.pdf.

and storage of citizenship information by the various Bureau systems that handle personally identifiable census data.

48. When the Department of Commerce and the Census Bureau produce a PIA concerning Bureau activities, the PIA is published on a webpage titled *U.S. Census Bureau Privacy Impact Assessments (PIAs) and Privacy Threshold Analysis (PTA)*.[83] The webpage lists Census Bureau systems and divisions that collect, process, and/or store personally identifiable information.[84] The page also provides links to (1) the most recent Privacy Impact Assessment for each system, and (2) the most recent Privacy Threshold Analysis for each system.[85]

49. At least five of the CEN systems and divisions identified on the Census Bureau's PIA webpage (CEN05, CEN08, CEN11, CEN13, and CEN18) will be used to collect, process, and/or store personally identifiable information obtained through the 2020 Census, including citizenship data.

50. The decision to collect personal data concerning citizenship status triggers the Bureau's obligation to update and publish the Privacy Impact Assessment for each of these five CENs. This is so for six reasons: (1) the Bureau is creating "a new collection of information" that will be digitally stored or transmitted "in an identifiable form";[86] (2) the introduction of citizenship data is a "factor[] affecting the collection and handling of information in identifiable form";[87] (3) the introduction of citizenship data is a change to a system that "creates new privacy risks";[88] (4)

---

[83] Dep't of Commerce, Office of Privacy & Open Gov't, *U.S. Census Bureau Privacy Impact Assessments (PIAs) and Privacy Threshold Analysis (PTA)*, (Oct. 1, 2018), http://www.osec.doc.gov/opog/privacy/Census-pias.html.
[84] *See id.*
[85] *See id.*
[86] E-Government Act § 208(b)(1).
[87] *OMB Guidance* § II.B.4.
[88] *Id.* § II.B.2.

the Bureau is introducing "new information in identifiable form to a collection [that] raises the risks to personal privacy" (similar to "health or financial information");[89] (5) the Bureau's plans for the data include "significant new uses or exchanges of information in identifiable form";[90] and (6) the Bureau is newly "merg[ing], centraliz[ing], [and] match[ing]" citizenship data with data in "other databases[.]"[91] Each these conditions independently requires the publication or republication of a fully compliant PIA.

51. The Defendants' obligation to conduct a PIA for each of the five CENs is already ripe, as the Bureau has "beg[u]n to develop a new or significantly modified [information technology] system or information collection[.]"[92]

52. EPIC, by itself and on behalf of its Members, sought out the full and complete Privacy Impact Assessment that the Defendants must publish under section 208 of the E-Government Act for each of the five CENs, including all of the information and analysis mandated by section 208(b)(2)(B)(ii), by the *OMB Guidance*, by the *Commerce Department Privacy Compliance* guidelines, and by the Census *Bureau PIA Policy*. Although a recent PIA exists for each CEN, three of the five fail to mention citizenship data at all, while the other two include zero analysis of how the collection of personal data concerning citizenship status would affect the privacy of census respondents. EPIC, by itself and on behalf of its Members, was unable to obtain—and was thereby denied—the information that it sought about each CEN and about the collection of personal data concerning citizenship status.

[89] *Id.* § II.B.2.i.
[90] *Id.* § II.B.2.g.
[91] *Id.* § II.B.2.d.
[92] *Id.* § II.C.2.

53. EPIC, by itself and on behalf of its Members, sought out the full and complete Privacy Impact Assessment required for "CEN05 Field Systems Major Application" by visiting the Census Bureau PIA webpage.[93] CEN05 is a "major information system" that "plans, organizes, coordinates, and carries out the Census Bureau's field data collection program for sample surveys, special censuses, the Economic Census, and the Decennial census."[94] As such, the system is slated to collect personal data concerning citizenship status for the 2020 Census.

54. Nonetheless, the most recent PIA for CEN05 fails to even acknowledge that the system would handle citizenship information, let alone analyze the privacy impact of that data collection.[95] As a result, EPIC, by itself and on behalf of its Members, was unable to obtain—and was thereby denied—information concerning CEN05 to which EPIC is legally entitled under the E-Government Act and implementing authorities.

55. EPIC, by itself and on behalf of its Members, also sought the full and complete Privacy Impact Assessment required for "CEN08 Decennial Information Technology Division (DITD)" by visiting the Census Bureau PIA webpage.[96] CEN08 is a Census Bureau division and major information system "consist[ing] of both general support systems and major applications," including applications that "process response data from census tests and 2020 Census operations[.]"[97] As such, the system is slated to process personal data concerning citizenship status from the 2020 Census.

---

[93] U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN05 Field Systems Major Application System* (approved June 22, 2018), http://www.osec.doc.gov/opog/privacy/Census%20PIAs/CEN05_PIA_SAOP_Approved.pdf.
[94] *Id.* at 1.
[95] *See generally id.*
[96] U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN08 Decennial Information Technology Division (DITD)* (approved Sep. 28, 2018), http://www.osec.doc.gov/opog/privacy/Census%20PIAs/CEN08_PIA_SAOP_Approved.pdf.
[97] *Id.* at 1.

56. Although the most recent PIA for CEN08 acknowledges the existence of the citizenship question through a single word ("Citizenship"),[98] the PIA entirely fails to analyze the implications of collecting this unique form of data.[99] The Bureau's one-word modification of a previous PIA, following the decision to add the most controversial question on the 2020 Census, is utterly inadequate and violates the Bureau's obligation to produce a PIA that is "commensurate with the size of the information system being assessed, the sensitivity of information that is in an identifiable form in that system, and the risk of harm from unauthorized release of that information[.]"[100] As a result, EPIC, by itself and on behalf of its Members, was unable to obtain—and was thereby denied—information concerning CEN08 to which EPIC is legally entitled under the E-Government Act and implementing authorities.

57. EPIC, by itself and on behalf of its Members, also sought the full and complete Privacy Impact Assessment required for "CEN11 Demographic Census, Surveys, and Special Processing" by visiting the Census Bureau PIA webpage.[101] CEN11 is a major information system "comprised of components that support the Demographic Directorate business functions" and includes "a Commercial off the Shelf (COTS) product used by Census Bureau demographic programs for data access, transformation, reporting, and statistical analysis."[102] As such, the system is slated to process personal data concerning citizenship status from the 2020 Census.

---

[98] *Id.* at 5.

[99] *See generally id.* The first CEN08 PIA to acknowledge the existence of the citizenship question was approved earlier on June 26, 2018. U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN08 Decennial Information Technology Division (DITD)* (approved June 26, 2018). But neither the June 26 nor the current (September 28) version of the CEN08 PIA analyzes the implications of collecting personal data concerning citizenship status.

[100] E-Government Act § 208(b)(2)(B)(i).

[101] U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN11 Demographic Census, Surveys, and Special Processing* (approved June 22, 2018), http://www.osec.doc.gov/opog/privacy/Census%20PIAs/CEN11_PIA_SAOP_Approved.pdf.

[102] *Id.* at 1.

58. Although the most recent PIA for CEN08 acknowledges the existence of the citizenship question through a single word ("Citizenship"),[103] the PIA entirely fails to analyze the implications of collecting this unique form of data.[104] The Bureau's one-word modification of a previous PIA, following the decision to add the most controversial question on the 2020 Census, is utterly inadequate and violates the Bureau's obligation to produce a PIA that is "commensurate with the size of the information system being assessed, the sensitivity of information that is in an identifiable form in that system, and the risk of harm from unauthorized release of that information[.]"[105] As a result, EPIC, by itself and on behalf of its Members, was unable to obtain—and was thereby denied—information concerning CEN08 to which EPIC is legally entitled under the E-Government Act and implementing authorities.

59. EPIC, by itself and on behalf of its Members, also sought out the full and complete Privacy Impact Assessment required for "CEN 13 Center for Economic Studies (CES)" by visiting the Census Bureau PIA webpage.[106] CEN13 is a major information system whose "data holdings include census and survey data which may contain name, gender, age, date of birth etc. from across the Census Bureau[.]"[107] As such, the system is slated to store personal data concerning citizenship status from the 2020 Census.

60. Nonetheless, the most recent PIA for CEN13 fails to even acknowledge that the system would handle citizenship information, let alone analyze the privacy impact of storing that data.[108]

---

[103] *Id.* at 4.
[104] *See generally id.*
[105] E-Government Act § 208(b)(2)(B)(i).
[106] U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN 18 Enterprise Applications* (approved June 26, 2018), http://www.osec.doc.gov/opog/privacy/Census%20PIAs/CEN13_PIA_SAOP_Approved.pdf.
[107] *Id.* at 1.
[108] *See generally id.*

As a result, EPIC, by itself and on behalf of its Members, was unable to obtain—and was thereby denied—information concerning CEN13 to which EPIC is legally entitled under the E-Government Act and implementing authorities.

61. EPIC, by itself and on behalf of its Members, also sought out the full and complete Privacy Impact Assessment required for "CEN 18 Enterprise Applications" by visiting the Census Bureau PIA webpage.[109] CEN18 is a major information system "used to deliver applications to end users of the U.S. Census Bureau network."[110] The system maintains "survey and census information," including "personal names, personal addresses, personal contact information (telephone numbers, email address), business information, occupation, medical information, tax information, etc."[111] As such, the system is slated to store personal data concerning citizenship status from the 2020 Census.

62. Nonetheless, the most recent PIA for CEN18 fails to even acknowledge that the system would handle citizenship information, let alone analyze the privacy impact of storing that data.[112] As a result, EPIC, by itself and on behalf of its Members, was unable to obtain—and was thereby denied—information concerning CEN18 to which EPIC is legally entitled under the E-Government Act and implementing authorities.

63. Though Secretary Ross's plan to add the citizenship question to the 2020 Census is arguably the most consequential decision in the Census Bureau's recent history, the Department

[109] U.S. Dep't of Commerce, *Privacy Impact Assessment for the CEN 13 Center for Economic Studies (CES)* (approved June 26, 2018), http://www.osec.doc.gov/opog/privacy/Census%20PIAs/CEN18_PIA_SAOP_Approved.pdf.
[110] *Id.* at 1.
[111] *Id.*
[112] *See generally id.*

of Commerce and the Bureau have failed to conduct *any* of the privacy analysis required by the E-Government Act for a major collection of personally identifiable information.

Count I

Violation of APA: Unlawful Agency Action

64. Plaintiff asserts and incorporates by reference paragraphs 1–63.

65. By placing a citizenship question on the 2020 Census and initiating the process of collecting personal data concerning citizenship status, the Defendants have unlawfully begun to develop a new or significantly modified collection of information prior to creating, reviewing, and publishing the full and complete Privacy Impact Assessments required by section 208(b) of the E-Government Act of 2002.

66. In violating section 208(b) the E-Government Act, Defendants have taken agency actions that are arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law under 5 U.S.C. § 706(2)(a) and short of statutory right under 5 U.S.C. § 706(2)(c).

67. Defendants' decision to collect personal data concerning citizenship status through the 2020 Census, Defendants' placement of a citizenship question on the 2020 Census, and Defendants' initiation of the citizenship data collection process are final agency actions within the meaning of 5 U.S.C. § 704.

68. Plaintiff EPIC is adversely affected, aggrieved, and injured in fact by Defendants' actions. By beginning the process of collecting personal data concerning citizenship status without publishing the full and complete Privacy Impact Assessments required by section 208(b) of E-Government Act, Defendants have frustrated Plaintiff's longstanding mission to educate the public about the privacy implications of government databases that contain personally

identifiable information and—in particular—about the collection of personally identifiable information via the census.

69. Plaintiff EPIC is also adversely affected, aggrieved, and injured in fact by Defendants' actions through EPIC's Members. EPIC's Members, most of whom reside in the United States, are required by law to respond to the citizenship question that will be imminently posed to them on the 2020 Census. 13 U.S.C. § 221(a)–(b). By beginning the process of collecting citizenship data without publishing the full and complete Privacy Impact Assessments required by section 208(b) of the E-Government Act, Defendants have unlawfully denied EPIC's Members—and by extension, EPIC—a full assessment of how their privacy interests will be affected before the collection process is initiated.

70. Plaintiff has exhausted all applicable administrative remedies.

Count II

Violation of APA: Agency Action Unlawful Withheld

71. Plaintiff asserts and incorporates by reference paragraphs 1–63.

72. Defendants have failed to create, review, and publish the full and complete Privacy Impact Assessments required by section 208(b) of E-Government Act of 2002 for Defendants' decision to collect personal data concerning citizenship status through the 2020 Census, for Defendants' placement of a citizenship question on the 2020 Census, and for Defendants' initiation of the citizenship data collection process.

73. In failing to take the steps required by section 208(b) of E-Government Act of 2002, Defendants have unlawfully withheld or unreasonably delayed agency action in violation of 5 U.S.C. § 706(1).

74. Plaintiff EPIC is adversely affected, aggrieved, and injured in fact by Defendants' inaction. By failing to publish the full and complete Privacy Impact Assessments required by section 208(b) of the E-Government Act prior to beginning the process of collecting personal data concerning citizenship status, Defendants have frustrated Plaintiff's longstanding mission to educate the public about the privacy implications of government databases that contain personally identifiable information and—in particular—about the collection of personally identifiable information via the census.

75. Plaintiff EPIC is also adversely affected, aggrieved, and injured in fact by Defendants' actions through EPIC's Members. EPIC's Members, most of whom reside in the United States, are required by law to respond to the citizenship question that will be imminently posed to them on the 2020 Census. 13 U.S.C. § 221(a)–(b). By failing to publish the full and complete Privacy Impact Assessments required by section 208(b) of the E-Government Act prior to beginning the process of collecting citizenship data, Defendants have unlawfully denied EPIC's Members—and by extension, EPIC—a full assessment of how their privacy interests will be affected before the collection process is initiated.

76. Plaintiff has exhausted all applicable administrative remedies.

Count III

Claim for Declaratory Relief

77. Plaintiff asserts and incorporates by reference paragraphs 1–63.

78. Plaintiff is entitled under 28 U.S.C. § 2201(a) to a declaration of the rights and other legal relations of the parties with respect to the claims set forth in Counts I-II.

Requested Relief

WHEREFORE, Plaintiff requests this Court:

A. Hold unlawful and set aside the Defendants' decision to collect citizenship data through the 2020 Census, Defendants' placement of a citizenship question on the 2020 Census, and Defendants' initiation of the citizenship data collection process;

B. Order Defendants to suspend and revoke their decision collect citizenship data through the 2020 Census until the Defendants have conducted, reviewed, and published the full and complete Privacy Impact Assessments required by section 208(b) of E-Government Act of 2002;

C. Order Defendants to revoke and remove the citizenship question from the 2020 Census until the Defendants have conducted, reviewed, and published the full and complete Privacy Impact Assessments required by section 208(b) of E-Government Act of 2002;

D. Order Defendants to cease and desist from any action in furtherance of Defendants' plan to collect citizenship data through the 2020 Census until the Defendants have conducted, reviewed, and published the full and complete Privacy Impact Assessments required by section 208(b) of E-Government Act of 2002;

E. Order Defendants to conduct, review, and publish the full and complete Privacy Impact Assessments required by section 208(b) of E-Government Act of 2002 for the Census Bureau's collection of citizenship data;

F. Award EPIC costs and reasonable attorney's fees incurred in this action; and

G. Grant such other relief as the Court may deem just and proper.

Respectfully Submitted,

MARC ROTENBERG, D.C. Bar #422825
EPIC President and Executive Director

ALAN BUTLER, D.C. Bar #1012128
EPIC Senior Counsel

/s/ John Davisson
JOHN DAVISSON, D.C. Bar #1531914
EPIC Counsel

ELECTRONIC PRIVACY
INFORMATION CENTER
1718 Connecticut Avenue, N.W.
Suite 200
Washington, D.C. 20009
(202) 483-1140 (telephone)
(202) 483-1248 (facsimile)

Dated: November 20, 2018

UNITED STATES DEPARTMENT OF COMMERCE
The Secretary of Commerce
Washington, D.C. 20230

To: Karen Dunn Kelley, Under Secretary for Economic Affairs

From: Secretary Wilbur Ross

Date: March 26, 2018

Re: Reinstatement of a Citizenship Question on the 2020 Decennial Census Questionnaire

Dear Under Secretary Kelley:

As you know, on December 12, 2017, the Department of Justice ("DOJ") requested that the Census Bureau reinstate a citizenship question on the decennial census to provide census block level citizenship voting age population ("CVAP") data that are not currently available from government survey data ("DOJ request"). DOJ and the courts use CVAP data for determining violations of Section 2 of the Voting Rights Act ("VRA"), and having these data at the census block level will permit more effective enforcement of the Act. Section 2 protects minority population voting rights.

Following receipt of the DOJ request, I set out to take a hard look at the request and ensure that I considered all facts and data relevant to the question so that I could make an informed decision on how to respond. To that end, the Department of Commerce ("Department") immediately initiated a comprehensive review process led by the Census Bureau.

The Department and Census Bureau's review of the DOJ request – as with all significant Census assessments – prioritized the goal of obtaining *complete and accurate data*. The decennial census is mandated in the Constitution and its data are relied on for a myriad of important government decisions, including apportionment of Congressional seats among states, enforcement of voting rights laws, and allocation of federal funds. These are foundational elements of our democracy, and it is therefore incumbent upon the Department and the Census Bureau to make every effort to provide a complete and accurate decennial census.

At my direction, the Census Bureau and the Department's Office of the Secretary began a thorough assessment that included legal, program, and policy considerations. As part of the process, I also met with Census Bureau leadership on multiple occasions to discuss their process for reviewing the DOJ request, their data analysis, my questions about accuracy and response rates, and their recommendations. At present, the Census Bureau leadership are all career civil servants. In addition, my staff and I reviewed over 50 incoming letters from stakeholders, interest groups, Members of Congress, and state and local officials regarding reinstatement of a citizenship question on the 2020 decennial census, and I personally had specific conversations on

the citizenship question with over 24 diverse, well informed and interested parties representing a broad range of views. My staff and I have also monitored press coverage of this issue.

Congress has delegated to me the authority to determine which questions should be asked on the decennial census, and I may exercise my discretion to reinstate the citizenship question on the 2020 decennial census, especially based on DOJ's request for improved CVAP data to enforce the VRA. By law, the list of decennial census questions is to be submitted two years prior to the decennial census – in this case, no later than March 31, 2018.

The Department's review demonstrated that collection of citizenship data by the Census has been a long-standing historical practice. Prior decennial census surveys of the entire United States population consistently asked citizenship questions up until 1950, and Census Bureau surveys of sample populations continue to ask citizenship questions to this day. In 2000, the decennial census "long form" survey, which was distributed to one in six people in the U.S., included a question on citizenship. Following the 2000 decennial census, the "long form" sample was replaced by the American Community Survey ("ACS"), which has included a citizenship question since 2005. Therefore, the citizenship question has been well tested.

DOJ seeks to obtain CVAP data for census blocks, block groups, counties, towns, and other locations where potential Section 2 violations are alleged or suspected, and DOJ states that the current data collected under the ACS are insufficient in scope, detail, and certainty to meet its purpose under the VRA. The Census Bureau has advised me that the census-block-level citizenship data requested by DOJ are not available using the annual ACS, which as noted earlier does ask a citizenship question and is the present method used to provide DOJ and the courts with data used to enforce Section 2 of the VRA. The ACS is sent on an annual basis to a sample of approximately 2.6 percent of the population.

To provide the data requested by DOJ, the Census Bureau initially analyzed three alternatives: Option A was to continue the status quo and use ACS responses; Option B was placing the ACS citizenship question on the decennial census, which goes to every American household; and Option C was not placing a question on the decennial census and instead providing DOJ with a citizenship analysis for the entire population using federal administrative record data that Census has agreements with other agencies to access for statistical purposes.

Option A contemplates rejection of the DOJ request and represents the status quo baseline. Under Option A, the 2020 decennial census would not include the question on citizenship that DOJ requested and therefore would not provide DOJ with improved CVAP data. Additionally, the block-group level CVAP data currently obtained through the ACS has associated margins of error because the ACS is extrapolated based on sample surveys of the population. Providing more precise block-level data would require sophisticated statistical modeling, and if Option A is selected, the Census Bureau advised that it would need to deploy a team of experts to develop model-based methods that attempt to better facilitate DOJ's request for more specific data. But the Census Bureau did not assert and could not confirm that such data modeling is possible for census-block-level data with a sufficient degree of accuracy. Regardless, DOJ's request is based at least in part on the fact that existing ACS citizenship data-sets lack specificity and

completeness. Any future modeling from these incomplete data would only compound that problem.

Option A would provide no improved citizenship count, as the existing ACS sampling would still fail to obtain *actual*, complete number counts, especially for certain lower population areas or voting districts, and there is no guarantee that data could be improved using small-area modeling methods. Therefore, I have concluded that Option A is not a suitable option.

The Census Bureau and many stakeholders expressed concern that **Option B**, which would add a citizenship question to the decennial census, would negatively impact the response rate for non-citizens. A significantly lower response rate by non-citizens could reduce the accuracy of the decennial census and increase costs for non-response follow up ("NRFU") operations. However, neither the Census Bureau nor the concerned stakeholders could document that the response rate would in fact decline materially. In discussing the question with the national survey agency Nielsen, it stated that it had added questions from the ACS on sensitive topics such as place of birth and immigration status to certain short survey forms without any appreciable decrease in response rates. Further, the former director of the Census Bureau during the last decennial census told me that, while he wished there were data to answer the question, none existed to his knowledge. Nielsen's Senior Vice President for Data Science and the former Deputy Director and Chief Operating Officer of the Census Bureau under President George W. Bush also confirmed that, to the best of their knowledge, no empirical data existed on the impact of a citizenship question on responses.

When analyzing Option B, the Census Bureau attempted to assess the impact that reinstatement of a citizenship question on the decennial census would have on response rates by drawing comparisons to ACS responses. However, such comparative analysis was challenging, as response rates generally vary between decennial censuses and other census sample surveys. For example, ACS self-response rates were 3.1 percentage points less than self-response rates for the 2010 decennial census. The Bureau attributed this difference to the greater outreach and follow-up associated with the Constitutionally-mandated decennial census. Further, the decennial census has differed significantly in nature from the sample surveys. For example, the 2000 decennial census survey contained only eight questions. Conversely, the 2000 "long form" sample survey contained over 50 questions, and the Census Bureau estimated it took an average of over 30 minutes to complete. ACS surveys include over 45 questions on numerous topics, including the number of hours worked, income information, and housing characteristics.

The Census Bureau determined that, for 2013-2016 ACS surveys, nonresponses to the citizenship question for non-Hispanic whites ranged from 6.0 to 6.3 percent, for non-Hispanic blacks ranged from 12.0 to 12.6 percent, and for Hispanics ranged from 11.6 to 12.3 percent. However, these rates were comparable to nonresponse rates for other questions on the 2013 and 2016 ACS. Census Bureau estimates showed similar nonresponse rate ranges occurred for questions on the ACS asking the number times the respondent was married, 4.7 to 6.9 percent; educational attainment, 5.6 to 8.5 percent; monthly gas costs, 9.6 to 9.9 percent; weeks worked in the past 12 months, 6.9 to 10.6 percent; wages/salary income, 8.1 to 13.4 percent; and yearly property insurance, 23.9 to 25.6 percent.

The Census Bureau also compared the self-response rate differences between citizen and non-citizen households' response rates for the 2000 decennial census short form (which did not include a citizenship question) and the 2000 decennial census long form survey (the long form survey, distributed to only one in six households, included a citizenship question in 2000). Census found the decline in self-response rates for non-citizens to be 3.3 percent greater than for citizen households. However, Census was not able to isolate what percentage of decline was caused by the inclusion of a citizenship question rather than some other aspect of the long form survey (it contained over six times as many questions covering a range of topics). Indeed, the Census Bureau analysis showed that for the 2000 decennial census there was a significant drop in self response rates overall between the short and long form; the mail response rate was 66.4 percent for the short form and only 53.9 percent for the long form survey. So while there is widespread belief among many parties that adding a citizenship question could reduce response rates, the Census Bureau's analysis did not provide definitive, empirical support for that belief.

Option C, the use of administrative records rather than placing a citizenship question on the decennial census, was a potentially appealing solution to the DOJ request. The use of administrative records is increasingly part of the fabric and design of modern censuses, and the Census Bureau has been using administrative record data to improve the accuracy and reduce the cost of censuses since the early 20th century. A Census Bureau analysis matching administrative records with the 2010 decennial census and ACS responses over several more recent years showed that using administrative records could be more accurate than self-responses in the case of non-citizens. That Census Bureau analysis showed that between 28 and 34 percent of the citizenship self-responses for persons that administrative records show are non-citizens were inaccurate. In other words, when non-citizens respond to long form or ACS questions on citizenship, they inaccurately mark "citizen" about 30 percent of the time. However, the Census Bureau is still evolving its use of administrative records, and the Bureau does not yet have a complete administrative records data set for the entire population. Thus, using administrative records alone to provide DOJ with CVAP data would provide an incomplete picture. In the 2010 decennial census, the Census Bureau was able to match 88.6 percent of the population with what the Bureau considers credible administrative record data. While impressive, this means that more than 10 percent of the American population – some 25 million voting age people – would need to have their citizenship imputed by the Census Bureau. Given the scale of this number, it was imperative that another option be developed to provide a greater level of accuracy than either self-response alone or use of administrative records alone would presently provide.

I therefore asked the Census Bureau to develop a fourth alternative, **Option D**, which would combine Options B and C. Under Option D, the ACS citizenship question would be asked on the decennial census, and the Census Bureau would use the two years remaining until the 2020 decennial census to further enhance its administrative record data sets, protocols, and statistical models to provide more complete and accurate data. This approach would maximize the Census Bureau's ability to match the decennial census responses with administrative records. Accordingly, at my direction the Census Bureau is working to obtain as many additional Federal and state administrative records as possible to provide more comprehensive information for the population.

It is my judgment that Option D will provide DOJ with the most complete and accurate CVAP data in response to its request. Asking the citizenship question of 100 percent of the population gives each respondent the opportunity to provide an answer. This may eliminate the need for the Census Bureau to have to impute an answer for millions of people. For the approximately 90 percent of the population who are citizens, this question is no additional imposition. And for the approximately 70 percent of non-citizens who already answer this question accurately on the ACS, the question is no additional imposition since census responses by law may only be used anonymously and for statistical purposes. Finally, placing the question on the decennial census and directing the Census Bureau to determine the best means to compare the decennial census responses with administrative records will permit the Census Bureau to determine the inaccurate response rate for citizens and non-citizens alike using the entire population. This will enable the Census Bureau to establish, to the best of its ability, the accurate ratio of citizen to non-citizen responses to impute for that small percentage of cases where it is necessary to do so.

Consideration of Impacts I have carefully considered the argument that the reinstatement of the citizenship question on the decennial census would depress response rate. Because a lower response rate would lead to increased non-response follow-up costs and less accurate responses, this factor was an important consideration in the decision-making process. I find that the need for accurate citizenship data and the limited burden that the reinstatement of the citizenship question would impose outweigh fears about a potentially lower response rate.

Importantly, the Department's review found that limited empirical evidence exists about whether adding a citizenship question would decrease response rates materially. Concerns about decreased response rates generally fell into the following two categories – distrust of government and increased burden. First, stakeholders, particularly those who represented immigrant constituencies, noted that members of their respective communities generally distrusted the government and especially distrusted efforts by government agencies to obtain information about them. Stakeholders from California referenced the difficulty that government agencies faced obtaining any information from immigrants as part of the relief efforts after the California wildfires. These government agencies were not seeking to ascertain the citizenship status of these wildfire victims. Other stakeholders referenced the political climate generally and fears that Census responses could be used for law enforcement purposes. But no one provided evidence that reinstating a citizenship question on the decennial census would materially decrease response rates among those who generally distrusted government and government information collection efforts, disliked the current administration, or feared law enforcement. Rather, stakeholders merely identified residents who made the decision not to participate regardless of whether the Census includes a citizenship question. The reinstatement of a citizenship question will not decrease the response rate of residents who already decided not to respond. And no one provided evidence that there are residents who would respond accurately to a decennial census that did not contain a citizenship question but would not respond if it did (although many believed that such residents had to exist). While it is possible this belief is true, there is no information available to determine the number of people who would in fact not respond due to a citizenship question being added, and no one has identified any mechanism for making such a determination.

A second concern that stakeholders advanced is that recipients are generally less likely to respond to a survey that contained more questions than one that contained fewer. The former Deputy Director and Chief Operating Officer of the Census Bureau during the George W. Bush administration described the decennial census as particularly fragile and stated that any effort to add questions risked lowering the response rate, especially a question about citizenship in the current political environment. However, there is limited empirical evidence to support this view. A former Census Bureau Director during the Obama Administration who oversaw the last decennial census noted as much. He stated that, even though he believed that the reinstatement of a citizenship question would decrease response rate, there is limited evidence to support this conclusion. This same former director noted that, in the years preceding the decennial census, certain interest groups consistently attack the census and discourage participation. While the reinstatement of a citizenship question may be a data point on which these interest groups seize in 2019, past experience demonstrates that it is likely efforts to undermine the decennial census will occur again regardless of whether the decennial census includes a citizenship question. There is no evidence that residents who are persuaded by these disruptive efforts are more or less likely to make their respective decisions about participation based specifically on the reinstatement of a citizenship question. And there are actions that the Census Bureau and stakeholder groups are taking to mitigate the impact of these attacks on the decennial census.

Additional empirical evidence about the impact of sensitive questions on survey response rates came from the SVP of Data Science at Nielsen. When Nielsen added questions on place of birth and time of arrival in the United States (both of which were taken from the ACS) to a short survey, the response rate was not materially different than it had been before these two questions were added. Similarly, the former Deputy Director and COO of the Census during the George W. Bush Administration shared an example of a citizenship-like question that he believed would negatively impact response rates but did not. He cited to the Department of Homeland Security's 2004 request to the Census Bureau to provide aggregate data on the number of Arab Americans by zip code in certain areas of the country. The Census Bureau complied, and Census employees, including the then-Deputy Director, believed that the resulting political firestorm would depress response rates for further Census Bureau surveys in the impacted communities. But the response rate did not change materially.

Two other themes emerged from stakeholder calls that merit discussion. First, several stakeholders who opposed reinstatement of the citizenship question did not appreciate that the question had been asked in some form or another for nearly 200 years. Second, other stakeholders who opposed reinstatement did so based on the assumption that the data on citizenship that the Census Bureau collects through the ACS are accurate, thereby obviating the need to ask the question on the decennial census. But as discussed above, the Census Bureau estimates that between 28 and 34 percent of citizenship self-responses on the ACS for persons that administrative records show are non-citizens were inaccurate. Because these stakeholder concerns were based on incorrect premises, they are not sufficient to change my decision.

Finally, I have considered whether reinstating the citizenship question on the 2020 Census will lead to any significant monetary costs, programmatic or otherwise. The Census Bureau staff have advised that the costs of preparing and adding the question would be minimal due in large part to the fact that the citizenship question is already included on the ACS, and thus the citizenship question has already undergone the cognitive research and questionnaire testing required for new questions. Additionally, changes to the Internet Self-Response instrument, revising the Census Questionnaire Assistance, and redesigning of the printed questionnaire can be easily implemented for questions that are finalized prior to the submission of the list of questions to Congress.

The Census Bureau also considered whether non-response follow-up increases resulting from inclusion of the citizenship question would lead to increased costs. As noted above, this estimate was difficult to assess given the Census Bureau and Department's inability to determine what impact there will be on decennial census survey responses. The Bureau provided a rough estimate that postulated that up to 630,000 additional households may require NRFU operations if a citizenship question is added to the 2020 decennial census. However, even assuming that estimate is correct, this additional ½ percent increase in NRFU operations falls well within the margin of error that the Department, with the support of the Census Bureau, provided to Congress in the revised Lifecycle Cost Estimate ("LCE") this past fall. That LCE assumed that NRFU operations might increase by 3 percent due to numerous factors, including a greater increase in citizen mistrust of government, difficulties in accessing the Internet to respond, and other factors.

Inclusion of a citizenship question on this country's decennial census is not new – the decision to collect citizenship information from Americans through the decennial census was first made centuries ago. The decision to include a citizenship question on a national census is also not uncommon. The United Nations recommends that its member countries ask census questions identifying both an individual's country of birth and the country of citizenship. *Principals and Recommendations for Population and Housing Censuses (Revision 3)*, UNITED NATIONS 121 (2017). Additionally, for countries in which the population may include a large portion of naturalized citizens, the United Nations notes that, "it may be important to collect information on the method of acquisition of citizenship." *Id.* at 123. And it is important to note that other major democracies inquire about citizenship on their census, including Australia, Canada, France, Germany, Indonesia, Ireland, Mexico, Spain, and the United Kingdom, to name a few.

The Department of Commerce is not able to determine definitively how inclusion of a citizenship question on the decennial census will impact responsiveness. However, even if there is some impact on responses, the value of more complete and accurate data derived from surveying the entire population outweighs such concerns. Completing and returning decennial census questionnaires is required by Federal law, those responses are protected by law, and inclusion of a citizenship question on the 2020 decennial census will provide more complete information for those who respond. The citizenship data provided to DOJ will be more accurate with the question than without it, which is of greater importance than any adverse effect that may result from people violating their legal duty to respond.

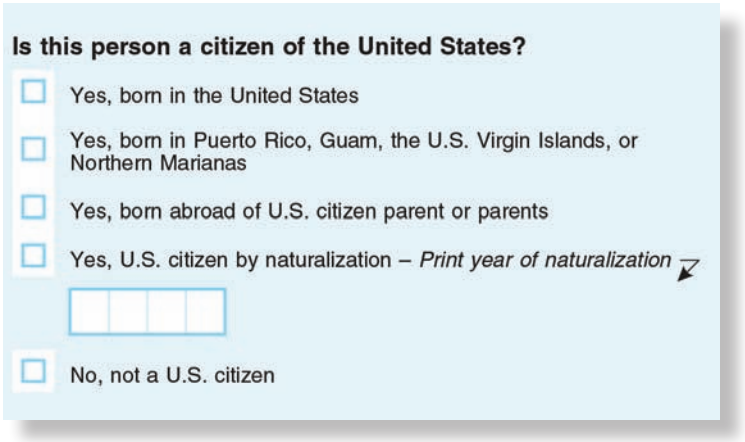
To conclude, after a thorough review of the legal, program, and policy considerations, as well as numerous discussions with the Census Bureau leadership and interested stakeholders, I have determined that reinstatement of a citizenship question on the 2020 decennial census is necessary to provide complete and accurate data in response to the DOJ request. To minimize any impact on decennial census response rates, I am directing the Census Bureau to place the citizenship question last on the decennial census form.

Please make my decision known to Census Bureau personnel and Members of Congress prior to March 31, 2018. I look forward to continuing to work with the Census Bureau as we strive for a complete and accurate 2020 decennial census.

CC: Ron Jarmin, performing the nonexclusive functions and duties of the Director of the Census Bureau

Enrique Lamas, performing the nonexclusive functions and duties of the Deputy Director of the Census Bureau

Citizenship asked since 1820.[1]



A QUESTION ABOUT A PERSON'S CITIZENSHIP IS USED TO CREATE STATISTICS ABOUT CITIZEN AND NONCITIZEN POPULATIONS.

These statistics are essential for enforcing the Voting Rights Act and its protections against voting discrimination.

[1] Citizenship asked 1820, 1830, 1870, and 1890 to present.

CITIZENSHIP DATA HELP COMMUNITIES:

Enforce Voting Rights Law

Knowing how many people reside in the community and how many of those people are citizens, in combination with other information, provides the statistical information that helps the government enforce Section 2 of the Voting Rights Act and its protections against discrimination in voting.

Understand Changes

Knowing how many citizens and noncitizens live in the United States, in combination with other information, is of interest to researchers, advocacy groups, and policymakers.

Simple Demographics Often Identify People Uniquely

Latanya Sweeney
Carnegie Mellon University
latanya@andrew.cmu.edu

This work was funded in part by H. John Heinz III School of Public Policy and Management at Carnegie Mellon University and by a grant from the U.S. Bureau of Census.

1. Abstract

In this document, I report on experiments I conducted using 1990 U.S. Census summary data to determine how many individuals within geographically situated populations had combinations of demographic values that occurred infrequently. It was found that combinations of few characteristics often combine in populations to uniquely or nearly uniquely identify some individuals. Clearly, data released containing such information about these individuals should not be considered anonymous. Yet, health and other person-specific data are publicly available in this form. Here are some surprising results using only three fields of information, even though typical data releases contain many more fields. It was found that 87% (216 million of 248 million) of the population in the United States had reported characteristics that likely made them unique based only on {5-digit ZIP, gender, date of birth}. About half of the U.S. population (132 million of 248 million or 53%) are likely to be uniquely identified by only {place, gender, date of birth}, where place is basically the city, town, or municipality in which the person resides. And even at the county level, {county, gender, date of birth} are likely to uniquely identify 18% of the U.S. population. In general, few characteristics are needed to uniquely identify a person.

2. Introduction

Data holders often collect person-specific data and then release derivatives of collected data on a public or semi-public basis after removing all explicit identifiers, such as name, address and phone number. Evidence is provided in this document that this practice of de-identifying data and of ad hoc generalization are not sufficient to render data anonymous because combinations of attributes often combine uniquely to re-identify individuals.

2.1. Linking to re-identify de-identified data

In this subsection, I will demonstrate how linking can be used to re-identify de-identified data. The National Association of Health Data Organizations (NAHDO) reported that 44 states have legislative mandates to collect hospital level data and that 17 states have started collecting ambulatory care data from hospitals, physicians offices, clinics, and so forth [1]. These data collections often include the patient's *ZIP* code, *birth date*, *gender*, and *ethnicity* but no explicit identifiers like *name* or *address*. The leftmost circle in Figure 1 contains some of the data elements collected and shared.

For twenty dollars I purchased the voter registration list for Cambridge Massachusetts and received the information on two diskettes [2]. The rightmost circle in Figure 1 shows that these data included the name, address, ZIP code, birth date, and gender of each voter. This information can be linked using *ZIP*, *birth date* and *gender* to the medical information, thereby linking diagnosis, procedures, and medications to particularly named individuals. The question that remains of course is how unique would such linking be.

In general I can say that the greater the number and detail of attributes reported about an entity, the more likely that those attributes combine uniquely to identify the entity. For example, in the voter list, there were 2 possible values for gender and 5 possible five-digit ZIP codes; birth dates were within a range of 365 days for 100 years. This gives 365,000 unique values, but there were only 54,805 voters.

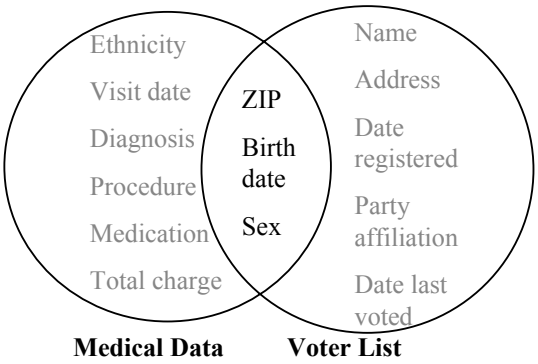


Figure 1 Linking to re-identify data

2.2. Publicly and semi-publicly available health data

As mentioned in the previous subsection, most states (44 of 50 or 88%) collect hospital discharge data [3]. Many of these states have subsequently distributed copies of these data to researchers, sold copies to industry and made versions publicly available. While there are many possible sources of patient-specific data, these represent a class of data collections that are often publicly and semi-publicly available.

#	Field description	Size
1	HOSPITAL ID NUMBER	12
2	PATIENT DATE OF BIRTH (MMDDYYYY)	8
3	SEX	1
4	ADMIT DATE (MMDYYYY)	8
5	DISCHARGE DATE (MMDDYYYY)	8
6	ADMIT SOURCE	1
7	ADMIT TYPE	1
8	LENGTH OF STAY (DAYS)	4
9	PATIENT STATUS	2
10	PRINCIPAL DIAGNOSIS CODE	6
11	SECONDARY DIAGNOSIS CODE - 1	6
12	SECONDARY DIAGNOSIS CODE - 2	6
13	SECONDARY DIAGNOSIS CODE - 3	6
14	SECONDARY DIAGNOSIS CODE - 4	6
15	SECONDARY DIAGNOSIS CODE - 5	6
16	SECONDARY DIAGNOSIS CODE - 6	6
17	SECONDARY DIAGNOSIS CODE - 7	6
18	SECONDARY DIAGNOSIS CODE - 8	6
19	PRINCIPAL PROCEDURE CODE	7
20	SECONDARY PROCEDURE CODE - 1	7
21	SECONDARY PROCEDURE CODE - 2	7
22	SECONDARY PROCEDURE CODE - 3	7
23	SECONDARY PROCEDURE CODE - 4	7
24	SECONDARY PROCEDURE CODE - 5	7
25	DRG CODE	3
26	MDC CODE	2
27	TOTAL CHARGES	9
28	ROOM AND BOARD CHARGES	9
29	ANCILLARY CHARGES	9
30	ANESTHESIOLOGY CHARGES	9
31	PHARMACY CHARGES	9
32	RADIOLOGY CHARGES	9
33	CLINICAL LAB CHARGES	9
34	LABOR-DELIVERY CHARGES	9
35	OPERATING ROOM CHARGES	9
36	ONCOLOGY CHARGES	9
37	OTHER CHARGES	9
38	NEWBORN INDICATOR	1
39	PAYER ID 1	9
40	TYPE CODE 1	1
41	PAYER ID 2	9
42	TYPE CODE 2	1
43	PAYER ID 3	9
44	TYPE CODE 3	1
45	PATIENT ZIP CODE	5
46	Patient Origin COUNTY	3
47	Patient Origin PLANNING AREA	3
48	Patient Origin HSA	2
49	PATIENT CONTROL NUMBER	
50	HOSPITAL HSA	2

Figure 2 IHCCCC Research Health Data

The Illinois Health Care Cost Containment Council (IHCCCC) is the organization in the State of Illinois that collects and disseminates health care cost data on hospital visits in Illinois. IHCCCC reports more than 97% compliance by Illinois hospitals in providing the information

[4]. Figure 2 contains a sample of the kinds of fields of information that are not only collected, but also disseminated.

Of the states mentioned in the NAHDO report, 22 of these states contribute to a national database called the State Inpatient Database (*SID*) sponsored by the Agency for Healthcare Research and Quality (AHRQ). A copy of each patient's hospital visit in these states is sent to AHRQ for inclusion in *SID*. Some of the fields provided in *SID* are listed in Figure 3 along with the compliance of the 13 states that contributed to *SID*'s 1997 data [5].

Field	Comments	#states	%states
Patient Age	years	13	100%
Patient Date of birth	month, year	5	38%
Patient Gender		13	100%
Patient Racial background		11	85%
Patient ZIP	5-digit	9	69%
Patient ID	encrypted (or scrambled)	3	23%
Admission date	month, year	8	62%
Admission day of week		12	92%
Admission source	emergency, court/law, etc	13	100%
Birth weight	for newborns	5	38%
Discharge date	month, year	7	54%
Length of stay		13	100%
Discharge status	routine, death, nursing home, etc	13	100%
Diagnosis Codes	ICD9, from 10 to30	13	100%
Procedure Codes	from 6 to 21	13	100%
Hospital ID	AHA#	12	92%
Hospital county		12	92%
Primary payer	Medicare, insurance, self-pay, etc	13	100%
Charges	from 1 to 63 categories	11	85%

Figure 3 Some data elements for AHRQ's State Inpatient Database (13 participating states)

State	Month and Year of Birth date	Age
Arizona	Yes	Yes
California		Yes
Colorado		Yes
Florida		Yes
Iowa	Yes	Yes
Massachusetts		Yes
Maryland		Yes
New Jersey		Yes
New York	Yes	Yes
Oregon	Yes	Yes
South Carolina		Yes
Washington		Yes
Wisconsin	Yes	Yes

Figure 4 Age information provided by states to SID

Figure 4 lists the states reported in Figure 3 that provide the *month and year of birth* and the *age* for each patient.

The remainder of this document provides experimental results from summary data that show how demographics often combine to make individuals unique or almost unique in data like these.

2.3. A single attribute

The frequency with which a single characteristic occurs in a population can help identify individuals based on unusual or outlying information. Consider a frequency distribution of birth years found in the list of registered voters. It is not surprising to see fewer people present with earlier birth years. Clearly, a person born in 1900 is unusual and by implication less anonymous in data.

2.4. More than one attribute

What may be more surprising is that combinations of characteristics can combine to occur even less frequently than the characteristics appear alone.

ZIP	Birth	Gender	Race
60602	7/15/54	m	Caucasian
60140	2/18/49	f	Black
62052	3/12/50	f	Asian

Figure 5 Data that looks anonymous

Consider Figure 5. If the three records shown were part of a large and diverse database of information about Illinois residents, then it may appear reasonable to assume that these three records would be anonymous. However, the 1990 federal census [6] reports that the ZIP (postal code) 60602 consisted primarily of a retirement community in the Near West Side of Chicago and therefore, there were very few people (less than 12) of an age under 65 living there. The ZIP code 60140 is the postal code for Hampshire, Illinois in Dekalb county and reportedly there were only two black women who resided in that town. Likewise, 62052 had only four Asian families. In each of these cases, the uniqueness of the combinations of characteristics found could help re-identify these individuals.

Race	Birth	Gender	ZIP	Problem
Black	09/20/65	m	02141	short of breath
Black	02/14/65	m	02141	chest pain
Black	10/23/65	f	02138	hypertension
Black	08/24/65	f	02138	hypertension
Black	11/07/64	f	02138	obesity
Black	12/01/64	f	02138	chest pain
White	10/23/64	m	02138	chest pain
White	03/15/65	f	02139	hypertension
White	08/13/64	m	02139	obesity
White	05/05/64	m	02139	short of breath
White	02/13/67	m	02138	chest pain
White	03/21/67	m	02138	chest pain

Figure 6 De-identified data

As another example, Figure 6 contains de-identified data. Each row contains information about a distinct person, so information about 12 people is reported. The table contains the following fields of information {*Race/Ethnicity*, *Date of Birth*, *Gender*, *ZIP*, *Medical Problem*}.

In Figure 6, there is information about an equal number of African Americans (listed as *Black*) as there are Caucasian Americans (listed as *White*) and an equal number of men (listed as *m*) as there are women (listed as *f*), but in combination, there appears only one Caucasian female.

2.5. Learned from the examples

These examples demonstrate that in general, the frequency distributions of combinations of characteristics have to be examined in combination with respect to the entire population in order to determine unusual values and cannot be generally predicted from the distributions of the characteristics individually. Of course, obvious predictions can be made from extreme distributions --such as values that do not appear in the data will not appear in combination either.

3. Background of definitions and terms

Definition (informal). Person-specific data Collections of information whose granularity of details are specific to an individual are termed *person-specific data*. More generally, in *entity-specific data*, the granularity of details is specific to an entity.

Example. Person-specific data

Figure 5 and Figure 6 provide examples of person-specific data. Each row of these tables contains information related to one person.

The idea of anonymous data is a simple one. The term "anonymous" means that the data cannot be linked or manipulated to confidently identify the individual who is the subject of the data.

Definition (informal). Anonymous data *Anonymous data* implies that the data cannot be manipulated or linked to confidently identify the entity that is the subject of the data.

Most people understand that there exist explicit identifiers, such as name and address, which can provide a direct means to communicate with the person. I term these *explicit identifiers*; see the informal definition below.

Definition (informal). Explicit identifier An *explicit identifier* is a set of data elements, such as {*name*, *address*} or {*name*, *phone number*}, for which there exists a direct communication method, such as email, telephone, postal mail, etc., where with no additional information, the designated person could be directly and uniquely contacted.

A common incorrect belief is that removing all explicit identifiers such as name, address and phone number from the data renders the result anonymous. I refer to this instead as *de-identified data*; see the informal definition below.

Definition (informal). De-identified data *De-identified data* result when all explicit identifiers, such as name, address, or phone number are removed, generalized or replaced with a made-up alternative.

Example. De-identified data

Figure 5 and Figure 6 provide examples of de-identified person-specific data. There are no explicit identifiers in these data.

Because a combination of characteristics can combine uniquely for an individual, it can provide a means of recognizing a person and therefore serve as an identifier. In the literature, such combinations were nominally introduced as *quasi-identifiers* [7] and *identificates* [3-58] with no supporting evidence provided as to how identifying specific combinations might be. Extending beyond the literature and its casual use in the literature, I term such a combination a quasi-identifier and informally define it below. I then examine specific quasi-identifiers found within publicly and semi-publicly available data and compute their general ability to uniquely associate with particular persons in the U.S. population.

Definition (informal). Quasi-identifier A quasi-identifier is a set of data elements in entity-specific data that in combination associates uniquely or almost uniquely to an entity and therefore can serve as a means of directly or indirectly recognizing the specific entity that is the subject of the data.

Example. Quasi-identifier

A quasi-identifier whose values are unique for all the records in Figure 6 is {*ZIP*, *gender*, *Birth*}.

In the next section, I will show that {*ZIP*, *gender*, *Birth*} is a unique quasi-identifier for most people in the U.S. population.

The term *table* is really quite simple and is synonymous with the casual use of the term data collection. It refers to data that are conceptually organized as a 2-dimensional array of rows (or records) and columns (or fields). A database is considered to be a set of one or more tables.

Definition (informal). Table, tuple and attribute A *table* conceptually organizes data as a 2-dimensional array of rows (or records) and columns (or fields). Each row (or record) is termed a *tuple*. A tuple contains a relationship among the set of values associated with an entity. Tuples within a table are not necessarily unique. Each column (also known as a field or data element) is called an *attribute* and denotes a field or semantic category of information that is a set of possible values; therefore, an attribute is also a domain. Attributes within a table are unique. So by observing a table, each row is an ordered n-tuple of values $<d_1, d_2, \ldots, d_n>$ such that each value d_j is in the domain of the j-th column, for j=1, 2, …, n where n is the number of columns.

In mathematical set theory, a relation corresponds with this tabular presentation; the only difference is the absence of column names. Ullman provides a detailed discussion of relational database concepts [9].

Examples of tables

Figure 5 provides an example of a person-specific table with attributes {*ZIP*, *Birth*, *Gender*, *Race*}. Each tuple concerns information about a single person. Figure 6 provides an example of a person-specific table with attributes {*Race*, *Birth*, *Gender*, *ZIP*, *Problem*}.

Unfortunately, the terminology with respect to data collections is not the same across communities and diverse communities have an interest in this work. In order to accommodate these different vocabularies, I provide the following thesaurus of interchangeable terms. In general, *data collection*, *data set* and *table* refer to the same representation of information though a data collection may have more than one table. The terms *record*, *row* and *tuple* all refer to same kind of information. Finally, the terms *data element*, *field*, *column* and *attribute* refer to the same kind of information. For brevity, from this point forward, I will use the more formal database terms of table, tuple and attribute. I do allow the tuples of a table to appear in a "sorted" order on occasion and such cases pose a slight deviation from its more formal meaning. These uses are explicitly noted.

4. Methods

4.1. Census Tables

Information from the 1990 US Census made available on the Web [10] and on CDROM [11] and from the U.S. Postal Service [12] was loaded into Microsoft Access and the following tables produced and used with Microsoft Excel.

1. ZIP census table provides 1990 federal census information summarized by each ZIP (postal code) in the United States.

2. Place census table provides 1990 federal census information summarized by place name (town, city, municipality, or postal facility name).

3. County census table provides 1990 federal census information summarized by US counties.

Figure 7 contains a list of attributes (or data elements) for each of these tables. The name and description of each attribute is listed and a "yes" appears in the column that associates the attribute to the ZIP, Place or County table in which the attribute appears. Information for all 50 states and the District of Columbia were provided. For example, values associated with the attribute *Tot_pop* in the ZIP table are the total numbers of individuals reported as living in each corresponding *ZIP*. Each tuple (or row) in the table corresponds to a unique ZIP.

Given a particular geographical specification such as ZIP, place or county, the number of people reported as residing in the noted geographical area is reported by age subdivision in the ZIP, Place and County tables. The age subdivisions are: under 12 years of age (denoted as *Aunder12*), between 12 and 18 years of age (denoted as *A12to18*), between 19 and 24 years of age (denoted as *A19to24*), between 25 and 34 years of age (denoted as *A25to34*), between 35 and 44 years of age (denoted as *A35to44*), between 45 and 54 years of age (denoted as *A45to54*),

between 55 and 64 years of age (denoted as *A55to64*) or more than 65 years of age (denoted as *A65Plus*).

Field	Description	ZIP	PLACE	COUNTY
StateID	State Code	yes	yes	yes
ZIP	5-digit ZIP	yes	**NO**	**NO**
Place	Name of Incorporated Place	**NO**	yes	**NO**
CoName	County Name	**NO**	**NO**	yes
Tot_Pop	Total Population	yes	yes	yes
AUnder12	Population Under Age 12 Years	yes	yes	yes
A12to18	Population Age 12-18 Years	yes	yes	yes
A19to24	Population Age 19-24 Years	yes	yes	yes
A25to34	Population Age 25-34 Years	yes	yes	yes
A35to44	Population Age 35-44 Years	yes	yes	yes
A45to54	Population Age 45-54 Years	yes	yes	yes
A55to64	Population Age 55-64 Years	yes	yes	yes
A65Plus	Population Age 65 Years and up	yes	yes	yes

Figure 7 1990 Census attributes in ZIP, Place, County tables

4.2. ZIPNameGIS Table

ZIP information provided from the U.S. Postal Service included place, which is a name of a town, city, municipality or postal facility uniquely assigned to a ZIP code. This information was loaded directly to provide the ZIPNameGIS table. The attributes (or data elements) for the ZIPNameGIS table are {*StateID*, *ZIP*, *State*, *POName*, *longitude*, *latitude*, *population*}.

The Place table was constructed by linking the ZIP table to the ZIPNameGIS table on *ZIP*. Results were then grouped by *POName* (respecting state designations) so that population information from multiple ZIP codes were grouped together by the city or town in which the ZIP code referred. Finally, the Place table was generated by collapsing these groupings into single entries that contained the sum of the population values reported for all ZIP codes corresponding to the same place.

During the process, 3 ZIP codes were found to cross state lines and therefore, be listed in two states. To avoid this duplication, the following assignments were made: (1) ZIP code 32530 refers to Pinetta in both Florida and Georgia. The Georgia entry was removed from Place; (2) ZIP code 42223 refers to Fort Campbell in both Kentucky and Tennessee. The Tennessee entry was removed from Place; and, (3) ZIP code 63673 refers to Saint Mary in both Illinois and Missouri. The Missouri entry was removed from Place.

4.2.1. Schemas of shared data

Figure 2 and Figure 3 contain descriptions of publicly and semi-publicly available hospital discharge data. Below are some quasi-identifiers found in those data that also appear in the census data. The experiments reported in this document estimate the uniqueness of values associated with these quasi-identifiers given the occurrences reported in the census data.

1. Illinois Research Health Data.

The Illinois Research Health Data (R_{rod}) is described in Figure 2. Among the attributes listed there, I consider QI_{rod} = {*date of birth*, *gender*, *5-digit ZIP*} to be a quasi-identifier within R_{rod}.

2. AHRQ's State Inpatient Database

The Agency for Healthcare Research and Quality's State Inpatient Database (R_{SID}) is described in part in Figure 3. Among the attributes listed there, I consider QI_{SID1} = {*month and year of birth*, *gender*, *5-digit ZIP*} to be a quasi-identifier within data released by some states and I consider QI_{SID2} = {*age*, *gender*, *5-digit ZIP*} to be a quasi-identifier within data released by other states.

4.3. Design and procedures

The experiments reported in the next section can be generally described in terms of values attributes can assume. Let $T(A_1,\ldots,A_n)$ be an entity-specific table and let Q_T be a quasi-identifier of T. Q_T is represented as a finite set of attributes $\{A_i,\ldots,A_j\} \subseteq \{A_1,\ldots,A_n\}$. I write $|A_m|$ to represent the finite number of values A_m can assume. So, the number of distinct possible values that be assigned to Q_T, written $|Q_T|$, is: $|Q_T| = |A_i| * |A_{i+1}| * \ldots * |A_j|$.

Example.
Given Q_{dob}={*date of birth*, *gender*}, then $|Q_{dob}| = 365 * 76 * 2 = 55{,}480$ because there are 365 days in a year, an expected lifetime of 76 years, and 2 genders.

In this document, I am concerned with a person-specific table $T(A_1,\ldots,Z,\ldots,A_n)$ that includes a *geographic attribute Z*. Values assigned to a *geographic attribute* are specific to the residences of people. Examples of geographic attributes include 5-digit ZIP codes, names of cities and towns, and names of counties in which people reside. Let *U* be the universe of all people and the person-specific table $Geo[z_i, A_r, \ldots, A_s)$ contain all or almost all of the people of *U* having $Z=z_i$. I say Geo_{zi} is a *population register* for z_i. And, $T[A_1,\ldots Z_i,\ldots A_n]$ is a pseudo-random sample drawn from $Geo[z_i, A_r, \ldots, A_s]$. Unique and unusual combinations of characteristics found in Geo with respect to z_i can be no less unique or unusual when recorded in T. Therefore, the probability distribution of combinations of characteristics found in Geo limits the values those combinations of characteristics can assume in T. Determining unique and unusual combinations of characteristics within a residential domain is a counting problem.

Theorem. Generalized Dirichlet drawer principle [13]
(also known as the Generalized pigeonhole principle)

If *N* objects are distributed in *k* boxes, then there is at least one box containing at least $\lceil N / k \rceil$ objects.

Proof.
Suppose that none of the boxes contain more than $\lceil N / k \rceil$ -1 objects. Then, the total number of objects is at most: $k * (\lceil N / k \rceil - 1) < k * (((N / k) + 1) - 1) = N$
This has the inequality $\lceil N / k \rceil < (N / k) + 1$
This is a contradiction because there are a total of *N* objects.

Example.
Given a random sample of 500 people, there are at least $\lceil 500 / 365 \rceil = 2$ people with the same birthday because there are 365 possible birthdays.

Let z_i be a 5-digit ZIP code. I write *population*(z_i) to denote the number of people who reside in z_i and *population*(z_i) ≡ |Geo_{zi}|. If *population*(Z_i) > |Q_{dob}|, then by the generalized pigeonhole principle, a tuple $t \in \mathsf{R}_{rod}$[*date of birth*, *gender*, z_i] would not uniquely correspond to one person. In these cases, I say t[A_1, …, *date of birth*, *gender*, z_i, ..., *An*] is not likely to be uniquely identifiable. On the other hand, if *population*(z_i) ≤ |Q_{dob}| then by the generalized pigeonhole principle, a tuple $t \in \mathsf{R}_{rod}$[*date of birth*, *gender*, z_i] would likely relate to only one person. In these cases, I say t[A_1, …, *date of birth*, *gender*, z_i, ..., *An*] is likely to be uniquely identifiable. This is the general approach to the experiments reported in the next section though each differs in terms of attribute specification.

4.3.1. Subdivision analyses

The analyses of the identifiability of geographically situated populations are based on age-based divisions within a geographic attribute. Let age subdivision *a* be either *Aunder12*, *A12to18*, *A19to24*, *A25to34*, *A35to44*, *A45to54*, *A55to64*, or *A65Plus*. The quasi-identifier Q_a has the same attributes as Q_{dob} but values which *date of birth* can assume are limited by *a*. That is, |Q_a| is the number of possible distinct values that can be assigned to Q_a. I say |Q_a| is the *threshold* for Q_{dob} with respect to age subdivision *a*.

Example.
Given Q_{dob}={*date of birth*, *gender*} and age subdivision *a* = *A19to24*, then |Q_a| = 365 * 2 * 6 = 4380 because there are 365 birthdays, 2 genders and 6 years between the ages of 19 to 24, inclusive.

Number of subjects uniquely identified in a subdivision of a geographical area (ID_{aZi})

Given a value for a geographic attribute, written z_i, and an age subdivision *a*, I write *population*(z_i, *a*) as the number of people residing in z_i with an age within *a*. The number of people considered uniquely identified by *a* and Z_i, written ID_{aZi}, is determined by the rule:

if *population*(z_i, *a*) ≥ |Q_a|, then ID_{aZi}= *population*(z_i, *a*)
else ID_{aZi} = 0.

By extension, the percentage of people residing in z_i considered uniquely identified (written ID_{zi}) with respect to the set of age subdivisions is computed as:

$$ID_{Zi} = \frac{population(z_i) - \sum_{a=AUnder12}^{A65Plus} ID_{aZi}}{population(z_i)}$$

4.3.2. Statistics on geographical areas

Statistics are reported on geographic regions. Given a geographic attribute *Z*, let $Region_Z$ = {z_i | $z_i \in Z$ } and *AgeDivs* = {*Aunder12*, *A12to18*, *A19to24*, *A25to34*, *A35to44*, *A45to54*, *A55to64*, *A65Plus*}. That is, $Region_Z$ is a set of values that can be assigned to the geographic

attribute Z and *AgeDivs* is a set of age subdivisions. $Region_Z$ is partitioned into NotIDSet and IDSet based on age subdivision a∈*AgeDivs* such that:

$$\text{NotIDSet}_{Za} = \{(z_i,a) \mid z_i \in Region_Z \text{ and } population(z_i, a) > |Q_a| \}$$
$$\text{IDSet}_{Za} = \{(z_i,a) \mid z_i \in Region_Z \text{ and } population(z_i, a) \le |Q_a| \}$$

The population of NotIDSet_Z is not considered uniquely identifiable by values of Q_{dob}. The population of IDSet_Z is considered uniquely identifiable by values of Q_{dob}. In the experiments, the following statistics are reported.

Maximum subpopulation(NotIDSet_{Za}) = max(*population*(z_1, *a*), …, *population*(z_y, *a*)), where $(z_i,a) \in \text{NotIDSet}_{Za}$

Maximum subpopulation(IDSet_{Za}) = max(*population*(z_1, *a*), …, *population*(z_y, *a*)), where $(z_i,a) \in \text{IDSet}_{Za}$

Minimum subpopulation(NotIDSet_{Za}) = min(*population*(z_1, *a*), …, *population*(z_y, *a*)), where $(z_i,a) \in \text{NotIDSet}_{Za}$

Minimum subpopulation(IDSet_{Za}) = min(*population*(z_1, *a*), …, *population*(z_y, *a*)), where $(z_i,a) \in \text{IDSet}_{Za}$

$$Average\ \ subpopulation(\text{NotIDSet}_{Za}) = \frac{\sum_{(z_i,a)\in NotIDSet} population(z_i,a)}{|\text{NotIDSet}_{Za}|}$$

$$Average\ \ subpopulation(\text{IDSet}_{Za}) = \frac{\sum_{(z_i,a)\in IDSet} population(z_i,a)}{|\text{IDSet}_{Za}|}$$

$$Number\ of\ geographical\ areas(\text{NotIDSet}_{Za}) = |\text{NOTIDSet}_{Za}|$$

$$Number\ of\ geographical\ areas(\text{IDSet}_{Za}) = |\text{IDSet}_{Za}|$$

$$Percentage\ of\ geographical\ areas(\text{NotIDSet}_{Za}) = \frac{|\text{NotIDSet}_{Za}|}{|\text{NotIDSet}_{Za}| + |\text{IDSet}_{Za}|}$$

$$Percentage\ of\ geographical\ areas(\text{IDSet}_{Za}) = \frac{|\text{IDSet}_{Za}|}{|\text{NotIDSet}_{Za}| + |\text{IDSet}_{Za}|}$$

State		AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
AL	4,040,587	699,554	425,425	369,639	652,466	585,299	422,565	363,033	522,606
AK	544,698	123,789	53,662	46,478	111,790	101,699	55,887	29,236	22,157
AZ	3,665,228	678,439	352,557	333,055	639,702	530,192	354,711	299,372	477,200
AR	2,350,725	410,665	246,486	197,424	361,268	328,397	244,096	212,573	349,816
CA	29,755,274	5,436,303	2,722,076	2,904,739	5,738,645	4,645,553	2,955,455	2,231,171	3,121,332
CO	3,293,771	599,278	305,595	282,268	617,333	570,797	340,276	249,924	328,300
CT	3,287,116	517,724	275,158	295,271	588,185	509,760	360,488	294,866	445,664
DE	666,168	113,963	58,980	64,726	119,782	100,110	68,367	59,570	80,670
DC	606,900	80,760	45,404	71,605	122,777	94,984	62,648	51,050	77,672
FL	12,686,788	1,931,088	1,041,486	1,010,156	2,102,614	1,778,994	1,283,728	1,235,820	2,302,902
GA	6,478,847	1,171,969	659,386	623,625	1,182,367	1,014,579	678,987	495,259	652,675
HI	1,108,229	195,278	98,594	104,537	203,466	178,406	109,493	93,778	124,677
ID	1,006,749	207,979	115,708	81,770	154,087	149,338	98,910	77,819	121,138
IL	11,429,942	2,012,780	1,102,499	1,021,458	2,003,217	1,702,509	1,179,345	974,035	1,434,099
IN	5,543,954	975,582	568,654	510,374	919,924	819,577	572,585	481,329	695,929
IA	2,776,442	487,879	271,630	240,359	430,947	397,287	272,959	249,594	425,787
KS	2,474,885	457,755	236,911	216,092	416,003	363,571	234,451	208,146	341,956
KY	3,673,969	626,236	383,356	337,585	610,721	549,204	380,791	320,712	465,364
LA	4,219,973	836,481	458,677	387,821	710,773	606,119	412,186	340,483	467,433
ME	1,226,626	210,082	117,015	104,754	205,713	194,139	123,745	108,198	162,980
MD	4,771,143	812,147	409,957	431,840	901,956	774,414	528,246	395,946	516,637
MA	6,011,978	933,306	506,033	613,116	1,104,645	914,852	605,951	514,398	819,677
MI	9,295,222	1,671,777	930,841	850,016	1,583,364	1,408,199	950,316	793,711	1,106,998
MN	4,370,288	815,963	409,705	377,084	783,562	666,480	428,315	343,315	545,864
MS	2,573,216	495,074	298,599	240,546	403,754	351,197	249,684	213,117	321,245

Figure 8 Population by state and age group, part 1

State		AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
MO	5,113,266	897,590	490,067	436,468	855,640	734,252	524,756	457,095	717,398
MT	799,065	150,406	83,457	57,351	123,913	128,067	81,522	67,930	106,419
NE	1,577,600	294,659	156,790	130,613	259,709	229,478	148,720	134,711	222,920
NV	1,201,833	208,695	100,891	102,609	223,599	192,324	138,893	107,621	127,201
NH	1,109,252	195,970	98,977	100,411	205,815	183,649	111,387	88,059	124,984
NJ	7,730,188	1,217,936	681,960	664,059	1,366,267	1,200,167	850,983	718,589	1,030,227
NM	1,515,069	307,898	160,598	123,983	259,975	229,577	149,712	120,808	162,518
NY	17,990,026	2,891,618	1,615,696	1,664,461	3,148,965	2,720,452	1,944,539	1,642,487	2,361,808
NC	6,628,637	1,074,691	637,603	662,849	1,152,229	1,008,277	705,099	585,832	802,057
ND	637,713	119,767	65,036	57,151	104,833	90,808	56,215	53,132	90,771
OH	10,846,581	1,899,661	1,064,732	957,750	1,805,063	1,619,291	1,115,355	978,701	1,406,028
OK	3,145,585	563,941	318,809	267,411	514,663	452,308	326,770	278,089	423,594
OR	2,842,321	495,834	265,630	225,488	455,371	476,343	297,101	235,423	391,131
PA	11,881,643	1,892,957	1,074,128	1,041,626	1,918,168	1,739,212	1,224,867	1,160,974	1,829,711
RI	1,003,211	155,439	86,271	102,680	174,149	146,571	97,958	89,156	150,987
SC	3,486,703	616,373	363,140	339,600	596,534	526,103	357,747	291,077	396,129
SD	695,133	137,110	71,070	56,976	109,919	96,063	61,962	59,623	102,410
TN	4,896,046	812,832	484,155	452,701	823,042	740,485	530,654	433,773	618,404
TX	16,984,748	3,320,887	1,776,426	1,578,004	3,118,515	2,548,657	1,649,538	1,284,825	1,707,896
UT	1,722,850	430,959	226,933	167,637	275,853	224,715	139,656	107,405	149,692
VT	562,758	99,365	53,099	53,049	95,880	92,804	57,274	45,118	66,169
VA	6,184,493	1,030,088	564,690	616,835	1,147,609	991,563	670,457	500,955	662,296
WA	4,866,692	878,141	444,693	417,468	861,441	804,413	504,238	380,725	575,573
WV	1,792,969	279,885	192,881	148,808	262,961	270,784	191,957	176,960	268,733
WI	4,891,452	887,426	472,270	437,743	825,056	726,753	478,819	412,492	650,893
WY	453,588	92,123	49,716	33,980	75,462	74,182	45,541	35,539	47,045
USA	248,418,140	43,454,102	23,694,112	22,614,049	43,429,692	37,582,954	25,435,905	21,083,554	31,123,772

Figure 9 Population by state and age group, part 2

Different experiments have different age and geographic attributes. See Figure 11 for a list of all 13 experiments identified as A through M. So, Q_{dob} and Z_i, as used above, are representative of several quasi-identifiers that have varying specifications. In experiment B through experiment E, $Z_i \in$ {ZIP codes in USA in which people reside}. In experiment F through experiment I, $Z_i \in$ {Cities, municipalities, towns and recognized post office names in the USA}. Finally, in experiment J through experiment M, $Z_i \in$ {Counties in the USA}. Similarly, in experiments B, F, and J, Q_{dob} = {*date of birth*, *gender*}. In experiments C, G and K, Q_{dob} = {*month and year of birth*, *gender*}. In experiments D, H and L, Q_{dob} = {*year of birth*, *gender*}. Finally, in experiments E, I and M, Q_{dob} = {*2 year age subdivision*, *gender*}.

For completeness, Figure 8 and Figure 9 report the total population per state of each age group. These values are used to compute percentages throughout this document unless otherwise noted.

4.4. Special data elements

This section compares age and year of birth values, as well as, 5-digit ZIP codes, places and counties.

4.4.1. Age versus Year of Birth

Values for an *age* attribute do not necessarily translate to known values for a *year of birth* attribute. There are two cases to consider. If there exists a date to which values for *age* can be referenced, then corresponding values for *year of birth* can be confidently computed. For example, in SID, states calculate the patient's age in years at the time of admission [14]. Because both the computed *age* and the *date of admission* are released, the patient's year of birth can be confidently determined. In experiment D, H and L, I examine *age* as providing a distinct year of birth, and so QI_{SID2} = {*age*, *gender*, *5-digit ZIP*} can be considered as QI_{SID2} = {*year of birth*, *gender*, *5-digit ZIP*}.

On the other hand, if values for *date of admission* were not released, values for *age* would be calendar year specific. In such cases, data are collected with respect to a particular calendar year (that is known) but not a particular day within that year. As a result, each value for *age* corresponds to two possible values for each person's *year of birth*. During any given calendar year, a person reports two ages. The first age occurs before the person's birthday and the second occurs on and after the person's birthday. Because each person's birthday can appear at any time during the calendar year (in contrast to societies in which everyone's "birthday", in terms of determining age, occurs on the same day), two values can be inferred for *year of birth* from a recorded value for *age*. In the experiment E, I and M, I examine {*2 yr age subdivision*, *gender*, *5-digit ZIP*} in which the birth year is within a known 2-year range.

4.4.2. Comparison of 5-digit ZIP codes, Places and Counties

Figure 10 shows a comparison of 5-digit ZIP codes, places and counties in the United States. There are a total of 29,343 ZIP codes, 25,688 places and 3,141 counties. The state having the largest number of counties was Texas (with 254). The District of Columbia had the fewest number of counties (with 1). The average number of counties per state was 62 and the standard deviation was 47.

State	Number 5-digit ZIPs	Number Places	Number Counties	State	Number digit ZIPs	Number Places	Number Counties
AL	567	511	67	MO	993	899	115
AK	195	183	25	MT	315	309	57
AZ	270	178	15	NE	572	518	93
AR	578	563	75	NV	104	66	17
CA	1,515	1,071	58	NH	218	212	10
CO	414	330	63	NJ	540	490	21
CT	263	224	8	NM	276	258	33
DE	53	46	3	NY	1,594	1,369	62
DC	24	2	1	NC	705	624	100
FL	804	463	67	ND	387	384	53
GA	636	561	159	OH	1,007	854	88
HI	80	70	5	OK	586	511	77
ID	244	233	44	OR	384	344	36
IL	1,236	1,147	102	PA	1,458	1,369	67
IN	675	597	92	RI	69	52	5
IA	922	889	99	SC	350	313	46
KS	713	646	105	SD	383	377	66
KY	810	772	120	TN	583	505	95
LA	469	408	64	TX	1,672	1,234	254
ME	410	408	16	UT	205	181	29
MD	419	378	24	VT	243	243	14
MA	473	404	14	VA	820	729	136
MI	875	768	83	WA	484	397	39
MN	877	809	87	WV	655	646	55
MS	363	342	82	WI	714	666	72
				WY	141	135	23
				USA	29,343	25,688	3,141
				max	1,672	1,369	254
				min	24	2	1
				avg	575	504	62
				stdev	401	337	47

Figure 10 Number of 5-digit ZIP codes, Places and Counties by State

5. Results

In the previous sections, I defined terminology and introduced the materials that will be used. In this section, I report on experiments I conducted to estimate the number of unique occurrences for various combinations of demographic attributes that are typically released in publicly and semi-publicly available data.

Experiment A: Uniqueness of {*ZIP*, *gender*, *date of birth*} assume uniform age distribution
Experiment B: Uniqueness of {*ZIP*, *gender*, *date of birth*} based on actual age distribution
Experiment C: Uniqueness of {*ZIP*, *gender*, *month and year of birth*}
Experiment D: Uniqueness of {*ZIP*, *gender*, *age*}
Experiment E: Uniqueness of {*ZIP*, *gender*, *2yr age range*}
Experiment F: Uniqueness of {*place/city*, *gender*, *date of birth*}
Experiment G: Uniqueness of {*place/city*, *gender*, *month and year of birth*}
Experiment H: Uniqueness of {*place/city*, *gender*, *age*}
Experiment I: Uniqueness of {*place/city*, *gender*, *2yr age range*}
Experiment J: Uniqueness of {*county*, *gender*, *date of birth*}
Experiment K: Uniqueness of {*county*, *gender*, *month and year of birth*}
Experiment L: Uniqueness of {*county*, *gender*, *age*}
Experiment M: Uniqueness of {*county*, *gender*, *2yr age range*}

Figure 11 List of 13 experiments

A total of 13 experiments were conducted [15]. These are identified below. Only experiment B, C, D, F and J are briefly reported in this document. Figure 32 contains a summary of results from all 13 experiments.

5.1. Experiment B: Uniqueness of {*ZIP*, *gender*, *date of birth*}

Recall, Illinois Research Health Data named ROD provides an example of shared data that contains demographic attributes; in particular, QI_{rod} = {*date of birth*, *gender*, *5-digit ZIP*}. This experiment shows that medical conditions included in these data can be attributed uniquely to one person in most cases.

5.1.1. Experiment B Design

Step 1. Use ZIP table for each of the 50 states and the District of Columbia. Step 2. Figure 12 contains the thresholds for Q={*gender*, *date of birth*} specific to each age subdivision. Step 3. Report statistical measurements computed from the table in step 1 using the thresholds determined in step 2. Figure 13 and Figure 14 report the results.

Q = {*gender*, *date of birth*}	
$\|Q_{AUnder12}\|$	= 2 * 365 * 12 = 8,760
$\|Q_{A12to18}\|$	= 2 * 365 * 7 = 5,110
$\|Q_{A19to24}\|$	= 2 * 365 * 6 = 4,380
$\|Q_{A25to34}\|$	= 2 * 365 * 10 = 7,300
$\|Q_{A35to44}\|$	= 2 * 365 * 10 = 7,300
$\|Q_{A45to54}\|$	= 2 * 365 * 10 = 7,300
$\|Q_{A55to64}\|$	= 2 * 365 * 10 = 7,300
$\|Q_{A65Plus}\|$	= 2 * 365 * 12 = 8,760

Figure 12 Number of possible values for each age subdivision {*gender*, *date of birth*}

5.1.2. Experiment B Results

Figure 13 and Figure 14 show the results from applying the 3 steps of experiment B to each state, the District of Columbia and the entire United States. The percentages computed for each locale appear in the column named "RANGE %ID_pop." The last row in Figure 14 reports the results of applying the 3 steps of experiment B to all ZIP codes in the United States. As shown, 87.1% of the population of the United States is likely to be uniquely identified by values of {*gender*, *date of birth*, *ZIP*} when age subdivisions are considered.

During the analysis of experiment B, many interesting ZIP codes were found. Here are a few. The ZIP code 11794 in the State of New York is small and extremely homogenous. 4666 of its total population of 5418 (or 86%) are in the age subdivision of 19 to 24. This is the home of the State University of New York at Sony Brook. The ZIP code 10475 in the State of New York reportedly has a larger population of 37077, but people are distributed somewhat evenly across the age subdivisions making the population in each range less than its corresponding threshold. The ZIP code 01701 in the Commonwealth of Massachusetts reportedly has a population of 65,001, which is the largest population for a ZIP code in the state. In experiment A, any person residing in that ZIP code would NOT have been considered likely to be uniquely identified by {*gender*, *date of birth*, *ZIP*}; however, only the subpopulation between the ages of 19 and 44 in

that ZIP code is large enough not to be considered uniquely identified by {*gender*, *date of birth*, *ZIP*}. Persons residing in that ZIP code, who are not in that age subdivision, are less common and considered likely to be uniquely identified by {*gender*, *date of birth*, *ZIP*} even though the population in the entire ZIP code is the largest in the state.

State	#ZIPs	Population	RANGE %population	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
AL	567	4,040,587	99%	100.0%	100.0%	89.7%	98.7%	100.0%	100.0%	100.0%	100.0%
AK	195	544,698	100%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
AZ	270	3,665,228	82%	82.3%	90.1%	67.4%	64.3%	88.8%	100.0%	100.0%	80.7%
AR	578	2,350,725	98%	97.8%	100.0%	87.1%	95.3%	100.0%	100.0%	100.0%	100.0%
CA	1,515	29,755,274	71%	62.4%	73.1%	54.9%	47.2%	70.0%	96.8%	99.6%	96.8%
CO	414	3,293,771	92%	89.7%	96.2%	85.0%	81.1%	92.1%	100.0%	100.0%	100.0%
CT	263	3,287,116	91%	94.3%	98.1%	76.1%	76.2%	88.9%	100.0%	100.0%	97.8%
DE	53	666,168	91%	100.0%	100.0%	72.0%	66.7%	100.0%	100.0%	100.0%	100.0%
DC	24	606,900	64%	62.0%	74.9%	32.5%	47.6%	55.3%	100.0%	84.9%	85.1%
FL	804	12,686,788	91%	93.9%	95.8%	87.5%	85.2%	94.3%	98.6%	99.2%	83.6%
GA	636	6,478,847	90%	90.4%	93.5%	80.4%	77.8%	87.6%	100.0%	100.0%	100.0%
HI	80	1,108,229	74%	62.5%	94.4%	56.7%	55.9%	71.9%	100.0%	100.0%	83.7%
ID	244	1,006,749	99%	100.0%	100.0%	85.6%	100.0%	100.0%	100.0%	100.0%	100.0%
IL	1,236	11,429,942	75%	73.0%	76.4%	59.2%	60.1%	73.9%	90.3%	93.9%	86.7%
IN	675	5,543,954	94%	94.3%	95.2%	80.4%	85.4%	94.7%	100.0%	100.0%	100.0%
IA	922	2,776,442	98%	100.0%	100.0%	78.9%	98.0%	100.0%	100.0%	100.0%	100.0%
KS	713	2,474,885	98%	100.0%	100.0%	83.1%	94.1%	100.0%	100.0%	100.0%	100.0%
KY	810	3,673,969	98%	100.0%	100.0%	85.7%	97.5%	98.6%	100.0%	100.0%	100.0%
LA	469	4,219,973	91%	89.8%	91.7%	80.4%	83.6%	93.0%	100.0%	100.0%	100.0%
ME	410	1,226,626	98%	100.0%	100.0%	86.3%	96.3%	100.0%	100.0%	100.0%	100.0%
MD	419	4,771,143	83%	84.8%	94.1%	79.2%	63.7%	80.2%	93.8%	100.0%	88.7%
MA	473	6,011,978	91%	95.7%	97.9%	73.5%	74.8%	92.8%	100.0%	100.0%	98.8%
MI	875	9,295,222	85%	80.5%	84.7%	72.5%	74.5%	83.2%	98.2%	99.1%	98.3%
MN	877	4,370,288	95%	96.2%	100.0%	81.8%	87.7%	97.4%	100.0%	100.0%	100.0%
MS	363	2,573,216	98%	98.2%	98.1%	88.3%	100.0%	97.8%	100.0%	100.0%	100.0%

Figure 13 Uniqueness of {*ZIP*, *Gender*, *Date of birth*} respecting age distribution, part 1

State	#ZIPs	Population	RANGE %population	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
MO	993	5,113,266	94%	94.4%	98.8%	86.9%	86.8%	92.1%	100.0%	100.0%	97.3%
MT	315	799,065	98%	100.0%	100.0%	78.9%	100.0%	100.0%	100.0%	100.0%	100.0%
NE	572	1,577,600	99%	100.0%	100.0%	90.2%	100.0%	100.0%	100.0%	100.0%	100.0%
NV	104	1,201,833	86%	79.5%	94.3%	79.5%	66.9%	88.3%	94.6%	100.0%	100.0%
NH	218	1,109,252	97%	100.0%	100.0%	94.1%	88.5%	100.0%	100.0%	100.0%	100.0%
NJ	540	7,730,188	92%	92.6%	93.1%	88.0%	79.8%	92.9%	99.1%	100.0%	94.1%
NM	276	1,515,069	88%	86.1%	89.0%	88.6%	71.6%	82.4%	100.0%	100.0%	100.0%
NY	1,594	17,990,026	76%	74.3%	77.3%	64.1%	60.0%	72.1%	88.3%	93.4%	85.5%
NC	705	6,628,637	94%	98.1%	96.4%	77.5%	86.4%	96.5%	98.8%	100.0%	100.0%
ND	387	637,713	96%	100.0%	100.0%	68.5%	91.9%	100.0%	100.0%	100.0%	100.0%
OH	1,007	10,846,581	92%	92.2%	94.7%	82.4%	82.5%	93.6%	100.0%	100.0%	98.5%
OK	586	3,145,585	97%	96.7%	100.0%	85.2%	93.5%	96.7%	100.0%	100.0%	100.0%
OR	384	2,842,321	97%	100.0%	100.0%	89.5%	90.6%	93.1%	100.0%	100.0%	100.0%
PA	1,458	11,881,643	91%	90.5%	94.0%	80.1%	82.2%	90.3%	99.3%	99.4%	94.3%
RI	69	1,003,211	92%	94.4%	100.0%	71.1%	84.2%	94.9%	100.0%	100.0%	94.2%
SC	350	3,486,703	91%	90.0%	95.1%	74.8%	79.5%	95.0%	97.9%	100.0%	100.0%
SD	383	695,133	96%	92.7%	100.0%	81.4%	91.6%	100.0%	100.0%	100.0%	100.0%
TN	583	4,896,046	93%	93.7%	94.8%	80.5%	87.1%	93.5%	100.0%	100.0%	100.0%
TX	1,672	16,984,748	88%	85.0%	89.1%	78.8%	76.5%	90.0%	100.0%	100.0%	100.0%
UT	205	1,722,850	87%	75.8%	80.0%	78.0%	90.2%	92.6%	100.0%	100.0%	100.0%
VT	243	562,758	98%	100.0%	100.0%	80.1%	100.0%	100.0%	100.0%	100.0%	100.0%
VA	820	6,184,493	87%	88.2%	91.6%	71.9%	75.5%	82.7%	97.8%	100.0%	100.0%
WA	484	4,866,692	92%	94.6%	100.0%	82.8%	82.5%	87.2%	100.0%	100.0%	100.0%
WV	655	1,792,969	97%	96.7%	96.4%	90.2%	95.7%	96.4%	100.0%	100.0%	96.5%
WI	714	4,891,452	92%	88.9%	97.7%	77.6%	86.4%	92.6%	100.0%	100.0%	100.0%
WY	141	453,588	98%	100.0%	100.0%	79.2%	100.0%	100.0%	100.0%	100.0%	100.0%
USA	29,343	248,418,140	87%	85.8%	90.2%	75.0%	75.1%	87.0%	97.8%	99.0%	95.3%

Figure 14 Uniqueness of {*ZIP*, *Gender*, *Date of birth*} respecting age distribution, part 2

Figure 15 plots the percentage of the population considered identifiable in each ZIP code in the United States based on experiment B's criteria. The horizontal axis represents the

population that resides in the ZIP code. The vertical axis represents the percentage of the population considered uniquely identified by values of *Q*={*date of birth*, *gender*, *5-digit ZIP*} for a particular ZIP code. The criteria for computing the percentage of the population considered identifiable in experiment B is based on binary decisions, where each decision considers whether a sufficient number of people in a particular age subdivision reside in a particular ZIP code. If so, that sub-population is not considered identifiable; otherwise, its entire sub-population is considered identifiable.

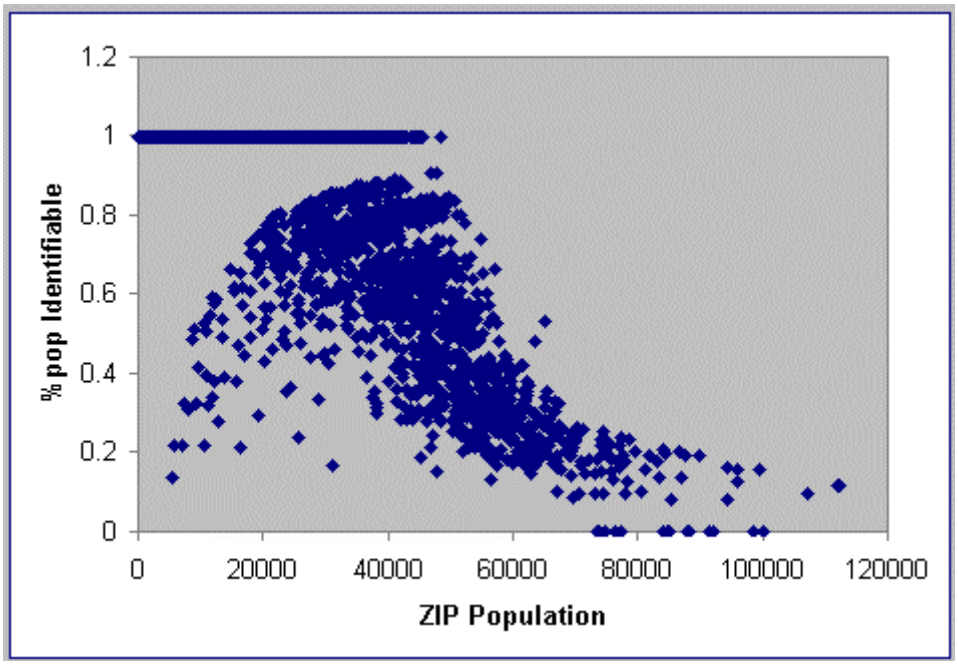


Figure 15 Percentage of Population Identifiable Based on Age subdivisions in ZIP Population

Most ZIP codes (27697 of 29212 or 95%) in the United States that have people listed as residing within them do not have enough people in any age subdivision to consider any such sub-population as identifiable. This is evidenced in Figure 15 by the appearance of dots where the *%pop identifiable* is 1. The largest population having *%pop identifiable* = 1 consists of 48,549 total people. There are very few ZIP codes (15 of 29212) in Figure 15 having sufficient numbers of people in each age subdivision that each such sub-population is not considered uniquely identifiable. This is evidenced in Figure 15 by the appearance of dots where the *%pop identifiable* is 0. The largest population having *%pop identifiable* = 0 has 99,995 people and the smallest has 73,321.

The ZIP code having the largest population, ZIP 60623 with 112,167 people, has a percentage of its population considered identifiable in Figure 15 as being only 11%. It is not 0% because there are insufficient numbers of people above the age of 55 living there despite the large number of people residing in the ZIP code. The point representing this ZIP code in Figure 15 is the rightmost point shown.

The lowest leftmost point shown in Figure 15 corresponds to ZIP 11794, which was discussed earlier. It has a total population of 5418 people and consists primarily of people between the ages of 19 and 24 (4666 of 5418 or 86%). Despite having a small population, the people residing there are very homogenous in terms of age and so the percentage of its population

considered identifiable based on experiment B's criteria is only 13%. It is clear from these examples that population size alone is not an absolute predictor of the identifiability of the people residing within. Care must be taken to model the population as precisely as possible to insure privacy protection.

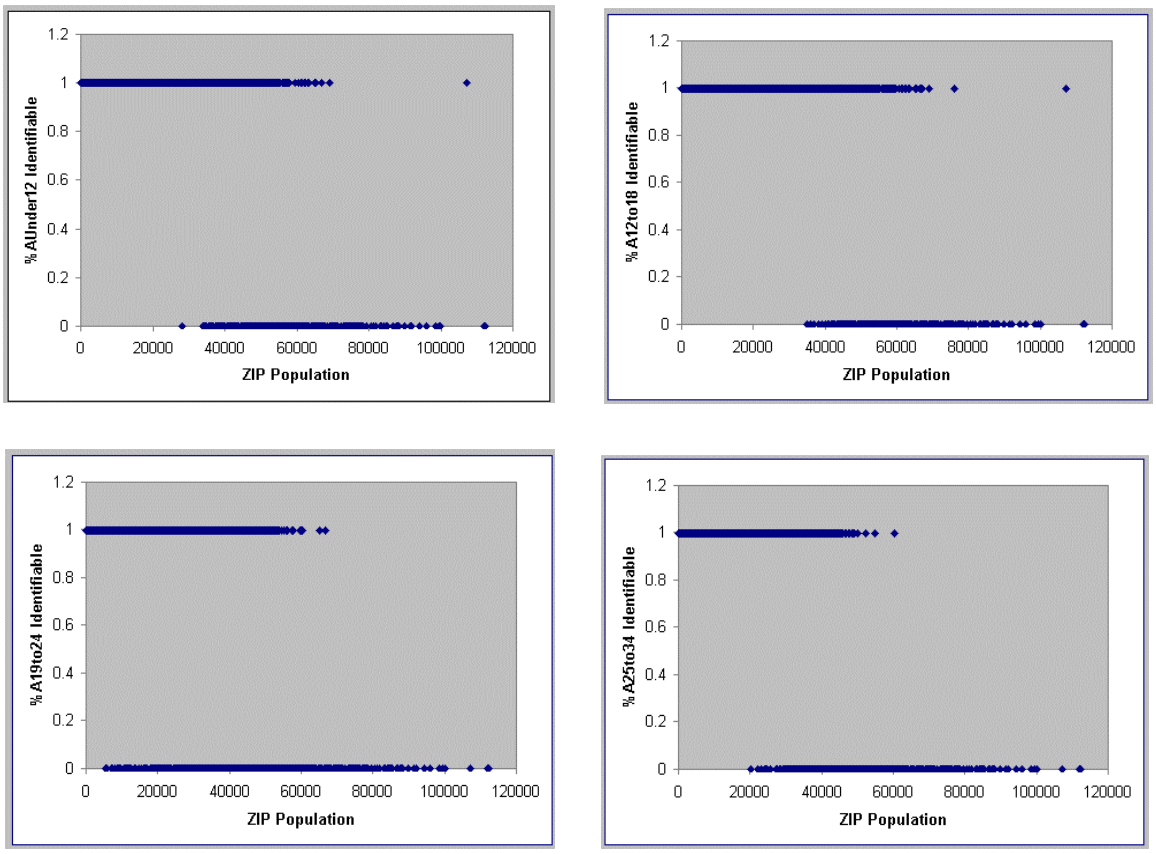


Figure 16 Percentage of Age-based Populations Identifiable within ZIP Population, Part 1

Recall the computation of the percentage of the population considered uniquely identified by values of *Q*={*date of birth*, *gender*, *5-digit ZIP*} for a particular ZIP code in experiment B is based on a composite of binary decisions. Each binary decision concerns the number of people residing within a specific ZIP code in a particular age subdivision. Figure 16 and Figure 17 show plots of the percentage of sub-populations considered identifiable in each ZIP code in the United States based on experiment B's criteria. The horizontal axis represents the population that resides in the ZIP code. The vertical axis represents the percentage of the population considered uniquely identified by values of *Q*={*date of birth*, *gender*, *5-digit ZIP*} for a particular ZIP code and a particular age subdivision. If a sufficient number of people within an age subdivision are reported as residing in a particular ZIP code, then that sub-population is considered identifiable; otherwise, the entire sub-population is not considered identifiable.

Figure 18 provides statistical highlights from the plots in Figure 16 and Figure 17. The topmost table provides statistics on ZIP codes in which the number of people within the noted age subdivision is less than or equal to the threshold for that subdivision. In these cases, the sub-population within the ZIP code is considered uniquely identifiable; that is, *%pop_Identifiable* = 1 for that age subdivision and ZIP code. The bottom table provides statistics in cases where *%pop_Identifiable* < 1. In these ZIP codes, the number of people within the noted age subdivision

is greater than the threshold for that subdivision; therefore, this subdivision is not considered uniquely identifiable.

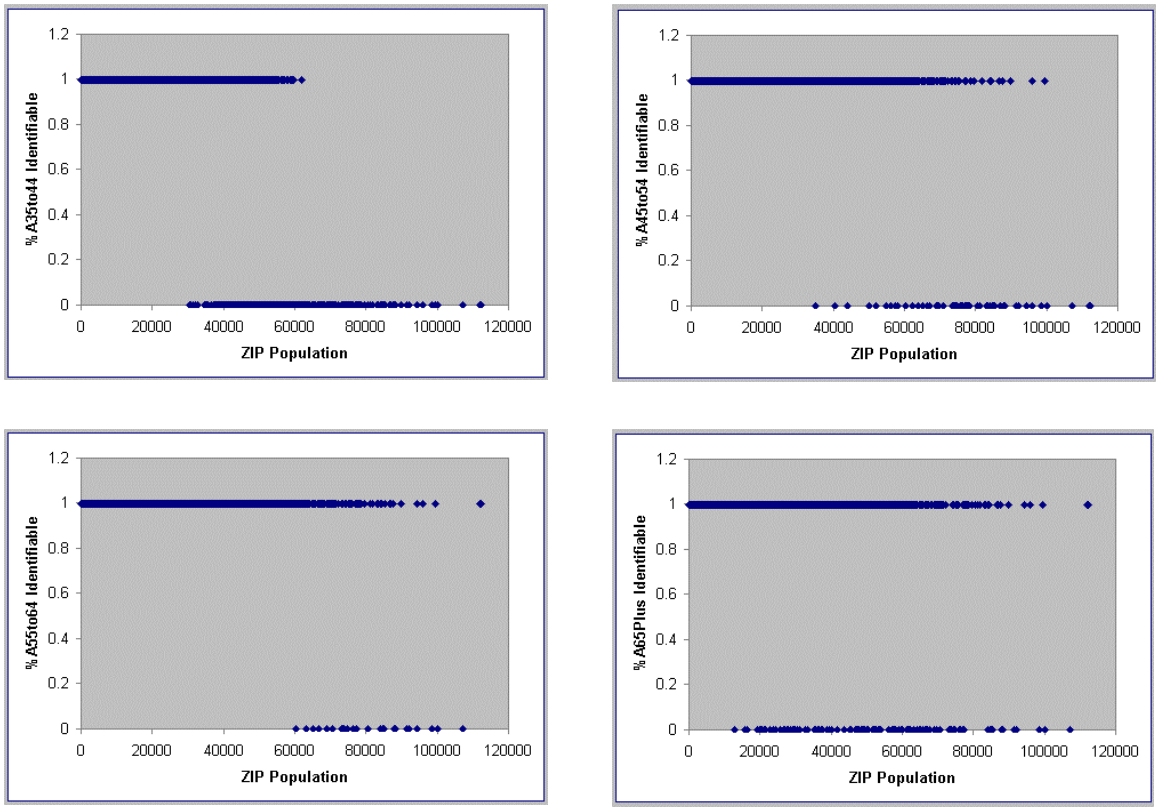


Figure 17 Percentage of Age-based Populations Identifiable within ZIP Population, Part 2

Sub-population considered uniquely identifiable (<= *threshold, IDSet*)

	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
Max ZIP population	107197	107197	66722	60388	62031	99420	112167	112167
Min ZIP population	1	1	1	1	1	1	1	1
Average ZIP population	7615	7873	7332	6911	7596	8358	8442	8311
standard deviation	19452	10915	10070	9227	10393	11938	12165	11956
Number of ZIP codes	28675	28860	28352	28105	28665	29148	29187	29081
Percentage ZIP codes	98.2%	98.8%	97.1%	96.2%	98.1%	99.8%	99.9%	99.6%

Sub-population NOT considered uniquely identifiable (> *threshold, NotIDSet*)

	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
Max ZIP population	112167	112167	112167	112167	112167	112167	107197	107197
Min ZIP population	28294	35092	5418	20211	30577	34860	60388	12890
Average ZIP population	55958	60254	47153	48944	56072	74798	80513	51313
standard deviation	12770	13036	17178	12681	13157	15961	12304	20367
Number of ZIP codes	537	352	860	1107	547	64	25	131
Percentage ZIP codes	1.8%	1.2%	2.9%	3.8%	1.9%	0.2%	0.1%	0.4%

Figure 18 Statistical highlights from Figure 16 and Figure 17

5.2. Experiment C: Uniqueness of {*ZIP*, *gender*, *month and year of birth*}

This experiment (referred to as experiment C) is motivated by the Agency for Healthcare Research and Quality's State Inpatient Database (R_{SID}), which is described in part in Figure 3. Among the attributes listed there, I consider QI_{SID1} = {*month and year of birth*, *gender*, *5-digit ZIP*} to be a quasi-identifier within data released by some states. This experiment attempts to characterize the identifiability of QI_{SID1}.

5.2.1. Experiment C Design

Step 1. Use ZIP table for each of the 50 states and the District of Columbia. Step 2. Figure 19 contains the thresholds for *Q*={*gender*, *month and year of birth*} specific to each age subdivision. Step 3. Report statistical measurements computed from the table in step 1 using the thresholds determined in step 2. Figure 20 and Figure 21 report the results.

Q3 = {*gender*, *month and year of birth*}		
$\|Q3_{AUnder12}\|$	= 2 * 12 * 12	= 288
$\|Q3_{A12to18}\|$	= 2 * 12 * 7	= 168
$\|Q3_{A19to24}\|$	= 2 * 12 * 6	= 144
$\|Q3_{A25to34}\|$	= 2 * 12 * 10	= 240
$\|Q3_{A35to44}\|$	= 2 * 12 * 10	= 240
$\|Q3_{A45to54}\|$	= 2 * 12 * 10	= 240
$\|Q3_{A55to64}\|$	= 2 * 12 * 10	= 240
$\|Q3_{A65Plus}\|$	= 2 * 12 * 12	= 288

Figure 19 Number of possible values for each age subdivision for {*gender*, *month and year of birth*}

5.2.2. Experiment C Results

Figure 20 and Figure 21 show the results of applying the 3 steps of experiment C to each state, the District of Columbia (as just reported) and the entire United States. The percentage of people residing in each locale likely to be uniquely identifiable based on {*gender*, *month and year of birth*, *ZIP*} appear in the column named "MonYr %ID_pop." For example, 18.1% of the population of Iowa (see Figure 20) and 26.5% of the population of North Dakota (see Figure 21) are likely to be uniquely identifiable based on {*gender*, *month and year of birth*, *ZIP*}.

The next to last row in Figure 21 labeled "USA" reports the results of applying the 3 steps of experiment C to all ZIP codes in the United States. As shown, 3.7% of the population of the United States is likely to be uniquely identified by values of {*gender*, *month and year of birth*, *ZIP*}. The last row in Figure 21 labeled "%ID_pop" displays the percentage of people in each age subdivision who are likely to be uniquely identified by values of {*gender*, *month and year of birth*, *ZIP*}. For example, it reports that 5% of the population of persons residing in the United States between the ages of 45 and 54 are likely to be uniquely identifiable based on {*gender*, *month and year of birth*, *ZIP*}.

Figure 22 plots the percentage of the population considered identifiable in each ZIP code in the United States based on experiment C's criteria. The horizontal axis represents the population that resides in the ZIP code. The vertical axis represents the percentage of the

population considered uniquely identified by values of QI_{SID1} = {*month and year of birth*, *gender*, *5-digit ZIP*} for a particular ZIP code. This is the same as the approach used in experiment B.

State	MonYr %ID_pop	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
AL	3.8%	22,253	11,325	10,982	18,197	19,285	22,443	24,806	24,254
AK	10.7%	12,416	6,542	4,826	9,045	7,633	6,253	5,522	6,063
AZ	1.4%	6,804	3,888	4,386	6,786	5,968	7,091	7,095	8,120
AR	11.4%	41,221	23,185	20,274	34,340	35,164	35,248	35,440	42,675
CA	0.8%	33,588	19,440	16,982	27,467	26,335	31,331	33,500	34,743
CO	3.7%	18,174	10,214	8,764	14,721	14,523	16,946	17,965	21,333
CT	1.2%	5,203	2,845	3,097	4,102	3,675	5,104	7,135	7,514
DE	0.9%	867	557	257	653	652	960	715	1,627
DC	0.2%	275	72	26	180	95	66	57	404
FL	0.6%	10,862	6,777	6,548	8,311	9,208	11,647	11,760	13,330
GA	2.7%	19,935	11,272	11,318	18,321	22,193	26,345	31,161	34,905
HI	1.6%	1,767	1,242	1,602	1,911	1,795	2,797	3,645	3,469
ID	8.9%	11,922	7,146	6,950	11,657	11,988	12,404	12,220	15,587
IL	4.4%	75,604	42,727	40,364	62,012	63,393	68,919	70,997	77,971
IN	4.0%	28,592	16,297	17,739	25,328	25,849	33,632	34,730	36,884
IA	18.1%	82,724	44,905	34,644	70,040	64,634	65,878	65,808	72,916
KS	12.1%	46,345	25,207	20,797	36,178	38,319	40,822	41,630	49,544
KY	8.3%	48,404	24,728	23,501	37,727	39,465	41,358	43,680	46,346
LA	2.8%	15,800	8,567	8,553	13,180	13,922	17,090	18,399	22,675
ME	15.5%	29,727	16,098	14,462	23,099	23,470	26,896	26,041	30,713
MD	2.1%	14,087	7,843	8,086	11,105	11,093	13,739	16,099	20,297
MA	1.1%	8,446	5,949	5,540	6,291	6,191	10,006	12,702	12,847
MI	2.4%	27,008	16,914	18,153	22,223	25,106	33,248	37,570	40,591
MN	9.0%	59,128	34,860	28,225	49,369	52,048	54,780	53,583	60,926
MS	4.4%	12,939	7,915	8,487	12,557	14,378	17,937	18,845	20,676

Figure 20 Uniqueness of {*ZIP*, *Gender*, *Month and year of birth*} respecting age distribution, part 1

Of the ZIP codes reported in Figure 22, about half (13,871 of 29,212 or 47%) have sufficient numbers of people in each age subdivision so that values of QI_{SID1} = {*month and year of birth*, *gender*, *5-digit ZIP*} are not likely to be uniquely identifying; in these cases, *%pop identifiable* = 0. Values of QI_{SID1} for about one third (9103 of 29212 or 31%) of the ZIP codes are considered uniquely identifying in all age subdivisions; in these cases, *%pop identifiable* = 1. The remaining ZIP codes (6238 of 29212 or 21%) have sub-populations in which values of QI_{SID1} are uniquely identifiable for some age subdivisions but not for others.

Figure 23 provides statistical highlights from the plot in Figure 22. The topmost table provides statistics on ZIP codes in which the number of people within the noted age subdivision is less than or equal to the threshold for that subdivision. In these cases, the sub-population within the ZIP code is considered uniquely identifiable; that is, *%pop_Identifiable* = 1 for that age subdivision and ZIP code. The bottom table provides statistics in cases where *%pop_Identifiable* < 1. In these ZIP codes, the number of people within the noted age subdivision is greater than the threshold for that subdivision; therefore, this subdivision is not considered uniquely identifiable. The method for computing these statistics was described earlier in the Methods section (on page 11).

State	MonYr %ID_pop	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
MO	8.2%	65,966	37,847	31,629	52,566	53,596	57,098	56,566	65,194
MT	15.5%	18,771	11,741	7,717	16,581	16,326	16,280	16,432	19,924
NE	18.2%	46,646	27,556	17,763	38,678	40,574	34,699	37,697	43,232
NV	2.0%	4,320	2,035	1,983	3,341	2,977	2,516	2,705	4,256
NH	7.5%	11,934	7,545	6,001	8,773	7,859	12,067	13,156	15,851
NJ	0.6%	6,760	4,693	3,510	3,811	4,642	5,846	8,238	8,142
NM	5.2%	11,169	6,307	5,208	10,048	10,235	10,844	11,340	14,141
NY	2.3%	54,792	33,243	31,443	45,160	49,560	61,882	68,223	76,979
NC	2.4%	22,064	11,906	10,595	17,177	16,987	23,559	27,726	31,714
ND	26.5%	28,362	16,090	9,492	22,535	22,563	20,666	22,226	27,314
OH	2.2%	28,645	14,449	18,930	24,301	24,283	37,395	43,814	47,838
OK	7.1%	32,749	20,178	16,901	26,174	29,484	29,507	32,320	35,238
OR	4.2%	18,614	9,286	8,839	15,741	14,495	15,766	15,778	19,684
PA	3.5%	58,144	32,516	32,758	47,305	45,996	62,507	66,894	75,584
RI	0.9%	1,085	642	500	764	1,417	1,025	1,487	1,996
SC	2.3%	9,342	5,171	5,813	8,643	8,309	12,372	13,670	16,738
SD	25.9%	27,699	17,147	11,054	25,496	24,375	22,171	23,721	28,405
TN	3.4%	24,172	12,553	13,053	18,105	19,074	22,832	25,898	30,553
TX	2.3%	51,615	29,794	30,883	45,082	50,060	58,173	62,784	68,838
UT	3.4%	8,496	4,844	4,042	7,026	7,447	8,832	8,293	10,307
VT	21.9%	19,797	11,196	8,334	16,536	17,312	16,075	16,093	18,066
VA	4.4%	41,345	23,241	20,634	30,706	33,035	35,263	40,117	47,007
WA	2.6%	18,736	11,083	9,104	14,925	15,043	17,563	19,665	21,650
WV	15.5%	43,535	25,866	21,381	36,753	37,676	34,584	35,731	42,582
WI	5.4%	32,406	21,664	21,855	31,257	30,297	40,576	43,567	44,714
WY	10.1%	8,492	3,943	2,743	6,058	5,943	6,251	5,893	6,684
USA	**3.7%**	1,329,747	759,051	676,728	1,098,342	1,125,947	1,269,289	1,351,139	1,529,041
%ID_pop		3.1%	3.2%	3.0%	2.5%	3.0%	5.0%	6.4%	4.9%

Figure 21 Uniqueness of {*ZIP*, *Gender*, *Month and year of birth*} respecting age distribution, part 2

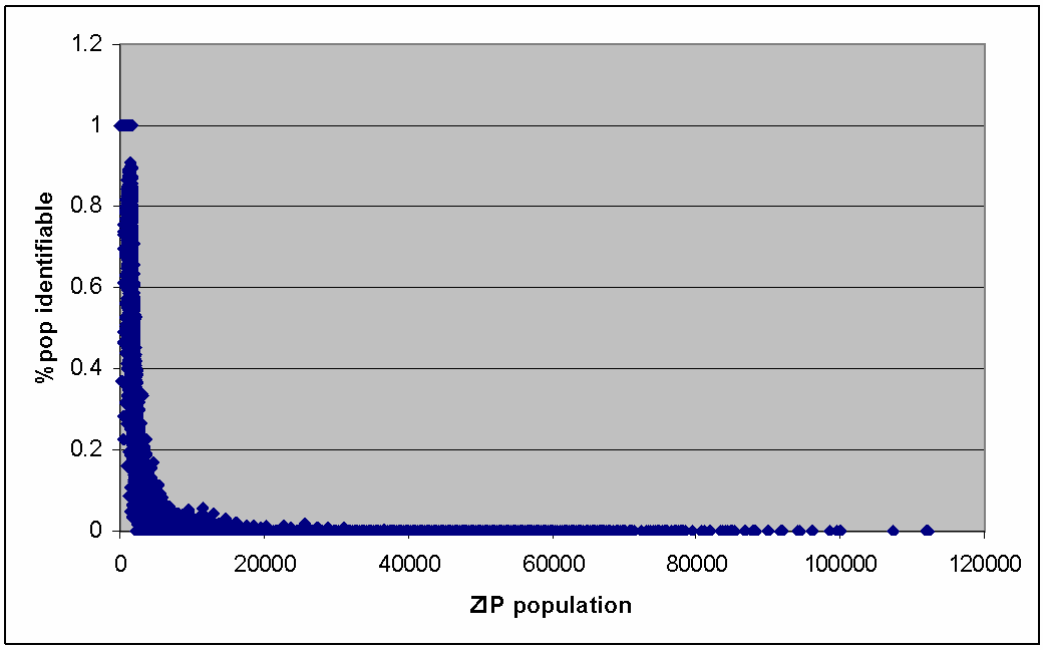


Figure 22 Percentage of Population Identifiable Based on Uniform Distribution of Ages in ZIP Population

The values reported as ZIP populations in Figure 22 are not the total number of people within the reported age subdivision residing in those ZIP codes but are just the numbers of people

residing in the ZIP code. For example, consider the values appearing in the "Aunder12" column in Figure 22. They report information about children under the age of 12 residing in 10,852 ZIP codes in the United States that had insufficient numbers of children to render corresponding values of QI_{SID1} = {*month and year of birth*, *gender*, *5-digit ZIP*} uniquely identifiable. Of these ZIP codes, the largest number of children of under the age of 12, residing in a ZIP code was 287. Some ZIP codes, who had people residing within them, had no children in this age. The average number of children in these ZIP codes was 123 with a standard deviation of 80.

Sub-population considered uniquely identifiable (<= *threshold, IDSet*)

	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
Max ZIP sub-population	287	167	143	239	239	239	239	287
Min ZIP sub-population	0	0	0	0	0	0	0	0
Average ZIP sub-population	123	71	53	101	102	96	95	118
standard deviation	80	47	40	66	66	66	66	80
Number of ZIP codes	10852	10725	12760	10883	11045	13202	14220	12905
Percentage ZIP codes	37.1%	36.7%	43.7%	37.3%	37.8%	45.2%	48.7%	44.2%

Sub-population NOT considered uniquely identifiable (> *threshold, NotIDSet*)

	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
Max ZIP sub-population	26914	15352	27123	24587	19543	15544	12205	25799
Min ZIP sub-population	288	168	144	240	240	240	240	288
Average ZIP sub-population	2294	1241	1333	2309	2007	1509	1316	1815
standard deviation	2530	1327	1690	2632	2096	1419	1174	1860
Number of ZIP codes	18360	18487	16452	18329	18167	16010	14992	16307
Percentage ZIP codes	62.9%	63.3%	56.3%	62.7%	62.2%	54.8%	51.3%	55.8%

Figure 23 Statistical highlights from Figure 20 and Figure 21

5.3. Experiment D: Uniqueness of {*ZIP*, *gender*, *age*}

In this experiment, I examine the identifiability of {*year of birth*, *gender*, *5-digit ZIP*} in the United States. Progressing through the results from the last three experiments, values referring to age became less specific and as expected, the values became less uniquely identifying. What may be surprising however is that these values remained uniquely identifying for some people.

The Agency for Healthcare Research and Quality's State Inpatient Database (SID; see Figure 3) motivated this experiment as well as experiment C. In addition to QI_{SID1} used in experiment C, SID also includes QI_{SID2} = {*age*, *gender*, *5-digit ZIP*} for some states in those data. Recall in section 4.4.1, I examine *age* as providing a distinct year of birth, and so QI_{SID2} = {*age*, *gender*, *5-digit ZIP*} can be considered as QI_{SID2} = {*year of birth*, *gender*, *5-digit ZIP*}.

5.3.1. Experiment D Design

Step 1. Use ZIP table for each of the 50 states and the District of Columbia. Step 2. Figure 24 contains the thresholds for Q={*gender*, *date of birth*} specific to each age subdivision. Step 3. Report statistical measurements computed from the table in step 1 using the thresholds determined in step 2. Figure 25 and Figure 26 report the results.

Q4 = {*gender*, *year of birth*}		
$\|Q4_{AUnder12}\|$	= 2 * 12	= 24
$\|Q4_{A12to18}\|$	= 2 * 7	= 14
$\|Q4_{A19to24}\|$	= 2 * 6	= 12
$\|Q4_{A25to34}\|$	= 2 * 10	= 20
$\|Q4_{A35to44}\|$	= 2 * 10	= 20
$\|Q4_{A45to54}\|$	= 2 * 10	= 20
$\|Q4_{A55to64}\|$	= 2 * 10	= 20
$\|Q4_{A65Plus}\|$	= 2 * 12	= 24

Figure 24 Number of possible values for each age subdivision for {*gender*, *year of birth*}

5.3.2. Experiment D Results

Figure 25 and Figure 26 show the results of applying the 3 steps of experiment D to each state, the District of Columbia (as just reported) and the entire United States. The percentage of people residing in each locale likely to be uniquely identifiable based on {*gender*, *year of birth*, *ZIP*} appears in the column named “BirthYr %ID_pop” and the number of people represented by the percentage appears in the column named "BirthYr #ID_pop". For example, 0.89% (or 5703 people) of the population of Iowa (see Figure 26) are likely to be uniquely identifiable by values of {*gender*, *year of birth*, *ZIP*}.

State	BirthYr %ID_pop	BirthYr Total	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
AL	0.02%	918	105	53	89	97	112	125	158	179
AK	0.70%	3,809	227	223	227	223	315	631	804	1,159
AZ	0.02%	638	68	31	23	53	98	98	96	171
AR	0.09%	2,121	452	138	264	208	248	312	349	150
CA	0.01%	4,229	541	319	362	461	336	540	678	992
CO	0.08%	2,752	287	224	346	336	201	447	426	485
CT	0.01%	474	69	55	36	52	30	108	63	61
DE	0.02%	158	18	13	21	28	36	5	10	27
DC	0.01%	46	6	-	-	-	-	-	16	24
FL	0.00%	512	76	63	9	5	43	90	121	105
GA	0.01%	780	83	29	91	101	56	120	182	118
HI	0.01%	165	28	11	9	33	42	12	20	10
ID	0.19%	1,943	259	148	205	255	258	310	248	260
IL	0.01%	1,401	167	111	148	141	123	246	255	210
IN	0.01%	746	82	27	54	88	84	89	131	191
IA	0.11%	3,106	278	305	647	182	249	583	535	327
KS	0.22%	5,482	575	446	924	571	594	1,017	750	605
KY	0.13%	4,722	671	309	280	528	448	697	966	823
LA	0.02%	870	118	48	75	118	84	135	169	123
ME	0.19%	2,296	293	217	190	287	228	280	331	470
MD	0.03%	1,275	152	119	96	156	179	187	194	192
MA	0.01%	499	83	50	51	35	25	58	100	97
MI	0.01%	920	124	133	134	151	71	133	120	54
MN	0.06%	2,709	365	214	439	421	265	326	335	344
MS	0.02%	462	54	23	21	39	26	57	136	106

Figure 25 Uniqueness of {*ZIP*, *Gender*, *Year of birth*} respecting age distribution, part 1

The next to last row in Figure 26 labeled "USA" reports the results of applying the 3 steps of experiment D to all ZIP codes in the United States. As shown, 0.04% (or 105,016 people) of the population of the United States is likely to be uniquely identified by values of {*gender*, *year of birth*, *ZIP*}. The last row in Figure 26 labeled "%ID_pop" displays the percentage of people in each age subdivision who are likely to be uniquely identified by values of {*gender*, *year of birth*, *ZIP*}. For example, it reports that 0.08% of the population of persons residing in the

United States between the ages of 55 and 64 are likely to be uniquely identified by values of {*gender*, *year of birth*, *ZIP*}.

State	BirthYr %ID_pop	BirthYr Total	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
MO	0.07%	3,403	451	320	402	312	371	549	531	467
MT	0.43%	3,465	399	263	405	433	362	534	492	577
NE	0.23%	3,560	241	241	717	325	387	676	455	518
NV	0.04%	439	77	35	39	47	47	62	57	75
NH	0.07%	777	154	62	106	56	81	111	100	107
NJ	0.01%	728	125	62	41	61	51	96	114	178
NM	0.22%	3,302	343	276	237	395	350	569	644	488
NY	0.03%	5,460	714	469	533	720	445	804	818	957
NC	0.02%	1,032	133	94	74	134	103	177	168	149
ND	0.89%	5,703	586	476	832	675	639	932	787	776
OH	0.00%	377	34	25	30	37	33	38	96	84
OK	0.06%	1,963	220	135	248	219	274	336	237	294
OR	0.07%	1,900	369	140	172	258	124	214	315	308
PA	0.03%	3,099	501	201	324	413	348	429	440	443
RI	0.01%	92	-	-	-	9	10	30	19	24
SC	0.01%	443	87	16	41	66	63	85	45	40
SD	0.63%	4,408	489	291	607	544	516	632	597	732
TN	0.02%	836	201	14	125	70	53	165	128	80
TX	0.03%	5,483	815	383	443	641	661	717	794	1,029
UT	0.08%	1,323	78	59	146	189	151	230	230	240
VT	0.20%	1,117	76	63	171	54	81	166	150	356
VA	0.06%	3,754	572	286	350	423	445	483	638	557
WA	0.03%	1,227	164	85	145	138	122	142	220	211
WV	0.30%	5,360	746	316	433	614	605	874	869	903
WI	0.02%	881	80	101	135	130	79	103	103	150
WY	0.41%	1,851	213	157	232	165	195	361	223	305
USA	**0.04%**	105,016	13,049	7,879	11,729	11,697	10,747	16,121	16,463	17,331
%ID_pop			0.03%	0.03%	0.05%	0.03%	0.03%	0.06%	0.08%	0.06%

Figure 26 Uniqueness of {*ZIP*, *Gender*, *Year of birth*} respecting age distribution, part 2

Most ZIP codes (25,705 of 29,212 or 88%) have sufficient numbers of people in each age subdivision so that values of QI_{SID2} = {*year of birth*, *gender*, *5-digit ZIP*} are not likely to be uniquely identifying; in these cases, *%pop identifiable* = 0. Values of QI_{SID2} for about one third (353 of 29212 or 1%) of the ZIP codes are considered uniquely identifying in all age subdivisions; in these cases, *%pop identifiable* = 1. The remaining ZIP codes (3154 of 29212 or 11%) have sub-populations in which values of QI_{SID2} are uniquely identifiable for some age subdivisions but not for all.

Figure 27 provides statistical highlights. The topmost table provides statistics on ZIP codes in which the number of people within the noted age subdivision is less than or equal to the threshold for that subdivision. In these cases, the sub-population within the ZIP code is considered uniquely identifiable; that is, *%pop_Identifiable* = 1 for that age subdivision and ZIP code. The bottom table provides statistics in cases where *%pop_Identifiable* < 1. In these ZIP codes, the number of people within the noted age subdivision is greater than the threshold for that subdivision; therefore, this subdivision is not considered uniquely identifiable.

Sub-population considered uniquely identifiable (<= *threshold, IDSet*)

	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
Max ZIP sub-population	24	14	12	20	20	20	20	24
Min ZIP sub-population	0	0	0	0	0	0	0	0
Average ZIP sub-population	11	6	5	10	9	10	9	11
standard deviation	8	5	4	7	7	7	7	8
Number of ZIP codes	1200	1342	2309	1210	1150	1651	1798	1584
Percentage ZIP codes	4.1%	4.6%	7.9%	4.1%	3.9%	5.7%	6.2%	5.4%

Sub-population NOT considered uniquely identifiable (> *threshold, NotIDSet*)

	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
Max ZIP sub-population	26914	15352	27123	24587	19543	15544	12205	25799
Min ZIP sub-population	25	15	13	21	21	21	21	25
Average ZIP sub-population	1551	850	840	1551	1339	922	768	1126
standard deviation	2291	1212	1460	2372	1914	1284	1057	1652
Number of ZIP codes	28012	27870	26903	28002	28062	27561	27414	27628
Percentage ZIP codes	95.9%	95.4%	92.1%	95.9%	96.1%	94.3%	93.8%	94.6%

Figure 27 Statistical highlights from Figure 25 and Figure 26

5.4. Experiment F: Uniqueness of {*place*/*city*, *gender*, *date of birth*}

This experiment examines the identifiability of {*date of birth*, *gender*, *place*}. While the number of places is expected to be less than the number of ZIP codes, the difference is not as dramatic as one would expect.

State	DOB %ID_pop	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
AL	74.31%	510,294	316,271	246,921	455,646	425,871	340,085	290,787	416,713
AK	67.62%	86,943	36,668	30,801	72,365	66,328	34,744	18,336	22,157
AZ	30.18%	207,821	117,371	79,857	154,789	150,173	121,318	116,660	158,254
AR	85.73%	355,634	221,013	144,471	278,355	286,099	217,119	197,637	314,833
CA	35.99%	1,705,016	1,032,675	785,915	1,266,384	1,411,260	1,494,618	1,350,466	1,663,710
CO	40.41%	221,248	124,459	100,826	189,908	196,192	164,375	145,456	188,554
CT	66.44%	355,973	208,871	144,966	296,959	320,087	299,897	249,481	307,714
DE	68.04%	78,966	40,675	32,116	63,018	69,766	49,625	52,013	67,054
DC	0.00%	-	-	26	-	-	-	-	-
FL	44.12%	866,146	523,124	416,970	743,419	783,719	697,379	690,365	875,685
GA	62.62%	737,096	425,884	331,861	601,348	569,614	501,763	393,910	495,444
HI	49.94%	89,975	69,406	41,139	80,566	82,216	68,636	55,413	66,056
ID	76.93%	147,599	93,691	50,482	116,729	113,067	83,114	66,524	103,285
IL	60.16%	1,205,138	698,921	490,199	976,815	965,017	842,088	731,266	966,748
IN	63.45%	610,004	362,468	272,124	485,926	499,979	431,504	366,413	488,978
IA	77.50%	375,417	218,025	141,276	310,173	302,724	238,696	219,669	345,691
KS	66.77%	295,043	167,547	111,512	236,104	229,189	182,750	160,132	270,086
KY	78.76%	513,045	319,232	234,139	451,331	419,197	325,073	277,950	369,257
LA	58.86%	474,999	271,968	196,903	380,395	336,651	278,656	233,811	310,514
ME	94.22%	201,167	117,015	82,913	184,342	184,857	123,745	108,198	153,502
MD	63.22%	542,516	299,174	256,363	432,696	456,506	379,792	307,456	341,639
MA	73.33%	738,432	409,915	351,483	610,144	673,586	526,058	440,426	658,804
MI	56.68%	912,385	535,570	393,345	760,515	737,677	656,494	551,937	720,202
MN	71.55%	582,951	327,576	213,712	462,644	439,233	358,955	299,529	442,243
MS	81.12%	386,515	232,392	164,750	307,447	278,994	231,718	197,189	288,516

Figure 28 Uniqueness of {*Place*, *Gender*, *Date of birth*} respecting age distribution, part 1

Step 1. Use ZIP table for each of the 50 states and the District of Columbia. Step 2. Figure 12 contains the thresholds for Q={*gender*, *date of birth*} specific to each age subdivision. Step 3. Report statistical measurements computed from the table in step 1 using the thresholds

determined in step 2. Figure 28 and Figure 29 report the results of applying the 3 steps of experiment F to each state, the District of Columbia and the entire United States.

The percentage of people residing in each locale likely to be uniquely identifiable by values of {*gender*, *date of birth*, *place*} appear in the column named "DOB %ID_pop." For example, 94.22% of the population of Maine (see Figure 28) and 74.99% of the population of Pennsylvania (see Figure 29) are likely to be uniquely identifiable by values of {*gender*, *date of birth*, *place*}. Vermont had the largest percentage of its population identifiable (98.12%). The District of Columbia had 0% identified. The state having the smallest percentage was Nevada with 26.48%. The average was 64.54% and the standard deviation was 17.88%.

State	DOB %ID_pop	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
MO	65.98%	575,534	345,340	253,443	490,825	454,370	395,626	347,346	509,243
MT	78.05%	111,323	63,624	30,390	86,536	94,856	73,526	67,930	95,497
NE	60.86%	173,370	100,557	63,607	137,330	136,238	98,945	92,145	157,885
NV	26.48%	48,890	29,379	17,274	44,040	48,251	49,077	36,910	44,428
NH	83.26%	164,556	84,043	75,108	158,945	156,196	94,268	79,579	110,913
NJ	75.46%	916,586	513,909	459,760	887,738	910,504	705,604	615,918	823,232
NM	58.82%	185,741	103,241	70,980	125,794	127,320	94,373	78,403	105,250
NY	50.89%	1,510,307	893,370	734,124	1,331,293	1,394,790	1,103,058	955,471	1,231,836
NC	66.99%	748,655	434,802	352,507	670,230	637,726	523,682	455,492	617,381
ND	89.24%	108,831	59,803	33,455	83,627	83,251	56,215	53,132	90,771
OH	65.65%	1,218,515	726,779	536,583	1,009,900	1,059,754	865,805	737,419	965,782
OK	64.24%	349,375	209,852	141,980	280,350	266,557	233,933	212,063	326,461
OR	64.29%	318,531	186,694	120,253	251,227	266,919	224,214	180,088	279,439
PA	74.99%	1,427,475	829,811	674,412	1,324,556	1,288,682	1,002,535	960,527	1,401,861
RI	55.57%	83,379	52,128	46,137	74,615	83,775	73,597	65,732	78,157
SC	67.65%	404,179	259,598	178,853	347,400	357,955	263,798	240,827	306,073
SD	81.02%	108,221	62,338	36,113	80,508	80,733	53,059	51,721	90,508
TN	64.98%	529,152	319,932	243,251	474,021	459,452	388,946	320,903	433,014
TX	44.27%	1,410,090	792,176	561,715	1,100,437	1,053,590	840,761	735,749	1,025,466
UT	56.43%	208,964	117,137	81,156	132,730	134,699	106,448	84,198	106,867
VT	98.12%	99,365	53,099	42,494	95,880	92,804	57,274	45,118	66,169
VA	58.50%	588,706	358,361	294,519	565,454	531,480	468,056	357,966	453,394
WA	53.56%	458,232	257,086	168,811	372,536	382,178	319,725	272,740	375,511
WV	90.95%	260,338	178,947	125,468	232,443	242,711	184,384	169,168	237,233
WI	68.27%	584,155	333,763	235,969	497,263	483,528	372,939	334,139	497,585
WY	79.05%	67,039	36,679	20,714	52,859	53,145	45,541	35,539	47,045
USA	**58.38%**	24,859,832	14,572,359	10,914,146	20,826,555	20,879,466	17,343,591	15,107,247	20,512,640
%ID_pop		57.2%	61.5%	48.3%	48.0%	55.6%	68.2%	71.7%	65.9%

Figure 29 Uniqueness of {*Place*, *Gender*, *Date of birth*} respecting age distribution, part 2

The next to last row in Figure 29 labeled "USA" reports the results of applying the 3 steps of experiment F to all places in the United States. As shown, 58.38% of the population of the United States is likely to be uniquely identified by values of {*gender*, *date of birth*, *place*}. The last row in Figure 29 labeled "%ID_pop" displays the percentage of people in each age subdivision who are likely to be uniquely identified by values of {*gender, date of birth*, *place*}. For example, it reports that 71.7% of the population of persons residing in the United States between the ages of 55 and 64 are likely to be uniquely identifiable based on {*gender*, *date of birth*, *place*}.

The place having the largest population was Chicago, Illinois, with 2,451,767 people. The place having the smallest population was Crooked Creek, Alaska that reports only one person of age 65 or more resides there. The average population for a place is 9,710 and the standard deviation is 44,149. There are a total of 25,585 places.

5.5. Experiment J: Uniqueness of {*county*, *gender*, *date of birth*}

This experiment examines the identifiability of {*date of birth*, *gender*, *county*}. Recall, there are a total of 29,343 ZIP codes, 25,688 places and 3,141 counties.

Step 1. Use ZIP table for each of the 50 states and the District of Columbia. Step 2. Figure 12 contains the thresholds for Q={*gender*, *date of birth*} specific to each age subdivision. Step 3. Report statistical measurements computed from the table in step 1 using the thresholds determined in step 2. Figure 30 and Figure 31 report the results of applying the 3 steps of experiment J to each state, the District of Columbia and the entire United States.

The percentage of people residing in each locale likely to be uniquely identifiable by values of {*gender*, *date of birth*, *county*} appear in the column named "DOB %ID_pop." For example, 58% of the population of Mississippi (see Figure 30) and 52% of the population of Nebraska (see Figure 31) are likely to be uniquely identifiable by values of {*gender*, *date of birth*, *county*}. Wyoming had the largest percentage of its population identifiable (75%). Connecticut, Delaware, the District of Columbia and New Jersey had 0% identified. The average was 28% and the standard deviation was 22%.

The next to last row in Figure 31 labeled "USA" reports the results of applying the 3 steps of experiment J to all counties in the United States. As shown, 18.1% of the population of the United States is likely to be uniquely identified by values of {*gender*, *date of birth*, *county*}. The last row in Figure 31 labeled "%ID_pop" displays the percentage of people in each age subdivision who are likely to be uniquely identified by values of {*gender, date of birth*, *county*}. For example, it reports that 25.84% of the population of persons residing in the United States between the ages of 55 and 64 are likely to be uniquely identifiable based on {*gender*, *date of birth*, *county*}.

State	DOB %ID_pop	DOB Total	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
AL	31%	1,239,261	203,418	119,887	100,859	147,718	149,568	145,639	165,943	206,229
AK	43%	231,537	39,695	25,282	18,424	47,769	29,382	31,533	17,293	22,159
AZ	5%	168,352	23,995	13,659	8,351	15,873	23,248	23,557	26,589	33,080
AR	55%	1,286,703	204,611	126,862	85,675	165,982	171,952	153,203	149,635	228,783
CA	2%	482,182	74,362	42,716	33,536	62,826	50,565	61,321	70,890	85,966
CO	16%	530,181	94,650	50,001	38,332	85,809	86,915	60,219	46,794	67,461
CT	0%	-	-	-	-	-	-	-	-	-
DE	0%	-	-	-	-	-	-	-	-	-
DC	0%	-	-	-	-	-	-	-	-	-
FL	5%	680,438	109,084	74,526	59,719	96,106	93,589	80,169	64,489	102,756
GA	36%	2,335,158	385,475	236,121	182,875	311,416	297,509	294,860	257,992	368,910
HI	2%	24,302	-	4,985	3,356	-	20	5,039	4,127	6,775
ID	50%	504,176	84,045	54,338	27,716	64,270	70,098	61,874	63,762	78,073
IL	15%	1,733,651	294,307	164,151	119,585	237,212	225,134	210,334	189,098	293,830
IN	33%	1,805,518	310,118	183,259	129,393	268,623	228,630	204,738	205,590	275,167
IA	57%	1,574,848	267,585	153,138	102,462	208,798	216,811	168,181	161,950	295,923
KS	45%	1,117,968	187,792	105,602	71,548	150,530	142,864	128,522	120,241	210,869
KY	55%	2,015,672	339,649	199,166	162,837	287,814	264,521	239,055	215,956	306,674
LA	26%	1,103,759	166,000	99,616	76,791	129,159	151,255	132,897	154,279	193,762
ME	24%	289,549	45,914	25,233	25,821	33,214	34,481	37,839	34,151	52,896
MD	6%	288,043	36,084	19,602	20,508	36,667	32,605	29,983	51,224	61,370
MA	1%	30,080	2,997	1,179	914	2,739	3,651	8,515	7,345	2,740
MI	14%	1,270,356	187,954	101,271	84,202	147,645	144,700	167,106	186,504	250,974
MN	35%	1,545,738	264,233	169,559	98,392	209,122	217,857	169,995	152,615	263,965
MS	58%	1,503,027	258,287	160,447	109,981	202,539	185,987	179,021	161,836	244,929

Figure 30 Uniqueness of {*County*, *Gender*, *Date of birth*} respecting age distribution, part 1

The county having the largest population was Los Angeles County in California, with 8,863,164 people. The county having the smallest population was Yellowstone County in

Montana where only 52 people reside. The average population for a county is 79,182 and the standard deviation is 263,813. There are a total of 3,141 counties.

State	DOB %ID_pop	DOB Total	AUnder12	A12to18	A19to24	A25to34	A35to44	A45to54	A55to64	A65Plus
MO	35%	1,777,250	299,822	177,231	117,495	235,542	226,741	199,449	201,418	319,552
MT	58%	461,847	79,868	50,920	26,385	52,524	57,155	54,121	58,303	82,571
NE	52%	827,590	148,680	81,858	48,439	114,003	115,534	80,505	86,132	152,439
NV	17%	205,707	39,060	19,336	13,922	34,721	32,791	23,853	18,433	23,591
NH	14%	158,917	18,460	14,248	10,218	17,034	17,684	29,898	27,505	23,870
NJ	0%	13,203	-	0	-	-	-	7,089	6,114	-
NM	31%	466,942	55,935	41,285	37,298	48,782	59,912	68,042	65,578	90,110
NY	4%	714,072	86,136	44,198	48,241	39,884	79,329	130,425	139,808	146,051
NC	26%	1,718,318	242,446	149,044	126,104	202,459	209,646	232,090	249,455	307,074
ND	63%	401,471	65,977	37,393	19,272	49,281	47,612	47,773	53,295	80,868
OH	14%	1,536,542	244,518	135,966	102,380	185,611	200,338	199,829	190,210	277,690
OK	44%	1,395,889	214,447	135,344	106,030	180,771	170,655	171,464	164,825	252,353
OR	16%	468,933	58,089	37,189	25,490	51,118	50,034	77,348	76,465	93,200
PA	7%	868,774	143,074	81,634	65,841	120,867	115,691	110,104	109,679	121,884
RI	4%	36,592	7,442	4,146	-	-	7,157	5,220	4,929	7,698
SC	23%	792,897	115,127	71,978	53,250	100,618	97,592	104,465	111,669	138,198
SD	73%	506,465	96,431	52,085	32,146	70,960	65,236	51,201	50,256	88,150
TN	37%	1,832,875	296,158	180,822	134,829	239,766	227,196	226,279	223,498	304,327
TX	19%	3,185,236	555,868	314,582	220,496	408,489	396,535	357,970	372,889	558,407
UT	17%	296,513	58,729	33,397	21,901	43,107	40,697	30,917	26,743	41,022
VT	59%	329,450	48,194	35,514	24,136	42,551	42,821	44,422	36,277	55,535
VA	35%	2,186,920	327,643	195,729	180,037	286,163	280,550	300,469	262,255	354,074
WA	11%	523,874	66,444	57,010	36,219	58,899	62,605	75,825	83,264	83,608
WV	59%	1,059,753	168,623	100,661	72,214	144,775	151,174	140,705	128,935	152,666
WI	25%	1,211,247	190,779	110,977	71,189	162,807	159,047	141,883	157,036	217,529
WY	75%	338,752	57,064	36,055	20,545	52,035	52,251	38,218	35,539	47,045
USA	**18.1%**	45,076,528	7,265,269	4,329,202	3,175,354	5,854,598	5,787,325	5,543,164	5,448,813	7,672,803
%ID_pop			16.72%	18.27%	14.04%	13.48%	15.40%	21.79%	25.84%	24.65%

Figure 31 Uniqueness of {*County*, *Gender*, *Date of birth*} respecting age distribution, part 2

6. Discussion

Figure 32 contains a summary of the results reported in the previous section. A description of each reported percentage is provided in the following paragraphs. These percentages demonstrate how combinations of characteristics can combine to narrow the number of possible people under consideration as the subject of de-identified person-specific data.

County	18.1	0.04	0.00004	0.00000*
Place	58.4	3.6	0.04	0.01
ZIP	87.1	3.7	0.04	0.01
	DOB	**Mon/Year**	**BirthYear**	**2yr Age**

Figure 32 Percentage of US population identified with gender as geography and age vary

Experiment B reported that 87.1% (216 million of 248 million) of the population in the United States had characteristics that were likely made them unique based only on {*5-digit ZIP*, *gender*, *date of birth*}. Experiment C reported that 3.7% of the population in the United States had characteristics that were likely made them unique based only on {*5-digit ZIP*, *gender*, *Month and year of birth*}. Experiment D reported that 0.04% of the population in the United States had characteristics that were likely made them unique based only on {*5-digit ZIP*, *gender*, Year *of birth*}. Experiment E reported that 0.01% of the population in the United States had characteristics that were likely made them unique based only on {*5-digit ZIP*, *gender*, *2year age range*}.

Experiment F reported that 58.4% of the population in the United States had characteristics that were likely made them unique based only on {*Place*, *gender*, *date of birth*}. Experiment G reported that 3.6% of the population in the United States had characteristics that were likely made them unique based only on {*Place*, *gender*, *Month and year of birth*}. Experiment H reported that 0.04% of the population in the United States had characteristics that were likely made them unique based only on {*Place*, *gender*, *Year of birth*}. Experiment I reported that 0.01% of the population in the United States had characteristics that were likely made them unique based only on {*Place*, *gender*, *2year age range*}.

Experiment J reported that 18.1% of the population in the United States had characteristics that were likely made them unique based only on {*County*, *gender*, *date of birth*}. Experiment K reported that 0.04% of the population in the United States had characteristics that were likely made them unique based only on {*County*, *gender*, *Month and year of birth*}. Experiment L reported that 0.00004% of the population in the United States had characteristics that were likely made them unique based only on {*County*, *gender*, *Year of birth*}. Experiment M reported that 0.00000% of the population in the United States had characteristics that were likely made them unique based only on {*County*, *gender*, *2year age range*}, but despite it being a very small number, it is not 0.*

As the number of possible values a quasi-identifier can assume decreases, the percentage of the population in the United States who had characteristics that were likely unique based on those values decreases. This is evidenced by each row in Figure 32. Moving from left to right within each row of Figure 32, the numbers of possible combinations decrease and the corresponding percentages decrease. Aggregating the geographical specification to county resulted in far fewer possible combinations than available with place or ZIP codes. This is evidenced within each column in Figure 32. Notice however that the differences between the number of places and the number of ZIP codes are not as dramatic, and as a result, neither are the corresponding percentages.

6.1. Predicting the number of people that can be identified in a release

It was already shown that de-identified releases of person-specific data that contain no explicit identifiers such as name, address or phone number, is not necessarily anonymous [16]. The maximum number of patients who could be identified in a public or semi-public release of health data is the number of patients who were hospitalized and whose information is therefore included in the data. Many possible combinations of attributes can combine to form a quasi-identifier useful for linking the de-identified data to explicitly identified data. The number of hospitalizations reported in the IHCCCC's R_{rod} data (see Figure 2) in one year is estimated to be 1 million based on the average statistic that 1:12 people are hospitalized each year.

However, the actual number of patients that could be re-identified in publicly and semi-publicly released health data is not necessarily every patient and the actual number is likely to differ among releases due to varying quasi-identifiers available. The results from the experiments reported in this document can help predict a minimum level of identifiability based on a combination of three demographics.

* In Loving County, Texas, 6 of 107 people are likely to be uniquely identified by values of {*gender*, *2yr age range*, *county*}. All of these 6 people are between the ages of 12 and 18 years.

6.1.1. Illinois Research Health Data

As shown in Figure 2, R_{rod} includes the full date of birth, gender, and the patient's 5-digit residential ZIP. Figure 13 reports that 75.3% of the population of Illinois is likely to be uniquely identified by {*5-digit ZIP*, *gender*, *date of birth*}. That corresponds to 753,000 patients being identified per year in R_{rod}.

6.1.2. AHRQ's State Inpatient Database

As shown in Figure 3, SID includes the month and year of birth, gender, and the patient's 5-digit residential ZIP for some states. Figure 33 estimates that 112,595 patients per year are likely to be uniquely identified by {*ZIP*, *Gender*, *Month and year of birth*} in SID. The five states known to report the month and year of the birth date of each patient to SID were introduced in **Error! Reference source not found.**. The populations for each of these states according to the 1990 Census data [17] were reported in Figure 8 and Figure 9. It is estimated that 1:12 people are hospitalized each year. These values are summarized in Figure 33.

State	Population	Hospitalized	Unique	PopID
AZ	3,665,228	305,436	1.4%	4,276
IA	2,776,442	231,370	18.1%	41,878
NY	17,990,026	1,499,169	2.3%	34,481
OR	2,842,321	236,860	4.2%	9,948
WI	4,891,452	407,621	5.4%	22,012
		Total per year		***112,595***

Figure 33 Estimated Uniqueness of {*ZIP*, *Gender*, *Month and year of birth*} in SID

State	Population	Hospitalized	Unique	PopID
AZ	3,665,228	305,436	0.02%	61
CA	29,755,274	2,479,606	0.01%	248
CO	3,293,771	274,481	0.08%	220
FL	12,686,788	1,057,232	0.00%	42
IA	2,776,442	231,370	0.11%	255
MA	6,011,978	500,998	0.01%	50
MD	4,771,143	397,595	0.03%	119
NJ	7,730,188	644,182	0.01%	64
NY	17,990,026	1,499,169	0.03%	450
OR	2,842,321	236,860	0.07%	166
SC	3,486,703	290,559	0.01%	29
WA	4,866,692	405,558	0.03%	122
WI	4,891,452	407,621	0.02%	82
		Total per year		***1,907***

Figure 34 Estimated Uniqueness of {*ZIP*, *Gender*, *Year of birth*} in SID

As shown in Figure 3, SID includes the year of birth (by way of age[18]), gender, and the patient's 5-digit residential ZIP for some states. Figure 34 estimates that 1,907 patients per year are likely to be uniquely identified by {*ZIP*, *Gender*, *Year of birth*} in SID. The 13 states known to report the year of the birth date of each patient to SID were introduced in **Error! Reference source not found.**. The populations for each of these states according to the 1990 Census data [19] were reported in Figure 8 and Figure 9. It is estimated that 1:12 people are hospitalized each year. These values are summarized in Figure 34.

There are many ways to misunderstand these values. These values are not to be considered an estimate of the uniqueness of R_{rod} or SID. There may exist other quasi-identifiers that may consist of more and different attributes that can link to other available data and thereby render the released health data even more identifiable. Such quasi-identifiers may use the hospital identifying number or discharge status or payment information. The estimates reported in this document are just approximations based on the demographic quasi identifiers stated. Therefore, these estimates should be viewed as a minimal estimate of the identifiability of these data. Clearly, these data are not anonymous.

6.2. Unique and unusual information found in data

A significant problem with producing anonymous data concerns unique and unusual information appearing within the data themselves. Instances of uniquely occurring characteristics found within the original data can be used by a reporter, private investigator and others to discredit the anonymity of the released data even when these instances are not unique in the general population. Unusual cases are often unusual in other sources of data as well making them easier to identify.

Importantly, close examination of the particulars of a database provides the best basis for determining uniquely identifying information and quasi-identifiers. In this document, I have examined outside information without examining the values of the released data themselves. The analysis is based on the fact that a combination of characteristics that makes one unique in a geographic population, for example, results in uniqueness in all other data that includes that geographic specification. An examination of the data however can reveal other kinds of unusual information that can be found in other sources of data making more patients easier to identify.

In an interview, for example, a janitor may recall an Asian patient whose last name was Chan and who worked as a stockbroker because the patient gave the janitor some good investing tips. Any single uniquely occurring value or group of values can be used to identify an individual. Remember that the unique characteristic may not be known beforehand. It could be based on diagnosis, treatment, birth year, visit date, or some other little detail or combination of details available to the memory of a patient or a doctor, or knowledge about the database from some other source.

As another example, consider the medical records of a pediatric hospital in which only one patient is older than 45 years of age. Suppose a de-identified version of the hospital's records is to be released for public-use that includes age and city of residence but not birth date or zip code. Many may believe the resulting data would be anonymous because there are thousands of people of age 45 living in that city. However, the rare occurrence of a 45 year-old pediatric patient at that facility can become a focal point for anyone seeking to discredit the anonymity of the data. Nurses, clerks and other hospital personnel will often remember unusual cases and in interviews may provide additional details that help identify the patient.

6.3. Future Work

Below are proposed projects of varying degrees of difficulties and skill requirements that extend this work.

In this document, I have demonstrated how combinations of characteristics can combine to narrow the number of possible people under consideration. However, knowing that there exist a one or a few people that share particular characteristics and explicitly identifying those people are not exactly the same. These combinations of characteristics must be linked to explicitly identified information to reveal the identities of the individuals. Further demonstrate the identifiability of these data by providing population registers to which the data could be linked to re-identify the noted individuals.

In an earlier document [20], privacy risk measures were computed on the data sets R_{rod} and SID based on the assumption that the entire populations within those data were identifiable. While that may be correct, use the findings reported in this document, which are based only on basic demographic attributes and do not include other attributes within those data that could be used for re-identification, and re-compute the measures of risk for those collections. Make an argument as to why these re-computed risk measurements should be considered "minimal" risk values.

7. References

1 National Association of Health Data Organizations, *NAHDO Inventory of State-wide Hospital Discharge Data Activities* (Falls Church: National Association of Health Data Organizations, May 2000).

2 Cambridge Voters List Database. *City of Cambridge, Massachusetts*. Cambridge: February 1997.

3 Supra note 1 NAHDO.

4 State of Illinois Health Care Cost Containment Council, *Data release overview.* (Springfield: State of Illinois Health Care Cost Containment Council, March 1998).

5 Agency for Healthcare Research and Quality, *Healthcare Cost and Utilization Project: Central Distributor* (April, 2000) available at http://www.ahcpr.gov/data/hcup/hcup-pkt.htm.

6 1990 U.S. Census Data, Database C90STF3B. *U.S. Bureau of the Census*. Available at http://venus.census.gov and http://www.census.gov. Washington: 1993.

7 T. Dalenius. Finding a needle in a haystack – or identifying anonymous census record. *Journal of Official Statistics,* 2(3):329-336, 1986.

8 G. Smith. Modeling security-relevant data semantics. In *Proceedings of the 1990 IEEE Symposium on Research in Security and Privacy,* May 1990.

9 J. Ullman. *Principles of Database and Knowledge Base Systems*. Computer Science Press, Rockville, MD. 1988.

10 Supra note 6 U.S. Bureau of the Census.

11 "Census Counts 90," *U.S. Bureau of the Census*. (Available on CDROM.) Washington: 1993.

12 "1996 National Five-Digit ZIP Code and Post Office Directory," *U.S. Postal Service*. Washington: 1996. Also available at http://www.usps.gov.

13 Brualdi, R.A., *Introductory Combinatorics*, North-Holland, New York, 1977.

14 Supra note 5 AHRQ.

15 L. Sweeney. *Inferences from unusual values in statistical data*. Carnegie Mellon Data Privacy Center Working Paper 3.

16 L. Sweeney. Weaving technology and policy together to maintain confidentiality. Journal of Law, Medicine and Ethics, 25(2-3):98-110, 1997.

17 Supra note 6 and note 11 U.S. Census.

18 Supra section 4.4.1 Age versus Year of Birth

19 Supra note 6 and note 11 U.S. Census.

20 L. Sweeney. *Towards all the data on all the people*. Formal publication forthcoming. Earlier version available as Carnegie Mellon Data Privacy Center Working Paper 2.

UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

January 19, 2018

MEMORANDUM FOR: Wilbur L. Ross, Jr.
Secretary of Commerce

Through: Karen Dunn Kelley
Performing the Non-Exclusive Functions and Duties of the Deputy Secretary

Ron S. Jarmin
Performing the Non-Exclusive Functions and Duties of the Director

Enrique Lamas
Performing the Non-Exclusive Functions and Duties of the Deputy Director

From: John M. Abowd
Chief Scientist and Associate Director for Research and Methodology

Subject: Technical Review of the Department of Justice Request to Add Citizenship Question to the 2020 Census

The Department of Justice has requested block-level citizen voting-age population estimates by OMB-approved race and ethnicity categories from the 2020 Census of Population and Housing. These estimates are currently provided in two related data products: the PL94-171 redistricting data, produced by April 1st of the year following a decennial census under the authority of 13 U.S.C. Section 141, and the Citizen Voting Age Population by Race and Ethnicity (CVAP) tables produced every February from the most recent five-year American Community Survey data. The PL94-171 data are released at the census block level. The CVAP data are released at the census block group level.

We consider three alternatives in response to the request: (A) no change in data collection, (B) adding a citizenship question to the 2020 Census, and (C) obtaining citizenship status from administrative records for the whole 2020 Census population.

We recommend either Alternative A or C. Alternative C best meets DoJ's stated uses, is comparatively far less costly than Alternative B, does not increase response burden, and does not harm the quality of the census count. Alternative A is not very costly and also does not harm the quality of the census count. Alternative B better addresses DoJ's stated uses than Alternative A. However, Alternative B is very costly, harms the quality of the census count, and would use substantially less accurate citizenship status data than are available from administrative sources.

Summary of Alternatives			
	Alternative A	***Alternative B***	***Alternative C***
Description	No change in data collection	Add citizenship question to the 2020 Census (i.e., the DoJ request), all 2020 Census microdata remain within the Census Bureau	Leave 2020 Census questionnaire as designed and add citizenship from administrative records, all 2020 Census microdata and any linked citizenship data remain within the Census Bureau
Impact on 2020 Census	None	Major potential quality and cost disruptions	None
Quality of Citizen Voting-Age Population Data	Status quo	Block-level data improved, but with serious quality issues remaining	Best option for block-level citizenship data, quality much improved
Other Advantages	Lowest cost alternative	Direct measure of self-reported citizenship for the whole population	Administrative citizenship records more accurate than self-reports, incremental cost is very likely to be less than $2M, USCIS data would permit record linkage for many more legal resident noncitizens
Shortcomings	Citizen voting-age population data remain the same or are improved by using small-area modeling methods	Citizenship status is misreported at a very high rate for noncitizens, citizenship status is missing at a high rate for citizens and noncitizens due to reduced self-response and increased item nonresponse, nonresponse followup costs increase by at least $27.5M, erroneous enumerations increase, whole-person census imputations increase	Citizenship variable integrated into 2020 Census microdata outside the production system, Memorandum of Understanding with United States Citizen and Immigration Services required to acquire most up-to-date naturalization data

Approved: _____ Date: _____

John M. Abowd, Chief Scientist
and Associate Director for Research and Methodology

Detailed Analysis of Alternatives

The statistics in this memorandum have been released by the Census Bureau Disclosure Review Board with approval number CBDRB-2018-CDAR-014.

Alternative A: Make no changes

Under this alternative, we would not change the current 2020 Census questionnaire nor the planned publications from the 2020 Census and the American Community Survey (ACS). Under this alternative, the PL94-171 redistricting data and the citizen voting-age population (CVAP) data would be released on the current schedule and with the current specifications. The redistricting and CVAP data are used by the Department of Justice to enforce the Voting Rights Act. They are also used by state redistricting offices to draw congressional and legislative districts that conform to constitutional equal-population and Voting Rights Act nondiscrimination requirements. Because the block-group-level CVAP tables have associated margins of error, their use in combination with the much more precise block-level census counts in the redistricting data requires sophisticated modeling. For these purposes, most analysts and the DoJ use statistical modeling methods to produce the block-level eligible voter data that become one of the inputs to their processes.

If the DoJ requests the assistance of Census Bureau statistical experts in developing model-based statistical methods to better facilitate the DoJ's uses of these data in performing its Voting Rights Act duties, a small team of Census Bureau experts similar in size and capabilities to the teams used to provide the Voting Rights Act Section 203 language determinations would be deployed.

We estimate that this alternative would have no impact on the quality of the 2020 Census because there would be no change to any of the parameters underling the Secretary's revised life-cycle cost estimates. The estimated cost is about $350,000 because that is approximately the cost of resources that would be used to do the modeling for the DoJ.

Alternative B: Add the question on citizenship to the 2020 Census questionnaire

Under this alternative, we would add the ACS question on citizenship to the 2020 Census questionnaire and ISR instrument. We would then produce the block-level citizen voting-age population by race and ethnicity tables during the 2020 Census publication phase.

Since the question is already asked on the American Community Survey, we would accept the cognitive research and questionnaire testing from the ACS instead of independently retesting the citizenship question. This means that the cost of preparing the new question would be minimal. We did not prepare an estimate of the impact of adding the citizenship question on the cost of reprogramming the Internet Self-Response (ISR) instrument, revising the Census Questionnaire Assistance (CQA), or redesigning the printed questionnaire because those components will not be finalized until after the March 2018 submission of the final questions. Adding the citizenship question is similar in scope and cost to recasting the race and ethnicity questions again, should that become necessary, and would be done at the same time. After the 2020 Census ISR, CQA and printed questionnaire are in final form, adding the citizenship question would be much more expensive and would depend on exactly when the implementation decision was made during the production cycle.

For these reasons, we analyzed Alternative B in terms of its adverse impact on the rate of voluntary cooperation via self-response, the resulting increase in nonresponse followup (NRFU), and the consequent effects on the quality of the self-reported citizenship data. Three distinct analyses support the conclusion of an adverse impact on self-response and, as a result, on the accuracy and quality of the 2020 Census. We assess the costs of increased NRFU in light of the results of these analyses.

B.1. Quality of citizenship responses

We considered the quality of the citizenship responses on the ACS. In this analysis we estimated item nonresponse rates for the citizenship question on the ACS from 2013 through 2016. When item nonresponse occurs, the ACS edit and imputation modules are used to allocate an answer to replace the missing data item. This results in lower quality data because of the statistical errors in these allocation models. The analysis of the self-responses responses is done using ACS data from 2013-2016 because of operational changes in 2013, including the introduction of the ISR option and changes in the followup operations for mail-in questionnaires.

In the period from 2013 to 2016, item nonresponse rates for the citizenship question on the mail-in questionnaires for non-Hispanic whites (NHW) ranged from 6.0% to 6.3%, non-Hispanic blacks (NHB) ranged from 12.0% to 12.6%, and Hispanics ranged from 11.6 to 12.3%. In that same period, the ISR item nonresponse rates for citizenship were greater than those for mail-in questionnaires. In 2013, the item nonresponse rates for the citizenship variable on the ISR instrument were NHW: 6.2%, NHB: 12.3% and Hispanic: 13.0%. By 2016 the rates increased for NHB and especially Hispanics. They were NHW: 6.2%, NHB: 13.1%, and Hispanic: 15.5% (a 2.5 percentage point increase). Whether the response is by mail-in questionnaire or ISR instrument, item nonresponse rates for the citizenship question are much greater than the comparable rates for other demographic variables like sex, birthdate/age, and race/ethnicity (data not shown).

B.2. Self-response rate analyses

We directly compared the self-response rate in the 2000 Census for the short and long forms, separately for citizen and noncitizen households. In all cases, citizenship status of the individuals in the household was determined from administrative record sources, not from the response on the long form. A noncitizen household contains at least one noncitizen. Both citizen and noncitizen households have lower self-response rates on the long form compared to the short form; however, the decline in self-response for noncitizen households was 3.3 percentage points greater than the decline for citizen households. This analysis compared short and long form respondents, categories which were randomly assigned in the design of the 2000 Census.

We compared the self-response rates for the same household address on the 2010 Census and the 2010 American Community Survey, separately for citizen and noncitizen households. Again, all citizenship data were taken from administrative records, not the ACS, and noncitizen households contain at least one noncitizen resident. In this case, the randomization is over the selection of household addresses to receive the 2010 ACS. Because the ACS is an ongoing survey sampling fresh households each month, many of the residents of sampled households completed the 2010 ACS with the same reference address as they used for the 2010 Census. Once again, the self-response rates were lower in the ACS than in the 2010 Census for both citizen and noncitizen households. In this 2010 comparison, moreover, the decline in self-response was 5.1 percentage points greater for noncitizen households than for citizen households.

In both the 2000 and 2010 analyses, only the long-form or ACS questionnaire contained a citizenship question. Both the long form and the ACS questionnaires are more burdensome than the shortform. Survey methodologists consider burden to include both the direct time costs of responding and the indirect costs arising from nonresponse due to perceived sensitivity of the topic. There are, consequently, many explanations for the lower self-response rates among all household types on these longer questionnaires. However, the only difference between citizen and noncitizen households in our studies was the presence of at least one noncitizen in noncitizen households. It is therefore a reasonable inference that a question on citizenship would lead to some decline in overall self-response because it would make the 2020 Census modestly more burdensome in the direct sense, and potentially much more burdensome in the indirect sense that it would lead to a larger decline in self-response for noncitizen households.

B.3. Breakoff rate analysis

We examined the response breakoff paradata for the 2016 ACS. We looked at all breakoff screens on the ISR instrument, and specifically at the breakoffs that occurred on the screens with the citizenship and related questions like place of birth and year of entry to the U.S. Breakoff paradata isolate the point in answering the questionnaire where a respondent discontinues entering data—breaks off—rather than finishing. A breakoff is different from failure to self-respond. The respondent started the survey and was prepared to provide the data on the Internet Self-Response instrument, but changed his or her mind during the interview.

Hispanics and non-Hispanic non-whites (NHNW) have greater breakoff rates than non-Hispanic whites (NHW). In the 2016 ACS data, breakoffs were NHW: 9.5% of cases while NHNW: 14.1% and Hispanics: 17.6%. The paradata show the question on which the breakoff occurred. Only 0.04% of NHW broke off on the citizenship question, whereas NHNW broke off 0.27% and Hispanics broke off 0.36%. There are three related questions on immigrant status on the ACS: citizenship, place of birth, and year of entry to the United States. Considering all three questions Hispanics broke off on 1.6% of all ISR cases, NHNW: 1.2% and NHW: 0.5%. A breakoff on the ISR instrument can result in follow-up costs, imputation of missing data, or both. Because Hispanics and non-Hispanic non-whites breakoff much more often than non-Hispanic whites, especially on the citizenship-related questions, their survey response quality is differentially affected.

B.4. Cost analysis

Lower self-response rates would raise the cost of conducting the 2020 Census. We discuss those increased costs below. They also reduce the quality of the resulting data. Lower self-response rates degrade data quality because data obtained from NRFU have greater erroneous enumeration and whole-person imputation rates. An erroneous enumeration means a census person enumeration that should not have been counted for any of several reasons, such as, that the person (1) is a duplicate of a correct enumeration; (2) is inappropriate (e.g., the person died before Census Day); or (3) is enumerated in the wrong location for the relevant tabulation (https://www.census.gov/coverage_measurement/definitions/). A whole-person census imputation is a census microdata record for a person for which all characteristics are imputed.

Our analysis of the 2010 Census coverage errors (Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States, Memo G-01) contains the relevant data. That study found that when the 2010 Census obtained a valid self-response (219 million persons),

the correct enumeration rate was 97.3%, erroneous enumerations were 2.5%, and whole-person census imputations were 0.3%. All erroneous enumeration and whole-person imputation rates are much greater for responses collected in NRFU. The vast majority of NRFU responses to the 2010 Census (59 million persons) were collected in May. During that month, the rate of correct enumerations was only 90.2%, the rate of incorrect enumeration was 4.8%, and the rate of whole-person census imputations was 5.0%. June NRFU accounted for 15 million persons, of whom only 84.6% were correctly enumerated, with erroneous enumerations of 5.7%, and whole-person census imputations of 9.6%. (See Table 19 of 2010 Census Memorandum G-01. That table does not provide statistics for all NRFU cases in aggregate.)

One reason that the erroneous enumeration and whole-person imputation rates are so much greater during NRFU is that the data are much more likely to be collected from a proxy rather than a household member, and, when they do come from a household member, that person has less accurate information than self-responders. The correct enumeration rate for NRFU household member interviews is 93.4% (see Table 21 of 2010 Census Memorandum G-01), compared to 97.3% for non-NRFU households (see Table 19). The information for 21.0% of the persons whose data were collected during NRFU is based on proxy responses. For these 16 million persons, the correct enumeration rate is only 70.1%. Among proxy responses, erroneous enumerations are 6.7% and whole-person census imputations are 23.1% (see Table 21).

Using these data, we can develop a cautious estimate of the data quality consequences of adding the citizenship question. We assume that citizens are unaffected by the change and that an additional 5.1% of households with at least one noncitizen go into NRFU because they do not self-respond. We expect about 126 million occupied households in the 2020 Census. From the 2016 ACS, we estimate that 9.8% of all households contain at least one noncitizen. Combining these assumptions implies an additional 630,000 households in NRFU. If the NRFU data for those households have the same quality as the average NRFU data in the 2010 Census, then the result would be 139,000 fewer correct enumerations, of which 46,000 are additional erroneous enumerations and 93,000 are additional whole-person census imputations. This analysis assumes that, during the NRFU operations, a cooperative member of the household supplies data 79.0% of the time and 21.0% receive proxy responses. If all of these new NRFU cases go to proxy responses instead, the result would be 432,000 fewer correct enumerations, of which 67,000 are erroneous enumerations and 365,000 are whole-person census imputations.

For Alternative B, our estimate of the incremental cost proceeds as follows. Using the analysis in the paragraph above, the estimated NRFU workload will increase by approximately 630,000 households, or approximately 0.5 percentage points. We currently estimate that for each percentage point increase in NRFU, the cost of the 2020 Census increases by approximately $55 million. Accordingly, the addition of a question on citizenship could increase the cost of the 2020 Census by at least $27.5 million. It is worth stressing that this cost estimate is a lower bound. Our estimate of $55 million for each percentage point increase in NRFU is based on an average of three visits per household. We expect that many more of these noncitizen households would receive six NRFU visits.

We believe that $27.5 million is a conservative estimate because the other evidence cited in this report suggests that the differences between citizen and noncitizen response rates and data quality will be amplified during the 2020 Census compared to historical levels. Hence, the decrease in self-response for citizen households in 2020 could be much greater than the 5.1 percentage points we observed during the 2010 Census.

Alternative C: Use administrative data on citizenship instead of add the question to the 2020 Census

Under this alternative, we would add the capability to link an accurate, edited citizenship variable from administrative records to the final 2020 Census microdata files. We would then produce block-level tables of citizen voting age population by race and ethnicity during the publication phase of the 2020 Census using the enhanced 2020 Census microdata.

The Census Bureau has conducted tests of its ability to link administrative data to supplement the decennial census and the ACS since the 1990s. Administrative record studies were performed for the 1990, 2000 and 2010 Censuses. We discuss some of the implications of the 2010 study below. We have used administrative data extensively in the production of the economic censuses for decades. Administrative business data from multiple sources are a key component of the production Business Register, which provides the frames for the economic censuses, annual, quarterly, and monthly business surveys. Administrative business data are also directly tabulated in many of our products.

In support of the 2020 Census, we moved the administrative data linking facility for households and individuals from research to production. This means that the ability to integrate administrative data at the record level is already part of the 2020 Census production environment. In addition, we began regularly ingesting and loading administrative data from the Social Security Administration, Internal Revenue Service and other federal and state sources into the 2020 Census data systems. In assessing the expected quality and cost of Alternative C, we assume the availability of these record linkage systems and the associated administrative data during the 2020 Census production cycle.

C.1. Quality of administrate record versus self-report citizenship status

We performed a detailed study of the responses to the citizenship question compared to the administrative record citizenship variable for the 2000 Census, 2010 ACS and 2016 ACS. These analyses confirm that the vast majority of citizens, as determined by reliable federal administrative records that require proof of citizenship, correctly report their status when asked a survey question. These analyses also demonstrate that when the administrative record source indicates an individual is not a citizen, the self-report is "citizen" for no less than 23.8% of the cases, and often more than 30%.

For all of these analyses, we linked the Census Bureau's enhanced version of the SSA Numident data using the production individual record linkage system to append an administrative citizenship variable to the relevant census and ACS microdata. The Numident data contain information on every person who has ever been issued a Social Security Number or an Individual Taxpayer Identification Number. Since 1972, SSA has required proof of citizenship or legal resident alien status from applicants. We use this verified citizenship status as our administrative citizenship variable. Because noncitizens must interact with SSA if they become naturalized citizens, these data reflect current citizenship status albeit with a lag for some noncitizens.

For our analysis of the 2000 Census long-form data, we linked the 2002 version of the Census Numident data, which is the version closest to the April 1, 2000 Census date. For 92.3% of the 2000 Census long-form respondents, we successfully linked the administrative citizenship variable. The 7.7% of persons for whom the administrative data are missing is comparable to the item non-response for self-responders in the mail-in pre-ISR-option ACS. When the administrative data indicated that the 2000 Census respondent was a citizen, the self-response was citizen: 98.8%. For this same group, the long-form response was

noncitizen: 0.9% and missing: 0.3%. By contrast, when the administrative data indicated that the respondent was not a citizen, the self-report was citizen: 29.9%, noncitizen: 66.4%, and missing: 3.7%.

In the same analysis of 2000 Census data, we consider three categories of individuals: the reference person (the individual who completed the census form for the household), relatives of the reference person, and individuals unrelated to the reference person. When the administrative data show that the individual is a citizen, the reference person, relatives of the reference person, and nonrelatives of the reference person have self-reported citizenship status of 98.7%, 98.9% and 97.2%, respectively. On the other hand, when the administrative data report that the individual was a noncitizen, the long-form response was citizen for 32.9% of the reference persons; that is, reference persons who are not citizens according to the administrative data self-report that they are not citizens in only 63.3% of the long-form responses. When they are reporting for a relative who is not a citizen according to the administrative data, reference persons list that individual as a citizen in 28.6% of the long-form responses. When they are reporting for a nonrelative who is not a citizen according to the administrative data, reference persons list that individual as a citizen in 20.4% of the long-form responses.

We analyzed the 2010 and 2016 ACS citizenship responses using the same methodology. The 2010 ACS respondents were linked to the 2010 version of the Census Numident. The 2016 ACS respondents were linked to the 2016 Census Numident. In 2010, 8.5% of the respondents could not be linked, or had missing citizenship status on the administrative data. In 2016, 10.9% could not be linked or had missing administrative data. We reached the same conclusions using 2010 and 2016 ACS data with the following exceptions. When the administrative data report that the individual is a citizen, the self-response is citizen on 96.9% of the 2010 ACS questionnaires and 93.8% of the 2016 questionnaires. These lower self-reported citizenship rates are due to missing responses on the ACS, not misclassification. As we noted above, the item nonresponse rate for the citizenship question has been increasing. These item nonresponse data show that some citizens are not reporting their status on the ACS at all. In 2010 and 2016, individuals for whom the administrative data indicate noncitizen respond citizen in 32.7% and 34.7% of the ACS questionnaires, respectively. The rates of missing ACS citizenship response are also greater for individuals who are noncitizens in the administrative data (2010: 4.1%, 2016: 7.7%). The analysis of reference persons, relatives, and nonrelatives is qualitatively identical to the 2000 Census analysis.

In all three analyses, the results for racial and ethnic groups and for voting age individuals are similar to the results for the whole population with one important exception. If the administrative data indicate that the person is a citizen, the self-report is citizen at a very high rate with the remainder being predominately missing self-reports for all groups. If the administrative data indicate noncitizen, the self-report is citizen at a very high rate (never less than 23.8% for any racial, ethnic or voting age group in any year we studied). The exception is the missing data rate for Hispanics, who are missing administrative data about twice as often as non-Hispanic blacks and three times as often as non-Hispanic whites.

C.2. Analysis of coverage differences between administrative and survey citizenship data

Our analysis suggests that the ACS and 2000 long form survey data have more complete coverage of citizenship than administrative record data, but the relative advantage of the survey data is diminishing. Citizenship status is missing for 10.9 percent of persons in the 2016 administrative records, and it is missing for 6.3 percent of persons in the 2016 ACS. This 4.6 percentage point gap between administrative and survey missing data rates is smaller than the gap in 2000 (6.9 percentage points) and 2010 (5.6

percentage points). Incomplete (through November) pre-production ACS data indicate that citizenship item nonresponse has again increased in 2017.

There is an important caveat to the conclusion that survey-based citizenship data are more complete than administrative records, albeit less so now than in 2000. The methods used to adjust the ACS weights for survey nonresponse and to allocate citizenship status for item nonresponse assume that the predicted answers of the sampled non-respondents are statistically the same as those of respondents. Our analysis casts serious doubt on this assumption, suggesting that those who do not respond to either the entire ACS or the citizenship question on the ACS are not statistically similar to those who do; in particular, their responses to the citizenship question would not be well-predicted by the answers of those who did respond.

The consequences of missing citizenship data in the administrative records are asymmetric. In the Census Numident, citizenship data may be missing for older citizens who obtained SSNs before the 1972 requirement to verify citizenship, naturalized citizens who have not confirmed their naturalization to SSA, and noncitizens who do not have an SSN or ITIN. All three of these shortcomings are addressed by adding data from the United States Citizen and Immigration Services (USCIS). Those data would complement the Census Numident data for older citizens and update those data for naturalized citizens. A less obvious, but equally important benefit, is that they would permit record linkage for legal resident aliens by allowing the construction of a supplementary record linkage master list for such people, who are only in scope for the Numident if they apply for and receive an SSN or ITIN. Consequently, the administrative records citizenship data would most likely have both more accurate citizen status and fewer missing individuals than would be the case for any survey-based collection method. Finally, having two sources of administrative citizenship data permits a detailed verification of the accuracy of those sources as well.

C.3. Cost of administrative record data production

For Alternative C, we estimate that the incremental cost, except for new MOUs, is $450,000. This cost estimate includes the time to develop an MOU with USCIS, estimated ingestion and curation costs for USCIS data, incremental costs of other administrative data already in use in the 2020 Census but for which continued acquisition is now a requirement, and staff time to do the required statistical work for integration of the administrative-data citizenship status onto the 2020 Census microdata. This cost estimate is necessarily incomplete because we have not had adequate time to develop a draft MOU with USCIS, which is a requirement for getting a firm delivery cost estimate from the agency. Acquisition costs for other administrative data acquired or proposed for the 2020 Census varied from zero to $1.5M. Thus the realistic range of cost estimates, including the cost of USCIS data, is between $500,000 and $2.0M

Questions on the Jan 19 Draft Census Memo on the DoJ Citizenship Question Reinstatement Request

1. **With respect to Alternatives B and C, what is the difference, if any, between the time when the data collected under each alternative would be available to the public?**

   Since the collection of this data, whether from administrative records or from an enumerated question, occurs prior to the creation of the Microdata Detail File (MDF) from which all tabulations will be performed, there is no difference in the timing of when the data collected under either alternative B or C could be made available to the public. The exact date for completion of the MDF is still being determined as the 2020 Census schedule is matured. However, the 2020 Census is working towards publishing the first post-apportionment tabulation data products as early as the first week of February 2021.

2. **What is the "2020 Census publication phase" (page 1 of the Detailed Analysis for Alternative B) versus Alternative C? Would there be any difference?**

   The 2020 Census publication phase is a broad window stretching from the release of the apportionment counts by December 31, 2020 through the last data product or report published in FY 2023, the final year of decennial funding for the 2020 Census. However, as stated in the answer to question 1, these data could be made available to the public on the same schedule as any other post-apportionment tabulated data product regardless of whether alternative B or C is used in its collection.

3. **What is the non-response rate for: (A) each question on the 2000 and 2010 Decennial Census short form and (B) each question on the 2010 ACS and most recent ACS?**

   The table below shows the item non-response (INR) rate for each question on the 2000 and 2010 Decennial Census short form. This is the percentage of respondents who did not provide an answer to an item.

   Item Nonresponse Rates for 2000 and 2010 Short Form Person Questions

	Relationship	Sex	Age	Hispanic Origin	Race	Tenure
2010	1.5	1.5	3.5	3.9	3.3	4.5
2000	1.3	1.1	3.7	3.1	2.9	4.1

   Source: Rothhaas, Lestina and Hill (2012) Tables

   Notes and Soucre:
   Rothhaas, C., Lestina, F. and Hill, J. (2012) "2010 Decennial Census Item Nonresponse and Imputation Assessment Report" 2010 Census Program for Evaluations and Experiments, January 24, 2012.

From report:

The INR rate is essentially the proportion of missing responses before pre-editing or imputation procedures for a given item (i.e., the respondent did not provide an answer to the item). For INR, missing values are included in the rates, but inconsistent responses (i.e., incompatible with other responses) are considered non-missing responses.

Online link to 2010 report that has 2000 information as well.
https://www.census.gov/2010census/pdf/2010_Census_INR_Imputation_Assessment.pdf

See attached spreadsheet for the item allocation rates by questions for the ACS for 2010, 2013, and 2016.

4. **What was the total survey response rate (i.e., percentage of complete questionnaires) for the 2000 long form and the 2000 short form? Of the incomplete long forms, what percentage left the citizenship question blank? Of the completed long forms, what percentage (if known) contained incorrect responses to the citizenship question?**

We do not have measures of total survey response rates from the 2000 long form and 2000 short form available at this time. The mail response rate in 2000 was 66.4 percent for short forms and 53.9 percent for long forms. No analysis that we were aware of was conducted on the incomplete long forms that left the citizenship question blank. The Census 2000 Content Reinterview Survey showed low inconsistency of the responses to the citizenship question. Only 1.8 percent of the respondents changed answers in the reinterview.

Source for 2000 mail response rates:
https://www.census.gov/pred/www/rpts/A.7.a.pdf

Source for 2000 Content Reinterview Survey. Page 32 source.
https://www.census.gov/pred/www/rpts/B.5FR_RI.PDF

5. **For the 2000 long and short forms, what was the percentage unanswered (left blank) for each question (i.e., what percentage of the responses for each question (sex, race, ethnicity, income, citizenship, etc.) were left blank)?**

For the 2000 shortform, the table in question 3a provides the percentage unanswered for each question.

For the 2000 longform, Griffin, Love and Obenski (2003) summarized the Census 2000 longform responses. Allocation rates for individual items in Census 2000 were computed, but because of the magnitude of these data, summary allocation measures were derived.

These rates summarize completeness across all data items for occupied units (households) and are the ratio of all population and housing items that had values allocated to the total number of population and housing items required to have a response. These composite measures provide a summary picture of the completeness of all data. Fifty-four population items and 29 housing items are included in these summary measures. The analysis showed that 9.9 percent of the population question items and 12.5 percent of the housing unit question items required allocation. Allocation involves using statistical procedures, such as within-household or nearest neighbor matrices, to impute missing values.

https://ww2.amstat.org/sections/srms/Proceedings/y2003/Files/JSM2003-000596.pdf

6. **What was the incorrect response rate for the citizenship question that was asked on the Long Form during the 2000 Decennial Census? Does the response rate on the 2000 Long Form differ from the incorrect response rate on the citizenship question for the ACS?**

In the 2000 long form, 2.3 percent of persons have inconsistent answers, 89.4 percent have consistent answers, and 8.2 percent have missing citizenship data in the SSA Numident and/or the 2000 long form. Among persons with nonmissing citizenship data in the SSA Numident and/or the 2000 long form, 2.6 percent have inconsistent answers and 97.4 percent have consistent answers.

In the 2010 ACS, 3.1 percent of persons have inconsistent answers, 86.0 percent have consistent answers, and 10.8 percent have missing citizenship data in the SSA Numident and/or the 2010 ACS. Among persons with nonmissing citizenship data in the SSA Numident and/or the 2010 ACS, 3.6 percent have inconsistent answers and 96.4 percent have consistent answers.

In the 2016 ACS, 2.9 percent of persons have inconsistent answers, 81.2 percent have consistent answers, and 15.9 percent have missing citizenship data in the SSA Numident and/or the 2016 ACS. Among persons with nonmissing citizenship data in the SSA Numident and/or the 2016 ACS, 3.5 percent have inconsistent answers and 96.5 percent have consistent answers.

These ACS and 2000 Census long form rates are based on weighted data.

This shows that inconsistent response rates are higher in the 2010 and 2016 ACS than in the 2000 long form.

7. **What is the incorrect response rate on other Decennial or ACS questions for which Census has administrative records available (for example, age, sex or income)?**

Table 7a shows the agreement rates between the 2010 Census response and the SSA Numident for persons who could be linked and had nonmissing values, and Table 7b shows

the agreement rates between the 2010 ACS and the SSA Numident. Gender has low disagreement (0.4-0.5 percent), and white alone (0.9 percent), black alone (1.7-2 percent), and age (2.1 percent) also have low disagreement rates. Disagreement rates are greater for other races (e.g., 46.4-48.6 percent for American Indian or Alaska Native alone). Hispanic origin is not well measured in the Numident, because it contains a single race response, one of which is Hispanic.

Table 7a. Demographic Variable Agreement Rates Between the 2010 Census and the SSA Numident

2010 Census Response	Percent Agreement with SSA Numident
Hispanic	54.2
Not Hispanic	99.7
White Alone	99.1
Black Alone	98.3
American Indian or Alaska Native Alone	51.4
Asian Alone	84.3
Native Hawaiian or Other Pacific Islander Alone	74.4
Some Other Race Alone	17.7
Age	97.9
Gender	99.4

Source: Rastogi, Sonya, and Amy O'Hara, 2012, "2010 Census Match Study," 2010 Census Planning Memoranda Series No. 247.

Table 7b. Demographic Variable Agreement Rates Between the 2010 Census and the SSA Numident

2010 ACS Response	Percent Agreement with SSA Numident
White Alone	99.1
Black Alone	98.0
American Indian or Alaska Native Alone	53.6
Asian Alone	82.9
Native Hawaiian or Other Pacific Islander Alone	72.9
Some Other Race Alone	17.2
Age 0-2 Date of Birth	95.2
Age 3-17 Date of Birth	95.6
Age 18-24 Date of Birth	95.2
Age 25-44 Date of Birth	95.8
Age 45-64 Date of Birth	95.9
Age 65-74 Date of Birth	96.5
Age 75 and older Date of Birth	92.7
Male	99.5
Female	99.5

Source: Bhaskar, Renuka, Adela Luque, Sonya Rastogi, and James Noon, 2014, "Coverage and Agreement of Administrative Records and 2010 American Community Survey Demographic Data," CARRA Working Paper #2014-14.

Abowd and Stinson (2013) find correlations of 0.75-0.89 between Survey of Income and Program Participation (SIPP) and SSA Detailed Earnings Record annual earnings between 1990-1999.[1]

8. How does the Census presently handle responses on the (A) Decennial Census and (B) the ACS when administrative records available to the Census confirm that the response on the Decennial Census or ACS is incorrect? Is the present Census approach to incorrect responses based on practice/policy or law (statute or regulation)?

We have always based the short form Decennial Census and the ACS on self-response, and while we have procedures in place to address duplicate or fraudulent responses, we do not check the accuracy of the answers provided to the specific questions on the Census questionnaire. This is a long established practice at the Census Bureau that has been thoroughly tested and in place since 1970, when the Census Bureau moved to a mail-out/respond approach to the Decennial Census. Title 13 of the U.S. Code allows the Census Bureau to use alternative data sources, like administrative records, for a variety of purposes, and we are using data in new ways in the 2020 Census. While this includes the use of administrative records data to fill in areas where a respondent does not provide an answer, we have not explored the possibility of checking or changing responses that a responding household has provided in response to the questionnaire.

9. Please explain the differences between the self-response rate analysis and the breakoff rate analysis. The range of breakoff rates between groups was far smaller than the range of self-response rates between groups.

Self-response means that a household responded to the survey by mailing back a questionnaire or by internet, and a sufficient number of core questions were answered so that an additional field interview was not required.

A breakoff occurs when an internet respondent stops answering questions prior to the end of the questionnaire. In most cases the respondent answers the core questions before breaking off, and additional fieldwork is not required. The breakoff rates are calculated separately by which question screen was the last one reached before the respondent stopped answering altogether.

The share of Hispanic respondents who broke off at some point before the end of the questionnaire (17.6 percent) is much higher than for non-Hispanic whites (9.5 percent).

[1] Abowd, John M., and Martha H. Stinson, 2013, "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data," Review of Economics and Statistics, Vol. 95(55), pp. 1451-1467.

Spreading the overall breakoff rates over 134 screens in the questionnaire works out to quite small rates per screen. It works out to an average breakoff rate of 0.131 percent per screen for Hispanics and 0.066 percent for non-Hispanic whites.

10. The NRFU numbers are comparatively small – approximately one additional household for NRFU per Census enumerator. Is this really a significant source of concern?

Yes, this is a significant concern. First, it gives rise to incremental NRFU cost of at least $27.5 million. This is a lower bound becaues it assumes the households that do not self-respond because we added a question on citizenship have the same follow-up costs as an average U.S. household. They won't because these households overwhelmingly contain at least one noncitzen, and that is one of our acknowledged hard-to-count subpopulations.

11. Given that the breakoff rate difference was approximately 1 percent, why did Census choose to use the 5.1 percent number for assessing the cost of Alternative B?

If a household breaks off an internet response at the citizenship, place of birth, or year of entry screens, this means it would have already responded to the core questions. This would not trigger follow-up fieldwork and thus would not involve additional fieldwork costs. In contrast, if a household does not mail back a questionnaire or give an internet response, fieldwork will be necessary and additional costs will be incurred. Thus, the 5.1 percent number for differential self-response is more appropriate for estimating the additional fieldwork cost of adding a citizenship question.

12. Alternative C states that Census would use administrative data from the Social Security Administration, Internal Revenue Service, and "other federal and state sources." What are the other sources?

In addition to continuing the acquisition of the Social Security Administration and Internal Revenue Service data, the Census Bureau is in discussion with the U.S. Citizen and Immigration Services (USCIS) staff to acquire additional citizenship data.

13. Is Census confident that administrative data will be able to be used to determine citizenship for all persons (e.g., not all citizens have social security numbers)?

We are confident that Alternative C is viable and that we have already ingested enough high-quality citizenship administrative data from SSA and IRS. The USCIS data are not required. They would, however, make the citizenship voting age tabulations better, but the administrative data we've got are very good and better than the data from the 2000 Census and current ACS. The type of activities required for Alternative C already occur daily and routinely at the Census Bureau. We have been doing this for business data products,

including the Economic Censuses, for decades. We designed the 2020 Census to use this technology too.

14. For Alternative C, the memo says, "we assume the availability of these record linkage systems and associated administrative data" – does Census already have in place access to this data or would this need to be negotiated? If negotiated, for which data sets specifically?

The Census Bureau has longstanding contractual relationships with the Social Security Administration and the Internal Revenue Service that authorize the use of data for this project. For new data acquired for this project (i.e., USCIS) we would estimate a six-month development period to put a data acquisition agreement in place. That agreement would also include terms specifying the authorized use of data for this project.

15. Are there any privacy issues / sensitive information prohibitions that might prevent other agencies from providing such data?

There are no new privacy or sensitivity issues associated with other agencies providing citizenship data. We have received such information in the past from USCIS. We are currently authorized to receive and use the data from SSA and IRS that are discussed in Alternative C.

16. How long would Census expect any negotiation for access to data take? How likely is it that negotiations would be successful? Are MOA's needed/required?

Current data available to the Census Bureau provide the quality and authority to use that are required to support this project. Additional information potentially available from USCIS would serve to supplement/validate those existing data. We are in early discussions with USCIS to develop a data acquisition agreement and at this time have no indications that this acquisition would not be successful.

17. What limitations would exist in working with other agencies like IRS, Homeland Security, etc. to share data?

The context for sharing of data for this project is for a one-way sharing of data from these agencies to the Census Bureau. Secure file transfer protocols are in-place to ingest these data into our Title 13 protected systems. For those data already in-place at the Census Bureau to support this project, provisions for sharing included in the interagency agreement restrict the Census Bureau from sharing person-level microdata outside the Census Bureau's Title 13 protections. Aggregates that have been processed through the Bureau's disclosure avoidance procedures can be released for public use.

18. If Alternative C is selected, what is Census's backup plan if the administrative data cannot be completely collected and utilized as proposed?

The backup plan is to use all of the administrative data that we currently have, which is the same set that the analyses of Alternative C used. We have verified that this use is consistent with the existing MOUs. We would then use estimation and modeling techniques similar to those used for the Small Area Income and Poverty Estimates (SAIPE) to impute missing citizenship status for those persons for whom we do not have administrative records. These models would also include estimates of naturalizations that occurred since the administrative data were ingested.

19. Does Census have any reason to believe that access to existing data sets would be curtailed if Alternative C is pursued?

No we do not believe that any access to existing data sets would be curtailed if we pursue Alternative C.

20. Has the proposed Alternative C approach ever been tried before on other data collection projects, or is this an experimental approach? If this has been done before, what was the result and what were lessons learned?

The approach in Alternative C has been routinely used in processing the economic censuses for several decades. The Bureau's Business Register was specifically redesigned for the 2002 Economic Census in order to enhance the ingestion and use of administrative records from the IRS and other sources. The data in these administrative records are used to substitute for direct responses in the economic censuses for the unsampled entities. They are also used as part of the review, edit, and imputation systems for economic censuses and surveys. On the household side, the approach in Alternative C was used extensively to build the residential characteristics for OnTheMap and OnTheMap for Emergency Management.

21. Is using sample data and administrative records sufficient for DOJ's request?

The 2020 Census data combined with Alternative C are sufficient to meet DoJ's request. We do not anticipate using any ACS data under Alternative C.

22. Under Alternative C, If Census is able to secure interagency agreements to provide needed data sets, do we know how long it would take to receive the data transmission from other agencies and the length of time to integrate all that data, or is that unknown?

With the exception of the USCIS data, the data used for this project are already integrated into the 2020 Census production schema. In mid-to late 2018, we plan to acquire the USCIS data and with those data and our existing data begin to develop models and business rules to select citizenship status from the composite of sources and attach that characteristic to

each U.S. person. We expect the development and refinement of this process to continue into 2019 and to be completed by third quarter calendar year 2019.

23. Cross referencing Census decennial responses with numerous governmental data sets stored in various databases with differing formats and storage qualities sounds like it could be complicated. Does Census have an algorithm in place to efficiently combine and cross reference such large quantities of data coming from many different sources? What cost is associated with Alternative C, and what technology/plan does Census have in place to execute?

Yes, the 2018 Census End-to-End test will be implementing processing steps to be able to match Census responses to administrative record information from numerous governmental data sets. The Census Bureau has in place the Person Identification Validation System to assign Protected Identification Keys to 2020 Census responses. The required technology for linking in the administrative records is therefore part of the 2020 Census technology. This incremental cost factored into the estimate for Alternative C is for integrating the citizenship variable specifically, since that variable is not currently part of the 2020 Census design. No changes are required to the production Person Identification Validation system to integrate the administrative citizenship data.

24. For section C-1 of the memo, when did Census do the analyses of the incorrect response rates for non-citizen answers to the long form and ACS citizenship question? Were any of the analyses published?

The comparisons of ACS, 2000 Decennial Census longform and SSA Numident citizenship were conducted in January 2018. This analysis has not been published.

25. Has Census corrected the incorrect responses it found when examining non-citizen responses? If not, why not?

In the American Community Survey (ACS), and the short form Decennial Census, we do not change self-reported answers. The Decennial Census and the ACS are based on self-response and we accept the responses provided by households as they are given. While we have procedures in place to address duplicate or fraudulent responses, we do not check the accuracy of the answers provided to the specific questions on the Census questionnaires. This is a long established process at the Census Bureau that has been thoroughly tested and in place since 1970, when the Census Bureau moved to a mail-out/respond approach to the Decennial Census.

26. Has the Department of Justice ever been made aware of inaccurate reporting of ACS data on citizenship, so that they may take this into consideration when using the data?

Not exactly. The Census Bureau is in close, regular contact with the Department of Justice (DOJ) regarding their data requirements. Our counterparts at DOJ have a solid understanding of survey methodology and the quality of survey data, and they are aware of the public documentation on sampling and accuracy surrounding the ACS. However, the specific rate of accuracy regarding responses to the ACS question on citizenship has never been discussed.

27. Why has the number of persons who cannot be linked increased from 2010 to 2016?

The linkage between the ACS and administrative data from the SSA Numident and IRS ITIN tax filings depends on two factors: (a) the quality of the personally identifiable information (PII) on the ACS response and (b) whether the ACS respondent is in the SSN/ITIN universe.

With respect to the quality of the PII on the ACS, there may be insufficient information on the ACS due to item nonresponse or proxy response for the person to allow a successful match using the production record linkage system. There may also be more than one record in the Numident or ITIN IRS tax filings that matches the person's PII. Finally, there may be a discrepancy between the PII provided to the ACS and the PII in the administrative records.

Alternatively, the person may not be in the Numident or ITIN IRS tax filing databases because they are out of the universe for those administrative systems. This happens when the person is a citizen without an SSN, or when the person is a noncitizen who has not obtained an SSN or ITIN.

Very few of the unlinked cases are due to insufficient PII in the ACS or multiple matches with administrative records. The vast majority of unlinked ACS persons have sufficient PII, but fail to match any administrative records sufficiently closely. This means that most of the nonmatches are because the ACS respondent is not in the administrative record universe.

The incidence of ACS persons with sufficient PII but no match with administrative records increased between 2010 and 2016. One contributing factor is that the number of persons linked to ITIN IRS tax filings in 2016 was only 39 percent as large as in 2010, suggesting that either fewer of the noncitizens in the 2016 ACS had ITINs, or more of them provided PII in the ACS that was inconsistent with their PII in IRS records.

28. Independent of this memo, what action does Census plan to take in response to the analyses showing that non-citizens have been incorrectly responding to the citizenship question?

The Census Bureau does not have plans to make any changes to procedures in the ACS. However, we will continue to conduct thorough evaluations and review of census and survey data. The ACS is focusing our research on the potential use of administrative records

in the survey. For instance, we are exploring whether we can use IRS data on income to reduce the burden of asking questions on income on the ACS. We are concentrating initially on questions that are high burden, e.g., questions that are difficult to answer or questions that are seen as intrusive.

29. Did Census make recommendations the last time a question was added?

Since the short form Decennial Census was established in 2010, the only requests for new questions we have received have been for the ACS. And, in fact, requests for questions prior to 2010 were usually related to the Decennial Census Long Form. We always work collaboratively with Federal agencies that request a new question or a change to a question. The first step is to review the data needs and the legal justification for the new question or requested changes. If, through this process, we determine that the request is justified, we work with the other agencies to test the question (cognitive testing and field testing). We also work collaboratively on the analysis of the results from the test which inform the final recommendation about whether or not to make changes or add the question.

30. Does not answering truthfully have a separate data standard than not participating at all?

We're not sure what you're asking here. Please clarify the question.

31. What was the process that was used in the past to get questions added to the decennial Census or do we have something similar where a precedent was established?

Because no new questions have been added to the Decennial Census (for nearly 20 years), the Census Bureau did not feed bound by past precedent when considering the Department of Justices' request. Rather, the Census Bureau is working with all relevant stakeholders to ensure that legal and regulatory requirements are filled and that questions will produce quality, useful information for the nation. As you are aware, that process is ongoing at your direction.

32. Has another agency ever requested that a question be asked of the entire population in order to get block or individual level data?

Not to our knowledge. However, it is worth pointing out that prior to 1980 the short form of the Decennial Census included more than just the 10 questions that have been on the short form since 1990.

33. Would Census linking of its internal data sets, with other data sets from places like IRS and Homeland Security, have an impact on participation as well (i.e., privacy concerns)?

The potential that concerns about the use of administrative records could have an impact on participation has always been a concern of ours, and it's a risk that we're managing on our risk register. We've worked closely with the privacy community throughout the decade, and we established a working group on our National Advisory Committee to explore this issue. We've also regularly briefed the Congress about our plans. At this stage in the decade there does not appear to be extensive concerns among the general public about our approach to using administrative records in the Nonresponse Operation or otherwise. We will continue to monitor this issue.

34. Would Alternative C require any legislation? If so, what is the estimated time frame for approval of such legislation?

No.

35. Census publications and old decennial surveys available on the Census website show that citizenship questions were frequently asked of the entire population in the past. Citizenship is also a question on the ACS. What was the justification provided for citizenship questions on the (A) short form, (B) long form, and (C) ACS?

In 1940, the Census Bureau introduced the use of a short form to collect basic characteristics from all respondents, and a long form to collect more detailed questions from only a sample of respondents. Prior to 1940, census questions were asked of everyone, though in some cases only for those with certain characteristics. For example, in 1870, a citizenship question was asked, but only for respondents who were male and over the age of 21.

Beginning in 2005, all the long-form questions – including a question on citizenship -- were moved to the ACS. 2010 was the first time we conducted a short-form only census. The citizenship question is included in the ACS to fulfill the data requirements of the Department of Justice, as well as many other agencies including the Equal Employment Opportunities Commission, the Department of Health and Human Services, and the Social Security Administration.

ACS Item Allocation Rates for United States: 2016, 2013, 2010

Title	2016	2013	2010
Overall housing allocation rate occupied and vacant housing units	4.9	5.6	5.2
Overall person allocation rate total population	9.5	8.4	5.8
Vacancy status vacant housing units	3.9	3.5	2.9
Tenure occupied housing units	1.2	1.3	1.2
Units in structure occupied and vacant housing units	1.5	1.5	1.5
Year moved in occupied housing units	3	3	3.4
Month moved in occupied housing units into which households move in the last two years	0.7	0.7	0.7
Year built occupied and vacant housing units	18.2	17.1	16.2
Lot size occupied and vacant single family and mobile homes	3.9	3.9	4.2
Agricultural sales occupied and vacant single family and mobile homes with lot size greater than or equal to 1 acre	4	4.2	4.4
Business on property occupied and vacant single family and mobile homes	**	2.4	3
Number of rooms occupied and vacant housing units	5	5.5	5.2
Number of bedrooms occupied and vacant housing units	5.5	4.6	4.3
Running water occupied and vacant housing units	2.4	2.1	2
Flush toilet occupied and vacant housing units	**	2.2	2
Bathtub or shower occupied and vacant housing units	2.6	2.2	2
Sink with a faucet occupied and vacant housing units	2.6	2.2	2
Stove or range occupied and vacant housing units	3.1	2.8	2.5
Refrigerator occupied and vacant housing units	3.2	2.9	2.7
Telephone occupied housing units	1.5	1.2	1.1
Number of vehicles occupied housing units	1.2	1.4	1.3
Heating fuel occupied housing units	3.4	3.4	3.3
Monthly electricity cost occupied housing units	8.1	8.2	7.3
Monthly gas cost occupied housing units	9.6	9.9	9.8

Yearly water and sewer cost occupied housing units	8.5	8.8	8.1
Yearly other fuel cost occupied housing units	7.3	8.3	10.6
Yearly food stamp recipiency - household occupied housing units	1.7	1.7	1.3
Yearly real estate taxes owner-occupied housing units	16.7	18.5	16.3
Yearly property insurance owner-occupied housing units	23.9	25.6	23.2
Mortgage status owner-occupied housing units	2.2	2.5	2.1
Monthly mortgage payment owner-occupied housing units with a mortgage	10.5	12.4	10.7
Mortgage payment incl. real estate taxes owner-occupied housing units with a mortgage	6.2	6.9	(X)
Mortgage payment incl. insurance owner-occupied housing units with a mortgage	6.8	7.4	(X)
Second mortgage owner-occupied housing units	3.2	3.7	3.4
Home equity loan owner-occupied housing units	3.7	4.3	4.2
Other monthly mortgage payment(s) owner-occupied housing units with second mortgage or home equity loan	23.3	21.7	17.9
Property value owner-occupied housing units and vacant housing units for sale	11.6	12.9	12.3
Yearly mobile home costs occupied mobile homes and other units	21.7	21.5	19.9
Monthly condominium fee owner-occupied housing units	0.8	0.8	0.7
Monthly rent occupied housing units rented for cash rent and vacant housing units for rent	10.5	9.8	9.3
Meals included in rent occupied housing units rented for cash rent and vacant housing units for rent	2.1	2.1	2
Desktop/laptop/notebook computer occupied housing units	1.3	3.2	**
Handheld computer/smart mobile phone occupied housing units	**	3.3	**
Tablet or other portable wireless computer occupied housing units	1.6	**	**
Smartphone occupied housing units	1.6	**	**
Other computer occupied housing units	1.7	3.7	**
Household has internet access occupied housing units	3.3	4.4	**
Dial-up internet service occupied housing units with internet access	3.8	5.7	**
DSL internet service occupied housing units with internet access	**	5.7	**

Cable modem internet service occupied housing units with internet access	**	5.7	**
Fiber-optic internet service occupied housing units with internet access	**	5.7	**
Cellular data plan (formerly mobile broadband) occupied housing units with internet access	7.6	26.7	**
Satellite internet service occupied housing units with internet access	3.8	5.7	**
High speed internet service occupied housing units with internet accesss	3.8	**	**
Some other internet service occupied housing units with internet access	3.8	5.7	**
Race total population	1.5	1.6	1.5
Hispanic origin total population	1.8	2.1	1.8
Sex total population	0.1	0.1	0.1
Age total population	1.7	1.6	1.3
Relationship total household population	1.2	1.1	1.2
Marital status total population 15 years and over	5.3	4.8	3
Married past 12 months total population 15 years and over, except those never married	6.9	6.6	4.7
Widowed past 12 months total population 15 years and over, except those never married	7.4	7	4.5
Divorced past 12 months total population 15 years and over, except those never married	7.4	7	4.5
Times married total population 15 years and over, except those never married	8.1	7.8	5.1
Year last married total population 15 years and over, except those never married	13.5	13.3	11.4
Place of birth total population	9.1	8.6	6.5
Citizenship total population	6	5.2	2.7
Year of naturalization total population naturalized citizens	22.5	22.5	16.6
Year of entry total population not born in US	14.8	13.2	10.3
Speaks another language at home total population 5 years and over	6.8	5.9	3.4
Language spoken total population 5 years and over who speak another language at home	8.3	7	5.7
English ability total population 5 years and over who speak another language at home	7.1	5.9	4
School enrollment total population 3 years and over	6.7	6	3.7
Grade level attending			

total population 3 years and over enrolled	10.2	8.9	6
Educational attainment total population 3 years and over	8.5	8	5.6
Field of degree total population 25 years and over with a bachelor's degree or higher	13.5	12.4	9.8
Mobility status total population 1 years and over	7.2	6.5	4
Migration state/foreign county total population 1 years and over movers	13.2	11.3	7.1
Migration county total population 1 years and over movers within US	14.6	12.5	8.3
Migration minor civil division total population 1 years and over movers within US	14.2	12.1	8.4
Migration place total population 1 years and over movers within US	15	12.9	8.8
Health insurance thru employer/union total population	10.7	9	6.2
Health insurance purchased directly total population	11.3	9.7	6.9
Health insurance through Medicare total population	9.5	8.1	5.2
Health insurance through Medicaid total population	12.2	10.5	7.9
Health insurance through TRICARE total population	12.5	10.8	8.1
Health insurance through VA total population	12.3	10.7	8.1
Health ins. thru Indian Health Service total population	12.8	11.1	8.5
Visual difficulty total population	7.1	6.1	3.4
Hearing difficulty total population	6.8	5.9	3.2
Physical difficulty total population 5 years and over	7.5	6.7	3.5
Difficulty remembering total population 5 years and over	7.5	6.7	3.5
Difficulty dressing total population 5 years and over	7.5	6.7	3.5
Difficulty going out total population 16 years and over	7.3	6.5	3.4
Grandchildren living in home noninstitutionalized population 30 years and over	1.1	1	0.9
Responsibility for grandchildren noninstitutionalized population 30 years and over who are grandparents with grandchildren in the home	17.7	15.7	12
Months responsible for grandchildren noninstitutionalized population 30 years and over who are grandparents with grandchildren in the home that have responsibility	17.2	16.1	14.9
Fertility status female total population 15-50	7.8	6.7	3.7

Veteran status total population 17 years and over	7.3	6.8	3.8
Periods of military service total population 17 years and over on active duty now or previously	9.7	9.3	6.3
Service-connected disability rating total population 17 years and over, except those who never served in the Armed Forces	6.8	6.6	3.9
Service-connected disability rating value total population 17 years and over with a service-connected disability	0.2	0.2	0.7
Employment status recode noninstitutionalized population 16 years and over	8.7	8.1	5.1
When last worked noninstitutionalized population 16 years and over	9.6	9.1	5.7
Weeks worked in the past 12 months noninstitutionalized population 16 years and over who worked in the past 12 months	10.6	9.7	6.9
Hours worked per week noninstitutionalized population 16 years and over who worked in the past 12 months	11.9	10.8	7.7
Place of work state/foreign county noninstitutionalized population 16 years and over at work last week	11.8	10.4	6.3
Place of work county noninstitutionalized population 16 years and over at work last week	12.5	11	7
Place of work minor civil division noninstitutionalized population 16 years and over at work last week	3.6	3.3	2.1
Place of work place noninstitutionalized population 16 years and over at work last week	13.1	11.6	7.6
Transportation to work noninstitutionalized population 16 years and over at work last week	9.6	8.8	5.7
Carpool size noninstitutionalized population 16 years and over at work last week who drive to work	10.9	9.9	6.8
Time of departure noninstitutionalized population 16 years and over at work last week who don't work at home	20.2	18.5	12.8
Commuting time noninstitutionalized population 16 years and over at work last week who don't work at home	14.5	13.3	9.7
Class of worker total population 16 years and over who worked in the last 5 years	11.7	10.7	7.2
Industry total population 16 years and over who worked in the last 5 years	12.7	11.4	7.8
Occupation total population 16 years and over who worked in the last 5 years	13.4	11.8	8.1
Wages/salary income total population 15 years and over	19.1	19	16
Self-employment income			

total population 15 years and over	10.5	9.3	5.9
Interest, dividends, etc. income total population 15 years and over	15.2	12.6	8.8
Social security or railroad retirement total population 15 years and over	14.5	12.3	8.9
Supplemental security income total population 15 years and over	12.7	10.3	6.7
Public assistance total population 15 years and over	13.2	10.5	6.8
Retirement income total population 15 years and over	13.6	11.1	7.5
Other income total population 15 years and over	13.2	10.8	7.4
Some or all income allocated total population 15 years and over	28.4	25.3	22.4

Source: ACS 1-year data. See following links for more information:

https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/item-allocation-rates/

https://www.census.gov/programs-surveys/acs/methodology/sample-size-and-data-quality/item-allocation-rates-definitions.html

Note:

** This item was not asked in this year.

Summary Analysis of the Key Differences Between Alternative C and Alternative D

This short note describes the Census Bureau's current assumptions about two alternatives to address the need for block level data on citizen voting age populations. The goal is to measure the citizenship status of all people enumerated in the 2020 Decennial Census. Both alternatives utilize administrative data on the citizenship status of individuals, however one option, Alternative D, proposes to also include the current American Community Survey (ACS) question on citizenship status on the 2020 Decennial Census short form.

In both alternatives described here, the methodology requires linking 2020 census response data and administrative records. However, as illustrated both alternatives would also need to assign/impute citizenship for a portion of the population. The Census Bureau will have to assign citizenship in cases of questionnaire non-response and item non-response. Additionally, it is important to note, that even when a self-response is available it is not always possible to link response data with administrative records data. Poor data quality (e.g., name and age) and nonresponse or incomplete 2020 Census responses mean that we will not have a direct measure of citizenship status for all residents enumerated in 2020. The Census Bureau will to need employ an imputation model for these cases.

One of the key differences between to the two alternatives described below is the number of cases requiring imputation. The other key difference is the impact of errors in the citizenship status reported on the 2020 Census.

In the most recent version of the 2020 Decennial Life Cycle Cost Estimate, the Census Bureau projects counting 330 million residents in 2020. Figure 1 summarizes how citizenship status will be measured under Alternative C that does not employ a citizenship question on the 2020 Census. Figure 2 summarizes how this will be done using both administrative records and a 2020 citizenship question under Alternative D.

Alternative C is a simplified process for assigning citizenship through direct linkage and modelling, without including the question on the 2020 Census. The Census Bureau will link the responses for the 330 million census records to administrative records that contain information on the citizenship status of individuals. The Census Bureau expects to successfully link and observe this status for approximately 295 million people. The Census Bureau would need to impute this status for approximately 35 million people under Alternative C whose 2020 responses cannot be linked to administrative data. Although the Census Bureau has fully developed and tested the imputation model, it has high confidence that an accurate model can be developed and deployed for this purpose. Further, we will most likely never possess a fully adequate truth deck to benchmark it to.

Measuring citizenship status is slightly more complex under Alternative D where all U.S. households will be given the opportunity to provide the citizenship status of each household member. Based on response data for the ACS citizenship and other response data research, we know that not all households that respond to the 2020 Census will answer this question, leaving the question blank or with otherwise invalid responses. Additionally, Alternative D, must also account for those households that do not respond at all or will have proxy responses. Due to these reasons, we estimate that we will get 2020 citizenship status responses for approximately 294.6 million people, a slightly higher estimate

than Alternative C. For the 35.4 million people without a 2020 citizenship response, the Census Bureau will employ the same methodology as in Alternative C, linking the 2020 Census responses to the administrative records. The Census Bureau estimates that it will be able to link these cases to administrative records where we observe citizenship status for approximately 21.5 million people. For the remaining 13.8 million will be imputed through a model as described above. Thus, there will be a need for imputing many cases across either alternative.

The Census Bureau will link the 294.6 million records from the 2020 Census with the administrative records. This will be done both for potential quality assurance purposes and to improve the quality of future modeling uses. Based on the current research from the ACS, the Census Bureau expects to successfully link approximately 272.5 million of these cases. Of these, 263 million will have citizenship statuses that agree across the 2020 response and administrative record. The Census Bureau estimates there will be 9.5 million cases where there is disagreement across the two sources. Historic Census Bureau practice is to use self-reported data in these situations. However, the Census Bureau now knows from linking ACS responses on citizenship to administrative data that nearly one third of noncitizens in the administrative data respond to the questionnaire indicating they are citizens, indicating that this practice should be revisited in the case of measuring citizenship. Finally, for those 22.2 million cases that do not link to administrative records (non-linkage occurs for the same data quality reasons discussed above), the Census Bureau will use the observed 2020 responses. Again, Census Bureau expect some quality issues with these responses. Namely, the Census Bureau estimates that just under 500 thousand noncitizens will respond as citizens.

The relative quality of Alternative C versus Alternative D will depend on the relative importance of the errors in administrative data, response data, and imputations. To be slightly more but not fully precise consider the following description of errors under both alternatives. First note that all possible measurement methods will have errors. Under Alternative C, there will be error in the administrative records, but we believe these to be relatively limited dues to the procedure following by SSA, USCIS and State. In both Alternative, the modeled cases will be subject to prediction error. Prediction error occur when the model returns the incorrect status of a case. As there are more models cases in Alternative C, prediction error will be a bigger issue there. Alternative D has an additional source or error, response error. This is where 2020 respondent give the incorrect status. Statisticians often hope these error are random and cancel out. However, we know from prior research that citizenship status responses are systematically biased for a subset of noncitizens. Response error is only an issue in alternative D. Unfortunately, the Census Bureau cannot quantify the relative magnitude of the errors across the alternatives at this time.

Figure 1

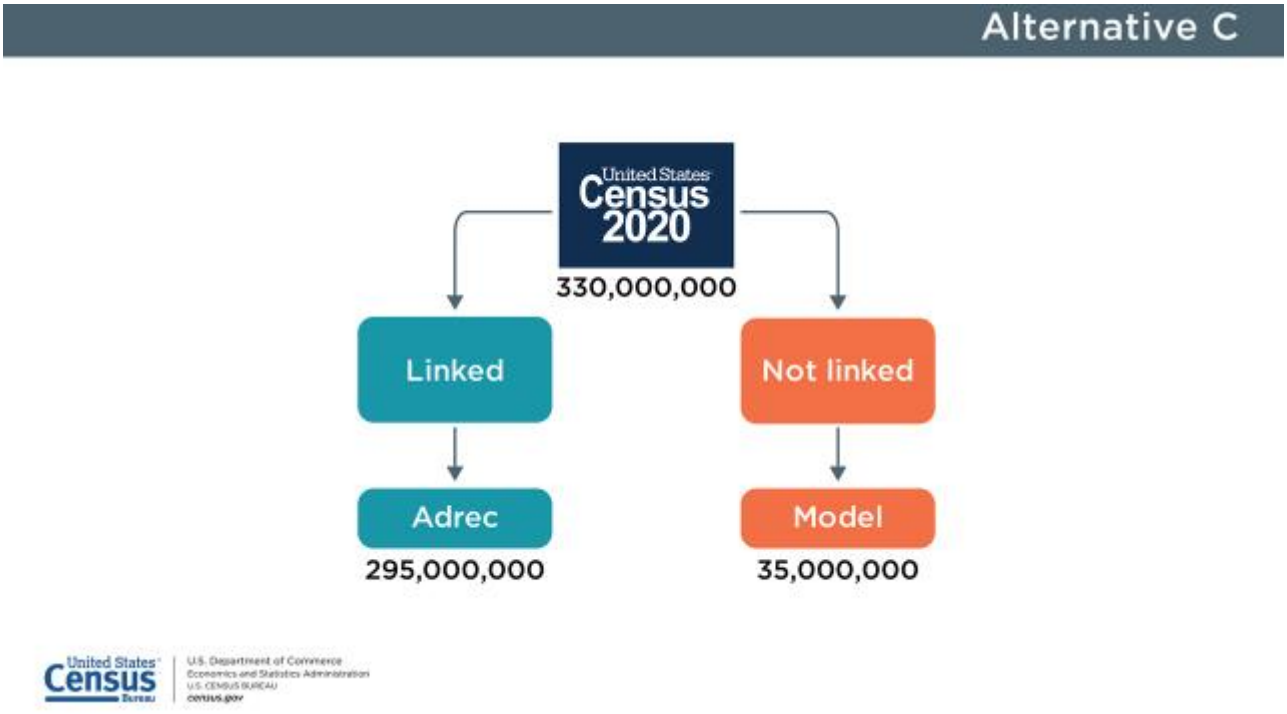
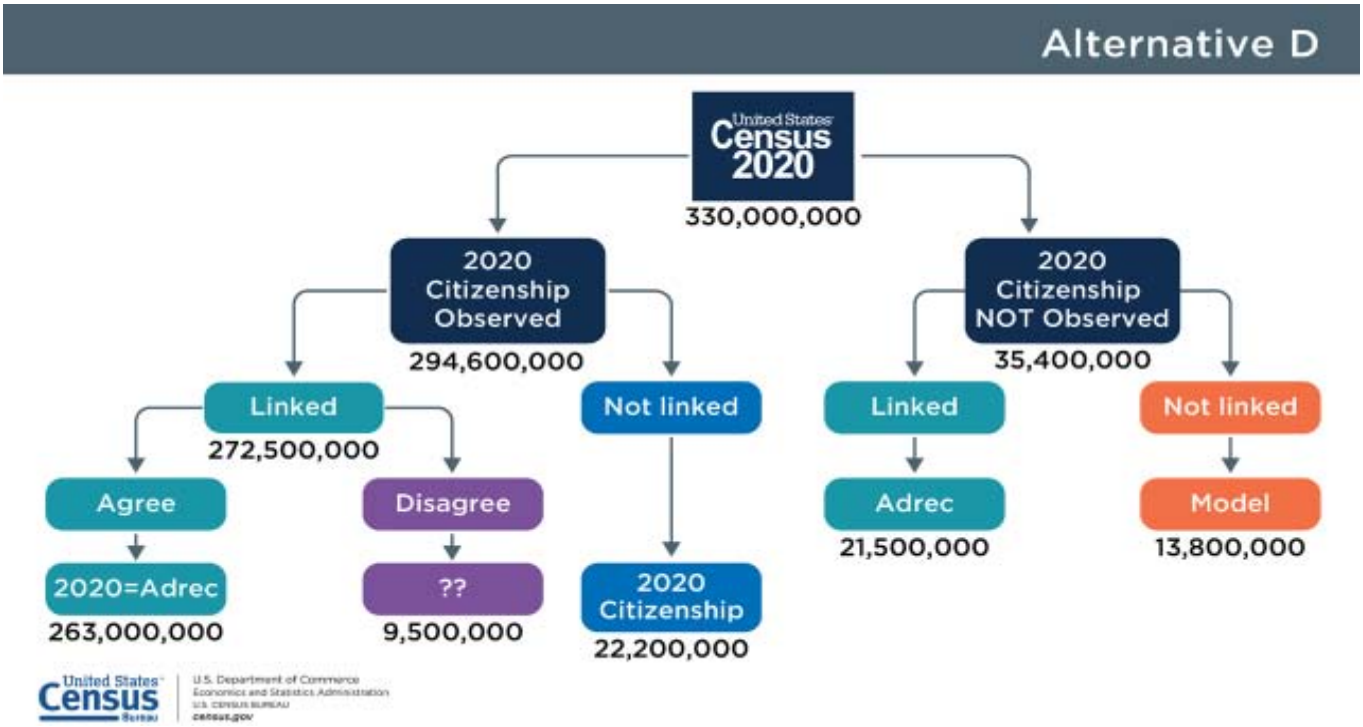

Alternative C
United States Census 2020
330,000,000
Linked
Not linked
Adrec
Model
295,000,000
35,000,000
United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Figure 2


Alternative D
United States Census 2020
330,000,000
2020 Citizenship Observed
294,600,000
2020 Citizenship NOT Observed
35,400,000
Linked
272,500,000
Not linked
Linked
Not linked
Agree
Disagree
Adrec
Model
21,500,000
13,800,000
2020=Adrec
263,000,000
??
9,500,000
2020 Citizenship
22,200,000
United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

March 1, 2018

MEMORANDUM FOR: Wilbur L. Ross, Jr.
Secretary of Commerce

Through: Karen Dunn Kelley
Performing the Non-Exclusive Functions and Duties of the Deputy Secretary

Ron S. Jarmin
Performing the Non-Exclusive Functions and Duties of the Director

Enrique Lamas
Performing the Non-Exclusive Functions and Duties of the Deputy Director

From: John M. Abowd
Chief Scientist and Associate Director for Research and Methodology

Subject: Preliminary analysis of Alternative D (Combined Alternatives B and C)

See attached.

Approved: _____ Date: _____
John M. Abowd, Chief Scientist
and Associate Director for Research and Methodology

Preliminary Analysis of Alternative D

At the Secretary's request we performed a preliminary analysis of combining Alternative B (asking the citizenship question of every household on the 2020 Census) and Alternative C (do not ask the question, link reliable administrative data on citizenship status instead) in the January 19, 2018 draft memo to the Department of Commerce into a new Alternative D. Here we discuss Alternative D, the weaknesses in Alternative C on its own, whether and how survey data could address these weaknesses, implications of including a citizenship question for using administrative data, and methodological challenges.

Description of Alternative D: Administrative data from the Social Security Administration (SSA), Internal Revenue Service (IRS), U.S. Citizenship and Immigration Services (USCIS), and the State Department would be used to create a comprehensive statistical reference list of current U.S. citizens. Nevertheless, there will be some persons for whom no administrative data are available. To obtain citizenship information for this sub-population, a citizenship question would be added to the 2020 Census questionnaire. The combined administrative record and 2020 Census data would be used to produce baseline citizenship statistics by 2021. Any U.S. citizens appearing in administrative data after the version created for the 2020 Census would be added to the comprehensive statistical reference list. There would be no plan to include a citizenship question on future Decennial Censuses or American Community Surveys. The comprehensive statistical reference list, built from administrative records and augmented by the 2020 Census answers would be used instead. The comprehensive statistical reference list would be kept current, gradually replacing almost all respondent-provided data with verified citizenship status data.

What are the weaknesses in Alternative C?

In the 2017 Numident (the latest available), 6.6 million persons born outside the U.S. have blank citizenship among those born in 1920 or later with no year of death. The evidence suggests that citizenship is not missing at random. Of those with missing citizenship in the Numident, a much higher share appears to be U.S. citizens than compared to those for whom citizenship data are not missing. Nevertheless, some of the blanks may be noncitizens, and it would thus be useful to have other sources for them.

A second question about the Numident citizenship variable is how complete and timely its updates are for naturalizations. Naturalized citizens are instructed to immediately apply for a new SSN card. Those who wish to work have an incentive to do so quickly, since having an SSN card with U.S. citizenship will make it easier to pass the E-Verify process when applying for a job, and it will make them eligible for government programs. But we do not know what fraction of naturalized citizens actually notify the SSA, and how soon after being naturalized they do so.

A third potential weakness of Numident citizenship is that some people are not required to have a Social Security Number (SSN), whether they are a U.S. citizen or not. It would also be useful to have a data source on citizenship that did not depend on the SSN application and tracking process inside SSA. This is why we proposed the MOU with the USCIS for naturalizations, and why we have now begun pursuing an MOU with the State Department for data on all citizens with passports.

IRS Individual Taxpayer Identification Numbers (ITIN) partially fill the gap in Numident coverage of noncitizen U.S. residents. However, not all noncitizen residents without SSNs apply for ITINs. Only those making IRS tax filings apply for ITINs. Once again, it would be useful to have a data source that did not depend on the ITIN process. The USCIS and State Department MOUs would provide an alternative source in this context as well.

U.S. Citizenship and Immigration Services (USCIS) data on naturalizations, lawful permanent residents, and I-539 non-immigrant visa extensions can partially address the weaknesses of the Numident. The USCIS data provide up-to-date information since 2001 (and possibly back to 1988, but with incomplete records prior to 2001). This will fill gaps for naturalized citizens, lawful permanent residents, and persons with extended visa applications without SSNs, as well as naturalized citizens who did not inform SSA about their naturalization. The data do not cover naturalizations occurring before 1988, as well as not covering and some between 1988-2000. USCIS data do not always cover children under 18 at the time a parent became a naturalized U.S. citizen. Such children automatically become U.S. citizens under the Child Citizenship Act of 2000. The USCIS receives notification of some, but not all, of these child naturalizations. Others inform the U.S. government of their U.S. citizenship status by applying for U.S. passports, which are less expensive than the application to notify the USCIS. USCIS visa applications list people's children, but those data may not be in electronic form.

U.S. passport data, available from the State Department, can help plug the gaps for child naturalizations, blanks on the Numident, and out-of-date citizenship information on the Numident for persons naturalized prior to 2001. Since U.S. citizens are not required to have a passport, however, these data will also have gaps in coverage.

Remaining citizenship data gaps in Alternative C include the following categories:

1. U.S. citizens from birth with no SSN or U.S. passport. They will not be processed by the production record linkage system used for the 2020 Census because their personally identifiable information won't find a matching Protected Identification Key (PIK) in the Person Validation System (PVS).

2. U.S. citizens from birth born outside the U.S., who do not have a U.S. passport, and either applied for an SSN prior to 1974 and were 18 or older, or applied before the age of 18 prior to 1978. These people will be found in PVS, but none of the administrative sources discussed above will reliably generate a U.S. citizenship variable.

3. U.S. citizens who were naturalized prior to 2001 and did not inform SSA of their naturalization because they originally applied for an SSN after they were naturalized, and it was prior to when citizenship verification was required for those born outside the U.S. (1974). These people already had an SSN when they were naturalized and they didn't inform SSA about the naturalization, or they didn't apply for an SSN. The former group have inaccurate data on the Numident. The latter group will not be found in PVS.

4. U.S. citizens who were automatically naturalized if they were under the age of 18 when their parents became naturalized in 2000 or later, and did not inform USCIS or receive a U.S. passport. Note that such persons would not be able to get an SSN with U.S. citizenship on the card without either a U.S. passport or a certificate from USCIS. These people will also not be found in the PVS.

5. Lawful permanent residents (LPR) who received that status prior to 2001 and either do not have an SSN or applied for an SSN prior to when citizenship verification was required for those born outside the U.S. (1974). The former group will not be found in PVS. The latter group has inaccurate data in Numident.

6. Noncitizen, non-LPR, residents who do not have an SSN or ITIN and who did not apply for a visa extension. These persons will not be found in PVS.

7. Persons with citizenship information in administrative data, but the administrative and decennial census data cannot be linked due to missing or discrepant PII.

Can survey data address the gaps in Alternative C?

One might think that survey data could help fill the above gaps, either when their person record is not linked in the PVS, and thus they have no PIK, or when they have a PIK but the administrative data lack up-to-date citizenship information. Persons in Category 6, however, have a strong incentive to provide an incorrect answer, if they answer at all. A significant, but unknown, fraction of persons without PIKs are in Category 6. Distinguishing these people from the other categories of persons without PIKs is an inexact science because there is no feasible method of independently verifying their non-citizen status. Our comparison of ACS and Numident citizenship data suggests that a large fraction of LPRs provide incorrect survey responses. This suggests that survey-collected citizenship data may not be reliable for many of the people falling in the gaps in administrative data. This calls into question their ability to improve upon Alternative C.

With Alternative C, and no direct survey response, the Census Bureau's edit and imputation procedures would make an allocation based primarily on the high-quality administrative data. In the presence of a survey response, but without any linked administrative data for that person, the edit would only be triggered by blank citizenship. A survey response of "citizen" would be accepted as valid. There is no scientifically defensible method for rejecting a survey response in the absence of alternative data for that respondent.

How might inclusion of a citizenship question on the questionnaire affect the measurement of citizenship with administrative data? Absent an in-house administrative data census, measuring citizenship with administrative data requires that persons in the Decennial Census be linked to the administrative data at the person level. The PVS system engineered into the 2020 Census does this using a very reliable technology. However, inclusion of a citizenship question on the 2020 Census questionnaire is very likely to reduce the self-response rate, pushing more households into Nonresponse Followup (NRFU). Not only will this likely lead to more incorrect enumerations, but it is also expected to increase the number of persons who cannot be linked to the administrative data because the NRFU PII is lower quality than the self-response data. In the 2010 Decennial Census, the percentage of NRFU persons who could be linked to administrative data rate was 81.6 percent, compared to 96.7 percent for mail responses. Those refusing to self-respond due to the citizenship question are particularly likely to refuse to respond in NRFU as well, resulting in a proxy response. The NRFU linkage rates were far lower for proxy responses than self-responses (33.8 percent vs. 93.0 percent, respectively).

Although persons in Category 6 will not be linked regardless of response mode, it is common for households to include persons with a variety of citizenship statuses. If the whole household does not self-

respond to protect the members in Category 6, the record linkage problem will be further aggravated. Thus, not only are citizenship survey data of suspect quality for persons in the gaps for Alternative C, collecting these survey data would reduce the quality of the administrative records when used in Alternative D by lowering the record linkage rate for persons with administrative citizenship data.

What methodological challenges are involved when combining these sources?

Using the 2020 Census data only to fill in gaps for persons without administrative data on citizenship would raise questions about why 100 percent of respondents are being burdened by a citizenship question to obtain information for the two percent of respondents where it is missing.

Including a citizenship question in the 2020 Census does not solve the problem of incomplete person linkages when producing citizenship statistics after 2020. Both the 2020 decennial record and the record with the person's future location would need to be found in PVS to be used for future statistics.

In sum, Alternative D would result in poorer quality citizenship data than Alternative C. It would still have all the negative cost and quality implications of Alternative B outlined in the draft January 19, 2018 memo to the Department of Commerce.

U.S. Department of Commerce
U.S. Census Bureau

Privacy Impact Assessment
for the
CEN05 Field Systems Major Application System
March 2018

Reviewed by: _____, Bureau Chief Privacy Officer
Robin J. Bachman

☑ Concurrence of Senior Agency Official for Privacy/DOC Chief Privacy Officer
☐ Non-concurrence of Senior Agency Official for Privacy/DOC Chief Privacy Officer

CATRINA PURVIS

Digitally signed by CATRINA PURVIS
DN: c=US, o=U.S. Government, ou=Department of Commerce, ou=Office of the Secretary, cn=CATRINA PURVIS, 0.9.2342.19200300.100.1.1=13001002875743
Date: 2018.06.22 13:49:29 -04'00'

Signature of Senior Agency Official for Privacy/DOC Chief Privacy Officer Date

U.S. Department of Commerce Privacy Impact Assessment
U.S. Census Bureau/CEN05 Field Systems Major Application System

Unique Project Identifier: 006-000401400

Introduction: System Description

Provide a description of the system that addresses the following elements:

(a) The CEN05 Field Systems Major Application System is a major information system managed by the Application Development Services Division (ADSD) in support of the Census Bureau Field Directorate.

(b) The IT system is housed at the Census Bureau's Bowie, MD computer center.

(c) CEN05 interconnects with other CENs. Desktop and laptop client services are provided by CEN17, server support is provided by CEN16, and Oracle 12c database support is provided by CEN18.

(d) How the system operates to achieve the purpose(s) identified in section 4:

The Field Directorate plans, organizes, coordinates, and carries out the Census Bureau's field data collection program for sample surveys, special censuses, the Economic Census, and the Decennial census. The CEN05 IT system maintains Personally Identifiable Information (PII) collected from respondents for these surveys and censuses such as name, address, contact information, race, gender, education, financial information, etc. In addition, Field Representative characteristics are also collected while surveys and census interviews are conducted.

Laptops are used to collect survey data in the field. An application assigns cases and monitors interviewing progress. Information systems covered by other CEN plans are used as routing mechanisms to transfer survey data from various survey sources, that includes but are not limited to, computer interviewing, telephone interviewing utilizing Census computer programs, internet self surveys, and paper surveys to the survey sponsors.

In addition to laptops, the Field Directorate has begun utilizing both tablet and other mobile computing devices to perform Census Tests that lead up to the 2020 Decennial Census to collect respondent information using Census Bureau issued mobile devices.

The Customer Experience Management (CEM) System is a centralized data store from five data sources currently being utilized by the Census Bureau and will deploy an enterprise dashboard for Census Bureau leadership.

This dashboard will provide insights into customer engagement for Census Bureau products & services, and will allow analysts to leverage data across data sources on a

holistic Business Intelligence (BI) platform. The system will not only eliminate manual processes but will:

1) Create an opportunity for a better understanding of patterns and trends of customer experiences that can lead to actionable improvement plans, and;
2) Establish a framework and foundation for other data integration, BI, and analytics efforts.

CEM will interface/collect information for various sources inside and outside of the Census Bureau and will contain PII data.

The Census Enterprise Data Collection and Processing initiative (CEDCaP) is a suite of systems and supporting infrastructure to handle data collection and processing for the nearly 100 surveys and three censuses conducted by the Census Bureau. CEN05 is part of this infrastructure and includes the *Enterprise Censuses and Surveys Enabling platform (ECaSE) which* will provide about half the data collection capabilities for CEDCaP

The ECaSE – ISR (Internet Self-Response) secure Internet data exchange system is a web-based framework for the design, delivery, and execution of surveys, censuses, and other data collection and data exchange efforts over the Internet. The enterprise-level application offers data collection areas the ability to reach a large number of potential respondents online, in a customizable manner to suit their business needs. ECaSE – ISR is developed to support increased demand for online data collections, including the extremely high loads associated with the 2020 Decennial Census.

ECaSE – OCS (Operational Control System) will serve as the standard tool to assign, control, track, and manage listing, survey and census workloads for the field workforce. ECaSE – OCS provides an enterprise application framework for this need, regardless of the interviewer-assisted mode used (phone or in person).

To establish a cohesive system boundary for the purposes of security assessment, the CEN05 IT system is comprised of three major areas: Collections, Business Support Processes, and Backend Processes. Each of these major components employs security control mechanisms that must be individually documented to ensure that the system as a whole is appropriately protected. As such, the system security plan is organized to reflect the implementation of technical controls for each subsystem component.

ADSD incorporated a Survey Field Identification Tool (sFIT) to aid in investigating situations where it is suspected that a Field Representative may be falsifying respondent information. The tool will be used by Contact Center and Regional Office (RO) representatives to indicate FR who are suspected of falsification and to facilitate and document the results of the investigations. The newly developed tool replaced the previous automated system and the paper 11-163 forms. sFIT collects and disseminates PII regarding a survey respondent and the Field Representative who is suspected of falsifying survey data. sFIT received its Authority to Operate (ATO) on 7/13/2017.

(e) How information in the system is retrieved by the user

There are many external sponsors but CEN05 does not have direct connections to any of them. Demographics Survey Division (DSD), Econ, etc., will give CEN05 the surveys that they have crafted for the external sponsors and the data collected for those surveys is placed into the CEN05 Master Control System (MCS) for the internal system to pick up. It will be the other internal sponsors' responsibility to vet the information, transform it into a format that the external sponsors can ingest and send it off. The sharing of the data should be on those systems with the external connection to the external sponsors. This is the same with the National Center for Health Statistics (NCHS) survey.

(f) How information is transmitted to and from the system

The CEN05 IT system includes a data warehouse that extracts and provides a view of survey data over time, data collection modes, and data collection operations. It aggregates data and creates canned reports. These reports are made available to stakeholders, approved individuals, and organizations to support optimization and coordination of decennial, current, and special surveys. The reports are developed by a special staff that was established through the Office of the Director to serve as an analytic team with specific, ongoing, responsibilities to develop analytic tools (charts and tables). These tools will be used by decennial and current survey field managers toward the goal of continuous improvement in survey operational efficiency. This group will both initiate and respond to issues related to survey performance indicators including cost, data quality, and data collection progress. This database interfaces with systems throughout the Census Bureau that contain PII, Business Identifiable Information (BII), and data collected and/or protected under Title 13 and Title 26.

(g) Any information sharing conducted by the system

The CEN05 IT system only shares PII and BII data within other CEN05 systems/components and with other Census Bureau IT systems. The data warehouse provides the ability for survey sponsors to view statistical data. It does not allow access to the information at the individual case level where the "raw" data resides. The data warehouse interacts with multiple systems within the Census Bureau network such as the system that contains information about businesses, which provides the mailing list for the Economic Census, and the primary sampling frame for virtually all other business surveys. The survey information about businesses is collected to track the movement of commodities from businesses throughout the United States. The data will be stored in its "raw" form and then transformed and stored in the data warehouse. The data warehouse will tabulate the data and display it in various reports. The data in its "raw" form will not be displayed. The data warehouse maintains PII, BII, Title 13 and Title 26 information.

(h) The specific programmatic authorities (statutes or Executive Orders) for collecting, maintaining, using, and disseminating the information

15 U.S.C. 301.
13 U.S.C. Chapter 5, 6(c), 8(b), 131, 132, 141, 161, 182, 193, 196
18 U.S.C. 2510-2521
15 CFR, Part 50.
26 U.S.C.

(i) The Federal Information Processing Standards (FIPS) 199 security impact category for the system

The Federal Information Processing Standard (FIPS) 199 security impact category for this system is Moderate.

Section 1: Status of the Information System

1.1 Indicate whether the information system is a new or existing system.

_____ This is a new information system.
_____ This is an existing information system with changes that create new privacy risks. *(Check all that apply.)*

Changes That Create New Privacy Risks (CTCNPR)					
a. Conversions		d. Significant Merging		g. New Interagency Uses	
b. Anonymous to Non-Anonymous		e. New Public Access		h. Internal Flow or Collection	
c. Significant System Management Changes		f. Commercial Sources		i. Alteration in Character of Data	
j. Other changes that create new privacy risks (specify):					

__X_ This is an existing information system in which changes do not create new privacy risks, and there is a SAOP approved Privacy Impact Assessment.

Section 2: Information in the System

2.1 Indicate what personally identifiable information (PII)/business identifiable information (BII) is collected, maintained, or disseminated. *(Check all that apply.)*

Identifying Numbers (IN)					
a. Social Security*	X	e. File/Case ID	X	i. Credit Card	
b. Taxpayer ID		f. Driver's License		j. Financial Account	
c. Employer ID		g. Passport		k. Financial Transaction	
d. Employee ID	X	h. Alien Registration		l. Vehicle Identifier	
m. Other identifying numbers (specify):					
*Explanation for the need to collect, maintain, or disseminate the Social Security number, including truncated form:					

1) Employee's full SSN info needed for cost reimbursement and other financial purposes.
2) The last 4 digits of the survey respondent's SSN helps is collected on behalf of the survey sponsor, The National Center for Health Statistics (NCHS). The justification for the necessity of collecting this information, taken from the latest approved Office of Management and Budget (OMB) Information Collection Request (ICR) supporting statement is below:

Social Security Number and Health Insurance Claim Number: The last four digits of the Social Security Number (SSN) is asked on the NHIS questionnaire to allow linkage with administrative and vital records, such as the National Death Index (NDI). The NDI is a computerized central file of death record information. It is compiled from data obtained by NCHS from the State vital statistics offices. The data contain a standard set of identi-fying information on decedents from 1979 to the present. Records are matched using Social Security Number and other variables such as name, father's surname, date of birth, sex, state of resi-dence, and marital status. Of these, Social Security Number is the most important identifier for successful matching. The last four digits has been shown to be nearly as effective for matching as the full number.

The Social Security Number is also used by the Medical Expenditure Panel Study to help track the location of respondents who have changed residence since their NHIS interview. Finding a correct address for respondents is essential to maintaining response levels at an acceptable level in linked surveys, and the Social Security Number is a key item for establishing a correct address.

Medicare beneficiaries are given a health insurance claim (HIC) number that is their (or their spouse's) SSN with an alphabetic prefix. The NHIS also asks for the last four digits of that number so that the NHIS data can be linked to Medicare claims information for purposes of statistical research.

General Personal Data (GPD)					
a. Name	X	g. Date of Birth	X	m. Religion	
b. Maiden Name	X	h. Place of Birth	X	n. Financial Information	X
c. Alias	X	i. Home Address	X	o. Medical Information	X
d. Gender	X	j. Telephone Number	X	p. Military Service	X
e. Age	X	k. Email Address	X	q. Physical Characteristics	
f. Race/Ethnicity	X	l. Education	X	r. Mother's Maiden Name	XX
s. Other general personal data (specify):					

Work-Related Data (WRD)					
a. Occupation	X	d. Telephone Number	X	g. Salary	X
b. Job Title	X	e. Email Address	X	h. Work History	XX
c. Work Address	X	f. Business Associates	X		
i. Other work-related data (specify):					

Distinguishing Features/Biometrics (DFB)					
a. Fingerprints		d. Photographs		g. DNA Profiles	
b. Palm Prints		e. Scars, Marks, Tattoos		h. Retina/Iris Scans	
c. Voice Recording/Signatures		f. Vascular Scan		i. Dental Profile	
j. Other distinguishing features/biometrics (specify):					

System Administration/Audit Data (SAAD)					
a. User ID	X	c. Date/Time of Access	X	e. ID Files Accessed	X
b. IP Address	X	d. Queries Run	X	f. Contents of Files	X

g. Other system administration/audit data (specify):

Other Information (specify)

2.2 Indicate sources of the PII/BII in the system. *(Check all that apply.)*

Directly from Individual about Whom the Information Pertains					
In Person	X	Hard Copy: Mail/Fax		Online	X
Telephone	X	Email	X		
Other (specify):					

Government Sources					
Within the Bureau	X	Other DOC Bureaus		Other Federal Agencies	X
State, Local, Tribal	X	Foreign			
Other (specify):					

Non-government Sources					
Public Organizations		Private Sector	X	Commercial Data Brokers	X
Third Party Website or Application					
Other (specify):					

2.3 Describe how the accuracy of the information in the system is ensured.

It will be the internal sponsors' responsibility to vet the information, transform it into a format that the external sponsors can ingest and send it off. The sharing of the data should be on those systems with the external connection to the external sponsors.

2.4 Is the information covered by the Paperwork Reduction Act?

	Yes, the information is covered by the Paperwork Reduction Act. Provide the OMB control number and the agency number for the collection.
X	No, the information is not covered by the Paperwork Reduction Act.

2.5 Indicate the technologies used that contain PII/BII in ways that have not been previously deployed. *(Check all that apply.)*

Technologies Used Containing PII/BII Not Previously Deployed (TUCPBNPD)			
Smart Cards		Biometrics	
Caller-ID		Personal Identity Verification (PIV) Cards	
Other (specify):			
X	There are not any technologies used that contain PII/BII in ways that have not been previously deployed.		

Section 3: System Supported Activities

3.1 Indicate IT system supported activities which raise privacy risks/concerns. *(Check all that apply.)*

Activities			
Audio recordings	X	Building entry readers	
Video surveillance		Electronic purchase transactions	
Other (specify):			
	There are not any IT system supported activities which raise privacy risks/concerns.		

Section 4: Purpose of the System

4.1 Indicate why the PII/BII in the IT system is being collected, maintained, or disseminated. *(Check all that apply.)*

Purpose			
For a Computer Matching Program		For administering human resources programs	
For administrative matters	X	To promote information sharing initiatives	
For litigation		For criminal law enforcement activities	
For civil enforcement activities		For intelligence activities	
To improve Federal services online		For employee or customer satisfaction	X
For web measurement and customization technologies (single-session)		For web measurement and customization technologies (multi-session)	
Other (specify):			

Section 5: Use of the Information

5.1 In the context of functional areas (business processes, missions, operations, etc.) supported by the IT system, describe how the PII/BII that is collected, maintained, or disseminated will be used. Indicate if the PII/BII identified in Section 2.1 of this document is in

reference to a federal employee/contractor, member of the public, foreign national, visitor or other (specify).

Census Bureau information shapes important policy and operational decisions that help improve the nation's social and economic conditions. We conduct the constitutionally mandated Census of Population and Housing every 10 years, which is used to apportion seats in the House of Representatives and informs congressional redistricting.

We also conduct a census of all business establishments and of all governmental units, known respectively as the Economic Census and the Census of Governments, every five years. The Economic Census is the benchmark used for measuring Gross Domestic Product (GDP) and other key indicators that guide public policy and business investment decisions.

In addition, we conduct several ongoing business and household surveys that provide the information in several of the Nation's key economic indicators and which is used to allocate over $400 billion in Federal funding annually.

The PII/BII collected for statistical purposes: The PII/BII maintained is from voluntary and mandatory surveys, census interviews, pilot tests and cognitive interviews collected from members of the public.

The PII collected for administrative purposes: We collect information about Census Bureau employees during the collection of respondent information. Field representative and interviewer characteristics obtained during census and survey interviews, pilot tests, and cognitive interviews are used for research and analytical studies to evaluate Census Bureau surveys and programs.

5.2 Describe any potential threats to privacy as a result of the bureau's/operating unit's use of the information, and controls that the bureau/operating unit has put into place to ensure that the information is handled, retained, and disposed appropriately. (For example: mandatory training for system users regarding appropriate handling of information, automatic purging of information in accordance with the retention schedule, etc.)

The Census Bureau Information technology systems employ a multitude of layered security controls to protect BII/PII at rest, during processing, as well as in transit. These NIST 800-53 controls, at a minimum, are deployed and managed at the enterprise level including, but not limited to the following:

- Intrusion Detection | Prevention Systems (IDS | IPS)
- Firewalls
- Mandatory use of HTTP(S) for Census Bureau Public facing websites
- Use of trusted internet connection (TIC)
- Anti-Virus software to protect host/end user systems
- Encryption of databases (Data at rest)
- HSPD-12 Compliant PIV cards
- Access Controls

The Census bureau Information technology systems also follow the National Institute of Standards and Technology (NIST) standards including special publications 800-53, 800-63, 800-37 etc. Any system within the Census Bureau that contains, transmits, or processes BII/PII has a current authority to operate (ATO) and goes through continuous monitoring on a yearly basis to ensure controls are implemented and operating as intended. The Census Bureau also deploys a Data Loss Prevention solution as well.

All Bureau employees and contractors undergo mandatory annual data stewardship training to include proper handling, dissemination, and disposal of BII/PII/Title 13/Title 26 data.

Section 6: Information Sharing and Access

6.1 Indicate with whom the bureau intends to share the PII/BII in the IT system and how the PII/BII will be shared. *(Check all that apply.)*

Recipient	How Information will be Shared		
	Case-by-Case	Bulk Transfer	Direct Access
Within the bureau	X	X	X
DOC bureaus			
Federal agencies			
State, local, tribal gov't agencies			
Public			
Private sector			
Foreign governments			
Foreign entities			
Other (specify):			

	The PII/BII in the system will not be shared.

6.2 Indicate whether the IT system connects with or receives information from any other IT systems authorized to process PII and/or BII.

X	Yes, this IT system connects with or receives information from another IT system(s) authorized to process PII and/or BII. Provide the name of the IT system and describe the technical controls which prevent PII/BII leakage:

	CEN03, CEN04 CBS, CEN06 NPC, CEN11, CEN13, CEN22, CEN35, CEN36 CEN05 uses a multitude of security controls mandated by the Federal Information Security Management Act of 2002 (FISMA) and various other regulatory control frameworks including the National Institute of Standards and Technology (NIST) special publication 800 series. These security controls include, but are not limited to the use of mandatory HTTPS for public facing websites, access controls, anti-virus solutions, enterprise auditing/monitoring, encryption of data at rest, and various physical controls at the Census Bureau facilities that house Information Technology systems. The Census Bureau also deploys an enterprise Data Loss Protection (DLP) solution as well.
	No, this IT system does not connect with or receive information from another IT system(s) authorized to process PII and/or BII.

6.3 Identify the class of users who will have access to the IT system and the PII/BII. *(Check all that apply.)*

Class of Users			
General Public		Government Employees	X
Contractors	X		
Other (specify):			

Section 7: Notice and Consent

7.1 Indicate whether individuals will be notified if their PII/BII is collected, maintained, or disseminated by the system. *(Check all that apply.)*

X	Yes, notice is provided pursuant to a system of records notice published in the Federal Register and discussed in Section 9.	
X	Yes, notice is provided by a Privacy Act statement and/or privacy policy. The Privacy Act statement and/or privacy policy can be found at:	
X	Yes, notice is provided by other means.	Specify how: as specified in some survey instrument(s), respondent letters, etc.
	No, notice is not provided.	Specify why not:

7.2 Indicate whether and how individuals have an opportunity to decline to provide PII/BII.

X	Yes, individuals have an opportunity to decline to provide PII/BII.	Specify how: For voluntary surveys or censuses, the respondent has an opportunity to decline to provide PII/BII.
X	No, individuals do not have an opportunity to decline to provide PII/BII.	Specify why not: For mandatory surveys or censuses, the respondent does not have an opportunity to decline to provide PII/BII

7.3 Indicate whether and how individuals have an opportunity to consent to particular uses of their PII/BII.

X	Yes, individuals have an opportunity to consent to particular uses of their PII/BII.	Specify how: For Employee Productivity Measurement Records the consent is required for employment
X	No, individuals do not have an opportunity to consent to particular	Specify why not: As identified in the applicable SORNs Census -3, -4, -5, and -7 the records are exempted from

	uses of their PII/BII.	notification, access, and contest requirements of the agency procedures (under via 5 U.S.C 552a(c)(3),(d), (e),(1), (e),(4),(G), (H) and (I), and (f). The data are maintained by the U.S. Census Bureau solely as statistical records as required under Title 13 U.S.C., and are not used in whole or in part in making any determination about an identifiable individual. This exemption is also made in accordance with the Department's rules which appear in 15 CFR part 4 subpart B.

7.4 Indicate whether and how individuals have an opportunity to review/update PII/BII pertaining to them.

X	Yes, individuals have an opportunity to review/update PII/BII pertaining to them.	Specify how: : As identified in the applicable SORN for Employee Productivity Measurement Records, these individual may contact the Associate Director for Field Operations for access to these records.
X	No, individuals do not have an opportunity to review/update PII/BII pertaining to them.	Specify why not: As identified in the applicable SORNs Census-3, -4, -5, and -7 the records are exempted from notification, access, and contest requirements of the agency procedures (under via 5 U.S.C 552a(c)(3),(d), (e),(1), (e),(4),(G), (H) and (I), and (f). The data are maintained by the U.S. Census Bureau solely as statistical records as required under Title 13 U.S.C., and are not used in whole or in part in making any determination about an identifiable individual. This exemption is also made in accordance with the Department's rules which appear in 15 CFR part 4 subpart B.

Section 8: Administrative and Technological Controls

8.1 Indicate the administrative and technological controls for the system. *(Check all that apply.)*

X	All users signed a confidentiality agreement or non-disclosure agreement.
X	All users are subject to a Code of Conduct that includes the requirement for confidentiality.
X	Staff (employees and contractors) received training on privacy and confidentiality policies and practices.
X	Access to the PII/BII is restricted to authorized personnel only.
X	Access to the PII/BII is being monitored, tracked, or recorded. Explanation:
X	The information is secured in accordance with FISMA requirements. Provide date of most recent Assessment and Authorization (A&A): 7/20/2017 ☐ This is a new system. The A&A date will be provided when the A&A package is approved.
X	The Federal Information Processing Standard (FIPS) 199 security impact category for this system is a moderate or higher.
X	NIST Special Publication (SP) 800-122 and NIST SP 800-53 Revision 4 Appendix J recommended security controls for protecting PII/BII are in place and functioning as intended; or have an approved Plan of Action and Milestones (POA&M).
X	A security assessment report has been reviewed for the supporting information system and it has been determined that there are no additional privacy risks.
X	Contractors that have access to the system are subject to information security provisions in their contracts required by DOC policy.
	Contracts with customers establish ownership rights over data including PII/BII.

	Acceptance of liability for exposure of PII/BII is clearly defined in agreements with customers.
X	Other (specify): Publications are approved by the Disclosure Review Board

8.2 Provide a general description of the technologies used to protect PII/BII on the IT system. *(Include data encryption in transit and/or at rest, if applicable).*

The Census Bureau Information technology systems employ a multitude of layered security controls to protect BII/PII at rest, during processing, as well as in transit. These NIST 800-53 controls, at a minimum, are deployed and managed at the enterprise level including, but not limited to the following:

- Intrusion Detection | Prevention Systems (IDS | IPS)
- Firewalls
- Mandatory use of HTTP(S) for Census Bureau Public facing websites
- Use of trusted internet connection (TIC)
- Anti-Virus software to protect host/end user systems
- Encryption of databases (Data at rest)
- HSPD-12 Compliant PIV cards
- Access Controls

The Census bureau Information technology systems also follow the National Institute of Standards and Technology (NIST) standards including special publications 800-53, 800-63, 800-37 etc. Any system within the Census Bureau that contains, transmits, or processes BII/PII has a current authority to operate (ATO) and goes through continuous monitoring on a yearly basis to ensure controls are implemented and operating as intended. The Census Bureau also deploys a Data Loss Prevention solution as well.

Section 9: Privacy Act

9.1 Indicate whether a system of records is being created under the Privacy Act, 5 U.S.C. § 552a. *(A new system of records notice (SORN) is required if the system is not covered by an existing SORN).*

As per the Privacy Act of 1974, "the term 'system of records' means a group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual."

X	Yes, this system is covered by an existing system of records notice (SORN). Provide the SORN name, number, and link. *(list all that apply)*: COMMERCE/CENSUS-2, Employee Productivity Measurement Records: http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-2.html COMMERCE/CENSUS-3, Special Censuses, Surveys, and Other Studies: http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-3.html COMMERCE/CENSUS-4, Economic Survey Collection: http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-4.html COMMERCE/CENSUS-5, Decennial Census Program:

	http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-5.html COMMERCE/CENSUS-7, Special Censuses of Population Conducted for State and Local Government: http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-.html
	Yes, a SORN has been submitted to the Department for approval on (date).
	No, this system is not a system of records and a SORN is not applicable.

Section 10: Retention of Information

10.1 Indicate whether these records are covered by an approved records control schedule and monitored for compliance. *(Check all that apply.)*

X	There is an approved record control schedule. Provide the name of the record control schedule: GRS 3.1: General Technology Management Records; GRS 3.2: Information Systems Security Records; GRS 4.1: Records Management Records; GRS 4.2: Information Access and Protection Records; GRS 4.3: Input Records, Output Records, and Electronic Copies, N1-29-89-5 American Housing Survey, NC1-29-79-7 Demographic Fields Area, NC1-29-80-6 Demographic and Economic area Divisions - Secondary Use Sampling Records
	No, there is not an approved record control schedule. Provide the stage in which the project is in developing and submitting a records control schedule:
X	Yes, retention is monitored for compliance to the schedule.
	No, retention is not monitored for compliance to the schedule. Provide explanation:

10.2 Indicate the disposal method of the PII/BII. *(Check all that apply.)*

Disposal			
Shredding		Overwriting	X
Degaussing		Deleting	X
Other (specify):			

Section 11: NIST Special Publication 800-122 PII Confidentiality Impact Level

11.1 Indicate the potential impact that could result to the subject individuals and/or the organization if PII were inappropriately accessed, used, or disclosed. *(The PII Confidentiality Impact Level is not the same as the Federal Information Processing Standards (FIPS) 199 security impact category.)*

	Low – the loss of confidentiality, integrity, or availability could be expected to have a limited adverse

	effect on organizational operations, organizational assets, or individuals.
	Moderate – the loss of confidentiality, integrity, or availability could be expected to have a serious adverse effect on organizational operations, organizational assets, or individuals.
X	High – the loss of confidentiality, integrity, or availability could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, or individuals.

11.2 Indicate which factors were used to determine the above PII confidentiality impact levels. *(Check all that apply.)*

X	Identifiability	Provide explanation: Individual data elements directly identifying unique individuals.
X	Quantity of PII	Provide explanation: A severe or catastrophic number of individuals affected by loss, theft, or compromise. Severe or catastrophic collective harm to individuals, harm to the organization's reputation, or cost to the organization in addressing a breach.
X	Data Field Sensitivity	Provide explanation: Data fields, alone or in combination, are directly usable in other contexts and make the individual or organization vulnerable to harms, such as identity theft, embarrassment, loss of trust, or costs.
X	Context of Use	Provide explanation: Disclosure of the PII is likely to result in severe or catastrophic harm to the individual or organization.
X	Obligation to Protect Confidentiality	Provide explanation: Organization or Mission- specific privacy laws, regulations, mandates, or organizational policy apply that add more restrictive requirements to government- wide or industry-specific requirements. Violations may result in severe civil or criminal penalties. PII in this IT system is collected under the authority of Title 5.
X	Access to and Location of PII	Located on computers and other devices on a network controlled by the organization. Access limited to a multiple populations of the organization's workforce beyond the direct program or office that owns the information on behalf of the organization. Access only allowed by organization- owned equipment outside of the physical locations owned by the organization only with a secured connection (e.g., virtual private network (VPN)).
	Other:	Provide explanation:

Section 12: Analysis

12.1 Identify and evaluate any potential threats to privacy that exist in light of the information collected or the sources from which the information is collected. Also, describe the choices that the bureau/operating unit made with regard to the type or quantity of information collected and the sources providing the information in order to prevent or mitigate threats to privacy. (For example: If a decision was made to collect less data, include a discussion of this decision; if it is necessary to obtain information from sources other than the individual, explain why.)

The biggest potential threat to privacy is the potential loss of a Field Representative's CAPI laptop. This threat is mitigated by the use of full disk encryption of the hard drive or solid state drive which renders the data inaccessible in the event that laptop is lost, damaged, or stolen. Due to the quantity, nature, and scope of the multiple surveys collected and processed through CEN05 (including the upcoming Decennial Census), there are no plans to reduce the quantity or type of data being collected, nor the sources of that data.

12.2 Indicate whether the conduct of this PIA results in any required business process changes.

	Yes, the conduct of this PIA results in required business process changes. Explanation:
X	No, the conduct of this PIA does not result in any required business process changes.

12.3 Indicate whether the conduct of this PIA results in any required technology changes.

	Yes, the conduct of this PIA results in required technology changes. Explanation:
X	No, the conduct of this PIA does not result in any required technology changes.

Version Number: 01-2017

U.S. Department of Commerce
U.S. Census Bureau

Privacy Impact Assessment
for the
CEN08 Decennial Information Technology Division (DITD)

Reviewed by: _____, Bureau Chief Privacy Officer

☑ Concurrence of Senior Agency Official for Privacy/DOC Chief Privacy Officer

☐ Non-concurrence of Senior Agency Official for Privacy/DOC Chief Privacy Officer

CATRINA PURVIS

Digitally signed by CATRINA PURVIS
DN: c=US, o=U.S. Government, ou=Department of Commerce, ou=Office of the Secretary, cn=CATRINA PURVIS, 0.9.2342.19200300.100.1.1=13001002875743
Date: 2018.09.28 19:04:38 -04'00'

09/27/2018

Signature of Senior Agency Official for Privacy/DOC Chief Privacy Officer — Date

U.S. Department of Commerce Privacy Impact Assessment
U.S. Census Bureau/CEN08 DITD

Unique Project Identifier: 006-000400400
Introduction: System Description

Provide a description of the system that addresses the following elements:
The response must be written in plain language and be as comprehensive as necessary to describe the system. Please answer each question (a) through (i)separately.

(a) Whether it is a general support system, major application, or other type of system

CEN08 Decennial Information Technology Division (DITD) consists of both general support systems and major applications:

Major Applications
CEN08 Decennial manages the development and implementation of major decennial census applications utilized by the Decennial Census Program. These applications process response data from census tests and 2020 Census operations, and perform quality assurance mechanisms for various census operations. These applications also facilitate the acquisition of software and integration services required to support the U.S. Census Bureau (USCB) during the preparation and actual Decennial Census operations.

Some examples of the information collected, maintained, and/or disseminated within CEN08 Decennial are names, addresses, gender, age, date of birth, race, email, education, telephone number and salary.

The CEN08 Decennial IT system monitors the cost, schedule, and technical performance milestones for each software system or application utilized for decennial census purposes. The CEN08 IT system manages the development and implementation of software and systems necessary to support collection, processing, and tabulation of census data.

General Support System
CEN08 DITD general support system consists of:

1) an external vendor general support system called the Third Party Fingerprinting solution that is managed by Indrasoft. The U.S. Census Bureau (USCB) employs hundreds of thousands of temporary workers to perform data collection activities via a non-competitive Schedule A hiring authority from the Office of Personnel Management (OPM) in support of the Decennial Census testing in Fiscal Year (FY) 2018 and 2020 Census. As part of the recruitment and security process, the USCB requires that these selectees undergo fingerprinting to determine their suitability for employment. In addition, contractors that provide services in support of the 2020 Decennial Census, such as Census Questionnaire Assistance (CQA) contractor candidates, may also be

fingerprinted. To support fingerprinting for the 2020 Census, the USCB will use the Third Party Fingerprinting solution to capture and transmit fingerprints to USCB and conduct identity proofing for these temporary hires and contractors.

2) CEN08 DITD also consists of another external vendor general support IT system called the Recruiting and Assessment (R&A) solution that is managed by Cornerstone OnDemand. R&A is a FedRAMP-approved IT system that allows the Census Bureau to have a recruiting & selection tool and a learning management tool in one. FedRAMP is a government-wide program that provides a standardized approach to security assessment, authorization, and continuous monitoring for cloud products and services.

(b) System location

Decennial Applications – Bowie Computer Center (BCC) and AWS GovCloud located in Oregon.

Third Party Fingerprinting – AWS U.S. East/West located in US East (Ohio), US East (N. Virginia), US West (N. California), and US West (Oregon) and physical fingerprinting capture sites across the United States.

R&A - Unified Talent Management Suite (CUTMS) Cloud located in El Segundo, CA and Ashburn, VA.

(c) Whether it is a standalone system or interconnects with other systems (identifying and describing any other systems to which it interconnects)

Decennial Applications – Shares information internally within the Census Bureau with CEN07 GEO, CEN05 Field, CEN11 DEMO, CEN18 CDL, CEN19 CEDSCI, CEN21 DAPPS, CEN30 ACS, and CEN36 ADEP.

Third Party Fingerprinting – Shares information internally within the Census Bureau with CEN21 CHEC and CEN18 SOA.

R&A – Shares information internally within the Census Bureau with *CEN21 DAPPS*

(d) The way the system operates to achieve the purpose(s) identified in Section 4

The CEN08 DITD provides updates and unit (e.g., a home, a building, or miscellaneous structure) status information to various divisions within the Census Bureau that maintain address information (e.g., street addresses, and status and control information for households

and other living quarters). In addition, CEN08 systems confirm receipt of response data. They also provide validation and acknowledgment of the data received from various IT systems.

Temporary hires and contractors looking to support the 2020 Census submit their job applications through the R&A system. R&A securely delivers the submitted application data and associated attachments to DAPPS for processing and selecting.

To support fingerprinting for the 2020 Census, the USCB will use the Third Party Fingerprinting solution to capture and transmit fingerprints to USCB and conduct identity proofing for temporary hires and contractors (selectees). These selectees will provide their fingerprints at one of the Third Party Fingerprinting physical capture sites.

(e) How information in the system is retrieved by the user

Information in the CEN08 DITD systems are retrieved by using PII information identified in Section 2 below by authorized users using internal web applications, secure databases, and managed file transfer servers. Authorized Census Hiring Employment Check (CHEC) users pull selectee fingerprint files from the Third Party Fingerprinting solution and forward to the FBI for processing.

(f) How information is transmitted to and from the system

Information is transferred to and from CEN08 DITD systems via authorized manual and/or automated connections.

User fingerprints are captured on Third Party Fingerprinting physical sites which is uploaded to authorized AWS U.S. East/West. Files are encrypted and transferred using the service-oriented architecture (SOA) via the Enterprise Service Bus (ESB), which then sends it over to CHEC within the U.S Census Bureau. The Enterprise Service Bus is a configuration-based, policy-driven enterprise service bus. It provides highly scalable and reliable service-oriented integration, service management, and traditional message brokering across heterogeneous IT environments. It combines intelligent message brokering with routing and transformation of messages, along with service monitoring and administration in a unified software product.

(g) Any information sharing conducted by the system

CEN08 DITD systems share information internally with approved Census Bureau systems on an as needed basis.

Fingerprints are shared between the fingerprinting solution, CHEC and FBI.

R&A Applicant information is shared internally with the Decennial Applicant, Personnel and Payroll System (DAPPS).

(h) The specific programmatic authorities (statutes or Executive Orders) for collecting, maintaining, using, and disseminating the information

Title 13, U.S.C. Section 6c
Title 13, U.S.C. Section 141
Title 13, U.S.C. Section 193
44 U.S.C. Section 3101
41 U.S.C. 433(d)
5 U.S.C. 301
5 U.S.C. 1302, 3109, 3301, 3302, 3304, 3305, 3306, 3307, 309, 3313, 3317, 3318, 3319, 3326, 4103, 4723, 5532, and 5533, and Executive Order 9397
Executive Order 12107
Executive Order 12564

(i) The Federal Information Processing Standards (FIPS) 199 security impact category for the system

Moderate

Section 1: Status of the Information System

1.1 Indicate whether the information system is a new or existing system.

_____ This is a new information system.
__x__ This is an existing information system with changes that create new privacy risks. *(Check all that apply.)*

Changes That Create New Privacy Risks (CTCNPR)					
a. Conversions		d. Significant Merging		g. New Interagency Uses	
b. Anonymous to Non-Anonymous		e. New Public Access		h. Internal Flow or Collection	
c. Significant System Management Changes		f. Commercial Sources		i. Alteration in Character of Data	
j. Other changes that create new privacy risks (specify): New Third Party Fingerprinting solution.					

____ This is an existing information system in which changes do not create new privacy risks, and there is not a SAOP approved Privacy Impact Assessment.

Section 2: Information in the System

2.1 Indicate what personally identifiable information (PII)/business identifiable information (BII) is collected, maintained, or disseminated. *(Check all that apply.)*

Identifying Numbers (IN)					
a. Social Security*		e. File/Case ID		i. Credit Card	
b. Taxpayer ID		f. Driver's License		j. Financial Account	
c. Employer ID		g. Passport		k. Financial Transaction	
d. Employee ID	x	h. Alien Registration		l. Vehicle Identifier	
m. Other identifying numbers (specify):					
*Explanation for the need to collect, maintain, or disseminate the Social Security number, including truncated form:					

General Personal Data (GPD)					
a. Name	x	g. Date of Birth	x	m. Religion	
b. Maiden Name		h. Place of Birth		n. Financial Information	
c. Alias		i. Home Address	x	o. Medical Information	
d. Gender	x	j. Telephone Number	x	p. Military Service	
e. Age	x	k. Email Address	x	q. Physical Characteristics	
f. Race/Ethnicity	x	l. Education	x	r. Mother's Maiden Name	
s. Other general personal data (specify): Citizenship					

Work-Related Data (WRD)					
a. Occupation		d. Telephone Number		g. Salary	x
b. Job Title		e. Email Address		h. Work History	
c. Work Address		f. Business Associates			
i. Other work-related data (specify):					

Distinguishing Features/Biometrics (DFB)					
a. Fingerprints	x	d. Photographs		g. DNA Profiles	
b. Palm Prints		e. Scars, Marks, Tattoos		h. Retina/Iris Scans	
c. Voice Recording/Signatures		f. Vascular Scan		i. Dental Profile	
j. Other distinguishing features/biometrics (specify):					

System Administration/Audit Data (SAAD)					
a. User ID	x	c. Date/Time of Access	x	e. ID Files Accessed	x
b. IP Address	x	d. Queries Run	x	f. Contents of Files	
g. Other system administration/audit data (specify): Audit Logs					

Other Information (specify)

2.2 Indicate sources of the PII/BII in the system. *(Check all that apply.)*

Directly from Individual about Whom the Information Pertains					
In Person	x	Hard Copy: Mail/Fax	x	Online	x
Telephone	x	Email	x		
Other (specify):					

Government Sources					
Within the Bureau	x	Other DOC Bureaus		Other Federal Agencies	x
State, Local, Tribal		Foreign			
Other (specify):					

Non-government Sources					
Public Organizations		Private Sector	x	Commercial Data Brokers	x
Third Party Website or Application			x		
Other (specify):					

2.3 Describe how the accuracy of the information in the system is ensured.

CEN08 DITD uses a multitude of security controls mandated by the Federal Information Security Management Act of 2002 (FISMA) and various other regulatory control frameworks including the National Institute of Standards and Technology (NIST) special publication 800 series. These security controls include, but are not limited to data validation controls to ensure accuracy of information.

Selectee information is verified for accuracy when individuals schedule their fingerprints by verification against other forms of identification. Further, background checks are performed by the FBI to validate name, credit, and criminal history.

2.4 Is the information covered by the Paperwork Reduction Act?

	Yes, the information is covered by the Paperwork Reduction Act. Provide the OMB control number and the agency number for the collection.
x	No, the information is not covered by the Paperwork Reduction Act.

2.5 Indicate the technologies used that contain PII/BII in ways that have not been previously deployed. *(Check all that apply.)*

Technologies Used Containing PII/BII Not Previously Deployed (TUCPBNPD)			
Smart Cards		Biometrics	x
Caller-ID		Personal Identity Verification (PIV) Cards	
Other (specify):			

	There are not any technologies used that contain PII/BII in ways that have not been previously deployed.

Section 3: System Supported Activities

3.1 Indicate IT system supported activities which raise privacy risks/concerns. *(Check all that apply.)*

Activities			
Audio recordings		Building entry readers	
Video surveillance		Electronic purchase transactions	
Other (specify):			

x	There are not any IT system supported activities which raise privacy risks/concerns.

Section 4: Purpose of the System

4.1 Indicate why the PII/BII in the IT system is being collected, maintained, or disseminated. *(Check all that apply.)*

Purpose			
For a Computer Matching Program		For administering human resources programs	x
For administrative matters		To promote information sharing initiatives	
For litigation		For criminal law enforcement activities	x
For civil enforcement activities		For intelligence activities	
To improve Federal services online		For employee or customer satisfaction	
For web measurement and customization technologies (single-session)		For web measurement and customization technologies (multi-session)	
Other (specify): For Statistical Purposes (i.e. Censuses/Surveys)			

Section 5: Use of the Information

5.1 In the context of functional areas (business processes, missions, operations, etc.) supported by the IT system, describe how the PII/BII that is collected, maintained, or disseminated will be used. Indicate if the PII/BII identified in Section 2.1 of this document is in reference to a federal employee/contractor, member of the public, foreign national, visitor or other (specify).

The PII collected, maintained, and/or disseminated by the CEN08 IT system is in reference to members of the public. Data collection is used to produce national statistical information.

Third Party Fingerprinting is capturing selectee fingerprint data on behalf of the U.S Census Bureauto hire selectees to help conduct the 2020 Census operations. The third party vendor is mandated to only utilize FedRAMP authorized solutions. The vendor does not directly submit the fingerprint information to the FBI, rather the information is securely sent to the U.S Census Bureau for processing and submission to the FBI.

5.2 Describe any potential threats to privacy as a result of the bureau's/operating unit's use of the information, and controls that the bureau/operating unit has put into place to ensure that the information is handled, retained, and disposed appropriately. (For example: mandatory training for system users regarding appropriate handling of information, automatic purging of information in accordance with the retention schedule, etc.)

CEN08 DITD systems adhere to the Information Technology Security Program Policy as it relates to handling, retaining, and disposing collected information. Census Bureau information technology systems employ a multitude of layered security controls to protect PII at rest, during processing, as well as in transit. These NIST 800-53 controls, at a minimum, are deployed and managed at the enterprise level including, but not limited to the following:

- Intrusion Detection | Prevention Systems (IDS | IPS)
- Firewalls
- Mandatory use of HTTP(S) for Census Public facing websites
- Use of trusted internet connection (TIC)
- Anti-Virus software to protect host/end user systems
- Encryption of databases (Data at rest)
- HSPD-12 Compliant PIV cards
- Access Controls

Census Bureau information technology systems also follow the National Institute of Standards and Technology (NIST) standards including special publications 800-53, 800-63, 800-37 etc. Any system within the Census Bureau that contains, transmits, or processes PII has a current authority to operate (ATO) and goes through continuous monitoring on a yearly basis to ensure controls are implemented and operating as intended. The Census Bureau also deploys a Data Loss Prevention (DLP) solution as well.

Fingerprints are retained for 120 days and then disposed of following NIST sanitation guidance. All individuals that handle PII are required to complete annual Data Stewardship Awareness training.

Section 6: Information Sharing and Access

6.1 Indicate with whom the bureau intends to share the PII/BII in the IT system and how the PII/BII will be shared. *(Check all that apply.)*

Recipient	How Information will be Shared		
	Case-by-Case	Bulk Transfer	Direct Access
Within the bureau	x	x	
DOC bureaus			
Federal agencies		x	
State, local, tribal gov't agencies			
Public			
Private sector			
Foreign governments			
Foreign entities			
Other (specify):			

	The PII/BII in the system will not be shared.

6.2 Indicate whether the IT system connects with or receives information from any other IT systems authorized to process PII and/or BII.

x	Yes, this IT system connects with or receives information from another IT system(s) authorized to process PII and/or BII. Provide the name of the IT system and describe the technical controls which prevent PII/BII leakage: Decennial Applications – Shares information internally within the Census Bureau with CEN07 GEO, CEN05 Field, CEN11 DEMO, CEN18 CDL, CEN19 CEDSCI, CEN21 DAPPS, CEN30 ACS, and CEN36 ADEP. Third Party Fingerprinting – Shares information internally within the Census Bureau with CEN21 CHEC and CEN18 SOA. R&A – Shares information internally within the Census Bureau with CEN21 DAPPS CEN08 uses a multitude of security controls mandated by the Federal Information Security Management Act of 2002 (FISMA) and various other regulatory control frameworks including the National Institute of Standards and Technology (NIST) special publication 800 series. These security controls include, but are not limited to the use of mandatory HTTPS for public facing websites, access controls, anti-virus solutions, enterprise auditing/monitoring, encryption of data at rest, and various physical controls at Census Bureau facilities that house Information Technology systems. The Census Bureau also deploys an enterprise Data Loss Protection (DLP) solution as well.
	No, this IT system does not connect with or receive information from another IT system(s) authorized to process PII and/or BII.

6.3 Identify the class of users who will have access to the IT system and the PII/BII. *(Check all that apply.)*

Class of Users

General Public		Government Employees	x
Contractors	x		
Other (specify):			

Section 7: Notice and Consent

7.1 Indicate whether individuals will be notified if their PII/BII is collected, maintained, or disseminated by the system. *(Check all that apply.)*

x	Yes, notice is provided pursuant to a system of records notice published in the Federal Register and discussed in Section 9.	
x	Yes, notice is provided by a Privacy Act statement and/or privacy policy. The Privacy Act statement and/or privacy policy can be found at: http://www.census.gov/about/policies/privacy/privacy-policy.html	
	Yes, notice is provided by other means.	Specify how:
	No, notice is not provided.	Specify why not:

7.2 Indicate whether and how individuals have an opportunity to decline to provide PII/BII.

x	Yes, individuals have an opportunity to decline to provide PII/BII.	Specify how: Employment with the U.S. Census Bureau is voluntary.
x	No, individuals do not have an opportunity to decline to provide PII/BII.	Specify why not: According to Title 13, Section 221 (Census, Refusal or neglect to answer questions; false answers) of the United States Code, persons who fail or refuse to respond to the mail-back census form, or refuse to respond to a follow-up Census Bureau taker can be fined up to $100. Persons who knowingly provide false information to the Census Bureau can be fined up to $500. For Census Bureau employees who access systems in CEN08 providing PII is a requirement for employment.

7.3 Indicate whether and how individuals have an opportunity to consent to particular uses of their PII/BII.

x	Yes, individuals have an opportunity to consent to particular uses of their PII/BII.	Specify how: Employment with the U.S. Census Bureau is voluntary. Temporary hires and contractors have to consent to the U.S. Census Bureau uses of their PII.
x	No, individuals do not have an opportunity to consent to particular uses of their PII/BII.	Specify why not: For records covered by SORN Census-5, Decennial Census Programs, there are no access and consent requirements since the data is collected for statistical purposes only. For Census Bureau employees consent to particular uses of PII is a requirement for employment.

7.4 Indicate whether and how individuals have an opportunity to review/update PII/BII pertaining to them.

x	Yes, individuals have an opportunity to review/update PII/BII pertaining to them.	Specify how: Selectee information is verified for accuracy when individuals schedule their fingerprints.
x	No, individuals do not have an opportunity to review/update PII/BII pertaining to them.	Specify why not: For records covered by SORN Census-5, Decennial Census Programs, there is no opportunity to review/update data unless the Census Bureau contacts the respondent for an update on their information. For Census Employees, employees have access to PII via the appropriate Human Resources applications that reside outside of the CEN08 IT system.

Section 8: Administrative and Technological Controls

8.1 Indicate the administrative and technological controls for the system. *(Check all that apply.)*

x	All users signed a confidentiality agreement or non-disclosure agreement.
x	All users are subject to a Code of Conduct that includes the requirement for confidentiality.
x	Staff (employees and contractors) received training on privacy and confidentiality policies and practices.
x	Access to the PII/BII is restricted to authorized personnel only.
x	Access to the PII/BII is being monitored, tracked, or recorded. Explanation: Only authorized government/contractor personnel are allowed to access PII within a system. Authorizations for users occur yearly, at a minimum in accordance with applicable Bureau, Agency, and Federal policies/guidelines. In additional to IT system processes that handle PII, all manual extractions for PII are logged and recorded per Department of Commerce Policy, the NIST 800-53 Appendix J Privacy Control Catalog, and specifically NIST control AU-03, Content of Audit records.
x	The information is secured in accordance with FISMA requirements. Provide date of most recent Assessment and Authorization (A&A): July 25, 2018 [ ] This is a new system. The A&A date will be provided when the A&A package is approved.
x	The Federal Information Processing Standard (FIPS) 199 security impact category for this system is a moderate or higher.
x	NIST Special Publication (SP) 800-122 and NIST SP 800-53 Revision 4 Appendix J recommended security controls for protecting PII/BII are in place and functioning as intended; or have an approved Plan of Action and Milestones (POA&M).
x	A security assessment report has been reviewed for the supporting information system and it has been determined that there are no additional privacy risks.
x	Contractors that have access to the system are subject to information security provisions in their contracts required by DOC policy.
	Contracts with customers establish ownership rights over data including PII/BII.
	Acceptance of liability for exposure of PII/BII is clearly defined in agreements with customers.
	Other (specify):

8.2 Provide a general description of the technologies used to protect PII/BII on the IT system. *(Include data encryption in transit and/or at rest, if applicable).*

Census Bureau information technology systems employ a multitude of layered security controls to protect PII at rest, during processing, as well as in transit. These NIST 800-53 controls, at a minimum, are deployed and managed at the enterprise level including, but not limited to the following:

- Intrusion Detection | Prevention Systems (IDS | IPS)
- Firewalls
- Mandatory use of HTTP(S) for Census Public facing websites
- Use of trusted internet connection (TIC)
- Anti-Virus software to protect host/end user systems
- Encryption of databases (Data at rest)
- HSPD-12 Compliant PIV cards
- Access Controls

Census Bureau information technology systems also follow the National Institute of Standards and Technology (NIST) standards including special publications 800-53, 800-63, 800-37 etc. Any system within the Census Bureau that contains, transmits, or processes PII has a current authority to operate (ATO) and goes through continuous monitoring on a yearly basis to ensure controls are implemented and operating as intended. The Census Bureau also deploys a Data Loss Prevention solution as well.

Section 9: Privacy Act

9.1 Indicate whether a system of records is being created under the Privacy Act, 5 U.S.C. § 552a. *(A new system of records notice (SORN) is required if the system is not covered by an existing SORN).*

As per the Privacy Act of 1974, "the term 'system of records' means a group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual."

x	Yes, this system is covered by an existing system of records notice (SORN). Provide the SORN name, number, and link. *(list all that apply)*: COMMERCE/CENSUS-5, Decennial Census Program- http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-5.html COMMERCE/DEPT-13, Investigative & Security Records- http://osec.doc.gov/opog/PrivacyAct/SORNs/DEPT-13.html COMMERCE/DEPT-18, Employees Personnel Files Not Covered by Notices of Other Agencies- http://www.osec.doc.gov/opog/PrivacyAct/SORNs/DEPT-18.html OPM SORN GOVT-5, Recruiting, Examining and Placement Records- https://www.opm.gov/information-management/privacy-policy/sorn/opm-sorn-govt-5-recruiting-examining-and-placement-records.pdf
	Yes, a SORN has been submitted to the Department for approval on (date).
	No, this system is not a system of records and a SORN is not applicable.

Section 10: Retention of Information

10.1 Indicate whether these records are covered by an approved records control schedule and monitored for compliance. *(Check all that apply.)*

X	There is an approved record control schedule. Provide the name of the record control schedule: Nl-29-05-01, N1-29-10-5, GRS 3.1, GRS 5.6 item 181
	No, there is not an approved record control schedule. Provide the stage in which the project is in developing and submitting a records control schedule:
X	Yes, retention is monitored for compliance to the schedule.
	No, retention is not monitored for compliance to the schedule. Provide explanation:

10.2 Indicate the disposal method of the PII/BII. *(Check all that apply.)*

Disposal			
Shredding	X	Overwriting	X
Degaussing	X	Deleting	X
Other (specify):			

Section 11: NIST Special Publication 800-122 PII Confidentiality Impact Level

11.1 Indicate the potential impact that could result to the subject individuals and/or the organization if PII were inappropriately accessed, used, or disclosed. *(The PII Confidentiality Impact Level is not the same as the Federal Information Processing Standards (FIPS) 199 security impact category.)*

	Low – the loss of confidentiality, integrity, or availability could be expected to have a limited adverse effect on organizational operations, organizational assets, or individuals.
	Moderate – the loss of confidentiality, integrity, or availability could be expected to have a serious adverse effect on organizational operations, organizational assets, or individuals.
X	High – the loss of confidentiality, integrity, or availability could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, or individuals.

11.2 Indicate which factors were used to determine the above PII confidentiality impact levels. *(Check all that apply.)*

X	Identifiability	PII collected can be directly and indirectly used to identify individuals.
X	Quantity of PII	The collection is for the decennial census, therefore, a severe or substantial number of individuals would be affected if there was loss, theft or compromise of the data. This could affect decennial 2020 Census response rates and have a long term effect on the Nation's population count. Severe collective harm to the USCB's

		reputation, or cost to the USCB in addressing a breach.
X	Data Field Sensitivity	The PII, alone or in combination, may be relevant in some other contexts and may, in those contexts, make the individuals or the Census Bureau vulnerable to harm.
X	Context of Use	Disclosure of PII in this IT system or the PII itself may result in severe harm to the individual or organization.
X	Obligation to Protect Confidentiality	PII collected is required to be protected in accordance with 5, U.S.C (552a) and 13, U.S.C, section 9.
X	Access to and Location of PII	The PII is located on computers (including laptops) and on a network, and IT systems controlled by the Census Bureau. Access is limited to those with a need-to-know including the Census Bureau regional offices and survey program offices, etc. Access is allowed by Census Bureau-owned equipment outside of the physical locations owned by the Census Bureau only with a secure connection. Backups are stored at Census Bureau-owned facilities. PII is also located on U.S. Census Bureau authorized vendor systems. Access is limited to those with a need-to-know for authorized U.S. Census Bureau contractors and employees.
	Other:	Provide explanation:

Section 12: Analysis

12.1 Identify and evaluate any potential threats to privacy that exist in light of the information collected or the sources from which the information is collected. Also, describe the choices that the bureau/operating unit made with regard to the type or quantity of information collected and the sources providing the information in order to prevent or mitigate threats to privacy. (For example: If a decision was made to collect less data, include a discussion of this decision; if it is necessary to obtain information from sources other than the individual, explain why.)

A third party vendor is capturing selectee fingerprint data and PII on behalf of the U.S Census Bureau. The third party vendor is mandated to only utilize authorized systems and FedRAMP solutions. The vendor does not directly submit the fingerprint information to the FBI, rather the information is securely sent to the U.S Census Bureau for processing and submission to the FBI.

12.2 Indicate whether the conduct of this PIA results in any required business process changes.

	Yes, the conduct of this PIA results in required business process changes. Explanation:
x	No, the conduct of this PIA does not result in any required business process changes.

12.3 Indicate whether the conduct of this PIA results in any required technology changes.

	Yes, the conduct of this PIA results in required technology changes. Explanation:
x	No, the conduct of this PIA does not result in any required technology changes.

U.S. Department of Commerce
U.S. Census Bureau

Privacy Impact Assessment
for the
CEN11 Demographic Census, Surveys, and Special Processing

Reviewed by: _____ 4/19/18 , Bureau Chief Privacy Officer

☑ Concurrence of Senior Agency Official for Privacy/DOC Chief Privacy Officer

☐ Non-concurrence of Senior Agency Official for Privacy/DOC Chief Privacy Officer

CATRINA PURVIS

Digitally signed by CATRINA PURVIS
DN: c=US, o=U.S. Government, ou=Department of Commerce, ou=Office of the Secretary, cn=CATRINA PURVIS, 0.9.2342.19200300.100.1.1=13001002875743
Date: 2018.06.22 13:55:40 -04'00'

Signature of Senior Agency Official for Privacy/DOC Chief Privacy Officer Date

U.S. Department of Commerce Privacy Impact Assessment
U.S. Census Bureau/Demographic Census, Surveys, and Special Processing

Unique Project Identifier: 006-000400500

Introduction: System Description

Provide a description of the system that addresses the following elements:
The response must be written in plain language and be as comprehensive as necessary to describe the system.

(a) a general description of the information in the system

The U.S. Census Bureau's CEN11 Demographic Census, Surveys, and Special Processing System is a system comprised of components that support the Demographic Directorate business functions. The component within CEN11 that contains PII is a Commercial off the Shelf (COTS) product used by Census Bureau demographic programs for data access, transformation, reporting, and statistical analysis. Some of the demographic data maintained in this system are Name, date of birth, telephone number, occupation, military service information (if applicable), medical information, financial information etc.

(b) a description of a typical transaction conducted on the system

The survey data for demographic programs is collected using a multi-mode approach made up of:

- Face-to-face or telephone induction interviews
- Field Representatives (FR)
- Web-based respondents

The information is collected by using:

Web-based respondents use a web-based application instrument that resides on the Census Bureau network via CEN15 Centurion. Respondents use their personal computers to access Centurion.

Introduction interviews are collected using electronic instruments – Computer-Assisted Personal Interviewing (CAPI) (a CEN05 component).

Once the information is collected by the instruments, the information is stored in a repository for use.

(c) any information sharing conducted by the system

There is PII being shared as follows:

- The Census Bureau provides access to staff at Department of Housing and Urban

Development (HUD) with Special Sworn Status (SSS) for the American Housing Survey (AHS) via the Census Bureau Virtual Desktop Infrastructure (VDI).

- The Census Bureau provides some Bureau of Labor Statistics (BLS) Staff access to Current Population Survey (CPS) and Consumer Expenditure Survey (CES) data on a CEN11 Server in the DMZ. BLS Staff have Special Sworn Status (SSS) to have Census accounts to access the server. Consumer Expenditure staff at BLS access CPS data for weighting purposes.
- Census Bureau sends a file containing respondent addresses to BLS for the Telephone Point of Purchase Survey (TPOPS).
- The Census Bureau sends data files to the National Center for Health Statistics (NCHS) for the National Ambulatory Medical Care Survey (NAMCS), the National Hospital Ambulatory Medical Care Survey (N(H)AMCS) and National Health Interview Survey (NHIS). Data transfers are conducted through the Centers for Disease Control and Prevention (CDC) Secure Access Management Services (SAMS).
- The Census Bureau sends data files to the National Center for Education Statistics (NCES) for the National Household Education Survey (NHES), the School Survey on Crime and Safety (SSOCS), the Schools and Staffing Survey (SSS), the Private School Survey (PSS), the National Teacher and Principal Survey (NTPS), the Teacher Follow-up Survey (TFS), the Principal Follow-Up Survey (PFS) and the Beginning Teacher Longitudinal Survey (BTLS). Data transfers are conducted through the Institute of Education Sciences (IES) Members Site.

(d) a citation of the legal authority to collect PII and/or BII

13 USC sections 8(b), 182 and 18 U.S.C. 2510-2521

(e) the Federal Information Processing Standard (FIPS) 199 security impact category for the system

This is categorized as a moderate system.

Section 1: Status of the Information System

1.1 Indicate whether the information system is a new or existing system.

____ This is a new information system.
____ This is an existing information system with changes that create new privacy risks. *(Check all that apply.)*

Changes That Create New Privacy Risks (CTCNPR)					
a. Conversions		d. Significant Merging		g. New Interagency Uses	
b. Anonymous to Non-Anonymous		e. New Public Access		h. Internal Flow or Collection	
c. Significant System Management Changes		f. Commercial Sources		i. Alteration in Character of Data	
j. Other changes that create new privacy risks (specify):					

__x_ This is an existing information system without changes that create new privacy risks.

Section 2: Information in the System

2.1 Indicate what personally identifiable information (PII)/business identifiable information (BII) is collected, maintained, or disseminated. *(Check all that apply.)*

Identifying Numbers (IN)					
a. Social Security*	x	e. File/Case ID		i. Credit Card	
b. Taxpayer ID	x	f. Driver's License		j. Financial Account	
c. Employer ID	x	g. Passport		k. Financial Transaction	
d. Employee ID		h. Alien Registration		l. Vehicle Identifier	
m. Other identifying numbers (specify):					

*Explanation for the need to collect, maintain, or disseminate the Social Security number, including truncated form: The last 4-digits of the SSN is used in admin records matching to the National Health Interview Survey (NHIS).

The justification for the necessity of collecting this information, taken from the latest approved OMB ICR supporting statement is below:

Social Security Number and Health Insurance Claim Number: The last four digits of the Social Security Number (SSN) is asked on the NHIS questionnaire to allow linkage with administrative and vital records, such as the National Death Index (NDI). The NDI is a computerized central file of death record information. It is compiled from data obtained by NCHS from the State vital statistics offices. The data contain a standard set of identi-fying information on decedents from 1979 to the present. Records are matched using Social Security Number and other variables such as name, father's surname, date of birth, sex, state of resi-dence, and marital status. Of these, Social Security Number is the most important identifier for successful matching. The last four digits has been shown to be nearly as effective for matching as the full number.

The Social Security Number is also used by the Medical Expenditure Panel Study to help track the location of respondents who have changed residence since their NHIS interview. Finding a correct address for respondents is essential to maintaining response levels at an acceptable level in linked surveys, and the Social Security Number

is a key item for establishing a correct address. Medicare beneficiaries are given a health insurance claim (HIC) number that is their (or their spouse's) SSN with an alphabetic prefix. The NHIS also asks for the last four digits of that number so that the NHIS data can be linked to Medicare claims information for purposes of statistical research.

General Personal Data (GPD)					
a. Name	x	g. Date of Birth	x	m. Religion	
b. Maiden Name	x	h. Place of Birth	x	n. Financial Information	x
c. Alias	x	i. Home Address	x	o. Medical Information	x
d. Gender	x	j. Telephone Number	x	p. Military Service	x
e. Age	x	k. Email Address	x	q. Physical Characteristics	x
f. Race/Ethnicity	x	l. Education	x	r. Mother's Maiden Name	
s. Other general personal data (specify): Citizenship					

Work-Related Data (WRD)					
a. Occupation	x	d. Telephone Number	x	g. Salary	x
b. Job Title	x	e. Email Address	x	h. Work History	x
c. Work Address	x	f. Business Associates			
i. Other work-related data (specify):					

Distinguishing Features/Biometrics (DFB)					
a. Fingerprints		d. Photographs		g. DNA Profiles	
b. Palm Prints		e. Scars, Marks, Tattoos		h. Retina/Iris Scans	
c. Voice Recording/Signatures	x	f. Vascular Scan		i. Dental Profile	
j. Other distinguishing features/biometrics (specify):					

System Administration/Audit Data (SAAD)					
a. User ID	x	c. Date/Time of Access	x	e. ID Files Accessed	
b. IP Address		d. Queries Run		f. Contents of Files	
g. Other system administration/audit data (specify):					

Other Information (specify)

2.2 Indicate sources of the PII/BII in the system. *(Check all that apply.)*

Directly from Individual about Whom the Information Pertains					
In Person	x	Hard Copy: Mail/Fax	x	Online	x
Telephone	x	Email	x		
Other (specify):					

Government Sources					
Within the Bureau	x	Other DOC Bureaus		Other Federal Agencies	
State, Local, Tribal	x	Foreign			
Other (specify):					

Non-government Sources					
Public Organizations		Private Sector		Commercial Data Brokers	x
Third Party Website or Application			x		
Other (specify):					

2.3 Indicate the technologies used that contain PII/BII in ways that have not been previously deployed. *(Check all that apply.)*

Technologies Used Containing PII/BII Not Previously Deployed (TUCPBNPD)			
Smart Cards		Biometrics	
Caller-ID		Personal Identity Verification (PIV) Cards	
Other (specify):			

x	There are not any technologies used that contain PII/BII in ways that have not been previously deployed.

Section 3: System Supported Activities

3.1 Indicate IT system supported activities which raise privacy risks/concerns. *(Check all that apply.)*

Activities			
Audio recordings	x	Building entry readers	
Video surveillance		Electronic purchase transactions	
Other (specify): For Computer Assisted Personal Interviewing (CAPI), the Demographic Programs Directorate (DEMO) uses Computer-audio recording interviewing (CARI) for select interviews for the Survey of Income and Program Participation (SIPP); future plans are to incorporate CARI for all demographic CAPI surveys. For Computer Assisted Telephone Interviewing (CATI) surveys, the NICE Sentinel 2.5 system is used to record selected interviews for Quality Assurance (QA) purposes. Both types of recordings contain PII.			

	There are not any IT system supported activities which raise privacy risks/concerns.

Section 4: Purpose of the System

4.1 Indicate why the PII/BII in the IT system is being collected, maintained, or disseminated. *(Check all that apply.)*

Purpose			
To determine eligibility		For administering human resources programs	
For administrative matters		To promote information sharing initiatives	
For litigation		For criminal law enforcement activities	
For civil enforcement activities		For intelligence activities	
To improve Federal services online		For employee or customer satisfaction	
For web measurement and customization technologies (single-session)		For web measurement and customization technologies (multi-session)	
Other (specify): For statistical purposes (i.e., Censuses/Surveys)			

Section 5: Use of the Information

5.1 In the context of functional areas (business processes, missions, operations, etc.) supported by the IT system, describe how the PII/BII that is collected, maintained, or disseminated will be used. Indicate if the PII/BII identified in Section 2.1 of this document is in reference to a federal employee/contractor, member of the public, foreign national, visitor or other (specify).

PII is used to process numerous national statistical surveys and other demographic data programs.

The data is used to calculate, process, and manipulate the statistical data input for the purpose of creating statistical information and reports (i.e. Annual household and group quarters population estimates by age, sex, race, and origin for counties).

The information that is collected in this system is from members of the public.

Section 6: Information Sharing and Access

6.1 Indicate with whom the bureau intends to share the PII/BII in the IT system and how the PII/BII will be shared. *(Check all that apply.)*

Recipient	How Information will be Shared		
	Case-by-Case	Bulk Transfer	Direct Access
Within the bureau	x	x	x

DOC bureaus			
Federal agencies		x	
State, local, tribal gov't agencies			
Public			
Private sector			
Foreign governments			
Foreign entities			
Other (specify):			

	The PII/BII in the system will not be shared.

6.2 Indicate whether the IT system connects with or receives information from any other IT systems authorized to process PII and/or BII.

x	Yes, this IT system connects with or receives information from another IT system(s) authorized to process PII and/or BII. Provide the name of the IT system and describe the technical controls which prevent PII/BII leakage: CEN11 shares information with CEN03, CEN07, CEN08, CEN13, CEN14, CEN18, CEN19, and CEN30. CEN11 receives information from CEN13, CEN30, and CEN35. The CEN11 IT system uses a multitude of security controls mandated by the Federal Information Security Management Act of 2002 (FISMA) and various other regulatory control frameworks including the National Institute of Standards and Technology (NIST) special publication 800 series. These security controls include, but are not limited to the use of mandatory HTTPS for public facing websites, access controls, anti-virus solutions, enterprise auditing/monitoring, encryption of data at rest, and various physical controls at Census Bureau facilities that house Information Technology systems. The Census Bureau also deploys an enterprise Data Loss Protection (DLP) solution as well.
	No, this IT system does not connect with or receive information from another IT system(s) authorized to process PII and/or BII.

6.3 Identify the class of users who will have access to the IT system and the PII/BII. *(Check all that apply.)*

Class of Users			
General Public		Government Employees	x
Contractors	x		
Other (specify):			

Section 7: Notice and Consent

7.1 Indicate whether individuals will be notified if their PII/BII is collected, maintained, or disseminated by the system. *(Check all that apply.)*

x	Yes, notice is provided pursuant to a system of records notice published in the Federal Register and discussed in Section 9.	
x	Yes, notice is provided by a Privacy Act statement and/or privacy policy. The Privacy Act statement and/or privacy policy can be found at: https://www.census.gov/about/policies/privacy/privacy-policy.html	
x	Yes, notice is provided by other means.	Specify how: Official correspondence letter or email from the Census Bureau to respondents.
	No, notice is not provided.	Specify why not:

7.2 Indicate whether and how individuals have an opportunity to decline to provide PII/BII.

x	Yes, individuals have an opportunity to decline to provide PII/BII.	Specify how: CEN11 surveys are voluntary. Individuals may refuse to participate in the survey or, if they do participate, they may refuse to answer specific questions.
	No, individuals do not have an opportunity to decline to provide PII/BII.	Specify why not:

7.3 Indicate whether and how individuals have an opportunity to consent to particular uses of their PII/BII.

	Yes, individuals have an opportunity to consent to particular uses of their PII/BII.	Specify how:
x	No, individuals do not have an opportunity to consent to particular uses of their PII/BII.	Specify why not: For records covered under SORNs Census-3 and SORN Census-7 the data is collected for statistical purposes and there is no opportunity to consent to uses of the data.

7.4 Indicate whether and how individuals have an opportunity to review/update PII/BII pertaining to them.

	Yes, individuals have an opportunity to review/update PII/BII pertaining to them.	Specify how:
x	No, individuals do not have an opportunity to review/update PII/BII pertaining to them.	Specify why not: For records covered under SORNs Census-3 and SORN Census-7 there is no access to the records since the data is collected for statistical purposes.

Section 8: Administrative and Technological Controls

8.1 Indicate the administrative and technological controls for the system. *(Check all that apply.)*

x	All users signed a confidentiality agreement or non-disclosure agreement.
x	All users are subject to a Code of Conduct that includes the requirement for confidentiality.
x	Staff (employees and contractors) received training on privacy and confidentiality policies and practices.
x	Access to the PII/BII is restricted to authorized personnel only.
x	Access to the PII/BII is being monitored, tracked, or recorded. Explanation:
x	The information is secured in accordance with FISMA requirements. Provide date of most recent Assessment and Authorization (A&A): 7/14/2017 ☐ This is a new system. The A&A date will be provided when the A&A package is approved.
x	The Federal Information Processing Standard (FIPS) 199 security impact category for this system is a moderate or higher.
x	NIST Special Publication (SP) 800-122 and NIST SP 800-53 Revision 4 Appendix J recommended security controls for protecting PII/BII are in place and functioning as intended; or have an approved Plan of Action and Milestones (POA&M).
x	Contractors that have access to the system are subject to information security provisions in their contracts required by DOC policy.
	Contracts with customers establish ownership rights over data including PII/BII.
	Acceptance of liability for exposure of PII/BII is clearly defined in agreements with customers.
x	Other (specify): Publications are approved by the Disclosure Review Board

8.2 Provide a general description of the technologies used to protect PII/BII on the IT system. *(Include data encryption in transit and/or at rest, if applicable).*

Census Bureau Information technology systems employ a multitude of layered security controls to protect PII/BII at rest during processing, as well as in transit. These NIST 800-53 controls, at a minimum, are deployed and managed at the enterprise level including, but limited to the following:

- Intrusion Detection / Prevention Systems (IDA/IPS)
- Firewalls
- Mandatory use of HTTP(S) for Census Public facing websites
- Use of trusted internet connection (TIC)
- Anti-Virus software to protect host/end user systems
- Encryption of databases (Data at rest)
- HSPD-12 Compliant PIV cards
- Access Controls

Census Bureau Information technology systems also follow the National Institute of Standards and Technology (NIST) standards, including special publications 800-53, 800-63, 800-37 etc. Any system within the Census that contains, transmits, or processes BII/PII has a current authority to operate (ATO) and goes through continuous monitoring on a yearly basis to ensure controls are implanted and operating as intended. Census also deploys a DLP solution as well.

Section 9: Privacy Act

9.1 Indicate whether a system of records is being created under the Privacy Act, 5 U.S.C. § 552a. *(A new system of records notice (SORN) is required if the system is not covered by an existing SORN).*

As per the Privacy Act of 1974, "the term 'system of records' means a group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual."

	Yes, this system is covered by an existing system of records notice (SORN). Provide the SORN name and number *(list all that apply)*:
x	Yes, a SORN has been submitted to the Department for approval on (date). 1. Census -3, Demographic Survey Collection (Census Bureau Sampling Frame) submitted on 7/22/16. 2. Census-7, Demographic Survey Collection (non-Census Bureau Sampling Frame) submitted on 7/22/16.
	No, a SORN is not being created.

Section 10: Retention of Information

10.1 Indicate whether these records are covered by an approved records control schedule and monitored for compliance. *(Check all that apply.)*

x	There is an approved record control schedule. Provide the name of the record control schedule: N1-29-99-5, N1-29-89-3, N1-29-87-3, N1-29-86-3, NC1-29-85-1, NC1-29-79-7, and GRS 3.1 GRS 3.2 GRS 4.1, GRS 4.3
	No, there is not an approved record control schedule. Provide the stage in which the project is in developing and submitting a records control schedule:
x	Yes, retention is monitored for compliance to the schedule.
	No, retention is not monitored for compliance to the schedule. Provide explanation:

10.2 Indicate the disposal method of the PII/BII. *(Check all that apply.)*

Disposal			
Shredding	x	Overwriting	x
Degaussing		Deleting	x
Other (specify):			

Section 11: NIST Special Publication 800-122 PII Confidentiality Impact Levels

11.1 Indicate the potential impact that could result to the subject individuals and/or the organization if PII were inappropriately accessed, used, or disclosed.

	Low – the loss of confidentiality, integrity, or availability could be expected to have a limited adverse effect on organizational operations, organizational assets, or individuals.
	Moderate – the loss of confidentiality, integrity, or availability could be expected to have a serious adverse effect on organizational operations, organizational assets, or individuals.
X	High – the loss of confidentiality, integrity, or availability could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, or individuals.

11.2 Indicate which factors were used to determine the above PII confidentiality impact levels. *(Check all that apply.)*

X	Identifiability	Provide explanation: PII/BII collected can be directly used to identify individuals
X	Quantity of PII	Provide explanation: The collection is for Census Bureau Censuses and surveys, therefore, a severe or catastrophic number of individuals would be affected if there was loss, theft or compromise of the data.
X	Data Field Sensitivity	Provide explanation: The PII/BII, alone or in combination, are directly usable in other contexts and make the individual or organization vulnerable to harms, such as identity theft, embarrassment, loss of trust, or costs.
X	Context of Use	Provide explanation: Disclosure of the act of collecting and using the PII/BII in this IT system or the PII/BII itself may result in severe or catastrophic harm to the individual or organization.
X	Obligation to Protect Confidentiality	Provide explanation: PII/BII collected is required to be protected in accordance with organization or mission- specific privacy laws, regulations, mandates, or organizational policy apply that add more restrictive requirements to government- wide or industry-specific requirements. Violations may result in severe civil or criminal penalties.
X	Access to and Location of PII	PII/BII is located on computers controlled by the Census Bureau or on mobile devices or storage media. Access is limited to certain populations of the Census Bureau's workforce and limited to Special Sworn Status individuals. Access is only allowed by organization-owned equipment outside of the physical locations, and only with a secured connection.
	Other:	Provide explanation:

Section 12: Analysis

12.1 Indicate whether the conduct of this PIA results in any required business process changes.

x	Yes, the conduct of this PIA results in required business process changes. Explanation: Next PTA will be update selection in question 2 to yes to address audio recordings and consistency with section 3.1 of the PIA.
	No, the conduct of this PIA does not result in any required business process changes.

12.2 Indicate whether the conduct of this PIA results in any required technology changes.

	Yes, the conduct of this PIA results in required technology changes. Explanation:
x	No, the conduct of this PIA does not result in any required technology changes.

Version Number: 01-2015

U.S. Department of Commerce
Bureau of the Census

Privacy Impact Assessment
for the
CEN13 Center for Economic Studies (CES)

Reviewed by: [signature] 4/19/18, Bureau Chief Privacy Officer

☑ Concurrence of Senior Agency Official for Privacy/DOC Chief Privacy Officer

☐ Non-concurrence of Senior Agency Official for Privacy/DOC Chief Privacy Officer

LISA MARTIN

Digitally signed by LISA MARTIN
DN: c=US, o=U.S. Government, ou=Department of Commerce, ou=Office of the Secretary, cn=LISA MARTIN, 0.9.2342.19200300.100.1.1=13001000105292
Date: 2018.06.26 14:07:58 -04'00'

For Dr. Catrina D. Purvis

Signature of Senior Agency Official for Privacy/DOC Chief Privacy Officer Date

U.S. Department of Commerce Privacy Impact Assessment
Bureau of the Census, CEN13 Center for Economic Studies (CES)

Unique Project Identifier: 006-000400700

Introduction: System Description

Provide a description of the system that addresses the following elements:

(a) a general description of the information in the system

The Center for Economic Studies IT System includes data maintained by the Federal Statistical Research Data Centers (FSRDCs), the Center for Economic Studies and the Center for Administrative Records Research and Applications (CARRA). The CEN13 CES IT system covers the personally identifiable information (PII) and Business identifiable information (BII) from each of the centers maintained by the system. The CES data holdings include census and survey data which may contain name, gender, age, date of birth etc. from across the Census Bureau, administrative records from other federal agencies, and proprietary data files from commercial vendors and some non-profits.

(b) a description of a typical transaction conducted on the system

For internal Census staff users, the Data Management System (DMS) is used to perform management and tracking functions for research proposal and active projects. The DMS is used to track the status and activity of all projects from initial conception through completion and close out.

For external FSRDC users, the CES Management System (CMS) is used to perform management and tracking functions for research proposals and active projects. The CMS is used to track the status and activity of all projects from initial conception through completion and close out. Data is available only to researchers who have received prior approval.

As an example, there may be a researcher wants to conduct an external project using the Census Bureau's 2012 Survey of Business Owners (SBO). Information about the researcher is collected to create an account in CMS; that account is associated with a project that is documented in CMS (e.g. datasets such as SBO), the date pertaining to the of the data (e.g. 2012), other researchers (i.e. research assistant), and the length of the project, (in this example three years). The CMS is used to track all required reviews for the proposal. Once the project and Special Sworn Status (SSS) for the individual is approved, the researcher is provided a badge to access the FSRDC facility and a user account to access the server. The researcher can then proceed to conduct their project. The CMS tracks ongoing activity during the life of the project: who works on the project, annual training for each person, annual reports, and disclosure requests. When the project is completed, the final report from the project and the archival of the project files.

(c) any information sharing conducted by the system

CES CEN13 has ISA ICD's with the following Census systems – data is shared with:

- ADEP ITO Associate Directorate for Economic Programs (CEN36)
- EAD Economic Census and Surveys and Special Processing (CEN03)
- GEO Geography (CEN07)
- DSD Demographic Census, Surveys and Special Processing (CEN11)

CES CEN13 has MOU's or contracts with the following systems external to Census:

- Agency for Healthcare Research and Quality
- Bureau of Labor Statistics
- Department of Energy
- National Center for Health Statistics
- Internal Revenue Service

CES also shares data within the Bureau, with designated Federal agencies, State, Local and Tribal governments and Non-Profit organizations as part of its mission.

(d) a citation of the legal authority to collect PII and/or BII:

13 U.S.C., Chapter 5, 8(b), 131, 132, and 182 and 13 U.S.C. 6

(e) the Federal Information Processing Standard (FIPS) 199 security impact category for the system is Moderate.

Section 1: Status of the Information System

1.1 Indicate whether the information system is a new or existing system.
Existing System, No New Security Risks.

____ This is a new information system.
____ This is an existing information system with changes that create new privacy risks. *(Check all that apply.)*

Changes That Create New Privacy Risks (CTCNPR)					
a. Conversions		d. Significant Merging		g. New Interagency Uses	
b. Anonymous to Non-Anonymous		e. New Public Access		h. Internal Flow or Collection	
c. Significant System Management Changes		f. Commercial Sources		i. Alteration in Character of Data	
j. Other changes that create new privacy risks (specify):					

_X_ This is an existing information system without changes that create new privacy risks.

Section 2: Information in the System

2.1 Indicate what personally identifiable information (PII)/business identifiable information (BII) is collected, maintained, or disseminated. *(Check all that apply.)*

Identifying Numbers (IN)					
a. Social Security*	X	e. File/Case ID	X	i. Credit Card	
b. Taxpayer ID	X	f. Driver's License	X	j. Financial Account	X
c. Employer ID	X	g. Passport		k. Financial Transaction	X
d. Employee ID	X	h. Alien Registration	X	l. Vehicle Identifier	
m. Other identifying numbers (specify):					
*SSNs are often included in administrative records datasets acquired in support of the Census Bureau's Title 13 authority to collect these data. When SSNs are present in these data they serve as one of several components used in a matching or look-up process to assign an anonymized protected identification key (PIK) to the record.					

General Personal Data (GPD)					
a. Name	X	g. Date of Birth	X	m. Religion	
b. Maiden Name	X	h. Place of Birth	X	n. Financial Information	X
c. Alias	X	i. Home Address	X	o. Medical Information	X
d. Gender	X	j. Telephone Number	X	p. Military Service	X
e. Age	X	k. Email Address	X	q. Physical Characteristics	
f. Race/Ethnicity	X	l. Education	X	r. Mother's Maiden Name	X
s. Other general personal data (specify):					

Work-Related Data (WRD)					
a. Occupation	X	d. Telephone Number	X	g. Salary	X
b. Job Title	X	e. Email Address	X	h. Work History	X
c. Work Address	X	f. Business Associates	X		
i. Other work-related data (specify):					

Distinguishing Features/Biometrics (DFB)					
a. Fingerprints		d. Photographs		g. DNA Profiles	
b. Palm Prints		e. Scars, Marks, Tattoos		h. Retina/Iris Scans	
c. Voice Recording/Signatures		f. Vascular Scan		i. Dental Profile	
j. Other distinguishing features/biometrics (specify):					

System Administration/Audit Data (SAAD)					
a. User ID	X	c. Date/Time of Access		e. ID Files Accessed	
b. IP Address		d. Queries Run		f. Contents of Files	
g. Other system administration/audit data (specify):					

Other Information (specify)

2.2 Indicate sources of the PII/BII in the system. *(Check all that apply.)*

Directly from Individual about Whom the Information Pertains					
In Person	X	Hard Copy: Mail/Fax		Online	
Telephone	X	Email	X		
Other (specify):					

Government Sources					
Within the Bureau	X	Other DOC Bureaus	X	Other Federal Agencies	X
State, Local, Tribal	X	Foreign			
Other (specify):					

Non-government Sources					
Public Organizations		Private Sector	X	Commercial Data Brokers	X
Third Party Website or Application					
Other (specify): Some data is received from Non-profit organizations					

2.3 Indicate the technologies used that contain PII/BII in ways that have not been previously deployed. *(Check all that apply.)*

Technologies Used Containing PII/BII Not Previously Deployed (TUCPBNPD)			
Smart Cards		Biometrics	
Caller-ID		Personal Identity Verification (PIV) Cards	
Other (specify):			

X	There are not any technologies used that contain PII/BII in ways that have not been previously deployed.

Section 3: System Supported Activities

3.1 Indicate IT system supported activities which raise privacy risks/concerns. *(Check all that apply.)*

Activities			
Audio recordings		Building entry readers	
Video surveillance		Electronic purchase transactions	
Other (specify):			

X	There are not any IT system supported activities which raise privacy risks/concerns.

Section 4: Purpose of the System

4.1 Indicate why the PII/BII in the IT system is being collected, maintained, or disseminated. *(Check all that apply.)*

Purpose			
To determine eligibility		For administering human resources programs	
For administrative matters	X	To promote information sharing initiatives	
For litigation		For criminal law enforcement activities	
For civil enforcement activities		For intelligence activities	
To improve Federal services online		For employee or customer satisfaction	
For web measurement and customization technologies (single-session)		For web measurement and customization technologies (multi-session)	
Other (specify): Research, Improvement/support of Census Bureau programs through use of administrative and other non-survey data, Quality assurance, and statistical purposes.			

Section 5: Use of the Information

5.1 In the context of functional areas (business processes, missions, operations, etc.) supported by the IT system, 1) describe how the PII/BII that is collected, maintained, or disseminated will be used. 2) Indicate if the PII/BII identified in Section 2.1 of this document is in reference to a federal employee/contractor, member of the public, foreign national, visitor or other (specify).

Administration and research mission of the CEN13 program for statistical purposes:

For administrative matters:
The PII/BII is used for record linkage. SSNs are used to assign protected identification keys (PIKs) after which SSN and other PII are dropped from the file. After the PIK is assigned, ,files are linked only by PIK. This PII/BII covers members of the public, businesses, contractors and federal employees.

Research, Improvement/support of Census Bureau programs through use of administrative and other non-survey data, Quality assurance, and statistical purposes:
Record linkage using BII and PIKs facilitates research to improve and support existing Census Bureau programs and creation of beta data products. These products use innovative techniques that leverage existing data and reduce the burden on respondents. This PII/BII covers members of the public, businesses, contractors and federal employees.

Section 6: Information Sharing and Access

6.1 Indicate with whom the bureau intends to share the PII/BII in the IT system and how the PII/BII will be shared. *(Check all that apply.)*

Recipient	How Information will be Shared		
	Case-by-Case	Bulk Transfer	Direct Access
Within the bureau	X	X	
DOC bureaus		X	
Federal agencies		X	
State, local, tribal gov't agencies		X	
Public			
Private sector			
Foreign governments			
Foreign entities			
Other (specify): Data is made available to approved researchers on the RDC and CES internal servers, the researchers are Sworn Census employees, Contractors or Special Sworn Status.	X		

	The PII/BII in the system will not be shared.

6.2 Indicate whether the IT system connects with or receives information from any other IT systems authorized to process PII and/or BII.

X	Yes, this IT system connects with or receives information from another IT system(s) authorized to process PII and/or BII. Provide the name of the IT system and describe the technical controls which prevent PII/BII leakage: CES CEN13 has shares data with the following Census Bureau systems: • ADEP ITO: Associate Directorate for Economic Programs (CEN36) • EAD: Economic Census and Surveys and Special Processing (CEN03) • GEO: Geography (CEN07) • DSD: Demographic Census, Surveys and Special Processing (CEN11) CES CEN13 receives data from the following Census Bureau systems: • ACSO: American Community Survey (CEN30) • ITMD: Foreign Trade Division Applications (CEN34) • DSCMO-DSSD: Decennial (CEN08) • ASCO: American Community Survey (CEN30) CES also receives data from within the Bureau, designated Federal agencies, State, Local and Tribal governments and Non-Profit organizations as part of its mission. CEN13 uses a multitude of security controls mandated by the Federal Information Security Management Act of 2002 (FISMA) and various other regulatory control frameworks including the National Institute of Standards and Technology (NIST) special publication 800 series. These security controls include, but are not limited to the use of mandatory HTTPS for public facing websites, access controls, anti-virus solutions, enterprise auditing/monitoring, encryption of data at rest, and various physical controls at

	Census Bureau facilities that house Information Technology systems. The Census Bureau also deploys an enterprise Data Loss Protection (DLP) solution as well.
	No, this IT system does not connect with or receive information from another IT system(s) authorized to process PII and/or BII.

6.3 Identify the class of users who will have access to the IT system and the PII/BII. *(Check all that apply.)*

Class of Users			
General Public		Government Employees	X
Contractors	X		
Other (specify): Special Sworn Status employees of the Census Bureau			

Section 7: Notice and Consent

7.1 Indicate whether individuals will be notified if their PII/BII is collected, maintained, or disseminated by the system. *(Check all that apply.)*

X	Yes, notice is provided pursuant to a system of records notice published in the Federal Register and discussed in Section 9.	
X	Yes, notice is provided by a Privacy Act statement and/or privacy policy. The Privacy Act statement and/or privacy policy can be found at: https://www.census.gov/about/policies/privacy/privacy-policy.html	
	Yes, notice is provided by other means.	Specify how:
	No, notice is not provided.	Specify why not:

7.2 Indicate whether and how individuals have an opportunity to decline to provide PII/BII.

X	Yes, individuals have an opportunity to decline to provide PII/BII.	Specify how: When survey is voluntary individuals may decline to provide PII/BII; however in the case of aggregated secondary data there is no interaction with an individual. Survey data are not collected by CEN13 CES.
X	No, individuals do not have an opportunity to decline to provide PII/BII.	Specify why not: When survey is mandatory individuals may not decline to provide PII/BII, and in the case of aggregated secondary data, there is no interaction with an individual. Survey data are not collected by CEN13 CES.

7.3 Indicate whether and how individuals have an opportunity to consent to particular uses of their PII/BII.

X	Yes, individuals have an opportunity to consent to particular uses of their PII/BII.	Specify how: When survey is voluntary individuals may consent to particular uses of their PII/BII; however in the case of aggregated secondary data there is no interaction with an individual. Survey data are not collected by CEN13 CES.
X	No, individuals do not have an opportunity to consent to particular	Specify why not: In the case of aggregated secondary data, there is no interaction with an individual. Survey data are not

	uses of their PII/BII.	collected by CEN13 CES.

7.4 Indicate whether and how individuals have an opportunity to review/update PII/BII pertaining to them.

	Yes, individuals have an opportunity to review/update PII/BII pertaining to them.	Specify how:
X	No, individuals do not have an opportunity to review/update PII/BII pertaining to them.	Specify why not: In the case of aggregated secondary data there is no direct contact with the individual.

Section 8: Administrative and Technological Controls

8.1 Indicate the administrative and technological controls for the system. *(Check all that apply.)*

X	All users signed a confidentiality agreement or non-disclosure agreement.
X	All users are subject to a Code of Conduct that includes the requirement for confidentiality.
X	Staff (employees and contractors) received training on privacy and confidentiality policies and practices.
X	Access to the PII/BII is restricted to authorized personnel only.
X	Access to the PII/BII is being monitored, tracked, or recorded. Explanation: All individual activities within PII systems are logged, access is controlled by Access Control Lists(ACL) and all controls are reviewed in accordance with Audit and Accountability controls and Continuous Monitoring as specified in NIST 800-53 Revision-4. Only authorized government/contractor personnel are allowed to access PII/BII within a system. Authorizations for users occur yearly, at a minimum in accordance with applicable Bureau, Agency, and Federal policies/guidelines. In additional to system processes that handle PII/BII, all manual extractions for PII/BII are logged and recorded per Department of Commerce Policy, the NIST 800-53 Appendix J Privacy Control Catalog, and specifically NIST control AU-03, Content of Audit records.
X	The information is secured in accordance with FISMA requirements. Provide date of most recent Assessment and Authorization (A&A): July 13, 2017 ☐ This is a new system. The A&A date will be provided when the A&A package is approved.
X	The Federal Information Processing Standard (FIPS) 199 security impact category for this system is a moderate or higher.
X	NIST Special Publication (SP) 800-122 and NIST SP 800-53 Revision 4 Appendix J recommended security controls for protecting PII/BII are in place and functioning as intended; or have an approved Plan of Action and Milestones (POAM).
X	Contractors that have access to the system are subject to information security provisions in their contracts required by DOC policy.
X	Contracts with customers establish ownership rights over data including PII/BII.
X	Acceptance of liability for exposure of PII/BII is clearly defined in agreements with customers.
	Other (specify):

8.2 Provide a general description of the technologies used to protect PII/BII on the IT system. (Similar to 6.2, technical controls and leakage management)

The Census Bureau Information technology systems employ a multitude of layered security controls to protect PII/BII at rest, during processing, as well as in transit. These NIST 800-53 controls, at a minimum, are deployed and managed at the enterprise level including, but not limited to the following: • Intrusion Detection | Prevention Systems (IDS | IPS) • Firewalls • Mandatory use of HTTP(S) for Census Bureau Public facing websites • Use of trusted internet connection (TIC) • Anti-Virus software to protect host/end user systems • Encryption of databases (Data at rest) • HSPD-12 Compliant PIV cards • Access Controls The Census bureau Information technology systems also follow the National Institute of Standards and Technology (NIST) standards including special publications 800-53, 800-63, 800-37 etc. Any system within the Census Bureau that contains, transmits, or processes BII/PII has a current authority to operate (ATO) and goes through continuous monitoring on a yearly basis to ensure controls are implemented and operating as intended. The Census Bureau also deploys a DLP solution as well.

Section 9: Privacy Act

9.1 Indicate whether a system of records is being created under the Privacy Act, 5 U.S.C. § 552a. *(A new system of records notice (SORN) is required if the system is not covered by an existing SORN).*

As per the Privacy Act of 1974, "the term 'system of records' means a group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual."

X	Yes, this system is covered by an existing system of records notice (SORN). Provide the SORN name and number *(list all that apply)*: Census-8, Statistical Administrative Records System
X	Yes, a SORN has been submitted to the Department for approval on (date). Census - 4, Economic Survey Collection submitted on 8/24/16
	No, a SORN is not being created.

Section 10: Retention of Information

10.1 Indicate whether these records are covered by an approved records control schedule and monitored for compliance. *(Check all that apply.)*

X	There is an approved record control schedule. Provide the name of the record control schedule:

	GRS 3.1 General Technology Management Records, GRS 3.2 Information Systems Security Records; GRS 4.3 Input Records, Output Records and Electronic Copies. DAA-0029-2014-0005: Records of the Center for Administrative Records Research and Applications.
	No, there is not an approved record control schedule. Provide the stage in which the project is in developing and submitting a records control schedule:
X	Yes, retention is monitored for compliance to the schedule.
	No, retention is not monitored for compliance to the schedule. Provide explanation:

10.2 Indicate the disposal method of the PII/BII. *(Check all that apply.)*

Disposal			
Shredding	X	Overwriting	X
Degaussing		Deleting	X
Other (specify):			

Section 11: NIST Special Publication 800-122 PII Confidentiality Impact Levels

11.1 Indicate the potential impact that could result to the subject individuals and/or the organization if PII were inappropriately accessed, used, or disclosed.

	Low – the loss of confidentiality, integrity, or availability could be expected to have a limited adverse effect on organizational operations, organizational assets, or individuals.
	Moderate – the loss of confidentiality, integrity, or availability could be expected to have a serious adverse effect on organizational operations, organizational assets, or individuals.
X	High – the loss of confidentiality, integrity, or availability could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, or individuals.

11.2 Indicate which factors were used to determine the above PII confidentiality impact levels. *(Check all that apply.)*

X	Identifiability	Provide explanation: Individual data elements directly identifying unique individuals.
X	Quantity of PII	Provide explanation: A severe or catastrophic number of individuals affected by loss, theft, or compromise. Severe or catastrophic collective harm to individuals, harm to the organization's reputation, or cost to the organization in addressing a breach.
X	Data Field Sensitivity	Provide explanation: Data fields, alone or in combination, are directly usable in other contexts and make the individual or organization vulnerable to harms, such as identity theft, embarrassment, loss of trust, or costs.

X	Context of Use	Provide explanation: Disclosure of the act of collecting, and using the PII, or the PII itself is likely to result in severe or catastrophic harm to the individual or organization
X	Obligation to Protect Confidentiality	Provide explanation: Organization or Mission- specific privacy laws, regulations, mandates, or organizational policy apply that add more restrictive requirements to government- wide or industry-specific requirements. Violations may result in severe civil or criminal penalties.
X	Access to and Location of PII	Provide explanation: Located on computers and other devices on a network controlled by the organization. Access limited to a multiple populations of the organization's workforce beyond the direct program or office that owns the information on behalf of the organization. Access only allowed by organization- owned equipment outside of the physical locations owned by the organization only with a secured connection
	Other:	Provide explanation

Section 12: Analysis

12.1 Indicate whether the conduct of this PIA results in any required business process changes.

	Yes, the conduct of this PIA results in required business process changes. Explanation:
X	No, the conduct of this PIA does not result in any required business process changes.

12.2 Indicate whether the conduct of this PIA results in any required technology changes.

	Yes, the conduct of this PIA results in required technology changes. Explanation:
X	No, the conduct of this PIA does not result in any required technology changes.

U.S. Department of Commerce
U.S. Census Bureau

Privacy Impact Assessment
for the
CEN18 Enterprise Applications

Reviewed by: 4/19/18, Bureau Chief Privacy Officer

☑ Concurrence of Senior Agency Official for Privacy/DOC Chief Privacy Officer

☐ Non-concurrence of Senior Agency Official for Privacy/DOC Chief Privacy Officer

LISA MARTIN

Digitally signed by LISA MARTIN
DN: c=US, o=U.S. Government, ou=Department of Commerce, ou=Office of the Secretary, cn=LISA MARTIN, 0.9.2342.19200300.100.1.1=13001000105292
Date: 2018.06.26 14:19:48 -04'00'

Dr. Catrina D. Purvis

Signature of Senior Agency Official for Privacy/DOC Chief Privacy Officer Date

U.S. Department of Commerce Privacy Impact Assessment
U.S. Census Bureau CEN18 Enterprise Applications

Unique Project Identifier: 006-000401700

Introduction: System Description

(a) a general description of the information in the system;

The CEN18 Enterprise Applications system is the functional management framework used to deliver applications to end users of the U.S. Census Bureau network. The CEN18 Enterprise Applications system includes several applications that maintain or collect personally identifiable information (PII). These applications are:

- an enterprise-level data tracking system;
- a general support system for internal data management,
- a transaction-based system
- a relational database management system, and
- a concurrent analysis and estimation system (CAES),

Some examples of survey and census information maintained by this system are personal names, personal addresses, personal contact information (telephone numbers, email address), business information, occupation, medical information, tax information, etc. The systems reside in Census Bureau's Bowie Computing center. The components that collect PII/BII do not reside in the cloud.

(b) a description of a typical transaction conducted on the system;

The purpose of the enterprise-level data tracking system is to ensure data consistency, data integrity, and generate meaningful data information through data management, tracking, and reporting for Census Bureau collections.

The general support system for CEN18 provides internal data management within the Census Bureau collections. This system allows users to request access to datasets, and when approved, users are granted access to the datasets within a secure environment provisioned by the system. Census Bureau datasets are for internal use by employees, and are capable of containing protected or administrative information.

The transaction-based system within CEN18 serves as the primary mechanism for operational control across surveys for data collection. The system can be considered an operational brain that determines operational workflow based on pre-existing protocols.

The relational database management system stores and retrieves data as requested by other software applications. This system provides both a testing, development and production environment for optimum functionality.

The CAES serves as the enterprise-wide analytics platform for surveys and censuses. This system allows statisticians within census and survey projects to perform statistical models using census and survey response data, paradata, administrative records, and many other types of data. The system will receive PII including Identifying Numbers, General Personal Data, and Work-Related Data. The PII is received from other information systems that collect, maintain and disseminate Census and Survey data.

(c) any information sharing conducted by the system

The enterprise-level data tracking system does not share information.

The general support system shares information within the Census Bureau by querying indexed metadata and by sending email to data owners, administrators, and other application users.

The relational database management system does not share information.

The transaction-based system shares demographic survey, Decennial, and Economic Census information within the Census Bureau and with the Department of Commerce, that is used to determine new survey content, support electronic collections, for statistical purposes, and to create datasets for the Census Bureau.

The Concurrent Analysis and Estimation System (CAES) is an environment for use by researchers to make decisions during the data collection phase of a survey or a Census. The CAES system will provide the researcher with any data that they request as input and will output and send decision based data only to other systems. This could include things such as case level intervention codes, a stop work decision, or best time of day to contact respondents. CAES does not provide a mechanism for sharing PII/BII with other systems.

(d) a citation of the legal authority to collect PII and/or BII;

The legal authorities to collect PII and/or BII for CEN18 are:

5 U.S.C. 301
13 U.S.C. Chapter 9, and Sections: 6, 8(b), 9, 131, 132, 141, 161, 182, 193, 196
15 C.F.R. Part 30 and 19 C.F.R. Section 24.53
18 U.S.C. 2510-2521
26 U.S.C. 6103(j) and
Foreign Trade Statistical Regulations or its successor document, the Foreign Trade Regulations

(e) the Federal Information Processing Standard (FIPS) 199 security impact category for the system

The Federal Information Processing Standard (FIPS) 199 category for this system is Moderate.

Section 1: Status of the Information System

1.1 Indicate whether the information system is a new or existing system.

____ This is a new information system.
____ This is an existing information system with changes that create new privacy risks. *(Check all that apply.)*

Changes That Create New Privacy Risks (CTCNPR)					
a. Conversions		d. Significant Merging		g. New Interagency Uses	
b. Anonymous to Non-Anonymous		e. New Public Access		h. Internal Flow or Collection	
c. Significant System Management Changes		f. Commercial Sources		i. Alteration in Character of Data	
j. Other changes that create new privacy risks (specify):					

_x___ This is an existing information system without changes that create new privacy risks.

Section 2: Information in the System

2.1 Indicate what personally identifiable information (PII)/business identifiable information (BII) is collected, maintained, or disseminated. *(Check all that apply.)*

Identifying Numbers (IN)					
a. Social Security*		e. File/Case ID		i. Credit Card	X
b. Taxpayer ID	X	f. Driver's License	X	j. Financial Account	X
c. Employer ID	X	g. Passport		k. Financial Transaction	X
d. Employee ID	X	h. Alien Registration		l. Vehicle Identifier	X
m. Other identifying numbers (specify):					
*Explanation for the need to collect, maintain, or disseminate the Social Security number, including truncated form:					

General Personal Data (GPD)					
a. Name	X	g. Date of Birth	X	m. Religion	
b. Maiden Name	X	h. Place of Birth	X	n. Financial Information	X
c. Alias	X	i. Home Address	X	o. Medical Information	X
d. Gender	X	j. Telephone Number	X	p. Military Service	X
e. Age	X	k. Email Address	X	q. Physical Characteristics	X
f. Race/Ethnicity	X	l. Education	X	r. Mother's Maiden Name	X
s. Other general personal data (specify):					

Work-Related Data (WRD)

a. Occupation	X	d. Telephone Number	X	g. Salary	X
b. Job Title	X	e. Email Address	X	h. Work History	X
c. Work Address	X	f. Business Associates	X		
i. Other work-related data (specify):					

Distinguishing Features/Biometrics (DFB)					
a. Fingerprints		d. Photographs		g. DNA Profiles	
b. Palm Prints		e. Scars, Marks, Tattoos		h. Retina/Iris Scans	
c. Voice Recording/Signatures		f. Vascular Scan		i. Dental Profile	
j. Other distinguishing features/biometrics (specify):					

System Administration/Audit Data (SAAD)					
a. User ID	X	c. Date/Time of Access	X	e. ID Files Accessed	X
b. IP Address	X	d. Queries Run	X	f. Contents of Files	X
g. Other system administration/audit data (specify):					

Other Information (specify)

2.2 Indicate sources of the PII/BII in the system. *(Check all that apply.)*

Directly from Individual about Whom the Information Pertains					
In Person		Hard Copy: Mail/Fax		Online	
Telephone		Email			
Other (specify):					

Government Sources					
Within the Bureau	x	Other DOC Bureaus	x	Other Federal Agencies	
State, Local, Tribal		Foreign			
Other (specify):					

Non-government Sources					
Public Organizations		Private Sector		Commercial Data Brokers	
Third Party Website or Application					
Other (specify):					

2.3 Indicate the technologies used that contain PII/BII in ways that have not been previously deployed. *(Check all that apply.)*

Technologies Used Containing PII/BII Not Previously Deployed (TUCPBNPD)			
Smart Cards		Biometrics	
Caller-ID		Personal Identity Verification (PIV) Cards	
Other (specify):			

X	There are not any technologies used that contain PII/BII in ways that have not been previously deployed.

Section 3: System Supported Activities

3.1 Indicate IT system supported activities, which raise privacy risks/concerns. *(Check all that apply.)*

Activities			
Audio recordings		Building entry readers	
Video surveillance		Electronic purchase transactions	
Other (specify):			

X	There are not any IT system supported activities, which raise privacy risks/concerns.

Section 4: Purpose of the System

4.1 Indicate why the PII/BII in the IT system is being collected, maintained, or disseminated. *(Check all that apply.)*

Purpose			
To determine eligibility		For administering human resources programs	
For administrative matters	X	To promote information sharing initiatives	X
For litigation		For criminal law enforcement activities	
For civil enforcement activities		For intelligence activities	
To improve Federal services online		For employee or customer satisfaction	
For web measurement and customization technologies (single-session)		For web measurement and customization technologies (multi-session)	
Other (specify): For research and statistical purposes			

Section 5: Use of the Information

5.1 In the context of functional areas (business processes, missions, operations, etc.) supported by the IT system, describe how the PII/BII that is collected, maintained, or disseminated will be used. Indicate if the PII/BII identified in Section 2.1 of this document is in reference to a federal employee/contractor, member of the public, foreign national, visitor or other (specify).

The PII/BII maintained for administrative purposes: This IT system maintains first name, last name, address, email address, etc. to ensure that mandatory survey or statistical, information is ready for internal Census use. The information pertains and is in reference to federal employees/contractors conducting the surveys and the public.

The PII/BII maintained for statistical and research purposes: The data maintained by this IT system is collected from other IT systems that collect censuses and surveys (e.g., responses and statuses) and is used to direct data collection efforts. It is also used to inform program areas within the Census Bureau (responsible for survey and census questionnaire mail out) whom to send survey and census forms to. The IT system gathers response data from the data collection modes to send it to the survey and census processing IT systems in a standardized way. This information enables the Census Bureau to fulfill its legal obligation to provide mandated statistics. The information pertains to members of the public.

The PII/BII maintained for information sharing initiatives: This information is collected and shared within the Census Bureau and the Department of Commerce to create datasets for various types of censuses and surveys. This information enables the Census Bureau to fulfill its legal obligation to enhance its information sharing initiatives. The information pertains to members of the public.

Section 6: Information Sharing and Access

6.1 Indicate with whom the bureau intends to share the PII/BII in the IT system and how the PII/BII will be shared. *(Check all that apply.)*

Recipient	How Information will be Shared		
	Case-by-Case	Bulk Transfer	Direct Access
Within the bureau	X	X	X
DOC bureaus	X		
Federal agencies			
State, local, tribal gov't agencies			

Public			
Private sector			
Foreign governments			
Foreign entities			
Other (specify):			

	The PII/BII in the system will not be shared.

6.2 Indicate whether the IT system connects with or receives information from any other IT systems authorized to process PII and/or BII.

X	Yes, this IT system connects with or receives information from another IT system(s) authorized to process PII and/or BII. Provide the name of the IT system and describe the technical controls which prevent PII/BII leakage: CEN01, CEN03, CEN04, CEN05, CEN07, CEN08, CEN10, CEN11, CEN13, CEN16, CEN17, CEN24, CEN26, CEN33 The CEN18 IT system uses a multitude of security controls mandated by the Federal Information Security Management Act of 2002 (FISMA) and various other regulatory control frameworks including the National Institute of Standards and Technology (NIST) special publication 800 series. These security controls include, but are not limited to the use of mandatory HTTPS for public facing websites, access controls, anti-virus solutions, enterprise auditing/monitoring, encryption of data at rest, and various physical controls at Census facilities that house Information Technology systems. The Census Bureau also deploys an enterprise Data Loss Protection (DLP) solution as well.
	No, this IT system does not connect with or receive information from another IT system(s) authorized to process PII and/or BII.

6.3 Identify the class of users who will have access to the IT system and the PII/BII. *(Check all that apply.)*

Class of Users			
General Public		Government Employees	X
Contractors	X		
Other (specify):			

Section 7: Notice and Consent

7.1 Indicate whether individuals will be notified if their PII/BII is collected, maintained, or disseminated by the system. *(Check all that apply.)*

x	Yes, notice is provided pursuant to a system of records notice published in the Federal Register and discussed in Section 9.	
x	Yes, notice is provided by a Privacy Act statement and/or privacy policy. The Privacy Act statement and/or privacy policy can be found at: https://www.census.gov/about/policies/privacy/privacy-policy.html	
	Yes, notice is provided by other means.	Specify how:
	No, notice is not provided.	Specify why not:

7.2 Indicate whether and how individuals have an opportunity to decline to provide PII/BII.

	Yes, individuals have an opportunity to decline to provide PII/BII.	Specify how:
x	No, individuals do not have an opportunity to decline to provide PII/BII.	Specify why not: PII/BII is pulled from other Census Bureau Information Systems, therefore there is not an opportunity to decline to provide PII/BII at the CEN18 system level.

7.3 Indicate whether and how individuals have an opportunity to consent to particular uses of their PII/BII.

	Yes, individuals have an opportunity to consent to particular uses of their PII/BII.	Specify how:
x	No, individuals do not have an opportunity to consent to particular uses of their PII/BII.	Specify why not: PII/BII is pulled from other Census Bureau Information Systems, therefore there is not an opportunity to consent to particular uses of PII/BII at the CEN18 system level.

7.4 Indicate whether and how individuals have an opportunity to review/update PII/BII pertaining to them.

	Yes, individuals have an opportunity to review/update PII/BII pertaining to them.	Specify how:
x	No, individuals do not have an opportunity to review/update PII/BII pertaining to them.	Specify why not: PII/BII is pulled from other Census Bureau Information Systems, therefore there is not an opportunity to review/update PII/BII at the CEN18 system level.

Section 8: Administrative and Technological Controls

8.1 Indicate the administrative and technological controls for the system. *(Check all that apply.)*

x	All users signed a confidentiality agreement or non-disclosure agreement.
x	All users are subject to a Code of Conduct that includes the requirement for confidentiality.
x	Staff (employees and contractors) received training on privacy and confidentiality policies and practices.
x	Access to the PII/BII is restricted to authorized personnel only.
x	Access to the PII/BII is being monitored, tracked, or recorded. Explanation: Only authorized government/contractor personnel are allowed to access PII/BII within a system. Authorizations for users occur yearly, at a minimum in accordance with applicable Bureau, Agency, and Federal policies/guidelines. In addition, audit logs are in place and assessed per NIST control *AU-03, Content of Audit records*.
x	The information is secured in accordance with FISMA requirements. Provide date of most recent Assessment and Authorization (A&A): July 20, 2017 ☐ This is a new system. The A&A date will be provided when the A&A package is approved.
x	The Federal Information Processing Standard (FIPS) 199 security impact category for this system is a moderate or higher.

x	NIST Special Publication (SP) 800-122 and NIST SP 800-53 Revision 4 Appendix J recommended security controls for protecting PII/BII are in place and functioning as intended; or have an approved Plan of Action and Milestones (POAM).
x	Contractors that have access to the system are subject to information security provisions in their contracts required by DOC policy.
	Contracts with customers establish ownership rights over data including PII/BII.
	Acceptance of liability for exposure of PII/BII is clearly defined in agreements with customers.
	Other (specify):

8.2 Provide a general description of the technologies used to protect PII/BII on the IT system.

Census Bureau Information technology systems employ a multitude of layered security controls to protect BII/PII at rest, during processing, as well as in transit. These NIST 800-53 controls, at a minimum, are deployed and managed at the enterprise level including, but not limited to the following:

- Intrusion Detection | Prevention Systems (IDS | IPS)
- Firewalls
- Mandatory use of HTTP(S) for Census Public facing websites
- Use of trusted internet connection (TIC)
- Anti-Virus software to protect host/end user systems
- Encryption of databases (Data at rest)
- HSPD-12 Compliant PIV cards
- Access Controls

Census Bureau Information technology systems also follow the National Institute of Standards and Technology (NIST) standards including special publications 800-53, 800-63, 800-37 etc. Any system within the Census that contains, transmits, or processes BII/PII has a current authority to operate (ATO) and goes through continuous monitoring on a yearly basis to ensure controls are implemented and operating as intended. The Census Bureau also deploys a DLP solution as well.

Section 9: Privacy Act

9.1 Indicate whether a system of records is being created under the Privacy Act, 5 U.S.C. § 552a. *(A new system of records notice (SORN) is required if the system is not covered by an existing SORN).*

As per the Privacy Act of 1974, "the term 'system of records' means a group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual."

x	Yes, this system is covered by an existing system of records notice (SORN). Provide the SORN name and number *(list all that apply)*: COMMERCE/CENSUS-2, Employee Productivity Measurement Records: http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-2.html COMMERCE/CENSUS-3, Special Censuses, Surveys, and Other Studies: http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-3.html CENSUS-4, Economic Survey Collection:

	http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-4.html COMMERCE/CENSUS-5, Decennial Census Program: http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-5.html COMMERCE/CENSUS-7, Special Censuses of Population Conducted for State and Local Government: http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-7.html COMMERCE/CENSUS-8, Statistical Administration Records Systems: http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-8.html COMMERCE/CENSUS-9, Longitudinal Employer Household Dynamics System http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-9.html COMMERCE/CENSUS-12, Foreign Trade Statistics: http://www.osec.doc.gov/opog/PrivacyAct/SORNs/census-12.html
	Yes, a SORN has been submitted to the Department for approval on (date).
	No, a SORN is not being created.

Section 10: Retention of Information

10.1 Indicate whether these records are covered by an approved records control schedule and monitored for compliance. *(Check all that apply.)*

x	There is an approved record control schedule. Provide the name of the record control schedule: GRS 3.1 GRS 3.2 GRS 4.1, GRS 4.2, GRS 4.3 Demographic Directorate N1-29-99-5, N1-29-89-3, N1-29-87-3, N1-29-86-3, NC1-29-85-1, NC1-29-79-7 Economics Directorate N1-029-10-2, N1-029-10-3, N1-029-12-004, N1-029-10-4 Company Statistics Division N1-29-10-1 Economic Surveys Division N1-29-03-1NC1-29-80-15, NC1-29-79-4, NC1-29-78-15 NC1-29-78-8 Manufacturing and Construction Division NC1-29-81-10 Decennial Directorate N1-29-05-01, N1-29-10-5 American Community Survey DAA-0029-2015-0001
	No, there is not an approved record control schedule. Provide the stage in which the project is in developing and submitting a records control schedule:
	Yes, retention is monitored for compliance to the schedule.
	No, retention is not monitored for compliance to the schedule. Provide explanation:

10.2 Indicate the disposal method of the PII/BII. *(Check all that apply.)*

Disposal			
Shredding		Overwriting	
Degaussing		Deleting	x
Other (specify):			

Section 11: NIST Special Publication 800-122 PII Confidentiality Impact Levels

11.1 Indicate the potential impact that could result to the subject individuals and/or the organization if PII were inappropriately accessed, used, or disclosed.

	Low – the loss of confidentiality, integrity, or availability could be expected to have a limited adverse effect on organizational operations, organizational assets. or individuals.
	Moderate the loss of confidentiality, integrity, or availability could be expected to have a serious adverse effect on organizational operations, organizational assets, or individuals.
X	High – the loss of confidentiality, integrity, or availability could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, or individuals.

11.2 Indicate which factors were used to determine the above PII confidentiality impact levels. *(Check all that apply.)*

X	Identifiability	Provide explanation: PII/BII collected can be directly used to identify individuals
X	Quantity of PII	Provide explanation: The collection is for demographic, Economic Surveys, and other surveys, and therefore, a severe or catastrophic number of individuals would be affected if there was loss, theft or compromise of the data.
X	Data Field Sensitivity	Provide explanation: The PII/BII, alone or in combination, are directly usable in other contexts and make the individual or organization vulnerable to harms, such as identity theft, embarrassment, loss of trust, or costs.
X	Context of Use	Provide explanation: Disclosure of the act of collecting and using the PII/BII in this IT system or the PII/BII itself may result in severe or catastrophic harm to the individual or organization.
X	Obligation to Protect Confidentiality	Provide explanation: PII/BII collected is required to be protected in accordance Title 13 collections. Violations may result in severe civil or criminal penalties.
X	Access to and Location of PII	PII/BII is located on computers controlled by the Census Bureau or on mobile devices or storage media. Access is limited to certain populations of the Census Bureau's workforce and limited to Special Sworn Status individuals. Access is only allowed by organization-owned equipment outside of the physical locations, and only with a secured connection.
X	Identifiability	Provide explanation: PII/BII collected can be directly used to identify individuals

Section 12: Analysis

12.1 Indicate whether the conduct of this PIA results in any required business process changes.

	Yes, the conduct of this PIA results in required business process changes. Explanation:
x	No, the conduct of this PIA does not result in any required business process changes.

12.2 Indicate whether the conduct of this PIA results in any required technology changes.

	Yes, the conduct of this PIA results in required technology changes. Explanation:
x	No, the conduct of this PIA does not result in any required technology changes.

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

ELECTRONIC PRIVACY INFORMATION CENTER,

Plaintiff,

v.

UNITED STATES DEPARTMENT OF COMMERCE, et al.,

Defendants.

Civ. Action No. 18-2711

DECLARATION OF CHRISTOPHER WOLF

I, Christopher Wolf, declare as follows:

1. My name is Christopher Wolf. I am over 18 years old. The information in this declaration is based on my personal knowledge.
2. I reside in Washington, DC, United States of America.
3. I am Senior Counsel in a Washington, DC law firm, focusing on privacy law.
4. The Electronic Privacy Information Center ("EPIC") is a non-profit, public interest research center established in 1994 to focus public attention on emerging privacy and civil liberties issues. Central to EPIC's mission is oversight of government activities that impact individual privacy, free expression, and democratic values. EPIC has a particular interest in preserving legal privacy protections for personal data, including the obligation of federal agencies to conduct privacy impact assessments.
5. I am a member of EPIC. EPIC's members are distinguished experts in law, technology, and public policy. As a member of EPIC, I serve on the EPIC Advisory Board, participate

in the activities of EPIC, help finance the activities of EPIC, and provide leadership for EPIC. I also pay annual membership dues, as required by the amended EPIC bylaws.

6. I became an EPIC Member because I am concerned about protecting privacy, freedom of expression, and democratic values in the information age. As privacy lawyer, EPIC's work on these issues is of special importance to me.

7. EPIC works on my behalf to protect my privacy rights, and I rely on EPIC to obtain and disseminate information about government activities that threaten my privacy.

8. As a resident of the United States, I understand that the Census Bureau intends to ask for my citizenship status on the 2020 Census. By law, I am required to provide a truthful response to such a question.

9. The Census Bureau has indicated that it may disclose my citizenship status to other federal agencies, including for criminal law enforcement purposes.

10. The imminent collection and possible disclosure of my citizenship status by means of the 2020 Census is an unwarranted invasion of my privacy and would cause me irreparable harm by exposing my personal information. I do not consent to the collection or disclosure of my citizenship status.

11. The Census Bureau is required to produce and publish a comprehensive Privacy Impact Assessment ("PIA") when it begins to develop a new collection of personal data, such as the citizenship status of every person in the United States.

12. I understand that the Census Bureau intends to collect, process, and store personal data concerning citizenship status using five separate Bureau systems: CEN05, CEN08, CEN11, CEN13, and CEN18.

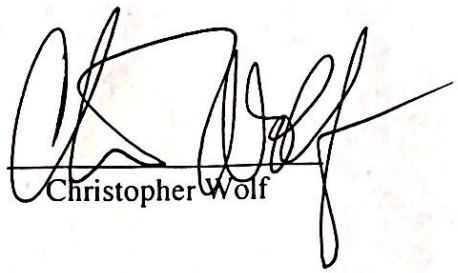
13. On my behalf, EPIC sought the five required Privacy Impact Assessments analyzing the collection, processing, and storage of personal data concerning citizenship status. I also sought the five required Privacy Impact Assessments myself by visiting the webpage where Census Bureau PIAs are published.[1]

14. For each of the five CEN systems, the Census Bureau failed to publish a Privacy Impact Assessment that adequately analyzed the collection, processing, and storage of personal data concerning citizenship status. I have therefore been denied information to which I am legally entitled concerning the privacy implications of the Census Bureau's citizenship question.

15. Moreover, I have been irreparably harmed by the Census Bureau's unlawful failure to publish the required Privacy Impact Assessments. I am unable to determine whether the Census Bureau has fully considered or addressed the risks to my privacy, even as the Bureau begins to develop a new collection of personal data that will contain my citizenship status.

I declare under penalty of perjury that, to the best of my knowledge, the foregoing is true and correct.

Executed on November 29, 2018



Christopher Wolf

[1] http://www.osec.doc.gov/opog/privacy/Census-pias.html.

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

ELECTRONIC PRIVACY INFORMATION CENTER, Plaintiff, v. UNITED STATES DEPARTMENT OF COMMERCE, et al., Defendants.	Civ. Action No. 18-2711

DECLARATION OF BRUCE SCHNEIER

I, Bruce Schneier, declare as follows:

1. My name is Bruce Schneier. I am over 18 years old. The information in this declaration is based on my personal knowledge.
2. I reside in Minneapolis, Minnesota, United States of America.
3. I am a Special Advisor to IBM Security. I am a fellow and lecturer at the Harvard Kennedy School. I am also an author and speaker on security and privacy topics.
4. The Electronic Privacy Information Center ("EPIC") is a non-profit, public interest research center established in 1994 to focus public attention on emerging privacy and civil liberties issues. Central to EPIC's mission is oversight of government activities that impact individual privacy, free expression, and democratic values. EPIC has a particular interest in preserving legal privacy protections for personal data, including the obligation of federal agencies to conduct privacy impact assessments.
5. I am a member of EPIC. EPIC's members are distinguished experts in law, technology, and public policy. As a member of EPIC, I serve on the EPIC Advisory Board, participate

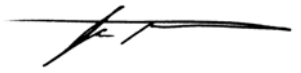
in the activities of EPIC, help finance the activities of EPIC, and provide leadership for EPIC. I also pay annual membership dues, as required by the amended EPIC bylaws.

6. I became an EPIC member because I am concerned about protecting privacy, freedom of expression, and democratic values in the information age. As someone who thinks, works, and writes about digital privacy, EPIC's work on these issues is of special importance to me.
7. EPIC works on my behalf to protect my privacy rights, and I rely on EPIC to obtain and disseminate information about government activities that threaten my privacy.
8. As a resident of the United States, I understand that the Census Bureau intends to ask for my citizenship status on the 2020 Census. By law, I am required to provide a truthful response to such a question.
9. The Census Bureau has indicated that it may disclose my citizenship status to other federal agencies, including for criminal law enforcement purposes.
10. The imminent collection and possible disclosure of my citizenship status by means of the 2020 Census is an unwarranted invasion of my privacy and would cause me irreparable harm by exposing my personal information. I do not consent to the collection or disclosure of my citizenship status.
11. The Census Bureau is required to produce and publish a comprehensive Privacy Impact Assessment ("PIA") when it begins to develop a new collection of personal data, such as the citizenship status of every person in the United States.
12. I understand that the Census Bureau intends to collect, process, and store personal data concerning citizenship status using five separate Bureau systems: CEN05, CEN08, CEN11, CEN13, and CEN18.

13. On my behalf, EPIC sought the five required Privacy Impact Assessments analyzing the collection, processing, and storage of personal data concerning citizenship status. I also sought the five required Privacy Impact Assessments myself by visiting the webpage where Census Bureau PIAs are published.[1]
14. For each of the five CEN systems, the Census Bureau failed to publish a Privacy Impact Assessment that adequately analyzed the collection, processing, and storage of personal data concerning citizenship status. I have therefore been denied information to which I am legally entitled concerning the privacy implications of the Census Bureau's citizenship question.
15. Moreover, I have been irreparably harmed by the Census Bureau's unlawful failure to publish the required Privacy Impact Assessments. I am unable to determine whether the Census Bureau has fully considered or addressed the risks to my privacy, even as the Bureau begins to develop a new collection of personal data that will contain my citizenship status.

I declare under penalty of perjury that, to the best of my knowledge, the foregoing is true and correct.

Executed on November 30, 2018



Bruce Schneier

[1] http://www.osec.doc.gov/opog/privacy/Census-pias.html.

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

ELECTRONIC PRIVACY INFORMATION CENTER, Plaintiff, v. UNITED STATES DEPARTMENT OF COMMERCE, et al., Defendants.	Civ. Action No. 18-2711

DECLARATION OF HARRY R. LEWIS

I, Harry R. Lewis, declare as follows:

1. My name is Harry R. Lewis. I am over 18 years old. The information in this declaration is based on my personal knowledge.
2. I reside in Brookline, Massachusetts United States of America.
3. I am Gordon McKay Professor of Computer Science at Harvard University.
4. The Electronic Privacy Information Center ("EPIC") is a non-profit, public interest research center established in 1994 to focus public attention on emerging privacy and civil liberties issues. Central to EPIC's mission is oversight of government activities that impact individual privacy, free expression, and democratic values. EPIC has a particular interest in preserving legal privacy protections for personal data, including the obligation of federal agencies to conduct privacy impact assessments.
5. I am a member of EPIC. EPIC's members are distinguished experts in law, technology, and public policy. As a member of EPIC, I serve on the EPIC Advisory Board, participate

in the activities of EPIC, help finance the activities of EPIC, and provide leadership for EPIC. I also pay annual membership dues, as required by the amended EPIC bylaws.

6. I became an EPIC member because I am concerned about protecting privacy, freedom of expression, and democratic values in the information age. As a professor of computer science, EPIC's work on these issues is of special importance to me.
7. EPIC works on my behalf to protect my privacy rights, and I rely on EPIC to obtain and disseminate information about government activities that threaten my privacy.
8. As a resident of the United States, I understand that the Census Bureau intends to ask for my citizenship status on the 2020 Census. By law, I am required to provide a truthful response to such a question.
9. The Census Bureau has indicated that it may disclose my citizenship status to other federal agencies, including for criminal law enforcement purposes.
10. The imminent collection and possible disclosure of my citizenship status by means of the 2020 Census is an unwarranted invasion of my privacy and would cause me irreparable harm by exposing my personal information. I do not consent to the collection or disclosure of my citizenship status.
11. The Census Bureau is required to produce and publish a comprehensive Privacy Impact Assessment ("PIA") when it begins to develop a new collection of personal data, such as the citizenship status of every person in the United States.
12. I understand that the Census Bureau intends to collect, process, and store personal data concerning citizenship status using five separate Bureau systems: CEN05, CEN08, CEN11, CEN13, and CEN18.

13. On my behalf, EPIC sought the five required Privacy Impact Assessments analyzing the collection, processing, and storage of personal data concerning citizenship status. I also sought the five required Privacy Impact Assessments myself by visiting the webpage where Census Bureau PIAs are published.[1]

14. For each of the five CEN systems, the Census Bureau failed to publish a Privacy Impact Assessment that adequately analyzed the collection, processing, and storage of personal data concerning citizenship status. I have therefore been denied information to which I am legally entitled concerning the privacy implications of the Census Bureau's citizenship question.

15. Moreover, I have been irreparably harmed by the Census Bureau's unlawful failure to publish the required Privacy Impact Assessments. I am unable to determine whether the Census Bureau has fully considered or addressed the risks to my privacy, even as the Bureau begins to develop a new collection of personal data that will contain my citizenship status.

I declare under penalty of perjury that, to the best of my knowledge, the foregoing is true and correct.

Executed on December 3, 2018

Harry R. Lewis

[1] http://www.osec.doc.gov/opog/privacy/Census-pias.html.

epic.org

Electronic Privacy Information Center
1718 Connecticut Avenue NW, Suite 200
Washington, DC 20009, USA

+1 202 483 1140
+1 202 483 1248
@EPICPrivacy
https://epic.org

COMMENTS OF THE ELECTRONIC PRIVACY INFORMATION CENTER to the

U.S. CENSUS BUREAU

2020 Census

83 FR 26643

August 7, 2018

The Electronic Privacy Information Center ("EPIC") submits these comments in response to the Census Bureau's notice[1] on the 2020 decennial census.

EPIC is a public interest research center established in 1994 to focus public attention on emerging privacy and civil liberties issues. The Census implicates numerous privacy issues.[2] We ask you to consider the risks of personal data being used for purposes that may undermine the integrity, reliability, and purpose of the U.S. census. EPIC specifically asks the Census Bureau to suspend the citizenship question from the 2020 census form until a thorough and updated Privacy Impact Assessment is conducted. So far, the Bureau has failed to demonstrate the data gathered from that particular question will not undermine the privacy rights of those who respond to the census.

EPIC has particular expertise with the misuse of census data after 9-11. And documents EPIC obtained under the Freedom of Information Act led to important reforms at the agency. We urge you not to take our comments lightly.

EPIC supports the work of the Census Bureau and the use of statistical analysis in policymaking and other government initiatives.[3] The Census is an essential part of understanding the changing demographics in America. The census helps ensure evidence-based policy decisions and census data is the source of much political and economic planning in the United States. However, it is of the utmost importance the individual privacy is respected. Every effort must be taken to ensure that the personal information of individuals and that census data is not used improperly.

---

[1] U.S. Census Bureau, *Proposed Information Collection; Comment Request; 2020 Census*, Notice, 83 FR 26643 (June 8, 2018), https://www.federalregister.gov/documents/2018/06/08/2018-12365/proposed-information-collection-comment-request-2020-census.

[2] *The Census and Privacy*, EPIC, https://epic.org/privacy/census/.

[3] EPIC testified before the Commission on Evidence-Based Policymaking and called for the Commission to adopt innovative privacy safeguards to protect personal data and make informed public policy decisions. Marc Rotenberg, Commission on Evidence-Based Policymaking: Privacy Perspectives, before the National Academies of Science, Sep. 9, 2016, https://epic.org/privacy/wiretap/RotenbergCEBP-9-16.pdf.

EPIC knows that the Census Bureau takes its confidentiality requirements seriously[4] and that some of the strictest privacy laws in the U.S. apply to census data.[5] But the addition of the citizenship question to the 2020 census raises new privacy issues. The Privacy Impact Assessment for the census indicates that the Bureau has not given proper consideration to these issues.

There has been greater concern about the confidentiality of 2020 census data than in previous decennial censuses. The Census Bureau conducted a study in 2017 that found respondents expressing new concerns including the "Muslim ban," the dissolution of DACA, and Immigration and Customs Enforcement.[6] The study found that these concerns were most pronounced among immigrant respondents.[7]

Part I discusses prior abuses of census data, to remind the Bureau that fears of 2020 census data being misused are not abstract. Part II explains why census data should never be used for enforcement purposes. Part III addresses the data integrity issues caused by the addition of the untested question. Part IV describes the Bureau's Privacy Impact Assessment and why it does not address the privacy risks raised by the addition of the citizenship question.

I. Past Misuses of Census Data

Despite strong census privacy laws, the U.S. has a sordid history of misusing census data to target minority groups. The most egregious misuse of census data was the role it played in the internment of Japanese-Americans during World War II.[8] In 1943 the Census Bureau complied with a request by the Treasury Secretary for the names and locations of all people of Japanese ancestry in the Washington, D.C., area.[9] The Bureau should remember this human rights abuse every time another agency requests census data.

After 9-11, EPIC pursued a Freedom of Information Act request about the potential misuse of census data. Documents obtained by EPIC revealed that the Census Bureau had provided the

---

[4] Ron Jarmin, *The U.S. Census Bureau's Commitment to Confidentiality*, Census Blog (May 7, 2018) https://census.gov/newsroom/blogs/director/2018/05/the_u_s_census_bure.html.

[5] 92 Stat. 915; Public Law 95-416. The Census Bureau cannot disclose "personally identifiable information about an individual to any other individual or agency until 72 years after it was collected for the decennial census."

[6] Center for Survey Measurement, MEMORANDUM FOR Associate Directorate for Research and Methodology (ADRM) Respondent Confidentiality Concerns (Sept. 20, 2017), https://www2.census.gov/cac/nac/meetings/2017-11/Memo-Regarding-Respondent-Confidentiality-Concerns.pdf.

[7] *Id*. at 1.

[8] JR Minkel, *Confirmed: The U.S. Census Bureau Gave Up Names of Japanese-Americans in WW II,* Scientific American (March 30, 2007), https://www.scientificamerican.com/article/confirmed-the-us-census-b/.

[9] W. Seltzer and M. Anderson, "Census Confidentiality under the Second War Powers Act (1942-1947)." Paper prepared for presentation at the annual meeting of the Population Association of America, New York, March 29-31, 2007, *available at* http://studylib.net/doc/7742798/census-confidentiality-under-the-second-war-powers.

Department of Homeland Security ("DHS") with census data on individuals of Arab ancestry.[10] In 2004 EPIC obtained documents revealing that the Census Bureau provided the DHS statistical data on people who identified themselves on the 2000 census as being of Arab ancestry. The special tabulations were prepared specifically for the law enforcement agency. There is no indication that the Department of Homeland Security requested similar information about any other ethnic groups.

One document,[11]obtained by EPIC, shows cities with populations of 10,000 or more and with 1,000 or more people who indicated they are of Arab ancestry. For each city, the tabulation provides total population, population of Arab ancestry, and percent of the total population which is of Arab ancestry. The tabulations were produced using data from the 2000 census long-form questionnaire, which goes to only a sample of the population. A second document[12] shows the number of census responses indicating Arab ancestry in certain zip codes throughout the country. The responses indicating Arab ancestry are subdivided into Egyptian, Iraqi, Jordanian, Lebanese, Moroccan, Palestinian, Syrian, Arab/Arabic, and Other Arab. Although this data was not personally identifiable, its disclosure to a law enforcement agency was unethical.

The reason DHS gave for requesting these tabulations was to determine which languages signs should be in at international airports.[13] Heavily redacted emails[14] between a Census Bureau analyst and a DHS official show that the Bureau gave the documents before the intended purpose for the data was known and that this explanation was given after the tabulations had already been disclosed. The ex-post-facto reason given by DHS seems pretextual.

As a result of these revelations, resulting from EPIC's FOIA litigation, the Census Bureau revised its policy on sharing statistical information about "sensitive populations" with law enforcement or intelligence agencies. Customs and Border Protection also changed its policy on requesting "information of a sensitive nature from the Census Bureau."[15]

EPIC's FOIA efforts led the Census Bureau to implement new procedures regarding the release of "potentially sensitive data to requesting law enforcement agencies and organizations or individuals."[16] EPIC supported those efforts but recognizes also the ongoing concerns about the

---

[10] *Department of Homeland Security Obtained Data on Arab Americans From Census Bureau*, EPIC, https://epic.org/privacy/census/foia/; Lynette Clemetson, *Homeland Security Given Data on Arab-Americans*, New York Times, Jul. 30, 2004, http://www.nytimes.com/2004/07/30/us/homeland-security-given-data-on-arab-americans.html.
[11] EPIC FOIA, Tabulation 1: "Places with 10,000 or More Population and with 1,000 or More Persons of Arab Ancestry: 2000" https://epic.org/privacy/census/foia/tab_1.pdf.
[12] EPIC FOIA, Tabulation 2: "People of Arab Ancestry by ZIP Code Tabulation Area: 2000" https://epic.org/privacy/census/foia/tab_2.pdf.
[13] EPIC FOIA, https://epic.org/privacy/census/foia/census_emails.pdf.
[14] *Id.*
[15] U.S. Customs and Border Protection, *Policy for Requesting Information of a Sensitive Nature from the Census Bureau*, Memorandum (Aug. 9, 2004), https://epic.org/privacy/census/foia/policy.pdf.
[16] Census Bureau News, ''U.S. Census Bureau Announces Policy Regarding Sensitive Data,'' press release CB04-145, August 30, 2004; Lynette Clemetson, *Census Policy On Providing Sensitive Data Is Revised*, New York Times, Aug. 31, 2004, http://www.nytimes.com/2004/08/31/us/census-policy-on-providing-sensitive-data-is-revised.html.

potential misuse of data collected by government agencies. Amid rising fears that some minority groups may be targeted by law enforcement agencies, your committee should ensure that the data collected by the federal government is not misused.

II. Census Data Should Never Be Used for Enforcement Purposes

Using census data to help enforce laws is a corruption of the decennial census's constitutional purpose. The Department of Justice ("DOJ") requested the citizenship question on the census would allow the agency to better enforce Section 2 of the Voting Rights Act, which bars the dilution of voting power of a minority group through redistricting. DOJ wants census-block-level data for locations where they suspect Section 2 violations have occurred.

The decennial census's statistical purpose is frustrated when other agencies ask it to collect data for other purposes. The DOJ's responsibility to enforce the Voting Rights Act is vital to ensuring fair elections. In its request to the Bureau, the DOJ called the decennial census the "most appropriate vehicle" for collecting data on the citizen voting-age population.[17] EPIC disagrees with this assertion. The decennial census was never intended to be a catch-all data collection to assist other federal agencies. In fact, the statutes concerning the privacy of census data are meant to expressly prohibit this.

If the Census Bureau gets into the business of collecting data because it will assist other federal agencies enforce laws, it will be difficult to stay true to its constitutional purpose of conducting impartial statistical analysis. And it will undermine the integrity, accuracy, and reliability of the census. As a former director of the Bureau succinctly put it thirteen years ago: "The Census Bureau cannot become a quasi-investigatory agency and still perform its basic responsibilities as a statistical agency."[18] The Bureau does not serve an investigatory function and the DOJ (or any other agency) should not expect it to.

III. Data Integrity

When a new question is added this late in preparations for the 2020 census, the burden of proof is on those proposing the new question to establish that it will not impact the integrity of the data collected. The fact that the 2020 census will have an additional question—regardless of what the question asks—is likely to yield lower response rates. As Secretary Ross told Congress: "One of the problems with adding questions is reduced response rates. It may seem counterintuitive, but the

---

[17] *Re: Request to Reinstate Citizenship Question on 2020 Census Questionnaire* (Dec. 12, 2017), https://www.documentcloud.org/documents/4340651-Text-of-Dec-2017-DOJ-letter-to-Census.html.

[18] *Counting the Vote: Should Only U.S. Citizens be Included in Apportioning Our Elected Representatives?*, Hearing Before Subcomm. on Federalism and the Census of the H. Comm. on Gov't Reform, 109th Cong. 77 (2005) (statement of former director of U.S. Census Bureau Kenneth Prewitt).

more things you ask in those forms, the less likely you are to get them in."[19] And the fact that the new question has not been tested will create further problems with the integrity of the 2020 census data.

Six former directors of the Census Bureau—who served under both Republican and Democrat administrations—wrote to Secretary Ross to warn him that adding a new question to the 2020 census without adequate testing "at this late point in the decennial planning process would put the accuracy of the enumeration and success of the census in all communities at grave risk."[20] The typical process for adding new questions takes multiple years and requires extensive testing. Not only did the citizenship question not go through this process, but it was added after the 2018 End—to—End Census Test was already underway, so it was not even tested in what is meant to be the "dress rehearsal" for the 2020 census.

The former directors said: "even small changes in survey question order, wording, and instructions can have significant, and often unexpected, consequences for the rate, quality, and truthfulness of response."[21] But this is a significant change: a new question on a sensitive topic. In addition to a lower response rate, the question would cause an increase in inaccurate responses because "[t]here would be little incentive for non-citizens to offer to the government their actual status."[22]

IV. Insufficiency of Privacy Impact Assessment

The Privacy Impact Assessment ("PIA") for the 2020 decennial census[23] indicates that the Census Bureau has not undertaken an appropriate analysis of the privacy risks of the citizenship question. "Citizenship" is listed as "other general personal data" that will be collected by the census.[24] But, where the Bureau is required to indicate the status of the information system, it selects: "This is an existing information system without changes that create new privacy risks."[25] There is no analysis of the new question.

---

[19] Rep. Jimmy Gomez Questions Secretary Ross at Oversight Hearing (Oct. 12, 2017), https://www.c-span.org/video/?c4685802/rep-jimmy-gomez-questions-secretary-ross-oversight-hearing?c4685802/rep-jimmy-gomez-questions-secretary-ross-oversight-hearing/video/.

[20] Letter of Former Directors of U.S. Census Bureau to The Honorable Wilbur L. Ross (Jan. 26, 2018), https://www.washingtonpost.com/r/2010-2019/WashingtonPost/2018/03/27/Editorial-Opinion/Graphics/DOJ_census_ques_request_Former_Directors_ltr_to_Ross.pdf.

[21] *Id.*

[22] Brief of Amici Curiae, *Evenwel v. Abbot*, at 25, http://www.scotusblog.com/wp-content/uploads/2015/10/Evenwel-FormerCensusBureauDirectorsBrief092515.pdf (Amici are former directors of the Census Bureau).

[23] U.S. Department of Commerce, U.S. Census Bureau, *Privacy Impact Assessment for the CEN08 Decennial Information Technology Division* (July 28, 2018) http://www.osec.doc.gov/opog/privacy/Census%20PIAs/CEN08_PIA_SAOP_Approved.pdf.

[24] *Id.* at 3.

[25] *Id.* at 2.

The PIA does not satisfy the Department of Commerce's own standards.[26] A PIA must be updated "where a system change creates new privacy risks." The Department of Commerce lists nine examples of new privacy risks, and the addition of the citizenship question and the stated purposes for doing so would create at least four of those risks.

The new question would alter the character of the data: "when new information in identifiable form added to a collection raises the risks to personal privacy (for example, the addition of health or financial information)."[27] Citizenship data is new information and it is identifiable because it is collected alongside the other information of the person filling out the census form. Immigration status added to the collection of the other demographic data collected by the census undeniably raises the risk to personal privacy. If that privacy were to be violated, that information could be used against the respondent for deportation or other purposes.

The PIA does not acknowledge the privacy risks raised by the new question and the Bureau should conduct a new PIA dealing specifically with the issues raised by the citizenship question.

Conclusion

EPIC recommends that the Census Bureau either remove the citizenship question or conducted a revised Privacy Impact Assessment, taking account of the recently added question, as is required by law. The question raises far-reaching privacy concerns and will also undermine the integrity, reliability and accuracy of the U.S. census.

EPIC has recently pursued related matters concerning the privacy of census data with the Census Bureau that were favorably resolved. We anticipate that will be the outcome here

Sincerely,

/s/ *Marc Rotenberg*
Marc Rotenberg
EPIC President

/s/ *Christine Bannan*
Christine Bannan
EPIC Administrative Law and Policy Fellow

[26] U.S. Department of Commerce, Office of Privacy and Open Government, *Privacy Compliance* http://www.osec.doc.gov/opog/privacy/compliance.html.
[27] *Id.* "Alteration in Character of Data" is the ninth example in the list of privacy risks.

epic.org

Electronic Privacy Information Center
1718 Connecticut Avenue NW, Suite 200
Washington, DC 20009, USA

+1 202 483 1140
+1 202 483 1248
@EPICPrivacy
https://epic.org

October 2, 2018

The Honorable Ron Johnson, Chairman
The Honorable Claire McCaskill, Ranking Member
U.S. Senate Committee on Homeland Security & Government Affairs
340 Dirksen Senate Office Building
Washington, DC 20510

Dear Chairman Johnson and Ranking Member McCaskill:

We write to you regarding the nomination hearing of Steven D. Dillingham to be Director of the Census"[1] EPIC is a public interest research center established in 1994 to focus public attention on emerging privacy and civil liberties issues. EPIC takes no position for or against the nominee. However, the Census implicates numerous privacy issues.[2] EPIC specifically asks the Census Bureau to suspend the citizenship question from the 2020 census form until a thorough and updated Privacy Impact Assessment is conducted. The Bureau has failed to demonstrate the data gathered from that particular question will not undermine the privacy rights of those who respond to the census.

EPIC supports the work of the Census Bureau and the use of statistical analysis in policymaking and other government initiatives.[3] The Census is an essential part of understanding the changing demographics in America. The census helps ensure evidence-based policy decisions and census data is the source of much political and economic planning in the United States. However, it is of the utmost importance the individual privacy is respected. Every effort must be taken to ensure that the personal information of individuals and that census data is not used improperly.

Through a Freedom of Information Act request EPIC has obtained documents regarding Secretary Ross's decision to add the citizenship question to the 2020 Census.[4] The documents obtained by EPIC reflect the varying opinions from lawmakers, scientists, and immigration groups

---

[1] *Nomination of Steven D. Dillingham to be Director of the Census*, 115th Cong. (2018), S. Comm. on Homeland Security & Governmental Affairs, https://www.hsgac.senate.gov/hearings/09/25/2018/nominations (October 3, 2018).
[2] *The Census and Privacy*, EPIC, https://epic.org/privacy/census/.
[3] EPIC testified before the Commission on Evidence-Based Policymaking and called for the Commission to adopt innovative privacy safeguards to protect personal data and make informed public policy decisions. Marc Rotenberg, Commission on Evidence-Based Policymaking: Privacy Perspectives, before the National Academies of Science, Sep. 9, 2016, https://epic.org/privacy/wiretap/RotenbergCEBP-9-16.pdf.
[4] FOIA Production 1, https://epic.org/foia/censusbureau/EPIC-18-03-22-Census-Bureau-FOIA-20180611-Production-1.pdf; FOIA Production 2, https://epic.org/foia/censusbureau/EPIC-18-03-22-Census-Bureau-FOIA-20180611-Production-2.pdf; FOIA Production 3, https://epic.org/foia/censusbureau/EPIC-18-03-22-Census-Bureau-FOIA-20180611-Production-3.pdf; FOIA Production 4, https://epic.org/foia/censusbureau/EPIC-18-03-22-Census-Bureau-FOIA-20180611-Production-4.pdf.

about the proposal. The documents also reveal that Kris Kobach, former Vice Chair of the now-defunct Presidential Advisory Commission on Election Integrity, urged Secretary Ross "on the direction of Steve Bannon" to add the citizenship question. According to an analysis conducted by the Census Bureau, the impact of asking about citizenship would be "very costly, harms the quality of the census count, and would use substantially less accurate citizenship data than are available" from other government resources.

I. Past Misuses of Census Data

There is substantial concern about the confidentiality of census data as a result of Secretary Ross's decision to add a citizenship question to the 2020 census. Despite strong census privacy laws, the U.S. has a sordid history of misusing census data to target minority groups. The most egregious misuse of census data was the role it played in the internment of Japanese-Americans during World War II.[5] In 1943 the Census Bureau complied with a request by the Treasury Secretary for the names and locations of all people of Japanese ancestry in the Washington, D.C., area.[6] The Bureau should remember this human rights abuse every time another agency requests census data.

EPIC has a strong interest in the government's use of Census data. After 9-11, EPIC pursued a Freedom of Information Act request about the transfer of Census data to the Department of Homeland Security. Documents obtained by EPIC revealed that the Census Bureau had provided the Department of Homeland Security ("DHS") with census data on individuals of Arab ancestry.[7] In 2004 EPIC obtained documents revealing that the Census Bureau provided the DHS statistical data on people who identified themselves on the 2000 census as being of Arab ancestry. The special tabulations were prepared specifically for the law enforcement agency. There is no indication that the Department of Homeland Security requested similar information about any other ethnic groups.

One document,[8] obtained by EPIC, shows cities with populations of 10,000 or more and with 1,000 or more people who indicated they are of Arab ancestry. For each city, the tabulation provides total population, population of Arab ancestry, and percent of the total population which is of Arab ancestry. The tabulations were produced using data from the 2000 census long-form questionnaire, which goes to only a sample of the population. A second document[9] shows the number of census responses indicating Arab ancestry in certain zip codes throughout the country. The responses

---

[5] JR Minkel, *Confirmed: The U.S. Census Bureau Gave Up Names of Japanese-Americans in WW II,* Scientific American (March 30, 2007), https://www.scientificamerican.com/article/confirmed-the-us-census-b/.

[6] W. Seltzer and M. Anderson, "Census Confidentiality under the Second War Powers Act (1942-1947)." Paper prepared for presentation at the annual meeting of the Population Association of America, New York, March 29-31, 2007, *available at* http://studylib.net/doc/7742798/census-confidentiality-under-the-second-war-powers.

[7] *Department of Homeland Security Obtained Data on Arab Americans From Census Bureau*, EPIC, https://epic.org/privacy/census/foia/; Lynette Clemetson, *Homeland Security Given Data on Arab-Americans*, New York Times, Jul. 30, 2004, http://www.nytimes.com/2004/07/30/us/homeland-security-given-data-on-arab-americans.html.

[8] EPIC FOIA, Tabulation 1: "Places with 10,000 or More Population and with 1,000 or More Persons of Arab Ancestry: 2000" https://epic.org/privacy/census/foia/tab_1.pdf.

[9] EPIC FOIA, Tabulation 2: "People of Arab Ancestry by ZIP Code Tabulation Area: 2000" https://epic.org/privacy/census/foia/tab_2.pdf.

indicating Arab ancestry are subdivided into Egyptian, Iraqi, Jordanian, Lebanese, Moroccan, Palestinian, Syrian, Arab/Arabic, and Other Arab. Although this data was not personally identifiable, its disclosure to a law enforcement agency was unethical.

The reason DHS gave for requesting these tabulations was to determine which languages signs should be in at international airports.[10] Heavily redacted emails[11] between a Census Bureau analyst and a DHS official show that the Bureau gave the documents before the intended purpose for the data was known and that this explanation was given after the tabulations had already been disclosed. The ex-post-facto reason given by DHS seems pretextual.

As a result of these revelations, resulting from EPIC's FOIA litigation, the Census Bureau revised its policy on sharing statistical information about "sensitive populations" with law enforcement or intelligence agencies. Customs and Border Protection also changed its policy on requesting "information of a sensitive nature from the Census Bureau."[12]

II. Census Data Should Never Be Used for Enforcement Purposes

Using census data to help enforce laws is a corruption of the decennial census's constitutional purpose. The Department of Justice ("DOJ") requested the citizenship question on the census would allow the agency to better enforce Section 2 of the Voting Rights Act, which bars the dilution of voting power of a minority group through redistricting. DOJ wants census-block-level data for locations where they suspect Section 2 violations have occurred.

The decennial census's statistical purpose is frustrated when other agencies ask it to collect data for other purposes. The DOJ's responsibility to enforce the Voting Rights Act is vital to ensuring fair elections. In its request to the Bureau, the DOJ called the decennial census the "most appropriate vehicle" for collecting data on the citizen voting-age population.[13] EPIC disagrees with this assertion. The decennial census was never intended to be a catch-all data collection to assist other federal agencies. In fact, the statutes concerning the privacy of census data are meant to expressly prohibit this.

If the Census Bureau gets into the business of collecting data because it will assist other federal agencies enforce laws, it will be difficult to stay true to its constitutional purpose of conducting impartial statistical analysis. And it will undermine the integrity, accuracy, and reliability of the census. As a former director of the Bureau succinctly put it thirteen years ago: "The Census Bureau cannot become a quasi-investigatory agency and still perform its basic responsibilities as a statistical agency."[14] The Bureau does not serve an investigatory function and the DOJ (or any other agency) should not expect it to.

---

[10] EPIC FOIA, https://epic.org/privacy/census/foia/census_emails.pdf.
[11] *Id.*
[12] U.S. Customs and Border Protection, *Policy for Requesting Information of a Sensitive Nature from the Census Bureau*, Memorandum (Aug. 9, 2004), https://epic.org/privacy/census/foia/policy.pdf.
[13] *Re: Request to Reinstate Citizenship Question on 2020 Census Questionnaire* (Dec. 12, 2017), https://www.documentcloud.org/documents/4340651-Text-of-Dec-2017-DOJ-letter-to-Census.html.
[14] *Counting the Vote: Should Only U.S. Citizens be Included in Apportioning Our Elected Representatives?*, Hearing Before Subcomm. on Federalism and the Census of the H. Comm. on Gov't Reform, 109th Cong. 77 (2005) (statement of former director of U.S. Census Bureau Kenneth Prewitt).

IV. Insufficiency of Privacy Impact Assessment

The Privacy Impact Assessment ("PIA") for the 2020 decennial census[15] indicates that the Census Bureau has not undertaken an appropriate analysis of the privacy risks of the citizenship question. "Citizenship" is listed as "other general personal data" that will be collected by the census.[16] But, where the Bureau is required to indicate the status of the information system, it selects: "This is an existing information system without changes that create new privacy risks."[17] There is no analysis of the new question.

The PIA does not satisfy the Department of Commerce's own standards.[18] A PIA must be updated "where a system change creates new privacy risks." The Department of Commerce lists nine examples of new privacy risks, and the addition of the citizenship question and the stated purposes for doing so would create at least four of those risks.

The new question would alter the character of the data: "when new information in identifiable form added to a collection raises the risks to personal privacy (for example, the addition of health or financial information)."[19] Citizenship data is new information and it is identifiable because it is collected alongside the other information of the person filling out the census form. Immigration status added to the collection of the other demographic data collected by the census undeniably raises the risk to personal privacy. If that privacy were to be violated, that information could be used against the respondent for deportation or other purposes. The PIA does not acknowledge the privacy risks raised by the new question and the Bureau should conduct a new PIA dealing specifically with the issues raised by the citizenship question.

EPIC looks forward to working with the Committee to ensure that the census data provides the maximum benefit to the American public while minimizing the privacy risks. We ask that this letter from EPIC be entered in the hearing record.

Sincerely,

/s/ *Marc Rotenberg*
Marc Rotenberg
EPIC President

/s/ *Caitriona Fitzgerald*
Caitriona Fitzgerald
EPIC Policy Director

/s/ *Christine Bannan*
Christine Bannan
EPIC Consumer Protection Counsel

---

[15] U.S. Department of Commerce, U.S. Census Bureau, *Privacy Impact Assessment for the CEN08 Decennial Information Technology Division* (July 28, 2018) http://www.osec.doc.gov/opog/privacy/Census%20PIAs/CEN08_PIA_SAOP_Approved.pdf.
[16] *Id.* at 3.
[17] *Id.* at 2.
[18] U.S. Department of Commerce, Office of Privacy and Open Government, *Privacy Compliance* http://www.osec.doc.gov/opog/privacy/compliance.html.
[19] *Id.* "Alteration in Character of Data" is the ninth example in the list of privacy risks.

United States Census 2010

This is the official form for all the people at this address.
It is quick and easy, and your answers are protected by law.

U.S. DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. CENSUS BUREAU

Use a blue or black pen.

Start here

The Census must count every person living in the United States on April 1, 2010.

Before you answer Question 1, count the people living in this house, apartment, or mobile home using our guidelines.

- Count all people, including babies, who live and sleep here most of the time.

The Census Bureau also conducts counts in institutions and other places, so:

- Do not count anyone living away either at college or in the Armed Forces.
- Do not count anyone in a nursing home, jail, prison, detention facility, etc., on April 1, 2010.
- Leave these people off your form, even if they will return to live here after they leave college, the nursing home, the military, jail, etc. Otherwise, they may be counted twice.

The Census must also include people without a permanent place to stay, so:

- If someone who has no permanent place to stay is staying here on April 1, 2010, count that person. Otherwise, he or she may be missed in the census.

1. How many people were living or staying in this house, apartment, or mobile home on April 1, 2010?

Number of people = ☐☐

2. Were there any additional people staying here April 1, 2010 that you did not include in Question 1? *Mark ✗ all that apply.*

- ☐ Children, such as newborn babies or foster children
- ☐ Relatives, such as adult children, cousins, or in-laws
- ☐ Nonrelatives, such as roommates or live-in baby sitters
- ☐ People staying here temporarily
- ☐ No additional people

3. Is this house, apartment, or mobile home — *Mark ✗ ONE box.*

- ☐ Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*
- ☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?
- ☐ Rented?
- ☐ Occupied without payment of rent?

4. What is your telephone number? *We may call if we don't understand an answer.*

Area Code + Number

☐☐☐ – ☐☐☐ – ☐☐☐☐

OMB No. 0607-0919-C: Approval Expires 12/31/2011.

Form **D-61** (1-15-2009)

5. Please provide information for each person living here. Start with a person living here who owns or rents this house, apartment, or mobile home. If the owner or renter lives somewhere else, start with any adult living here. This will be Person 1.

What is Person 1's name? *Print name below.*

Last Name ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

First Name ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ MI ☐

6. What is Person 1's sex? *Mark ✗ ONE box.*

☐ Male ☐ Female

7. What is Person 1's age and what is Person 1's date of birth?

Please report babies as age 0 when the child is less than 1 year old.
Print numbers in boxes.

Age on April 1, 2010	Month	Day	Year of birth
☐☐☐	☐☐	☐☐	☐☐☐☐

➜ **NOTE: Please answer BOTH Question 8 about Hispanic origin and Question 9 about race. For this census, Hispanic origins are not races.**

8. Is Person 1 of Hispanic, Latino, or Spanish origin?

- ☐ **No,** not of Hispanic, Latino, or Spanish origin
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

9. What is Person 1's race? *Mark ✗ one or more boxes.*

- ☐ White
- ☐ Black, African Am., or Negro
- ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*

☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

- ☐ Asian Indian
- ☐ Chinese
- ☐ Filipino
- ☐ Japanese
- ☐ Korean
- ☐ Vietnamese
- ☐ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
- ☐ Native Hawaiian
- ☐ Guamanian or Chamorro
- ☐ Samoan
- ☐ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*

☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

- ☐ Some other race — *Print race.*

☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

10. Does Person 1 sometimes live or stay somewhere else?

☐ No ☐ Yes — *Mark ✗ all that apply.*

- ☐ In college housing
- ☐ In the military
- ☐ At a seasonal or second residence
- ☐ For child custody
- ☐ In jail or prison
- ☐ In a nursing home
- ☐ For another reason

➜ **If more people were counted in Question 1, continue with Person 2.**

USCENSUSBUREAU

1. Print name of Person 2

Last Name

First Name MI

2. How is this person related to Person 1? *Mark ✗ ONE box.*

- ☐ Husband or wife
- ☐ Biological son or daughter
- ☐ Adopted son or daughter
- ☐ Stepson or stepdaughter
- ☐ Brother or sister
- ☐ Father or mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law or daughter-in-law
- ☐ Other relative
- ☐ Roomer or boarder
- ☐ Housemate or roommate
- ☐ Unmarried partner
- ☐ Other nonrelative

3. What is this person's sex? *Mark ✗ ONE box.*

☐ Male ☐ Female

4. What is this person's age and what is this person's date of birth?

Please report babies as age 0 when the child is less than 1 year old.

Print numbers in boxes.

Age on April 1, 2010 | Month | Day | Year of birth

➜ NOTE: Please answer BOTH Question 5 about Hispanic origin and Question 6 about race. For this census, Hispanic origins are not races.

5. Is this person of Hispanic, Latino, or Spanish origin?

- ☐ **No,** not of Hispanic, Latino, or Spanish origin
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

6. What is this person's race? *Mark ✗ one or more boxes.*

- ☐ White
- ☐ Black, African Am., or Negro
- ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
- ☐ Asian Indian
- ☐ Chinese
- ☐ Filipino
- ☐ Japanese
- ☐ Korean
- ☐ Vietnamese
- ☐ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
- ☐ Native Hawaiian
- ☐ Guamanian or Chamorro
- ☐ Samoan
- ☐ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*
- ☐ Some other race — *Print race.*

7. Does this person sometimes live or stay somewhere else?

☐ No ☐ Yes — *Mark ✗ all that apply.*

- ☐ In college housing
- ☐ In the military
- ☐ At a seasonal or second residence
- ☐ For child custody
- ☐ In jail or prison
- ☐ In a nursing home
- ☐ For another reason

➜ If more people were counted in Question 1 on the front page, continue with Person 3.

1. Print name of Person 3

Last Name

First Name MI

2. How is this person related to Person 1? *Mark ✗ ONE box.*

- ☐ Husband or wife
- ☐ Biological son or daughter
- ☐ Adopted son or daughter
- ☐ Stepson or stepdaughter
- ☐ Brother or sister
- ☐ Father or mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law or daughter-in-law
- ☐ Other relative
- ☐ Roomer or boarder
- ☐ Housemate or roommate
- ☐ Unmarried partner
- ☐ Other nonrelative

3. What is this person's sex? *Mark ✗ ONE box.*

☐ Male ☐ Female

4. What is this person's age and what is this person's date of birth?

Please report babies as age 0 when the child is less than 1 year old.

Print numbers in boxes.

Age on April 1, 2010 | Month | Day | Year of birth

➜ NOTE: Please answer BOTH Question 5 about Hispanic origin and Question 6 about race. For this census, Hispanic origins are not races.

5. Is this person of Hispanic, Latino, or Spanish origin?

- ☐ **No,** not of Hispanic, Latino, or Spanish origin
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

6. What is this person's race? *Mark ✗ one or more boxes.*

- ☐ White
- ☐ Black, African Am., or Negro
- ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
- ☐ Asian Indian
- ☐ Chinese
- ☐ Filipino
- ☐ Japanese
- ☐ Korean
- ☐ Vietnamese
- ☐ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
- ☐ Native Hawaiian
- ☐ Guamanian or Chamorro
- ☐ Samoan
- ☐ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*
- ☐ Some other race — *Print race.*

7. Does this person sometimes live or stay somewhere else?

☐ No ☐ Yes — *Mark ✗ all that apply.*

- ☐ In college housing
- ☐ In the military
- ☐ At a seasonal or second residence
- ☐ For child custody
- ☐ In jail or prison
- ☐ In a nursing home
- ☐ For another reason

➜ If more people were counted in Question 1 on the front page, continue with Person 4.

1. Print name of Person 4

Last Name

First Name MI

2. How is this person related to Person 1? *Mark ✗ ONE box.*

- ☐ Husband or wife
- ☐ Biological son or daughter
- ☐ Adopted son or daughter
- ☐ Stepson or stepdaughter
- ☐ Brother or sister
- ☐ Father or mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law or daughter-in-law
- ☐ Other relative
- ☐ Roomer or boarder
- ☐ Housemate or roommate
- ☐ Unmarried partner
- ☐ Other nonrelative

3. What is this person's sex? *Mark ✗ ONE box.*

☐ Male ☐ Female

4. What is this person's age and what is this person's date of birth?

Please report babies as age 0 when the child is less than 1 year old.
Print numbers in boxes.

Age on April 1, 2010 | Month | Day | Year of birth

➜ NOTE: Please answer BOTH Question 5 about Hispanic origin and Question 6 about race. For this census, Hispanic origins are not races.

5. Is this person of Hispanic, Latino, or Spanish origin?

- ☐ **No,** not of Hispanic, Latino, or Spanish origin
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

6. What is this person's race? *Mark ✗ one or more boxes.*

- ☐ White
- ☐ Black, African Am., or Negro
- ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
- ☐ Asian Indian
- ☐ Chinese
- ☐ Filipino
- ☐ Japanese
- ☐ Korean
- ☐ Vietnamese
- ☐ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
- ☐ Native Hawaiian
- ☐ Guamanian or Chamorro
- ☐ Samoan
- ☐ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*
- ☐ Some other race — *Print race.*

7. Does this person sometimes live or stay somewhere else?

☐ No ☐ Yes — *Mark ✗ all that apply.*

- ☐ In college housing
- ☐ In the military
- ☐ At a seasonal or second residence
- ☐ For child custody
- ☐ In jail or prison
- ☐ In a nursing home
- ☐ For another reason

➜ If more people were counted in Question 1 on the front page, continue with Person 5.

1. Print name of Person 5

Last Name

First Name MI

2. How is this person related to Person 1? *Mark ✗ ONE box.*

- ☐ Husband or wife
- ☐ Biological son or daughter
- ☐ Adopted son or daughter
- ☐ Stepson or stepdaughter
- ☐ Brother or sister
- ☐ Father or mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law or daughter-in-law
- ☐ Other relative
- ☐ Roomer or boarder
- ☐ Housemate or roommate
- ☐ Unmarried partner
- ☐ Other nonrelative

3. What is this person's sex? *Mark ✗ ONE box.*

☐ Male ☐ Female

4. What is this person's age and what is this person's date of birth?

Please report babies as age 0 when the child is less than 1 year old.
Print numbers in boxes.

Age on April 1, 2010 | Month | Day | Year of birth

➜ NOTE: Please answer BOTH Question 5 about Hispanic origin and Question 6 about race. For this census, Hispanic origins are not races.

5. Is this person of Hispanic, Latino, or Spanish origin?

- ☐ **No,** not of Hispanic, Latino, or Spanish origin
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

6. What is this person's race? *Mark ✗ one or more boxes.*

- ☐ White
- ☐ Black, African Am., or Negro
- ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
- ☐ Asian Indian
- ☐ Chinese
- ☐ Filipino
- ☐ Japanese
- ☐ Korean
- ☐ Vietnamese
- ☐ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
- ☐ Native Hawaiian
- ☐ Guamanian or Chamorro
- ☐ Samoan
- ☐ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*
- ☐ Some other race — *Print race.*

7. Does this person sometimes live or stay somewhere else?

☐ No ☐ Yes — *Mark ✗ all that apply.*

- ☐ In college housing
- ☐ In the military
- ☐ At a seasonal or second residence
- ☐ For child custody
- ☐ In jail or prison
- ☐ In a nursing home
- ☐ For another reason

➜ If more people were counted in Question 1 on the front page, continue with Person 6.

1. Print name of Person 6

Last Name

First Name MI

2. How is this person related to Person 1? *Mark ✗ ONE box.*

- ☐ Husband or wife
- ☐ Biological son or daughter
- ☐ Adopted son or daughter
- ☐ Stepson or stepdaughter
- ☐ Brother or sister
- ☐ Father or mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law or daughter-in-law
- ☐ Other relative
- ☐ Roomer or boarder
- ☐ Housemate or roommate
- ☐ Unmarried partner
- ☐ Other nonrelative

3. What is this person's sex? *Mark ✗ ONE box.*

☐ Male ☐ Female

4. What is this person's age and what is this person's date of birth?

Please report babies as age 0 when the child is less than 1 year old.

Print numbers in boxes.

Age on April 1, 2010	Month	Day	Year of birth

➜ **NOTE: Please answer BOTH Question 5 about Hispanic origin and Question 6 about race. For this census, Hispanic origins are not races.**

5. Is this person of Hispanic, Latino, or Spanish origin?

- ☐ **No,** not of Hispanic, Latino, or Spanish origin
- ☐ Yes, Mexican, Mexican Am., Chicano
- ☐ Yes, Puerto Rican
- ☐ Yes, Cuban
- ☐ Yes, another Hispanic, Latino, or Spanish origin — *Print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.*

6. What is this person's race? *Mark ✗ one or more boxes.*

- ☐ White
- ☐ Black, African Am., or Negro
- ☐ American Indian or Alaska Native — *Print name of enrolled or principal tribe.*
- ☐ Asian Indian
- ☐ Chinese
- ☐ Filipino
- ☐ Other Asian — *Print race, for example, Hmong, Laotian, Thai, Pakistani, Cambodian, and so on.*
- ☐ Japanese
- ☐ Korean
- ☐ Vietnamese
- ☐ Native Hawaiian
- ☐ Guamanian or Chamorro
- ☐ Samoan
- ☐ Other Pacific Islander — *Print race, for example, Fijian, Tongan, and so on.*
- ☐ Some other race — *Print race.*

7. Does this person sometimes live or stay somewhere else?

☐ No ☐ Yes — *Mark ✗ all that apply.*

- ☐ In college housing
- ☐ In the military
- ☐ At a seasonal or second residence
- ☐ For child custody
- ☐ In jail or prison
- ☐ In a nursing home
- ☐ For another reason

➜ **If more than six people were counted in Question 1 on the front page, turn the page and continue.**

➜ If more people live here, turn the page and continue.

Use this section to complete information for the rest of the people you counted in Question 1 on the front page. *We may call for additional information about them.*

Person 7

Last Name | **First Name** | **MI**

Sex
- ☐ Male
- ☐ Female

Age on April 1, 2010

Date of Birth — Month | Day | Year

Related to Person 1?
- ☐ Yes
- ☐ No

Person 8

Last Name | **First Name** | **MI**

Sex
- ☐ Male
- ☐ Female

Age on April 1, 2010

Date of Birth — Month | Day | Year

Related to Person 1?
- ☐ Yes
- ☐ No

Person 9

Last Name | **First Name** | **MI**

Sex
- ☐ Male
- ☐ Female

Age on April 1, 2010

Date of Birth — Month | Day | Year

Related to Person 1?
- ☐ Yes
- ☐ No

Person 10

Last Name | **First Name** | **MI**

Sex
- ☐ Male
- ☐ Female

Age on April 1, 2010

Date of Birth — Month | Day | Year

Related to Person 1?
- ☐ Yes
- ☐ No

Person 11

Last Name | **First Name** | **MI**

Sex
- ☐ Male
- ☐ Female

Age on April 1, 2010

Date of Birth — Month | Day | Year

Related to Person 1?
- ☐ Yes
- ☐ No

Person 12

Last Name | **First Name** | **MI**

Sex
- ☐ Male
- ☐ Female

Age on April 1, 2010

Date of Birth — Month | Day | Year

Related to Person 1?
- ☐ Yes
- ☐ No

Thank you for completing your official 2010 Census form.

FOR OFFICIAL USE ONLY

JIC1 | JIC2

U.S. Census Bureau
Privacy Impact Assessments (PIAs) and Privacy Threshold Analysis (PTA)

System Name	PTA	PIA	Approval Date
• CEN01 Data Communications		PIA	4/23/2015
• CEN02 Lenel		PIA	8/8/2016
• CEN03 Economic Census and Surveys and Special Processing		PIA	6/28/2016
• CEN04 Commerce Business Systems		PIA	3/4/2015
• CEN05 Field Systems Major Application System		PIA	6/29/2015
• CEN06 National Processing Center (NPC)		PIA	8/20/2014
• CEN08 Decennial		PIA	8/11/2014
• CEN09 Cloud Services		PIA	10/16/2016
• CEN11 Demographic Census, Surveys, and Special Processing		PIA	7/1/2014
• CEN12 AESDirect		PIA	9/17/2014
• CEN13 Center for Economic Studies		PIA	6/2/2015

• CEN15 Centurion	PIA	2/10/2015
• CEN17 Client Services General Support System	PIA	8/8/2016
• CEN18 Enterprise Applications	PIA	6/7/2016
• CEN20 Budget Division Applications	PIA	5/5/2016
• CEN21 Human Resources Division Applications	PIA	8/8/2016
• CEN25 Office of Information Security (OIS) Systems	PIA	10/1/2014
• CEN26 SharePoint	PIA	7/16/2015
• CEN30 Amercian Community Survey	PIA	3/24/2015
• CEN31 Administrative Systems Vol. II	PIA	4/27/2016
• CEN33 DataWeb	PIA	2/4/2015
• CEN34 Foreign Trade Division Applications	PIA	6/2/2014
• CEN35 Governments Division Applications	PIA	11/4/2014
• CEN36 integrated Computer Assisted Data Entry (iCADE), Census Image Retrieval Application (CIRA), and MOJO Enhanced Operational Control System	PIA	7/15/2015
• Automated Export System (AES)	PIA	8/11/2014

• Social Media and Web 2.0 Websites and Applications (Third Party Website with PII)	PIA	3/31/2015

Questions and Comments

Send Questions, Comments or Complaints on the Commerce Privacy program to CPO@doc.gov.

Office of Privacy and Open Government
Office of the Chief Financial Officer and Assistant Secretary for Administration
U.S. Department of Commerce

Page last updated: December 13, 2016

BYLAWS OF

THE ELECTRONIC PRIVACY INFORMATION CENTER ("EPIC")

As Adopted As of March 16, 2000,
As Amended May 1, 2007,
As Amended February 8, 2010,
As Amended February 1, 2015,
As Amended January 26, 2018

ARTICLE I
NAME AND PURPOSES

Section 1.01. Name. The name of the organization is ELECTRONIC PRIVACY INFORMATION CENTER.

Section 1.02. Purpose. The Corporation is organized for the charitable and educational purposes of promoting personal privacy and constitutional rights.

ARTICLE II
AUTHORITY AND DUTIES OF DIRECTORS

Section 2.01. Authority of Directors. The Board of Directors ("the Board") is the policy-making body and may exercise all the powers and authority granted to the Corporation by law.

Section 2.02. Number, Selection, and Tenure. The Board shall consist of no more than fifteen (15) nor less than seven (7) Directors, all of whom must be Members of the Corporation, elected in staggered terms of three (3) years, nominated from among its Members, and confirmed by the Board. A director may serve for any number of terms, consecutive or otherwise. Directors will be elected by the directors then in office. The Board will seek to promote pluralism and diversity among its membership.

Section 2.03. Compensation. Board members shall not be compensated for serving on the Board, but may be reimbursed for actual expenses incurred on behalf of the Corporation. Board members who also serve as employees of the Corporation may be compensated for their service as employees. No more than one (1) paid staff member shall serve as a voting member of the Board. A paid staff member serving on the Board shall not serve as the Chair or Treasurer.

Section 2.04. Resignation and Removal. Resignations are effective upon receipt by the Secretary of written notification, or receipt by the President or other officer if the Secretary is resigning. One or more Directors may be removed at a meeting called for that purpose, with or without cause, by such vote as would suffice for the Director's election.

Section 2.05. Vacancies. Vacancies existing by reason of resignation, death, incapacity or removal before the expiration of a term shall be filled by a majority vote of the remaining

directors.

Section 2.06. Meetings. The Board shall hold at least one regular meeting annually. Regular meetings shall be at such times and places as the Board shall determine. Special meetings may be called by any two directors with telephone or written notice.

If an officer or director fails to attend three (3) consecutive meetings of the Board or Executive Committee, the Executive Committee shall evaluate the officer's or director's contribution to the work of the Corporation, his or her reasons for not attending the meetings, as well as any other relevant factors, and if it appears to be in the best interest of the Corporation, may declare the position vacant.

Section 2.07. Quorum and Voting. A quorum shall consist of a majority of the total number of Board members in office. All decisions will be by majority vote of those present at a meeting at which a quorum is present.

Section 2.08. Action Without a Meeting. Any action required or permitted to be taken at a meeting of the Board (including amendment of these Bylaws or the Articles of Incorporation) or of any committee may be taken without a meeting if all the members of the Board or committee consent in writing to taking the action without a meeting and to approving the specific action. Such consents shall have the same force and effect as a unanimous vote of the Board or of the committee as the case may be.

Section 2.09. Participation in Meeting by Conference Telephone. Any or all members of the Board may participate in a meeting by conference telephone or similar communications equipment, so long as members participating in such meeting can hear one another.

Section 2.10. Committees. The Board may, by resolution adopted by a majority of the Directors in office, establish committees of the Board composed of at least two (2) Directors.

Other committees not having and exercising the authority of the Board in the management of the corporation may be designated and appointed by a resolution adopted by a majority of the directors present at a meeting at which a quorum is present. Such committees shall consist of at least two members.

The standing committees of the Board shall be an Executive Committee, composed of the officers, and the Program and Litigation Committee, Finances and Fundraising Committee, Management and Staffing Committee, and Engagement and Outreach Committee. The names of these committees may be changed by resolution of the Board.

For both types of committees, the Board may make provisions for appointment of the chair, establish procedures to govern their activities, and delegate authority as may be necessary or desirable for the efficient management of the property, affairs, business, and/or activities of the Corporation.

ARTICLE III
AUTHORITY AND DUTIES OF OFFICERS

Section 3.01. Officers. The officers of the Corporation shall be a President, a Chair, a Secretary, a Treasurer, and such other officers as the Board may designate. Two or more offices may be held by the same person, except the offices of secretary and president.

Section 3.02. Election of Officers; Terms of Office. Officers shall serve one (1) year terms. The President, the Secretary, and the Treasurer shall be elected by the Board at its annual meeting in each year. The terms of office shall expire at the next succeeding annual meeting and shall be filled by the Board, at a meeting or by action in writing pursuant to Section 2.08 for a term expiring at the next succeeding annual meeting. Officers shall be eligible for reelection.

Vacancies existing by reason of resignation, death, incapacity or removal before the expiration of a term shall be filled by the Board for the remainder of the unexpired term.

Section 3.03. Resignation. Resignations are effective upon receipt by the Secretary of a written notification, or receipt by the President if the Secretary is resigning.

Section 3.04. Removal. An officer may be removed by the Board at a meeting, or by action in writing pursuant to Section 2.08 whenever in the Board's judgment the best interests of the Corporation will be served thereby. The removal of a person from corporate office will not terminate or otherwise affect any contractual relationship between that individual and the corporation.

ARTICLE IV
INDEMNIFICATION

Section 4.01. Definitions. "Matter" shall mean any actual or threatened civil, criminal, or administrative action, arbitration proceeding, claim, suit, proceeding, or appeals therefrom, or any criminal, administrative, or Congressional (or other body's) investigation, hearing, or other proceeding.
"Eligible Person" shall mean any person who at any time was or is a director, a member or any committee or subcommittee, an officer, employee, or volunteer of the corporation.

Section 4.02. Right to Indemnification. Any Eligible Person made a party to or involved in a Matter by reason of his or her position with or service to the corporation shall**,** to the fullest extent permitted by law, be indemnified by the corporation against all liabilities and all expenses reasonably incurred by him or her arising out of or in connection with such Matter, except in relation to Matters as to which (i) the Eligible Person failed to act in good faith and for a purpose which he or she reasonably believed to be in the best interests of the corporation, or (ii) in the case of a criminal Matter, the person had reasonable cause to believe that his or her conduct was unlawful, or (iii) the person shall be adjudged to be liable for willful misconduct in the performance of a duty, or in the case of a Matter settled by agreement, the settlement shall be predicated on such a liability.

Section 4.03. Limitation on Right of Indemnification. Except where an Eligible Person has been successful on the merits with respect to such Matter, any indemnification hereunder shall be made only after (i) the Board (acting by a quorum consisting of Directors who were not involved in such Matter) determines that the Eligible Person met the applicable indemnification standard set forth in section 4.02 above; or (ii) in the absence of a quorum, a finding is rendered in a written opinion by independent legal counsel that the person or persons met the applicable indemnification standard set forth in section 4.02 above.

Section 4.04. Other Rights. The right of indemnification provided hereunder shall not be deemed exclusive of any other right to which any person may be entitled in addition to the indemnification provided hereunder. This indemnification shall in the case of the death of the person entitled to indemnification, inure to the benefit of his or her heirs, executors or other lawful representative.

Section 4.05. Interim Indemnification. The corporation shall, with respect to a Matter described in section 4.02, advance attorneys fees as interim indemnification to any Eligible Person if the following conditions are satisfied: (i)(a) the Board (acting by a quorum consisting of Directors who are not involved in such litigation) determines that the Eligible Person is likely to meet the applicable indemnification standard set forth in section 4.02 above, or (b) in the absence of such a quorum, a finding is rendered in a written opinion by independent legal counsel that the Eligible Person is likely to meet the applicable indemnification standard set forth in section 4.02 above; and (ii) the Eligible Person (a) requests interim indemnification, (b) agrees to repay the interim indemnification promptly upon a determination unfavorable to him or her under section 4.03, and (c) deposits a bond or equivalent security.

Section 4.06. Insurance. The Board may authorize the purchase of and maintain insurance on behalf of any Eligible Person against any liability asserted against or incurred by him which arises out of such person's status in such capacity, or out of acts taken in such capacity, whether or not the Corporation would have the power to indemnify the person against that liability under law.

ARTICLE V
MEMBERS

Section 5.01. Qualifications. The Corporation shall designate as Members, following nomination by the current Members and a vote of the Board, distinguished experts in law, technology, and public policy.

Section 5.02. Advisory Board. Members shall sit on the Advisory Board, where they shall provide guidance for the work of the Corporation, participate in the activities of the Corporation, offer support for the Corporation, and provide leadership for the Corporation

Section 5.03. Dues. Members shall pay dues in an amount and on a schedule established by the Board. There shall be provisions for lifetime memberships.

Section 5.04. Service on Committees. Members may serve on Board committees.

Section 5.05. Annual Meeting. There shall be an annual meeting, open to all Members of the Corporation.

Section 5.05. Annual Evaluation. The Members shall provide an annual evaluation of the Corporation, which shall be reported to the Board of Directors at the Annual Meeting.

ARTICLE VI
FINANCIAL ADMINISTRATION

Section 6.01. Fiscal Year. The fiscal year of the Corporation shall be January 1 -December 31 but may be changed by resolution of the Board.

Section 6.02. Checks, Drafts, Etc. All checks, orders for the payment of money, and insurance certificates shall be signed or endorsed by an officer or officers or agent or agents of the Corporation and in a manner as shall from time to time be determined by resolution of the Board or of any committee to which such authority has been expressly delegated by the Board.

Section 6.03. Contracts. Unless the Board determines otherwise by resolution, the President, Secretary, Treasurer, Board Chair and Chair of the Executive Committee shall all be authorized to execute contracts on behalf of the corporation.

These individuals may, with written notice to the Board, delegate this authority to employees or volunteers subject to limitations upon the delegated authority as may be necessary or expedient for running the affairs of the corporation. Unless otherwise expressly determined by the Board, no other individuals shall be authorized to bind the corporation to any contract, including the chair of any committee other than the Executive Committee.

Section 6.04. Deposits and Accounts. All funds of the Corporation, not otherwise employed, shall be deposited in general or special accounts in the banks, trust companies, or other depositories as the Board or any committee to which such authority has been delegated by the Board may select, or as may be selected by any officer or officers or agent or agents of the Corporation, to whom such power may be delegated by the Board. For the purpose of deposit and for the purpose of collection for that account of the Corporation, checks, drafts, and other orders of the Corporation may be endorsed, assigned, and delivered on behalf of the Corporation by any officer or agent of the Corporation.

Section 6.05. Annual Financial Statements. Complete financial statements shall be presented to and reviewed by the Board after the close of each fiscal year.

ARTICLE VII
CONFLICTS OF INTEREST

Section 7.01. Disclosure of Financial Interests. To identify possible conflicts of interest, all directors, officers, and members of any committee exercising Board- delegated powers must

disclose to the Board, or to the members of such committee, the existence of any financial interest in any entity with which s/he knows or has reason to know the Corporation or any legally related organization has or is negotiating a transaction or arrangement, and all material facts related to that interest. Financial interests include any direct or indirect relationship, through business, investment, or family, such as actual or potential ownership or investment interests or compensation arrangements. Directors shall also disclose any fiduciary duty to a person or entity other than the Corporation that might jeopardize the director's ability to exercise independent judgment and act in the best interests of the Corporation. The fact that a director, officer, or committee member is also a director or officer or member of a not-for-profit organization that obtains or seeks funds from institutions or individuals from which the Corporation also obtains or seeks funds shall not by itself be deemed to be a conflict of interest.

Section 7.02. Determination of Conflicts of Interest. After the interested person has delivered all relevant information and has retired from the room, the Board or committee must determine whether or not the financial interest creates a conflict of interest that merits recusal of the interested Director from consideration of the matter.

Section 7.03. Resolution of Conflicts of Interest. If the Board determines that a conflict of interest does exist, it must ensure that the interested director(s) do not participate in final decision making with regard to the transaction. The Board may approve the transaction or arrangement, or some alternative if it determines it: a) is in the organization's best interests and for its own benefit; b) is fair and reasonable to the organization; and c) is the most advantageous transaction or arrangement the organization can obtain with reasonable efforts under the circumstances.

Section 7.04. Violation of Conflict of Interest Policy. If an officer, director, or member of a committee with Board-delegated powers violates this conflict of interest policy, the Board, in order to protect the Corporation's best interests, may take appropriate disciplinary action against the interested person. Such action may include formal reprimand, cancellation of the transaction or arrangement generating the conflict, suspension of employment, and/or removal from the Board.

Section 7.05. Distribution of Conflict of Interest Policy. All officers, directors, and members of committees with Board-delegated powers shall receive a copy of the Conflict of Interest Policy, as it appears in these By-laws. All officers, directors, and members of committees with Board-delegated powers shall sign an annual statement declaring that the person: received a copy of the policy; has read and understands the policy; and agrees to comply with the policy.

ARTICLE VIII
COMPENSATION

The Board shall adopt a policy establishing procedures for reviewing and setting financial compensation to any individual exercising substantial influence over the corporation to ensure that such compensation is no more than reasonable and does not otherwise result in an excess benefit to the person, and requiring that adequate documentation be maintained to support the basis for setting such compensation.

ARTICLE IX
RECORD KEEPING

The Secretary or his or her designee shall keep or cause to be kept adequate minutes of all Board or committee meetings, and all meetings of committees with Board-designated powers reflecting at a minimum the names of those in attendance, any resolutions passed and the outcomes of any votes taken. When potential conflicts of interests are discussed, the minutes shall include: the names of the persons who disclosed financial interests; the nature of the financial interests; whether or not the Board determined that a conflict existed; the names of the persons present for the discussions and votes related to the relevant transaction or arrangement; the content of those discussions, including any alternative transactions or arrangements; and a record of the vote. At the request of any participating Board member, the records of such discussions and individual votes may be kept sealed, with only the outcome reported publicly.

ARTICLE X
ANNUAL REPORTS

An annually updated written account of the Corporation's purposes, structure, programs and financial condition shall be published and made publicly available. The annual report shall contain: a description of the Corporation's purpose(s); descriptions of its overall programs, activities and accomplishments; a statement of its eligibility to receive deductible contributions; information about the governing body and structure, including identification of officers, directors, and chief administrative personnel; and the audited financial statements or, at a minimum, a comprehensive financial summary that reflects all revenue, reports expenses by program, management and fund-raising categories, and reports year-end balances.

ARTICLE XI
AMENDMENT OF BYLAWS

These Bylaws may be amended by a majority vote of the Board, provided seven (7) days' prior notice is given of the proposed amendment or provided all members of the Board waive such notice, or by unanimous consent in writing without a meeting pursuant to Section 2.08.

PREPARED STATEMENT OF

RON JARMIN, PhD.
PERFORMING THE NON-EXCLUSIVE FUNCTIONS AND DUTIES OF THE DIRECTOR
U.S. CENSUS BUREAU

and

EARL COMSTOCK
DIRECTOR, OFFICE OF POLICY AND STRATEGIC PLANNING
U.S. DEPARTMENT OF COMMERCE

Before the Committee on Oversight and Government Reform
U.S. House of Representatives

"Progress Report on the 2020 Census"

8 May 2018

Good afternoon, Chairman Gowdy, Ranking Member Cummings, and members of the Committee. We welcome this opportunity to update you on the status of the 2020 Census. Before we begin, we would like to thank Congress for the Census Bureau's recent appropriation, which underscores Congress' commitment to a successful 2020 Census. The Department of Commerce and Census Bureau share your commitment and today we will provide a brief review of our progress to conduct a complete and accurate census.

2018 End-to-End Census Test

We are in the midst of the 2018 End-to-End Census Test. This is our final major field test before the 2020 Census. We are testing the interfaces between 44 IT systems critical for the 2020 Census, and their integration with the 24 major operations that are part of the 2018 End-to-End Census Test.

Tomorrow, we begin the Nonresponse Follow-up Operations for the 2018 End-to-End Census Test. Nonresponse Follow-up is an eight-week field operation to gather responses from households that have not yet responded by Internet, telephone, or paper questionnaire. Approximately 56 percent of households in the test did not respond to the Census questionnaire, so we will send approximately 900 enumerators to knock on doors and gather the information. This key operation involves 31 systems with appropriate interfaces, and will provide important feedback as we prepare to conduct the 2020 Census.

The 2018 End-to-End Census Test's Nonresponse Follow-up operation began in August 2017 with the Address Canvassing operation in three locations: Pierce County, Washington; the Bluefield-Beckley-Oak Hill, West Virginia area; and, Providence County, Rhode Island. These locations allowed the Census Bureau to test systems with and without Internet connectivity and to study critical address list development operations in a wide range of geographical situations including rural and mountainous areas. Importantly, in areas with low Internet connectivity Census Bureau field staff were able to receive their assignments and submit their payroll and operational data at the beginning or end of their shifts. By design, staff began and ended their shifts in areas with Internet connectivity so that they could do this. They did not require Internet access while they were actually conducting the listing operation. We were able to successfully integrate with the operational control system in field conditions. We also integrated with the other decennial systems effectively, such as the system we are using to provide directional information to staff ensuring they complete their work in the most efficient manner possible. Additionally, we tested the implementation of the independent quality control component, which ensures the proper disposition of cases in real time providing us with important information when cases fail to pass quality control.

The 2017 Address Canvassing operation revealed areas where our systems and operations need to improve, especially with regard to training. Some of these areas are being addressed as we continue to refine our training modules. Address listing is a complicated process and we have identified ways to improve technical training. The West Virginia site, in particular, experienced connectivity issues. The challenge of unreliable wireless signals is one of the reasons that site was chosen. Because of connectivity issues, some of the employees hired to carry out the address listing operation had difficulty completing the self-paced online training at their home. We continue to examine the experiences of our listers so that we can enhance this functionality and improve both systems and operations.

Just as we learned a great deal from the Address Canvassing operation, we are learning from the operations conducted in Providence. We will continue to make the necessary adjustments to our systems and operations in response to what we observe and experience during the 2018 End-to-End Census Test.

We opted to conduct the 2018 End-to-End Test in Providence County because it represents an ideal location for testing data collection operations. It offers many different situations and provides challenges that we will face across the country in 2020, and its demographics mirror those of the nation. Providence County has a population of over 600,000, more than a quarter-million housing units, and historically hard-to-count populations. Providence County provides an opportunity to test all of the systems and operations planned for the 2018 End-to-End Census Test. Providence County's urban areas include high vacancy rates, and neighborhoods and housing that have undergone considerable conversion for many years. This allows us to validate our address canvassing operation (for example, identifying split or converted housing units). There also is a solid presence of Group Quarters, which are residences that contain multiple unrelated residents, allowing us to conduct that operation.

The test's self-response phase began with a series of mailings that were sent to housing units beginning in mid-March. We received a high number of responses through the Internet—almost

64 percent of the total responses. Households also are responding by telephone (6 percent) or returning the paper questionnaire by mail (30 percent). However, consistent with our planned approach for the 2020 Census, households only received a questionnaire on the first mailing if the household does not have strong Internet connectivity or if the household is considered unlikely to use the Internet. Households with older people often fall into this latter category. Regardless, every household that had not responded after three mailings received a paper questionnaire on the fourth mailing.

The remaining peak operations in Providence, Rhode Island, are Group Quarters, Update Leave, and Nonresponse Follow-up. The Group Quarters operation is designed to enumerate people who live in places such as college dormitories, skilled nursing facilities, and prisons. The Update Leave operation is designed for areas that do not have a city-style address, such as "100 Main Street," where the address and the geographic location of the housing units are linked. Areas without city-style addresses are most common in rural areas where mailing addresses are grouped together (e.g., multiple mail boxes grouped together at the end of a rural road) and not tied to the geographic location of the actual housing unit (e.g., Post Office Boxes). During Update Leave, field staff will update our address list and leave a questionnaire packet at each household in this area. Those who do not respond are included in the Nonresponse Follow-up operation where we collect information by sending a census enumerator to interview the household.

All of the planned innovations for the 2020 Census are coming together in the 2018 End-to-End Census Test and the lessons learned will provide a firm foundation for success in 2020. We are looking closely at data from the Address Canvassing operation to make sure that our blend of in-office and in-field Address Canvassing meets our standards for an accurate and comprehensive address list. The Internet and telephone response modes are being thoroughly tested, as is the Paper Data Capture operation. We also are making it easier for people to respond in real time with a smart phone or a tablet. Field staff will collect information using hand-held devices and their work will be managed efficiently and effectively by leveraging automated processes. Finally, we will continue to examine the use of administrative records to inform final determinations about our ability to improve the efficiency of the Nonresponse Follow-up operation using information that people have already provided to the government. During and after the test, we will adjust the systems and operations based on what we learned to make sure that both are ready for the 2020 Census.

Systems Readiness

We developed a comprehensive schedule for developing and integrating the key systems for the 2018 End-to-End Census Test. Each system has its own well-defined scope, requirements, schedule, and costs, and each is overseen by experienced project management teams. As we enter the peak operations of the test, 40 out of the 44 systems supporting the test have been deployed. No system will be released without completing the necessary integration testing and security authorizations. All 44 are on track to be fully integrated and deployed when they are needed to support operations in the test.

The Census Bureau has been working closely with the Government Accountability Office (GAO) to ensure that the Census Bureau and the GAO share a consistent understanding of the status of systems readiness, and we both are monitoring final development and testing of the remaining systems needed for the 2018 End-to-End Census Test. All of our systems for the 2020 Census are being integrated through a contracted solution that the Census Bureau refers to as the "technical integrator."

The Census Bureau's Office of Information Security, under the guidance of the Department of Commerce's Chief Information Officer, is working with the technical integrator to ensure that the systems are secure and are authorized to operate (ATO) before going live. The ATO process is critical because it ensures that cybersecurity standards are addressed and risks are minimized for all systems based on federal policies and procedures. GAO is reviewing our progress to ensure that our processes and procedures within our Systems Engineering and Integration framework are consistent with GAO's best practices. We value GAO's independent assessment and are appreciative of their continued assessment and support.

With respect to the 2018 End-to-End Census Test and looking forward to the 2020 Census, the Census Bureau is maintaining a well-defined schedule and framework for releasing systems for the remainder of the 2018 Test and then for the 2020 Census. Lessons learned from the 2018 Test will be incorporated to enhance and improve the systems. We also have a well-developed process for conducting the scalability tests on our systems during 2018. The technical integrator is enabling the scalability test using infrastructure in the cloud and the data center. Conducting the scalability tests is a major milestone this year, and we are eager to meet this milestone and fine tune our systems to scale to predicted peak loads in 2020.

Finally, we are engaging with the private sector and the federal government intelligence community in the area of cybersecurity to ensure the information we collect is protected, that we can withstand the threat of cyber attacks, and, if necessary, contain a threat in a way that sustains service and maintains public trust.

2020 Operations Already Underway: LUCA, In-Office Address Canvassing, and Delivery of the 2020 Census Questions

Even as we are in the midst of the 2018 End-to-End Census Test, key operations for the 2020 Census are already up and running. Today, we would like to summarize our progress on the Local Update of Census Addresses (LUCA) program, In-Office Address Canvassing, and the delivery of the 2020 Census planned questions to Congress.

LUCA, which began last year, provides tribal, state, and local governments an opportunity to review and comment on the Census Bureau's address list, while following the requirements of Title 13. More than eleven thousand governmental units, 45 states, the District of Columbia, and the Commonwealth of Puerto Rico, as well as major counties and cities, and tribal governments, have registered for the 2020 Census LUCA program. These registrations exceed those for the 2010 Census program and we are pleased that they cover more than 98 percent of both the country's housing units and the population. The Census Bureau has begun sending these governments their address packages so that they can conduct their review. As of today,

approximately 1200 hundred governments have returned their packages to the Census Bureau, and of these about 30 percent of the submissions have not made any changes to the address list.

The In-Office Address Canvassing operation is one of the important innovations in the 2020 Census. Since 2015, we have been updating the most recent Census Bureau address list with new information from the United States Postal Service and data from tribal, state, and local governments and third parties (i.e., commercial vendors). We also review satellite imagery to help determine where address changes are occurring. Based on these changes, the Census Bureau will develop a plan for capturing those changes. This plan will include In-Field Address Canvassing where address updates cannot be obtained or verified or in areas undergoing rapid change. The number of addresses requiring In-Field Address Canvassing is expected to be approximately 30 percent of the total number of addresses, compared to 100 percent in 2010.

The Census Bureau is required by Section 141(f) of Title 13 of the U.S. Code to submit the subjects proposed for the next census to Congress no later than three years before April 1st of the upcoming decennial year. A document fulfilling that requirement was submitted to Congress on March 28, 2017. Section 141(f) also requires the questions proposed to be included in the next census be submitted to Congress no later than two years before Census Day, which will be April 1, 2020. The document that fulfilled this requirement for the 2020 Census, and the American Community Survey, was delivered to Congress on March 29, 2018. Pursuant to the Paperwork Reduction Act of 1995, later this year the Census Bureau will provide the public an opportunity to comment on the proposed questions, and then send the proposed questions to OMB.

As background, we would like to provide an overview of the process for determining the content on the 2020 Census and the American Community Survey. The Department of Commerce and the Census Bureau conduct a rigorous legal, technical, program, and policy review of each question to determine whether it should be included for the Decennial Census Program, which includes the 2020 Census and the American Community Survey. When it passed Title 13, Congress delegated to the Secretary of Commerce the authority to determine content for the Decennial Census Program.

Throughout each decade, regular content reviews are conducted to ensure that the information collected through the Decennial Census Program is required by federal programs. In December 2017, the Department of Justice requested that the Census Bureau reinstate a citizenship question on the 2020 Census to provide census block level citizenship voting age population data, which DOJ finds critical to its enforcement of Section 2 of the Voting Rights Act. This request underwent a thorough legal, technical, and policy review. On March 26, 2018, the Secretary of Commerce directed the Census Bureau to reinstate the citizenship question on the 2020 Census.

For the 2020 Census, the Census Bureau will use the citizenship question currently used on the American Community Survey. A question on citizenship has been asked on the American Community Survey each year since 2005 and also was asked on the long form censuses in 1970, 1980, 1990, and 2000.

The Census Bureau currently is taking steps to make the necessary operational adjustments to all data collection and processing systems to include the new question. The Census Bureau also is

taking steps to further enhance its administrative record data sets, protocols, and statistical models to provide more complete and accurate data. As is our standard practice, the Census Bureau will conduct research and develop this administrative records use strategy in an open and transparent manner, consulting with expert groups and stakeholders.

In addition to citizenship, the 2020 Census will ask questions about age, sex, race, Hispanic origin, relationship, and tenure. Some operational questions, such as name and telephone number, also are asked to better administer the data collection process and to ensure greater accuracy of the data collected. Contact information is not part of the published data and all data collected is carefully protected, as mandated by federal law, to maintain confidentiality and respect the personal information of respondents.

The document *Questions Planned for the 2020 Census and American Community Survey*, which was delivered to Congress on March 29, 2018, includes images of each question, information about why we ask each question, and the federal and selected local community uses of the data.

American Community Survey data are critical for communities and businesses nationwide. The American Community Survey is the nation's premier survey, providing the only source of comparable data for all of America's communities, and will continue to cover more than 35 topics, for example, school enrollment, veteran status and period of service, home value, and computer and Internet use. In 2019 and 2020, the Census Bureau will implement several changes to make it easier for respondents to answer the survey questions and improve the quality of the American Community Survey data.

The 2020 Census also is conducted in the Island Areas of American Samoa, the Commonwealth of the Northern Mariana Islands, Guam, and the U.S. Virgin Islands. For the 2020 Island Areas Censuses, we plan to use a modified version of the American Community Survey form that better meets the needs of the Island Areas. For example, we do not plan to ask about home heating fuel but do plan to ask about sewage disposal and source of water. This innovative approach of using existing data collection and processing systems will enable the Census Bureau to more quickly and efficiently tabulate and publish data from the 2020 Census for the Island Areas.

Enumerating Traditionally Undercounted Populations

To ensure a complete and accurate census we also must encourage people to respond. As in past decennial censuses, we will mount a robust Communications and Partnership Program to encourage everyone to respond, including those who are traditionally hard to count. People may be hard to count for different reasons, such as those who are highly mobile, homeless, living in remote areas, living in gated communities, or other reasons.

For the 2020 Census, we are building a research foundation for our Integrated Partnership and Communications Program that is stronger than in 2010. We recently conducted the Census Barriers, Attitudes and Motivators Survey (CBAMS), which consisted of two components. First, a quantitative survey was mailed out to 50,000 households. We expect to receive data from this survey in the summer, which will give us important information about the major population

groups in the country. The second component consists of qualitative focus groups that will allow us to hone in on the attitudes and motivators of smaller population groups and populations that cannot be well represented in the quantitative survey, including Native Americans, Alaska Natives, and people who speak languages other than English or are not proficient in English.

The information we receive from CBAMS will be combined with other data at the Census Bureau that we are using to develop models projecting the likelihood of people to respond. Along with the information from CBAMS, this information will provide us with the strongest research foundation we have ever had. These research efforts will help us ensure that our messaging is as effective as possible and that our advertising and partnership efforts are well-directed. Finally, with the support of the recent appropriation, we are looking at other important communications activities and building out our partnership staff earlier than planned.

We recognize that communications and partnerships are at the heart of reaching traditionally hard-to-count populations. Extensive advertising in media consumed by specific populations is critical. In addition to traditional media, we will be active in digital media and on the Web. Our advertising and partnership support materials will be in multiple languages and we will develop and implement a Statistics in Schools program that will help young people understand the importance of responding to the census so that they can take the message home to their families.

We also plan to strengthen our national and local partnerships. We are currently reaching out to major corporations and national organizations, two years ahead of Census Day, so that they can build support for the 2020 Census into their business plans. This is the earliest we have ever started this effort. In addition, seasoned partnership specialists, many with decades of experience, have been working since January 2017 to help tribal, state, and local governments develop Complete Count Committees, which bring leaders and government officials together to develop plans to support the 2020 Census. As we ramp up to 1,000 partnership specialists in FY 2019, our goal is to exceed the 248,000 community partners we had in 2010. Our partners are the trusted voices in communities across the country. They include schools, hospitals, clinics, legal aid centers, faith-based organizations, and businesses large and small. Census partners help people understand that responding to the census is safe and important, and that their responses are confidential and protected by law.

Our efforts to reach traditionally undercounted populations do not stop with communications and partnerships. Through our language program, we will make it possible to respond to the Census in 12 languages, in addition to English, covering nearly 99 percent of the population, or 87 percent of those with limited English proficiency. Language support materials will be provided in a total of 59 languages, and we are developing templates to help our partners carry our messaging in the languages and dialects spoken by small population groups throughout the country.

To reach people who live in different places and situations, the Integrated Partnership and Communications Program will be the most tailored in our history. In fact, most of the operations we have mentioned —from our field operations, to our communication and partnership efforts, to our language program—are designed to reach people who live in different places or situations. We tailor our operations for rural areas and for areas that have experienced natural disasters.

With that in mind, we also would like to share that we recently made the decision to extend the Update Leave operation that we described earlier across the entire Commonwealth of Puerto Rico. This will help us ensure that our address list is correct and that we are adapting to any changes caused by Hurricane Maria. We will make similar changes to other areas as necessary, including the Gulf States impacted by flooding, and those parts of the Western states that experienced wildfires. We also conduct a specific operation in remote Alaska that starts in January of 2020 before the tundra thaws and many people leave their villages. These are only some of the many other examples of the steps we take to adapt our operations to the unique places where people live.

Finally, we are asking federal, state, and local leaders to encourage their constituents to stand up and be counted. We ask that leaders remind the public that the answers they provide on a census form are confidential and protected by law. It is important that community leaders ensure the public knows that census answers may not be used for law enforcement or any other purpose that would reveal a person's identity or how an individual responded to a question. The Census Bureau is bound by Title 13 of the United States Code. This law not only provides authority for the work we do, but also provides strong protection for the information we collect from individuals and businesses.

Anyone who handles census data swears an oath for life to keep those data confidential. Under Title 13, private information is never published. It is against the law to disclose or publish private information. The information the Census Bureau collects cannot be used for any reason except to produce statistics, and violations of Title 13 are punishable by fines and up to five years in prison. The Census Bureau trains every employee, including its field enumerators, on the importance of protecting private information and the importance of the oath, as well as the fact that penalties for violating this law are applicable for a lifetime. Public encouragement and reassurance of this law and our commitment to protecting confidentiality would help achieve our goal of a complete and accurate census in 2020.

2020 Census Activities Supported by the FY 2019 President's Budget

Looking forward, the FY 2019 President's Budget requested $3.8 billion for the Census Bureau, which includes $3.1 billion to support the 2020 Census. As noted above, the Census Bureau appreciates Congress' support for the decennial census and the inclusion of an additional $1.1 billion beyond the program's estimate for 2018 (including contingency funds) in the 2018 Omnibus appropriations act. These funds will help provide financial certainty to the program as we transition from FY 2018 to FY 2019, and represent a down payment on the $3.1 billion requested in 2019 to continue all preparatory activities and existing field operations.

As we move into 2019, we will undertake a wide range of simultaneous activities designed to finalize preparations for the 2020 Census as outlined in the President's budget. Key activities for the 2020 Census will take place in 2019. We will make final refinements for all systems to ensure they meet all requirements, are secured, are tested, and are seamlessly integrated. The field offices and other field infrastructure must be stood up nationwide in 2019, with the first 40 Area Census Offices (ACO) opening between January and March of 2019, and the remaining 208 offices opening in the summer of 2019. The ACOs will house the managers, staff, materials,

and equipment needed to support the hundreds of thousands of Census Bureau employees conducting local census operations.

Partnership Specialists will work throughout 2019 to build the network of more than 300,000 census partners who will be the trusted voices to encourage communities across the nation to respond to the 2020 Census. In October 2018, we will begin the full-scale development and implementation of all aspects of the communications program, including the establishment of the media spend plans, messaging, and the creative treatments of the advertising campaign. This work lays the foundation for our first advertising buys for the 2020 Census in May 2019.

In 2019, we will conduct the first major field operation for the 2020 Census, which is the In-Field Address Canvassing we mentioned earlier. In this operation, we send field staff to check the accuracy of the addresses in areas that are the most difficult to canvass and cannot be updated by the In-Office Address Canvassing Operation. Approximately 76,000 address canvassing listers and supervisors will be trained and sent into the field to complete this difficult and important work from mid-August to early October of 2019. Recruitment for this operation will take place in the spring of 2019, with training beginning in summer 2019. The operation itself will begin in late summer and stretch to early fall.

The physical printing of the majority of the paper materials for the 2020 Census will begin in June 2019. We also will finalize and secure all of the operations and systems related to the use of administrative records and third-party data and ensure they are working together in preparation for use in the 2020 Census.

Both the 2020 Census and the American Community Survey are essential to the Census Bureau's mission to serve as the leading source of data about our nation’s people and economy. All of the censuses and surveys we conduct are important, and we appreciate Congress’ support for our work. With less than two years until Census Day, though, we think the Congress is appropriately focused on the 2020 Census, and as you can see, we are now moving ahead at full speed. Much work remains, but we are well-positioned to get it done so that we can conduct a complete and accurate 2020 Census. We look forward to answering your questions.

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

ELECTRONIC PRIVACY INFORMATION CENTER, Plaintiff, v. UNITED STATES DEPARTMENT OF COMMERCE, *et al.*, Defendants.	Civil Action No. 1:18-cv-02711 (DLF)

DECLARATION OF ROBIN J. BACHMAN

I, Robin J. Bachman, make the following Declaration pursuant to 28 U.S.C. § 1746, and state that under penalty of perjury the following is true and correct to the best of my knowledge and belief:

1. I am the Chief of the Policy Coordination Office (PCO) and the Chief Privacy Officer with the U.S. Census Bureau and have served in this capacity since May 19, 2014. I previously served at the Census Bureau as Chief of Congressional Affairs from 1999 to 2001. As the Chief Privacy Officer, I am responsible for providing guidance to the Census Bureau programs on matters concerning confidentiality, data stewardship and safeguards, privacy and privacy compliance, including compliance with Title 13, policy development, open government and the Paperwork Reduction Act (44 U.S.C. § 3501, *et seq.*) (PRA). I supervise the work of the Privacy Compliance Branch which is responsible for promoting the agency's privacy principles across the enterprise, adherence to the requirements of the Privacy Act (5 U.S.C. § 552a et seq.) and the E-Government Act of 2002 (Public Law 107-347), as well as other federal statutes and the privacy policies of the Department of Commerce and the Office of Management of Budget.

2. In connection with my job responsibilities, I am familiar with the case *EPIC v. Department of Commerce*, No. 18-2711 (D.D.C. filed Nov. 20, 2018).

3. The purpose of this declaration is to explain the Census Bureau's process for completing Privacy Impact Assessments (PIAs), and specifically the PIA for CEN08 systems – a primary system used to administer the decennial census. CEN08 is owned by the Decennial Information Technology Division (DITD); the system is used to manage the development and implementation of a number of major decennial census applications utilized by the Decennial Census Program, including the Census Schedule A Human Resources Recruiting and Payroll Systems (C-SHaRPS), the Control and Response Data System (CaRDS), and the Third-Party Fingerprinting System. The systems covered by the CEN08 PIA contain several categories of information, including data collected from decennial census respondents, such as name, address, and date of birth, citizenship, etc. The system also includes decennial census personnel data, such as fingerprints collected for background checks of individuals applying to be enumerators for the 2020 Decennial Census.

4. Section 208 of the E-Government Act requires federal agencies to conduct PIAs for information technology (IT) systems operated by or on behalf of the U.S. government that are collecting, maintaining, or disseminating personally identifiable information (PII). A PIA is an analysis of how information in identifiable form is collected, maintained, stored, and disseminated, in addition to examining and evaluating the privacy risks associated with these tasks and the protections and processes for handling information to mitigate those privacy risks. A PIA includes both the ongoing analysis of how personally identifying information (PII) is handled as well as a formal document detailing the process and the outcome of that process. *See* Office of Management and Budget, Circular No. A-130, Appendix II-10.

5. Section 208 of the Act requires the PIA to be conducted prior to:
    a. the development or procurement of information technology that collects, maintains, or disseminates information that is in an identifiable form, or,
    b. initiating, consistent with the Paperwork Reduction Act, a new electronic collection of information in identifiable form for 10 or more persons (excluding agencies, instrumentalities or employees of the federal government).

6. Since the PIA process is initiated prior to the development and procurement of IT systems or the collection of new (i.e. not previously collected) information using existing IT systems, the Census Bureau completes and, where practicable, publishes PIAs, including the CEN08 PIA, with the understanding that the PIAs are subject to change as program operations become more refined.

7. All Census Bureau PIAs are reviewed for possible updates no fewer than once a year. The Census Bureau updates PIAs to reflect changed information collection authorities, business processes, or other factors affecting the collection and handling of information

in identifiable form, in addition to where a system change creates new privacy risks[1], such as:

a. Conversions - when converting paper-based records to electronic systems;
b. Anonymous to Non-Anonymous - when functions applied to an existing information collection change anonymous information into information in identifiable form;
a. Significant System Management Changes - when new uses of an existing IT system, including application of new technologies, significantly change how information in identifiable form is managed in the system;
b. Significant Merging - when agencies adopt or alter business processes so that government databases holding information in identifiable form are merged, centralized, matched with other databases or otherwise significantly manipulated;
c. New Public Access - when user-authenticating technology (e.g., password, digital certificate, biometric) is newly applied to an electronic information system accessed by members of the public;
d. Commercial Sources - when agencies systematically incorporate into existing information systems databases of information in identifiable form purchased or obtained from commercial or public sources. (Merely querying such a source on an ad hoc basis using existing technology does not trigger the PIA requirement);
e. New Interagency Uses - when agencies work together on shared functions involving significant new uses or exchanges of information in identifiable form, such as the cross-cutting E-Government initiatives; in such cases, the lead agency should prepare the PIA;
f. Internal Flow or Collection - when alteration of a business process results in significant new uses or disclosures of information or incorporation into the system of additional items of information in identifiable form; or
g. Alteration in Character of Data - when new information in identifiable form added to a collection raises the risks to personal privacy.

8. Upon completion or update of a PIA, the Census Bureau is required to submit the PIA to the Department of Commerce's Office of Privacy and Open Government for final review by the Department's PIA Compliance Review Board (CRB). Upon receiving concurrence from the Department's CRB, the PIA is submitted to the Department of Commerce's Senior Agency Official for Privacy (SAOP) for final approval.

9. The Census Bureau's CEN08 PIA was updated twice in 2018, and subsequently published on the Commerce Department and Census Bureau's websites: once in early 2018 as part of the annual review process (attachment A PIA published June 26, 2018) and again during late summer (attachment B PIA published September 27, 2018).

[1] Office of Management and Budget (OMB) Memorandum M-03-22.

Another update to the CEN08 PIA is currently underway and is scheduled to be released in late February or early March of 2019.

10. On March 26, 2018, Commerce Secretary Wilbur Ross announced his decision to include a citizenship question on the 2020 Decennial Census questionnaire. On March 29, 2018, the Census Bureau delivered its planned questions for the 2020 Decennial Census to Congress, which included a citizenship question. The CEN08 PIA was updated and published in June 2018 to reflect the intent to add citizenship status to the personally identifying information (PII) data to be collected during the 2020 Decennial Census and to assess the PII confidentiality risk level of collecting the planned PII data. The confidentiality risk level was already rated as moderate based on National Institute of Standards and Technology (NIST) Special Publication 800-122 standards, prior to the addition of the citizenship question, and that rating did not change with the addition of the citizenship question.[2]

11. Census Bureau information technology systems employ a multitude of layered security controls to protect PII at rest, during processing, as well as in transit. These NIST Special Publication 800-53 controls, at a minimum, are deployed and managed at the enterprise level including, but not limited to the following:
    - Intrusion Detection | Prevention Systems (IDS | IPS)
    - Firewalls
    - Mandatory use of HTTP(S) for Census public facing websites
    - Use of trusted internet connection (TIC)
    - Anti-Virus software to protect host/end user systems
    - Encryption of databases (data at rest)
    - HSPD-12 Compliant PIV cards
    - Access controls

12. Census Bureau information technology systems also follow other NIST standards including special publications 800-63, 800-37 etc. Any system within the Census Bureau that contains, transmits, or processes PII has a current authority to operate (ATO) and goes through continuous monitoring on a yearly basis to ensure controls are implemented

[2] The potential impact is **MODERATE** if the loss of confidentiality, integrity, or availability could be expected to have a **serious adverse effect** on organizational operations, organizational assets, or individuals. A serious adverse effect means that, for example, the loss of confidentiality, integrity, or availability might (i) cause a significant degradation in mission capability to an extent and duration that the organization is able to perform its primary functions, but the effectiveness of the functions is significantly reduced; (ii) result in significant damage to organizational assets; (iii) result in significant financial loss; or (iv) result in significant harm to individuals that does not involve loss of life or serious life threatening injuries. Source: NIST Special Publication 800-122, Section 3.1. Impact Level Definitions.

and operating as intended. The Census Bureau also deploys a Data Loss Prevention (DLP) solution as well.

13. In July 2018, the CEN08 PIA was once again updated to reflect employment recruiting and hiring activities for the 2020 Decennial Census. The updated PIA was published on September 27, 2018. The collections of fingerprints and other personal information from potential Census employees were added to the PIA. The updated PIA noted in Section 4: Purpose of the System, that this information would be shared with other federal agencies for criminal background investigations. Background investigations are required for all 2020 Decennial Census hires. The collected information, including fingerprints, is shared with other federal agencies solely for conducting the background investigations of potential news hires and is not linked to 2020 Decennial Census questionnaire responses. The updated PIA also noted in Section 4 that collected information would be shared with other federal agencies for the purpose of administering human resources programs (i.e., personnel and payroll processing). These recruiting and hiring updates included additional sensitive PII which, in consideration of the overall system, changed the PII confidentiality risk level from moderate to high.[3]

14. Census Bureau IT systems share information with other Census Bureau IT systems for administrative and statistical purposes. As relevant here, CEN08 shares administrative information with other administrative IT systems such as CEN21 (Human Resources Application) for the purpose of processing employment applications, background investigations, payroll, and other personnel activities. The administrative information CEN08 shares with CEN21 and other IT systems is protected from unauthorized disclosure under the Privacy Act of 1974. The Decennial Census response data collected by CEN08 is also shared with several other internal Census IT systems, including CEN05 (Field Systems Major Application System), CEN11 (Demographic Census, Surveys, and Special Processing), and CEN13 (Center for Economic Studies) in support for producing household, demographic and economic aggregate statistical information. This information is protected from unauthorized disclosure under Title 13 United States Code. The information collected cannot be used to affect the rights, benefits, or privileges of the

[3] The potential impact is **HIGH** if the loss of confidentiality, integrity, or availability could be expected to have a **severe or catastrophic adverse effect** on organizational operations, organizational assets, or individuals. A severe or catastrophic adverse effect means that, for example, the loss of confidentiality, integrity, or availability might (i) cause a severe degradation in or loss of mission capability to an extent and duration that the organization is not able to perform one or more of its primary functions; (ii) result in major damage to organizational assets; (iii) result in major financial loss; or (iv) result in severe or catastrophic harm to individuals involving loss of life or serious life threatening injuries. Source: NIST Special Publication 800-122, Section 3.1. Impact Level Definitions.

individual respondent. CEN18 (Enterprise Applications) is a conduit for passing information from CEN08 to these other Census IT systems.

15. The CEN21 PIA was last updated on June 29, 2018, the CEN05 PIA was last updated on June 22, 2018, the CEN11 PIA was last updated on June 22, 2018, the CEN13 PIA was last updated on June 26, 2018, and the CEN18 PIA was last updated on June 26, 2018. Each of these PIAs will be reviewed within the next two months as part of our annual PIA review processes and updated, as appropriate.

16. The CEN08 PIA cites the Census Bureau's programmatic authority, which is Title 13 of the U.S. Code. Similarly, other PIAs for Census Bureau data collections and processing also cite this authority. Title 13 provides authority to conduct our work in addition to providing robust confidentiality protections. Section 9 of Title 13 not only requires that the Census Bureau maintain the confidentiality of the information it collects, but also mandates that the Census Bureau may only use the information it collects for statistical purposes, and the information cannot be used to a respondent's detriment. The Census Bureau cannot publish data that identifies a particular individual or establishment because of Title 13. Only "sworn individuals" are permitted to access confidential information, and the Census Bureau administers an "oath of nondisclosure" with all of its employees, as well as contractors. This oath of nondisclosure is a lifetime commitment to protect the confidentiality of the information collected, and an acknowledgement that violations of this law are a federal crime with serious penalties that could include a prison sentence up to five years, a fine of up to $250,000, or both.

17. The Census Bureau leverages its own Title 13 authority and obligations in coordination with other federal statutes and mandates for privacy, data security, transparency, and accountability, including the Privacy Act, the E-Government Act of 2002 (which directs PIAs), the Federal Information Security Modernization Act or FISMA, and the Paperwork Reduction Act as well as federal standards and guidance promulgated by the Office of Management and Budget (OMB) and the National Institute of Standards and Technology.

18. Although the legal requirement as found in Section 208(a) of the E-Government Act requires the PIA to be conducted for citizen-centered electronic government, the Department of Commerce has extended the privacy provisions of the E-Government Act to include personally identifiable information (PII) from Census employees, contractors, or potential employees (administrative data) that is collected, maintained, or disseminated on a IT system operated by or on behalf of the U.S. government.

19. In summary, this declaration details how the U.S. Census Bureau has complied with and intends to continue to comply with the requirements of the E-Government Act.

Jan. 30, 2019

Robin J. Bachman
Chief, Policy Coordination Office and
Chief Privacy Officer
Bureau of the Census

Attachments (2)

Version Number: 01-2015

U.S. Department of Commerce
U.S. Census Bureau

Privacy Impact Assessment
for the
CEN08 Decennial Information Technology Division (DITD)

Reviewed by: _____, Bureau Chief Privacy Officer

☑ Concurrence of Senior Agency Official for Privacy/DOC Chief Privacy Officer

☐ Non-concurrence of Senior Agency Official for Privacy/DOC Chief Privacy Officer

Dr. Catrina D. Purvis LISA MARTIN Digitally signed by LISA MARTIN
DN: c=US, o=U.S. Government, ou=Department of Commerce, ou=Office of the Secretary, cn=LISA MARTIN, 0.9.2342.19200300.100.1.1=13001000105292
Date: 2018.06.26 14:23:47 -04'00'

Signature of Senior Agency Official for Privacy/DOC Chief Privacy Officer Date

U.S. Department of Commerce Privacy Impact Assessment
U.S. Census Bureau/CEN08 DITD

Unique Project Identifier: 006-000400400

Introduction: System Description

Provide a description of the system that addresses the following elements:
The response must be written in plain language and be as comprehensive as necessary to describe the system.

(a) a general description of the information in the system

CEN08 Decennial manages the development and implementation of major decennial census applications utilized by the Decennial Census Program. These applications process response data from census tests and 2020 Census operations, and perform quality assurance mechanisms for various census operations. These applications also facilitate the acquisition of software and integration services required to support the U.S. Census Bureau during the preparation and actual Decennial Census operations.

Some examples of the information collected, maintained, and/or disseminated within CEN08 Decennial are names, addresses, gender, age, date of birth, race, email, education, telephone number and salary.

The CEN08 Decennial IT system monitors the cost, schedule, and technical performance milestones for each software system or application utilized for decennial census purposes. The CEN08 IT system manages the development and implementation of software and systems necessary to support collection, processing, and tabulation of census data.

(b) a description of a typical transaction conducted on the system

The CEN08 Decennial IT system provides updates and unit (e.g., a home, a building, or miscellaneous structure) status information to various divisions within the Census Bureau that maintain address information (e.g., street addresses, and status and control information for households and other living quarters).

In addition, CEN08 applications confirm receipt of response data. They also provide validation and acknowledgment of the data received from various IT systems.

(c) any information sharing conducted by the system

The CEN08 Decennial IT system shares information internally within the Census Bureau (to CEN07 and CEN11).

(d) a citation of the legal authority to collect PII and/or BII

Title 13, U.S.C. Section 6c
Title 13, U.S.C. Section 141
Title 13, U.S.C. Section 193
Title 26, U.S.C. 6103(j)

(e) the Federal Information Processing Standard (FIPS) 199 security impact category for the system

The Federal Information Processing Standard (FIPS) 199 security impact category for the system is Moderate.

Section 1: Status of the Information System

1.1 Indicate whether the information system is a new or existing system.

____ This is a new information system.
____ This is an existing information system with changes that create new privacy risks. *(Check all that apply.)*

Changes That Create New Privacy Risks (CTCNPR)					
a. Conversions		d. Significant Merging		g. New Interagency Uses	
b. Anonymous to Non-Anonymous		e. New Public Access		h. Internal Flow or Collection	
c. Significant System Management Changes		f. Commercial Sources		i. Alteration in Character of Data	
j. Other changes that create new privacy risks (specify):					

__X_ This is an existing information system without changes that create new privacy risks.

Section 2: Information in the System

2.1 Indicate what personally identifiable information (PII)/business identifiable information (BII) is collected, maintained, or disseminated. *(Check all that apply.)*

Identifying Numbers (IN)					
a. Social Security*		e. File/Case ID		i. Credit Card	
b. Taxpayer ID		f. Driver's License		j. Financial Account	
c. Employer ID		g. Passport		k. Financial Transaction	
d. Employee ID	X	h. Alien Registration		l. Vehicle Identifier	
m. Other identifying numbers (specify):					
*Explanation for the need to collect, maintain, or disseminate the Social Security number, including truncated form:					

General Personal Data (GPD)					
a. Name	X	g. Date of Birth	X	m. Religion	

b. Maiden Name		h. Place of Birth		n. Financial Information	
c. Alias		i. Home Address	X	o. Medical Information	
d. Gender	X	j. Telephone Number	X	p. Military Service	
e. Age	X	k. Email Address	X	q. Physical Characteristics	
f. Race/Ethnicity	X	l. Education		r. Mother's Maiden Name	
s. Other general personal data (specify): Citizenship					

Work-Related Data (WRD)					
a. Occupation		d. Telephone Number		g. Salary	
b. Job Title		e. Email Address		h. Work History	
c. Work Address		f. Business Associates			
i. Other work-related data (specify): GPS coordinates					
j.					

Distinguishing Features/Biometrics (DFB)					
a. Fingerprints		d. Photographs		g. DNA Profiles	
b. Palm Prints		e. Scars, Marks, Tattoos		h. Retina/Iris Scans	
c. Voice Recording/Signatures		f. Vascular Scan		i. Dental Profile	
j. Other distinguishing features/biometrics (specify):					

System Administration/Audit Data (SAAD)					
a. User ID	X	c. Date/Time of Access	X	e. ID Files Accessed	
b. IP Address	X	d. Queries Run	X	f. Contents of Files	
g. Other system administration/audit data (specify): Audit Logs					

Other Information (specify):

2.2 Indicate sources of the PII/BII in the system. *(Check all that apply.)*

Directly from Individual about Whom the Information Pertains					
In Person	X	Hard Copy: Mail/Fax	X	Online	X
Telephone	X	Email			
Other (specify):					

Government Sources					
Within the Bureau	X	Other DOC Bureaus		Other Federal Agencies	
State, Local, Tribal		Foreign			
Other (specify):					

Non-government Sources					
Public Organizations		Private Sector	X	Commercial Data Brokers	X
Third Party Website or Application					
Other (specify):					

2.3 Indicate the technologies used that contain PII/BII in ways that have not been previously deployed. *(Check all that apply.)*

Technologies Used Containing PII/BII Not Previously Deployed (TUCPBNPD)			
Smart Cards		Biometrics	
Caller-ID		Personal Identity Verification (PIV) Cards	
Other (specify):			

X	There are not any technologies used that contain PII/BII in ways that have not been previously deployed.

Section 3: System Supported Activities

3.1 Indicate IT system supported activities which raise privacy risks/concerns. *(Check all that apply.)*

Activities			
Audio recordings		Building entry readers	
Video surveillance		Electronic purchase transactions	
Other (specify):			

X	There are not any IT system supported activities, which raise privacy risks/concerns.

Section 4: Purpose of the System

4.1 Indicate why the PII/BII in the IT system is being collected, maintained, or disseminated. *(Check all that apply.)*

Purpose			
To determine eligibility		For administering human resources programs	
For administrative matters		To promote information sharing initiatives	
For litigation		For criminal law enforcement activities	
For civil enforcement activities		For intelligence activities	
To improve Federal services online		For employee or customer satisfaction	
For web measurement and customization technologies (single-session)		For web measurement and customization technologies (multi-session)	
Other (specify): For Statistical Purposes (i.e. Censuses/Surveys)			

Section 5: Use of the Information

5.1 In the context of functional areas (business processes, missions, operations, etc.) supported by the IT system, describe how the PII/BII that is collected, maintained, or disseminated

will be used. Indicate if the PII/BII identified in Section 2.1 of this document is in reference to a federal employee/contractor, member of the public, foreign national, visitor or other (specify).

The PII that is collected, maintained, or disseminated will be used for:

Statistical Purposes (i.e. Censuses/Surveys):

The U.S. Census Bureau collects, maintains, and disseminates PII data to produce national statistical information from the U.S. population. This information determines the distribution of Congressional seats to states, aids in the decisions about what community services to provide, determines the amount distributed of more than $400 billion in federal funds to local, state and tribal governments each year.

The PII collected, maintained, and/or disseminated by the CEN08 IT system is in reference to members of the public.

Section 6: Information Sharing and Access

6.1 Indicate with whom the bureau intends to share the PII/BII in the IT system and how the PII/BII will be shared. *(Check all that apply.)*

Recipient	How Information will be Shared		
	Case-by-Case	Bulk Transfer	Direct Access
Within the bureau	X		
DOC bureaus			
Federal agencies			
State, local, tribal gov't agencies			
Public			
Private sector			
Foreign governments			
Foreign entities			
Other (specify):			

	The PII/BII in the system will not be shared.

6.2 Indicate whether the IT system connects with or receives information from any other IT systems authorized to process PII and/or BII.

X	Yes, this IT system connects with or receives information from another IT system(s) authorized to process PII and/or BII. Provide the name of the IT system and describe the technical controls which prevent PII/BII leakage:

	CEN08 shares information with IT systems within the Census Bureau (i.e. CEN07 and CEN11). CEN08 receives information from a number of other IT systems within the Census Bureau (i.e. CEN05 and CEN07). CEN08 uses a multitude of security controls mandated by the Federal Information Security Management Act of 2002 (FISMA) and various other regulatory control frameworks including the National Institute of Standards and Technology (NIST) special publication 800 series. These security controls include, but are not limited to the use of mandatory HTTPS for public facing websites, access controls, anti-virus solutions, enterprise auditing/monitoring, encryption of data at rest, and various physical controls at Census Bureau facilities that house Information Technology systems. The Census Bureau also deploys an enterprise Data Loss Protection (DLP) solution as well.
	No, this IT system does not connect with or receive information from another IT system(s) authorized to process PII and/or BII.

6.3 Identify the class of users who will have access to the IT system and the PII/BII. *(Check all that apply.)*

Class of Users			
General Public		Government Employees	X
Contractors	X		
Other (specify):			

Section 7: Notice and Consent

7.1 Indicate whether individuals will be notified if their PII/BII is collected, maintained, or disseminated by the system. *(Check all that apply.)*

X	Yes, notice is provided pursuant to a system of records notice published in the Federal Register and discussed in Section 9.	
X	Yes, notice is provided by a Privacy Act statement and/or privacy policy. The Privacy Act statement and/or privacy policy can be found at: http://www.census.gov/about/policies/privacy/privacy-policy.html	
	Yes, notice is provided by other means.	Specify how:
	No, notice is not provided.	Specify why not:

7.2 Indicate whether and how individuals have an opportunity to decline to provide PII/BII.

	Yes, individuals have an opportunity to decline to provide PII/BII.	Specify how:
X	No, individuals do not have an opportunity to decline to provide PII/BII.	Specify why not: According to Title 13, Section 221 (Census, Refusal or neglect to answer questions; false answers) of the United States Code, persons who fail or refuse to respond to the mail-back census form, or refuse to respond to a follow-up Census Bureau taker can be fined up to $100. Persons who knowingly provide false information to the Census Bureau can be fined up to $500.

		For Census Bureau employees who access systems in CEN08 providing PII is a requirement for employment.

7.3 Indicate whether and how individuals have an opportunity to consent to particular uses of their PII/BII.

	Yes, individuals have an opportunity to consent to particular uses of their PII/BII.	Specify how:
X	No, individuals do not have an opportunity to consent to particular uses of their PII/BII.	Specify why not: For records covered by SORN Census-5, Decennial Census Programs, there are no access and consent requirements since the data is collected for statistical purposes only. For Census Bureau employees consent to particular uses of PII is a requirement for employment.

7.4 Indicate whether and how individuals have an opportunity to review/update PII/BII pertaining to them.

	Yes, individuals have an opportunity to review/update PII/BII pertaining to them.	Specify how:
X	No, individuals do not have an opportunity to review/update PII/BII pertaining to them.	Specify why not: For records covered by SORN Census-5, Decennial Census Programs, there is no opportunity to review/update data unless the Census Bureau contacts the respondent for an update on their information. For Census Employees, employees have access to PII via the appropriate Human Resources applications that reside outside of the CEN08 IT system.

Section 8: Administrative and Technological Controls

8.1 Indicate the administrative and technological controls for the system. *(Check all that apply.)*

X	All users signed a confidentiality agreement or non-disclosure agreement.
X	All users are subject to a Code of Conduct that includes the requirement for confidentiality.
X	Staff (employees and contractors) received training on privacy and confidentiality policies and practices.
X	Access to the PII/BII is restricted to authorized personnel only.
X	Access to the PII/BII is being monitored, tracked, or recorded. Explanation: Only authorized government/contractor personnel are allowed to access PII within a system. Authorizations for users occur yearly, at a minimum in accordance with applicable Bureau, Agency, and Federal policies/guidelines. In additional to IT system processes that handle PII, all manual extractions for PII are logged and recorded per Department of Commerce Policy, the NIST 800-53 Appendix J Privacy Control Catalog, and specifically NIST control AU-03, Content of Audit records.
X	The information is secured in accordance with FISMA requirements. Provide date of most recent Assessment and Authorization (A&A): July 25, 2017 ☐ This is a new system. The A&A date will be provided when the A&A package is approved.
X	The Federal Information Processing Standard (FIPS) 199 security impact category for this system is a moderate or higher.
X	NIST Special Publication (SP) 800-122 and NIST SP 800-53 Revision 4 Appendix J recommended security controls for protecting PII/BII are in place and functioning as intended; or have an approved Plan of Action and Milestones (POAM).
X	Contractors that have access to the system are subject to information security provisions in their contracts required by DOC policy.
	Contracts with customers establish ownership rights over data including PII/BII.
	Acceptance of liability for exposure of PII/BII is clearly defined in agreements with customers.
	Other (specify):

8.2 Provide a general description of the technologies used to protect PII/BII on the IT system.

Census Bureau information technology systems employ a multitude of layered security controls to protect PII at rest, during processing, as well as in transit. These NIST 800-53 controls, at a minimum, are deployed and managed at the enterprise level including, but not limited to the following:

- Intrusion Detection | Prevention Systems (IDS | IPS)
- Firewalls
- Mandatory use of HTTP(S) for Census Public facing websites
- Use of trusted internet connection (TIC)
- Anti-Virus software to protect host/end user systems
- Encryption of databases (Data at rest)
- HSPD-12 Compliant PIV cards
- Access Controls

Census Bureau information technology systems also follow the National Institute of Standards and Technology (NIST) standards including special publications 800-53, 800-63, 800-37 etc. Any system within the Census Bureau that contains, transmits, or processes PII has a current authority to operate (ATO) and goes through continuous monitoring on a yearly basis to ensure controls are implemented and operating as intended. The Census Bureau also deploys a DLP solution as well.

Section 9: Privacy Act

9.1 Indicate whether a system of records is being created under the Privacy Act, 5 U.S.C. § 552a. *(A new system of records notice (SORN) is required if the system is not covered by an existing SORN).*

As per the Privacy Act of 1974, "the term 'system of records' means a group of any records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual."

X	Yes, this system is covered by an existing system of records notice (SORN). Provide the SORN name and number (list all that apply): Census-5 Decennial Census Program
	Yes, a SORN has been submitted to the Department for approval on (date).
	No, a SORN is not being created.

Section 10: Retention of Information

10.1 Indicate whether these records are covered by an approved records control schedule and monitored for compliance. *(Check all that apply.)*

X	There is an approved record control schedule. Provide the name of the record control schedule: Nl-29-05-01, N1-29-10-5, GRS 3.1, GRS 4.3
	No, there is not an approved record control schedule. Provide the stage in which the project is in developing and submitting a records control schedule:
X	Yes, retention is monitored for compliance to the schedule.
	No, retention is not monitored for compliance to the schedule. Provide explanation:

10.2 Indicate the disposal method of the PII/BII. *(Check all that apply.)*

Disposal			
Shredding	X	Overwriting	X
Degaussing	X	Deleting	X
Other (specify):			

Section 11: NIST Special Publication 800-122 PII Confidentiality Impact Levels

11.1 Indicate the potential impact that could result to the subject individuals and/or the organization if PII were inappropriately accessed, used, or disclosed.

	Low – the loss of confidentiality, integrity, or availability could be expected to have a limited adverse effect on organizational operations, organizational assets, or individuals.
X	Moderate – the loss of confidentiality, integrity, or availability could be expected to have a serious adverse effect on organizational operations, organizational assets, or individuals.
	High – the loss of confidentiality, integrity, or availability could be expected to have a severe or catastrophic adverse effect on organizational operations, organizational assets, or individuals.

11.2 Indicate which factors were used to determine the above PII confidentiality impact levels. *(Check all that apply.)*

X	Identifiability	Provide explanation: PII collected can be indirectly used to identify individuals or if combined with other data elements may uniquely identify an individual.
X	Quantity of PII	Provide explanation: The collection is for the decennial census, therefore, a serious or substantial number of individuals would be affected if there was loss, theft or compromise of the data. Serious collective harm to the organization's reputation, or cost to the organization in addressing a breach.
X	Data Field Sensitivity	Provide explanation: The PII, alone or in combination, may be relevant in some other contexts and may, in those contexts, make the individuals or the Census Bureau vulnerable to harm.
X	Context of Use	Provide explanation: Disclosure of the act of collecting and using the PII in this IT system or the PII itself may result in serious harm to the individual or organization.
X	Obligation to Protect Confidentiality	Provide explanation: PII collected is required to be protected in accordance with 13 U.S.C .6c,141,193 and 26 U.S.C 6103 (j).
X	Access to and Location of PII	The PII is located on computers (including laptops) and on a network, and IT systems controlled by the Census Bureau. Access is limited to those with a need-to-know including the Census Bureau regional offices and survey program offices, etc. Access is allowed by Census Bureau-owned equipment outside of the physical locations owned by the Census Bureau only with a secure connection. Backups are stored at Census Bureau-owned facilities.
	Other:	Provide explanation:

Section 12: Analysis

12.1 Indicate whether the conduct of this PIA results in any required business process changes.

	Yes, the conduct of this PIA results in required business process changes. Explanation:
X	No, the conduct of this PIA does not result in any required business process changes.

12.2 Indicate whether the conduct of this PIA results in any required technology changes.

	Yes, the conduct of this PIA results in required technology changes. Explanation:
X	No, the conduct of this PIA does not result in any required technology changes.