

TABLE OF CONTENTS

INTRODUCTION 1

ARGUMENT..... 3

I. Fair Lines has narrowed its challenge in this litigation to Defendants’ withholding of GQCI totals by state, thereby waiving its right to challenge any other aspect of Defendants’ alleged conduct..... 3

II. FOIA Exemption 3 requires the Court to determine whether the requested information falls within the statutory scheme, and Defendants’ burden is merely to show that it logically does. 4

III. Defendants’ withholding of unobscured GQCI state totals is consistent with the Census Act’s statutory scheme, because the forced disclosure of such data is likely to undermine the DAS, which puts at risk the confidentiality of all census data. 6

A. Title 13’s confidentiality provisions encompass GQCI data. 6

1. Imputed data is derived from data furnished by census respondents..... 9

2. Even if imputed data were not derived from direct responses, Title 13’s confidentiality provisions would still apply to GQCI state totals for the 2020 Census because the alternative would directly contravene Congress’s intent. 11

3. The law supports Defendants’ common sense understanding of Title 13’s confidentiality provisions. 13

B. Defendants have shown that the forced disclosure of the requested data unobscured by disclosure avoidance techniques is likely to undermine the DAS and Fair Lines fails to rebut Defendants’ predictive judgement, which is entitled to deference..... 14

1. Fair Lines has provided no evidence to show that a destabilized DAS poses no risk to the confidentiality of census data..... 15

2. Fair Lines’ expert’s critique of the Census Bureau’s simulated reconstruction attack is irrelevant in this FOIA action. 17

3. In this highly technical area that calls for an agency’s predictive judgment, deference is owed to the agency. 20

CONCLUSION..... 23

TABLE OF AUTHORITIES

Cases

Alabama v. U.S. Dep’t of Commerce,
 No. 3:21-cv-211-RAH-ECM-KCN, 2021 WL 2668810 (M.D. Ala. June 29, 2021) 18

Am. Civil Liberties Union v. U.S. Dep’t of Defense,
 628 F.3d 612 (D.C. Cir. 2011) 20

Am. Wildlands v. Kempthorne,
 530 F.3d 991 (D.C. Cir. 2008) 19-20

**Baldrige v. Shapiro*,
 455 U.S. 345 (1982) *passim*

CIA v. Sims,
 471 U.S. 159 (1985) 5

CREW v. U.S. Dep’t of Justice,
 No. Civ. 05-2078, 2006 WL 1518964 (D.D.C. June 1, 2006) 22

Defs. of Wildlife & Ctr. for Biological Diversity v. Jewell,
 815 F.3d 1 (D.C. Cir. 2016) 19

Dep’t of Homeland Sec. v. Regents of the Univ. of California,
 140 S. Ct. 1891 (2020) 19

Dillon v. U.S. Dep’t of Justice,
 444 F. Supp. 3d 67 (D.D.C. 2020) 3

Ethyl Corp. v. EPA,
 541 F.2d 1 (D.C. Cir. 1976) 20

Flora v. United States,
 357 U.S. 63 (1958) 1

Food Marketing Institute v. Argus Leader Media,
139 S. Ct. 2356 (2019) 20

Int’l Fabricare Inst. v. EPA,
972 F.2d 384 (D.C. Cir. 1992) 20

Juluke v. Hodel,
811 F.2d 1553 (D.C. Cir. 1987) 7

**Larson v. Dep’t of State*,
565 F.3d 857 (D.C. Cir. 2009) 5, 14

Local 3, Int’l Bhd. Of Elec. Workers v. NLRB,
845 F.2d 1177 (2d Cir. 1988)..... 22

Marsh v. Or. Nat. Res. Council,
490 U.S. 360 (1989) 20

Mississippi v. EPA,
744 F.3d 1334 (D.C. Cir. 2013) 20

Ocasio v. Merit Systems Protection Board,
No. 17-5085, 2018 WL 1391868 (D.C. Cir. Mar. 14, 2018) 18

**Seymour v. Barabba*,
559 F.2d 806 (D.C. Cir. 1977) 13, 14

Troy Corp. v. Browner,
120 F.3d 277 (D.C. Cir. 1997) 20

United Techs. Corp. v. U.S. Dep’t of Defense,
601 F.3d 557 (D.C. Cir. 2010) 21

**Utah v. Evans*,
536 U.S. 452 (2002) 11, 12

<i>Voinche v. FBI</i> , 412 F. Supp. 2d 60 (D.D.C. 2006)	23
<i>Wheeler v. CIA</i> , 271 F.Supp.2d 132 (D.D.C.2003)	23
<i>Whitaker v. U.S. Dep’t of Justice</i> , No. 18-CV-01434 (APM), 2020 WL 6075681 (D.D.C. Oct. 15, 2020)	17
<i>Wolf v. CIA</i> , 569 F. Supp. 2d 1 (D.D.C. 2008)	19
Statutes	
5 U.S.C. § 552.....	4, 19, 21
5 U.S.C. § 706.....	19
13 U.S.C. § 1.....	2, 7
13 U.S.C. § 8.....	1, 7
13 U.S.C. § 9.....	1, 7
13 U.S.C. § 214.....	12

INTRODUCTION

When construing a statute, the duty of a court “is to give effect to the intent of Congress.” *Flora v. United States*, 357 U.S. 63, 65 (1958). There is no dispute that “[t]he . . . history of the Census Act reveals a congressional intent to protect the confidentiality of census information.” *Baldrige v. Shapiro*, 455 U.S. 345, 358 (1982). Congress’s intent is unmistakably reflected in the Census Act’s confidentiality provisions, 13 U.S.C. §§ 8(b) and 9(a). The Census Bureau is charged with safeguarding the confidentiality of census information. Pursuant to that responsibility, the Bureau has, for decades, employed disclosure avoidance techniques to defend against the release of any “publication whereby the data furnished by any particular establishment or individual . . . can be identified.” 13 U.S.C. § 9(a)(2). And it has assiduously avoided disclosing information “reported by, or on behalf of, any particular respondent.” *Id.* § 8(b). But computer technology and processing power have grown exponentially in recent years, substantially increasing the risk of reconstruction and re-identification attacks on census data. The Census Bureau has kept pace (and tried to stay a step ahead) by developing ever-more sophisticated disclosure avoidance systems to protect the confidentiality of census information. Still, as with any information security system, if the system’s foundation is undermined, the information being secured is rendered vulnerable. In this litigation, Defendants are withholding aggregated group quarter count imputation (GQCI) totals by state because the release of that data “unobscured”¹ would undermine the 2020 Census’s disclosure avoidance system (the “DAS”), which in turn would put in jeopardy the Census Bureau’s ability to protect the confidentiality of all census data disclosed by the public.

¹ The Census Bureau used the phrase “unobscured” in its opening brief, *see, e.g.*, Defs.’ Mem. in Support of Summary J. at 3, ECF No. 13-1, as shorthand for, as a technical matter, “released as enumerated without additional disclosure avoidance procedures beyond aggregation,” or, more informally, “released without going through appropriate disclosure avoidance review.” This same meaning applies for this reply/opposition brief.

That is consistent with Title 13's confidentiality provisions, and therefore Defendants' invocation of Exemption 3 in this case is valid.

Fair Lines America Foundation, Inc., argues in favor of a limited interpretation of Sections 8(b) and 9(a), whereby those provisions protect only the confidentiality of information obtained directly from a census respondent. Fair Lines argues that the state GQCI totals it seeks through its narrowed FOIA request were not obtained directly from a census respondent, but instead are aggregate data the Census Bureau developed in the absence of direct responses from individuals or establishments to census inquiries. Mem. in Support of Pl.'s Opp'n and Cross-Mot. for Summary J. at 3, ECF No. 14-1 ("Opp'n"). Fair Lines admits that "the disclosure of imputed statewide GQCI data may contravene 'the Census Bureau's established disclosure avoidance rules,'" *id.* at 29 n.21, but argues that this is irrelevant because Title 13's confidentiality provisions, in combination with FOIA, do not allow Defendants to withhold the requested information.

Fair Lines is wrong on both the facts and the law. As a factual matter, imputed data, whether group quarters count data or housing count data, derives, at least in part, from census respondents, who are defined as "individual[s], or other organization[s] or entit[ies] which reported information, or on behalf of which information was reported, in response to a questionnaire, inquiry or other request of the Bureau." 13 U.S.C. § 1 (defining respondent). And even if imputed data did not derive from census respondents, it would not matter because the publication of such data unobscured by the DAS would put in jeopardy the confidentiality of all 2020 Census data, which is exactly what Title 13's confidentiality provisions prohibit. The Census Act's language and legislative history, elucidated by Supreme Court precedent, support Defendants' common sense understanding that Title 13's confidentiality provisions forbid the disclosure of any data—

whether directly or indirectly derived from a census respondent—that could lead to the identification of any census respondent. And through agency declarations attested to by the Census Bureau’s Chief Scientist and Associate Director for Research and Methodology, Dr. John M. Abowd, Defendants have set forth in substantial detail the Census Bureau’s judgment that the forced disclosure of GQCI totals by state unobscured by the DAS would put at risk the confidentiality of all 2020 Census data. The predictive judgment of the Census Bureau is owed deference, particularly in the absence of any credible rebuttal from Fair Lines, who admits that disclosure might contravene the DAS.

Accordingly, for the reasons set forth in Defendants’ opening brief and below, Defendants’ Motion for Summary Judgment should be granted and Fair Lines’ Cross-Motion for Summary Judgment should be denied.

ARGUMENT

I. Fair Lines has narrowed its challenge in this litigation to Defendants’ withholding of GQCI totals by state, thereby waiving its right to challenge any other aspect of Defendants’ alleged conduct.

Fair Lines has narrowed its FOIA request at issue in this litigation only to “documents identifying the total population (number of individuals) imputed statewide by the Census Bureau for group quarters for each U.S. state.” Opp’n at 2 (citations omitted). Fair Lines also states in a footnote that its “agreement to narrow the scope of its request should in no way be construed as waiving its right to pursue the other information sought in its original request should it deem it necessary to do so later on.” *Id.* at 17 n.15. To the extent that Fair Lines is claiming the right to seek additional information at a later point in this litigation, it may not do so. *See Dillon v. U.S. Dep’t of Justice*, 444 F. Supp. 3d 67, 86 (D.D.C. 2020) (“Plaintiff’s contention that a FOIA

requester can explicitly narrow, test, and then broaden his request in this way after filing litigation suits without any firm basis in the statutory text or associated case law.”). Fair Lines did not file a partial motion for summary judgment – its cross-motion seeks a final determination of its claims. *See* Opp’n at 37. Nor does the summary judgment briefing schedule to which the parties agreed and which the Court ordered on August 6, 2021, contemplate piece-meal summary judgment briefing.

There is no dispute among the parties that Defendants are in possession of the narrowed information Fair Lines seeks and have the capability of producing it. Thus, Fair Lines has raised no challenge to, and therefore waived its right to challenge, Defendants’ search for responsive records. Accordingly, Defendants are entitled to summary judgment on the adequacy of its search.

All that remains then, is whether FOIA Exemption 3, in combination with Title 13’s confidentiality provisions, permit Defendants to withhold the GQCI state totals.

II. FOIA Exemption 3 requires the Court to determine whether the requested information falls within the statutory scheme, and Defendants’ burden is merely to show that it logically does.

The Supreme Court held in *Baldrige* that Sections 8(b) and 9(a) are non-disclosure statutes under FOIA Exemption 3. Exemption 3 permits the government to withhold information if that information is covered by a statute that either “requires that the matters be withheld from the public in such a manner as to leave no discretion on the issue,” 5 U.S.C. § 552(b)(3)(A)(i) (“Exemption 3(A)(i)”), or “establishes particular criteria for withholding or refers to particular types of matters to be withheld.” *Id.* § 552(b)(3)(A)(ii) (“Exemption 3(A)(i)”). In *Baldrige*, the Court did not

explicitly say which subpart of Exemption 3 applies, but its language suggests that both apply.² The Court not only stated that Title 13 leaves “[n]o discretion to the Census Bureau on whether or not to disclose the information referred to in Sections 8(b) and 9(a),” 455 U.S. at 355, but also made clear that Sections 8(b) and 9(a) refer to “particular types of matters to be withheld”—i.e., they protect information “only to the extent that the data is within the confidentiality provisions of [the Census] Act.” *Id.* at 353, 355. *Cf. CIA v. Sims*, 471 U.S. 159, 168 (1985) (determining that the National Security Act of 1947 satisfied Exemption 3(A)(ii) because it was limited to particular records, “only to the extent they contain ‘intelligence sources and methods’ or if disclosure would reveal otherwise protected information”).

Regardless of whether Exemption 3(A)(i) or (ii) applies, once a court deems Exemption 3 to apply, it must decide whether the information requested “falls within the [statutory] scheme.” *Larson v. Dep’t of State*, 565 F.3d 857, 865 (D.C. Cir. 2009). “If an agency’s statements supporting exemption contain reasonable specificity of detail as to demonstrate that the withheld information logically falls within the claimed exemption and evidence in the record does not suggest otherwise, . . . the court should not conduct a more detailed inquiry to test the agency’s judgment and expertise or to evaluate whether the court agrees with the agency’s opinions.” *Id.* This is particularly true where the question of whether the information logically falls within the claimed exemption depends on the predictive judgment of the agency. In that context, courts routinely defer to executive affidavits regarding the risk of harm, particularly where, as here, the

² Contrary to Fair Lines’ contention, *see* Opp’n at 23 n.17, Defendants did not argue in their opening brief that only one of the two Exemption 3 provisions applied, as is evident from the citations Fair Lines selected from Defendants’ opening brief.

Court is in a “poor position to second-guess’ the predictive judgments” of an agency tasked with the complex task of safeguarding the confidentiality of census data. *Id.* (citation omitted).

III. Defendants’ withholding of unobscured GQCI state totals is consistent with the Census Act’s statutory scheme, because the forced disclosure of such data is likely to undermine the DAS, which puts at risk the confidentiality of all census data.

Title 13’s confidentiality provisions broadly protect the confidentiality of data supplied directly or indirectly by census respondents. As explained below, such data includes imputed count information, which derives, at least in part, from information reported by census respondents. The Census Bureau has developed the DAS to protect the confidentiality of such data. The Census Bureau’s declarant, Associate Director for Research and Methodology and Chief Scientist Dr. John M. Abowd, has explained that the 2020 Census DAS depends on limiting the number of invariants in the system. *See* Second Abowd Decl. ¶¶ 54-61, ECF No. 13-3. Dr. Abowd has further detailed how the forced disclosure of the information requested unobscured by disclosure avoidance techniques would expose a chink in the 2020 Census DAS’s confidentiality armor, which in turn would logically leave census data vulnerable to re-identification and reconstruction attacks in violation of Title 13’s confidentiality provisions. *Id.*; *see also id.* ¶¶ 62-79. Fair Lines’ expert, Dr. Ruggles, disagrees, but his opinion is fatally undermined by Fair Lines’ admission that the disclosure of the requested data may undermine the DAS. Accordingly, the requested GQCI state totals fall within the statutory scheme of the Census Act’s confidentiality provisions, and Exemption 3 provides a valid reason for Defendants to withhold such data.

A. Title 13’s confidentiality provisions encompass GQCI data.

Section 9(a) prohibits the Secretary, or any other officer or employee of Defendants, from

“mak[ing] any publication whereby the data furnished by any particular establishment or individual under this title can be identified.” 13 U.S.C. § 9(a)(2). Section 8(b) permits the Secretary to “furnish copies of tabulations and other statistical material which do not disclose the information reported by, or on behalf of, any particular respondent.” *Id.* § 8(b). These provisions use broad language, which reflects Congress’s intent to strenuously protect the confidentiality of census data.

Fair Lines insists that Section 9(a) must be read in a limited way because a broad reading of Section 9(a) would “swallow up Section 8(b)’s allowance for disclosure of preliminary summary or tabulated data.” Opp’n at 21. But Fair Lines’ reading of the statute misses the mark. Section 9(a) refers to “any publication,” not only “tabulations and other statistical material,” as in Section 8(b). That Congress made Section 9(a) broader than Section 8(b) is logical because its primary intent was to maintain the confidentiality of census data; secondary to that goal was affording discretion to the Secretary to release certain data, but only if doing so would not contravene the primary goal by “disclos[ing] the information reported by, or on behalf of, any particular respondent.” 13 U.S.C. § 8(b). This language, as with Section 9(a), invokes Congress’s intention to allow the Secretary to protect the confidentiality of census data. It does not suggest, as Fair Lines would have it, that Congress intended for the publication of tables and other statistical materials to drive Defendants’ confidentiality decisions. Fair Lines’ limited interpretation would create an improbable statutory scheme whereby a “small tail”—i.e., Section 8(b)—would “wag . . . a large dog”—i.e., Section 9(a). *Juluke v. Hodel*, 811 F.2d 1553, 1561 (D.C. Cir. 1987).

Section 9(a) is best read as prohibiting the publication of any data that exposes data supplied by any census respondent to any census “questionnaire, inquiry, or other request.” 13

U.S.C. § 1(3). The statute's focus is not on the derivation of the data, but the consequences of the publication. If the publication of any data by Defendants would allow for the identification of data furnished by any census respondent, the Defendants are prohibited from releasing that data.

Fair Lines wrongly construes Section 9(a)(2) to prohibit Defendants from making a publication whereby the data furnished by any particular census respondent can be identified, but only if the published data comes directly by a particular respondent. Fair Lines argues that the data it seeks is not covered by Section 9(a) because it is "imputed" data, which, according to Fair Lines, is not derived directly from any particular respondent. Opp'n at 27 ("[B]y its very nature[, imputed] . . . data is not derived from individuals' responses to Census questions, but rather is created in the absence of such responses."). This is the lynchpin of Fair Lines' argument and the opinion of its expert, Dr. Steven Ruggles. *See* Decl. of Dr. Steven Ruggles, App'x A at 3, ECF No. 14-4 ("Most obviously, the counts of imputed group quarters case are not 'data furnished by any particular establishment or individual'" because "[t]he data were not furnished by anyone, which is the reason that the Census Bureau ha[s] to invent it by means of imputation"). For example, relying on Dr. Ruggles's declaration, Fair Lines argues that "disclosure of statewide imputed group quarters data poses no threat to confidentiality of individual responses, whether in combination with other data through the 'mosaic effect,' . . . or otherwise," because "[n]o computer is powerful enough to reverse-engineer individual data that did not exist in the first place." Opp'n at 28 (citing Ruggles Decl., App'x A at 2-6). But Fair Lines' argument is fatally flawed because imputed data is, in fact, derived, at least in part, from actual respondent data. Fair Lines' argument should also be rejected because it would require the disclosure of census data in direct contravention of Congress's intent in enacting Title 13's confidentiality provisions.

1. *Imputed data is derived from data furnished by census respondents.*

Baldrige explains why imputed data is covered by Title 13’s confidentiality provisions. In that case, the data requested via FOIA included housing “vacancy information contained in the updated master address registers maintained by the Bureau.” 455 U.S. at 351. The master address register at the time listed “addresses, householders’ names, number of housing units, type of census inquiry, and, where applicable, the vacancy status of the unit.” *Id.* at 350. It “was compiled initially from commercial mailing address lists and census postal checks, and was updated further through direct responses to census questionnaires, pre- and post-enumeration canvassing by census personnel, and in some instances by a crosscheck with the 1970 census data.” *Id.* Similar to this case, the parties requesting the data in *Baldrige* argued that the vacancy information in the master address register was not covered by Title 13’s confidentiality provisions because those provisions were “designed to prohibit disclosure of the identities of individuals who provide raw census data” and “protect[ed] raw data only if the individual respondent c[ould] be identified.” *Id.* at 356. The requesting parties argued that no individual could be identified by the release of vacancy information from the register. *Id.* at 355. The Court disagreed finding that the Census Act’s confidentiality provisions clearly protected more than the identity of individual respondents—it protected “the ‘information’ or ‘data’ compiled” during the census. *Id.* at 356. The Court concluded that “[a] list of vacant addresses” qualified as data compiled during the census (i.e., reported by, or on behalf of, any particular respondent) because it was initially “taken from prior censuses and mailing lists,” then “verified both by direct mailings and census enumerators who go to areas not responding,” and, “[a]s with all the census material, . . . updated from data obtained

from neighbors and others who spoke with the follow up census enumerators.” *Id.* at 358-59. Accordingly, the Court held that the list of vacant addresses was information that Congress intended to protect. *Id.* at 358.

The same is true here. When an individual does not respond to a census inquiry, the Census Bureau uses methods, including proxy interviewing and imputation, for obtaining information about that individual. Third Decl. of Dr. John M. Abowd ¶ 4 (attached hereto as Exhibit A). The Bureau does not create these records out of thin air; the imputed records are derived from actual census responses collected from similarly-placed individuals. *Id.* ¶ 7. For the 2020 Census, for group quarters (GQs), if the Census Bureau could not obtain a particular GQ’s count in the data collection, but had evidence the GQ was occupied, it inserted the GQ into the GQCI operation. *Id.* ¶ 6. The Census Bureau often had relevant information about the expected or maximum person count in the GQ from the GQ Advance Contact (an earlier 2020 Census operation) or prior Census Bureau surveys. *Id.* If so, the Bureau used this information—combined with other information from GQs of the same type and in the same state that responded—to impute a count for the GQ. *Id.* If not (that is, if such information on the nonresponding GQ was not available), the Bureau used other information from other federal records (for some college GQs) or from the data collected successfully from other GQs to impute a count. *Id.* Then, after imputing the count, the Bureau inserted characteristics for each imputed person in the GQ, based on characteristics from respondents in other GQs of the same type, often located nearby (“donor records”). *Id.* Thus, GQCI state totals are directly analogous to the information on vacancies the Supreme Court deemed confidential in *Baldrige*.

Utah v. Evans, 536 U.S. 452 (2002), the seminal decision upholding the Census Bureau’s use of imputation, provides further insight. The Court described imputation using the following example: “Imagine a librarian who wishes to determine the total number of books in a library.” *Id.* at 465. Now suppose “the librarian . . . tries to count every book one by one,” but finds “empty shelf spaces.” *Id.* The librarian may “‘impute[]’ to that empty shelf space the number of books (currently in use) that likely filled them . . . say, by measuring the size of nearby books and dividing the length of each empty shelf space by a number representing the average size of nearby books on the same shelf.” *Id.* In this example, the nearby book, whose size and location are directly observed, is akin to donor records used in imputation, whose information must be kept confidential. Because, as *Evans* shows, imputation in the census context works by taking information on behalf of missing respondents from direct responses, it is analogous to how the vacancy information was derived in *Baldrige*. Therefore, imputed information, like the information in *Baldrige*, is information Congress intended to protect when it enacted Title 13’s confidentiality provisions.

2. *Even if imputed data were not derived from direct responses, Title 13’s confidentiality provisions would still apply to GQCI state totals for the 2020 Census because the alternative would directly contravene Congress’s intent.*

Fair Lines’ construction of Title 13’s confidentiality provisions cannot be squared with Congress’ intent. According to Fair Lines, “because the plain language of Title 13’s protections do not extend to statewide tabulations of imputed data, that data is subject to mandatory disclosure under FOIA.” Opp’n at 19. But that cannot be the correct reading of Title 13’s confidentiality provisions. If taken to its logical extreme, Fair Lines’ argument would require the Court to order

Defendants to release the requested data even if the parties and the Court knew for an absolute certainty that the publication of such data would be guaranteed to lead to the re-identification of data furnished by every census respondent. Not only would that contravene Congress's intentions, but also it would force Defendants into the untenable position of facing a Court order that would expose them to criminal liability under 13 U.S.C. § 214.

Thus, while Fair Lines attempts to portray its request for GQCI state totals as a limited one, if the Court were to adopt its interpretation of Title 13's confidentiality provisions, all imputed data created by the Census Bureau would be left exposed to unobscured disclosure. Bear in mind that even Dr. Ruggles admits that "[i]t is universally accepted in the demographic research community that imputation improves the accuracy of the population count." Ruggles Decl., App'x A at 3. Thus, under Fair Lines' interpretation, a tool the Census Bureau has been using since the 1960s, *id.*, whose constitutionality was upheld in *Utah v. Evans* almost twenty years ago, *id.*, and which improves the accuracy of the census, *id.*, would suddenly become an unanticipated obstacle to the Census Bureau's ability to protect the confidentiality of census data. Dr. Ruggles downplays this concern by suggesting that in the past, "the Census Bureau has released detailed information—down to the block level—on the number of imputed persons in each locality." *Id.* This is not factually accurate. *See* Third Abowd Decl. ¶ 14. While a few tables of household count imputation data were released at the state level following the 2010 Census, Disclosure Review Board-approved disclosure avoidance methods in place in 2011 were applied to these tables. *Id.* ¶ 16. And, in any event, those methods were strengthened for the 2020 Census precisely because of the increased threat from reconstruction and re-identification attacks. *Id.* Fair Lines suggests that the Court ignore these broader concerns, *see* Opp'n at 28 n.19 & 29 n.21, but the Court cannot be blind

to the consequences of any ruling that would serve as a basis to hamstring Defendants' efforts to abide by Title 13's confidentiality provisions.

3. *The law supports Defendants' common sense understanding of Title 13's confidentiality provisions.*

Fair Lines argues that the case law supports its interpretation of Title 13's confidentiality provisions, but it is again mistaken. As Defendants argued in their opening brief, *see* Defs.' Br. in Support of Mot. for Summary J. at 18, ECF No. 13-1, the primary cases Fair Lines relies upon—*Baldrige* and *Seymour v. Barabba*, 559 F.2d 806 (D.C. Cir. 1977)—both interpret Title 13's confidentiality provisions within the context of Congress's "thrice emphatically expressed intent . . . to protect census information." *Barabba*, 559 F.2d at 809. Fair Lines cannot argue otherwise and does not attempt to do so. Rather, it takes selective quotes from these cases out of context to make it appear as though these cases adopted a limited understanding of Sections 8(b) and 9(a) when the opposite is true. For example, Fair Lines suggests that the D.C. Circuit in *Barabba* held that the Secretary could disclose "'tabulations and statistical materials of a numerical nature' in response to FOIA requests," as long they "exclud[ed] 'names and addresses of specific individuals or firms reporting data to the Census Bureau' for purposes of protecting privacy of individual respondents." Opp'n at 24 (quoting *Barabba*, 559 F.2d at 809). But that is a vast oversimplification of *Barabba*. There the court actually stated that the Secretary could release statistical data as long it did "not identify any person, corporation, or entity *in any way*." *Barabba*, 559 F.2d at 809 (emphasis added). Further, in *Baldrige*, the Supreme Court made clear that in enacting Sections 8(b) and 9(a), "Congress was concerned not solely with protecting the *identity*

of individuals,” but also “expressed its concern that confidentiality of *data* reported by individuals also be preserved.” *Baldrige*, 455 U.S. at 356 (emphasis added).

Fair Lines argues that neither *Baldrige* nor *Barabba* permit Defendants to withhold the requested information because Fair Lines is not asking for “raw data, nor even summaries or tabulations of such raw data.” Opp’n at 26. But neither *Baldrige* nor *Barraba* were decided in a world in which huge volumes of data can be compiled and manipulated to reconstruct or re-identify data supplied directly or indirectly by individual respondents to the Census. If they had been, it stands to reason that the courts in those cases still would have concluded that Title 13’s confidentiality provisions protect the publication of any information—whether raw data or otherwise—that could, “in any way” permit rogue actors to reconstruct or re-identify data supplied by individual respondents to the Census. *Barabba*, 559 F.2d at 809.

B. Defendants have shown that the forced disclosure of the requested data unobscured by disclosure avoidance techniques is likely to undermine the DAS and Fair Lines fails to rebut Defendants’ predictive judgement, which is entitled to deference.

To succeed on summary judgment, Defendants have the burden of demonstrating, with “reasonable specificity of detail . . . that the withheld information logically falls within the claimed exemption and evidence in the record does not suggest otherwise.” *Larson*, 565 F.3d at 865. Defendants have carried their burden. The Census Bureau’s established disclosure avoidance system for the 2020 Census is based on differential privacy. That is an uncontroverted fact. The 2020 Census DAS currently protects the confidentiality of census data that has already been released, including the redistricting data that Defendants released on August 12, 2021. Second Abowd Decl. ¶ 74, ECF No. 13-3. The Census Bureau has explained in detail through Dr. Abowd’s

declarations that forcing Defendants to produce the requested information unobscured “would severely compromise and weaken the confidentiality protections of the DAS, which would have cascading effects on the Census Bureau’s ability to meet its confidentiality obligations under Title 13.” *Id.* ¶ 57. Dr. Abowd explained that if the requested information is released unobscured, it will create an unplanned-for invariant. *Id.* Dr. Abowd further explained that “invariants defeat privacy protections and must be limited in order to protect the integrity of the system as a whole.” *Id.* ¶ 56. In addition, Dr. Abowd explained that “if the Court were to order disclosure in this case, the Census Bureau’s disclosure avoidance system would be exposed to challenge from other FOIA requests.” *Id.* ¶ 79.³

1. *Fair Lines has provided no evidence to show that a destabilized DAS poses no risk to the confidentiality of census data.*

Fair Lines not only fails to rebut this evidence, but it concedes, crucially, that the “disclosure of imputed statewide GQCI data may contravene ‘the Census Bureau’s established disclosure avoidance rules.’” Opp’n at 29 n.21. Because Fair Lines does not dispute that the court-

³ Fair Lines repeatedly refers to Dr. Abowd’s declaration as “conclusory” and argues as a result that it cannot support summary judgment for Defendants. *See, e.g.*, Opp’n at 28 n.20. It is self-evident that Dr. Abowd’s carefully detailed 79-paragraph-long declaration is anything but conclusory. Dr. Abowd’s declaration is meticulously supported with citations to numerous resources. Moreover, unlike Dr. Ruggles, a solitary academic whose opinion reflects only his own mind, Dr. Abowd’s declaration represents the Census Bureau’s institutional concerns. Contrary to Fair Lines’ assertions, Dr. Abowd does not have “unchecked authority to determine what information is protected from disclosure under Title 13.” Opp’n at 31 (citation omitted). Dr. Abowd is just one of twelve career senior executives who comprise the Data Stewardship Executive Policy Committee (DSEP). Third Abowd Decl. ¶ 1. DSEP has five supporting committees, one of which is the Disclosure Review Board (DRB). It is the DRB’s mission to ensure that the Census Bureau protects Title 13 respondent confidentiality. *Id.* ¶ 2. And it is the DRB, not Dr. Abowd alone, who reviews and clears all Census Bureau microdata, tabulation, and research result table releases under its purview for confidentiality protection. *Id.*

ordered disclosure of GQCI state totals may contravene the 2020 Census DAS, it cannot credibly rebut Defendants’ position that disclosure will risk compromising the integrity of the disclosure avoidance system. And while Dr. Ruggles opines that “[i]nformation about imputation . . . would not add to the list of invariants,” he directly contradicts himself—in the previous sentence, no less—by conceding that “under the 2020 disclosure avoidance rules, the only ‘true’ numbers released are (1) total population at the state level; (2) total number of housing units at the block level, but not the population at the block level; and (3) total number of group quarters for seven types at the block level, but not the population in group quarters.” Ruggles Decl., App’x A at 4-5; *see also* Third Abowd Decl. ¶ 21 (“[B]ecause the GQCI statistics cannot be derived from the current approved set of invariants . . . they would constitute 52 new invariants.”) Dr. Ruggles admits that “[a]side from these invariants, all other statistics released from the 2020 Census will have deliberate error introduced to the counts.” Ruggles Decl., App’x A at 5. Fair Lines seeks imputed group quarters totals that do not have deliberate error introduced to the counts.⁴

Further, Fair Lines recognizes that “other future FOIA requests may impermissibly risk disclosure of confidential individual data.” Opp’n at 28 n.19. Although Fair Lines argues that such a future event “has no bearing on this Court’s determination of whether Plaintiff’s Request

⁴ It is true that, even with the release of the GQCI information as an invariant, differentially private noise would still remain in the total GQ population counts. Third Abowd Decl. ¶ 22. But, as Dr. Abowd explains, the kind of uncertainty that would remain in the total GQ population counts if the GQCI information were released without differentially private noise, is not comparable to differentially private uncertainty. *Id.* Some uncertainty may remain, but it would not be easily quantified, and unlike with carefully structured differentially private noise, it would not be possible to promise that the additional disclosure risk from the published statistics remained small and controlled. *Id.* Sources of uncertainty that cannot quantifiably limit disclosure risk are not suitable for use as part of a disclosure avoidance system. *Id.*

does so,” *id.*, the Court absolutely has the right to take such a risk into consideration. In fact, that is the basis of the mosaic effect theory. *See Whitaker v. U.S. Dep’t of Justice*, No. 18-CV-01434 (APM), 2020 WL 6075681, at *5 (D.D.C. Oct. 15, 2020) (“A mosaic theory posits that separate disclosures of otherwise innocuous information could be assembled by a requester or other person to reveal . . . [the larger puzzle.] Thus, the only way to prevent anyone from constructing the broader ‘mosaic’ is to shield each individual piece from disclosure.”). Dr. Ruggles opines that the mosaic effect generally does not apply in this case, but provides no explanation for why that is true other than to mischaracterize a source Defendants cited to explain the concept’s broad support. *See Ruggles Decl.*, App’x A at 5 (opining incorrectly that OMB Memorandum M-13-13 has no application). He states that the mosaic effect is not relevant in this case because, he opines, “[t]here is no possible means by which the number of imputed cases could be used in combination with other statistics to allow for identification of an individual.” *Id.* But Dr. Ruggles provides no reasoning to support this statement. In contrast, Dr. Abowd explains that unobscured information about imputed persons directly compromises data from direct respondents because the Census Bureau usually substitutes data from another member of the group quarters facility, household or a nearby household (known as “donor records”) when imputing characteristic data. Third Abowd Decl. ¶ 20. Providing unobscured imputed information derived from donor records puts those nearby individual donor’s data at risk. *Id.*

2. *Fair Lines’ expert’s critique of the Census Bureau’s simulated reconstruction attack is irrelevant in this FOIA action.*

Fair Lines’ rebuttal evidence from its expert does not go to whether the integrity of the DAS is put at risk by the forced unobscured disclosure of the requested information, but whether

the Census Bureau's decision to install a differential privacy-based disclosure avoidance system was sound.⁵ Fair Lines' counsel attempted this sort of challenge through a preliminary injunction in a prior case he helped bring against Defendants. *See Alabama v. U.S. Dep't of Commerce*, No. 3:21-cv-211-RAH-ECM-KCN, 2021 WL 2668810, *1 (M.D. Ala. June 29, 2021) ("Plaintiffs requested a preliminary injunction against the Bureau's plan to use 'differential privacy' . . . in the processing of 2020 Census data, on the grounds that it violates the Census Act, the Administrative Procedure Act ("APA"), and the Individual Plaintiffs' due process and equal protection rights under the Fifth Amendment.") (internal citations omitted). In that case, Dr. Ruggles submitted an expert report containing many of the same conclusions he has supplied in his expert declaration in this case. *Compare* Decl. of Dr. Steven Ruggles, App'x A at 6, ECF No. 94-6, *Alabama v. U.S. Dep't of Commerce*, No. 3:21-cv-211-RAHECM-KCN (M.D. Ala.), *with* Ruggles's Decl., App'x A at 9, ECF No. 14-4. The court in that case did not address the merits of the plaintiffs' position because it denied plaintiffs' preliminary injunction motion for jurisdictional and prudential reasons. *Alabama*, 2021 WL 2668810, *4. The plaintiffs in *Alabama* did not appeal that determination and eventually voluntarily dismissed the action.

Fair Lines' counsel and Dr. Ruggles, through Fair Lines' FOIA request, are seeking another bite at that apple. But FOIA is not the proper vehicle through which an agency policy decision, such as Defendants' decision to employ a differential privacy-based disclosure avoidance system, can be challenged. *Cf. Ocasio v. Merit Systems Protection Board*, No. 17-5085, 2018 WL

⁵ This is clear from the portion of Dr. Ruggles's declaration in which he spends five pages arguing that "[d]ifferential privacy is a poor fit for the protection of census data." Ruggles Decl., App'x A at 18-23.

1391868, at *1 (D.C. Cir. Mar. 14, 2018) (per curiam) (rejecting a FOIA requestor’s attempt to use FOIA litigation “to challenge the decision by an administrative law judge in a Merit Systems Protection Board proceeding,” because “a FOIA action is not the appropriate vehicle to do so”); and *Wolf v. CIA*, 569 F. Supp. 2d 1, 11 (D.D.C. 2008) (noting the distinction between a FOIA action, which “assesses the sufficiency of the agency’s search for documents,” and a “proposed APA claim” regarding the agency’s compliance with recordkeeping laws, as being radically different in scope and nature). The Court has no authority in this case to order the Census Bureau to stop using its differential privacy-based disclosure avoidance system. FOIA merely affords courts “jurisdiction to enjoin [an] agency from withholding agency records and to order the production of any agency records improperly withheld from the complainant.” 5 U.S.C. § 552(a)(4)(B).

Thus, the Court must take as a given that the Census’s 2020 DAS is based on differential privacy. It is not the Court’s role in this FOIA action to make an assessment of whether the Census Bureau’s decision to use such a system was sound. That might serve as the basis of an APA arbitrary-and-capricious challenge under 5 U.S.C. § 706(2)(A), but such a challenge would have to be reviewed under the APA’s “narrow standard of review” under which “a court is not to substitute its judgment for that of the agency.” *Dep’t of Homeland Sec. v. Regents of the Univ. of California*, 140 S. Ct. 1891, 1905 (2020) (citations omitted). Under such a challenge, the Defendants would have the benefit of a “highly deferential [standard that] . . . presumes agency action to be valid.” *Defs. of Wildlife & Ctr. for Biological Diversity v. Jewell*, 815 F.3d 1, 9 (D.C. Cir. 2016) (internal punctuation omitted). That deference would be “particularly strong” as it is whenever “the [agency] is evaluating scientific data within its technical expertise.” *Am. Wildlands*

v. Kempthorne, 530 F.3d 991, 1000 (D.C. Cir. 2008) (citing *Int'l Fabricare Inst. v. EPA*, 972 F.2d 384, 389 (D.C. Cir. 1992)); *Troy Corp. v. Browner*, 120 F.3d 277, 283 (D.C. Cir. 1997) (courts “review scientific judgments of the agency ‘not as the chemist, biologist, or statistician that [they] are qualified neither by training nor experience to be, but as a reviewing court exercising [its] narrowly defined duty of holding agencies to certain minimal standards of rationality.’”) (quoting *Ethyl Corp. v. EPA*, 541 F.2d 1, 36 (D.C. Cir. 1976)). In fact, under an arbitrary and capricious standard, when the record reflects the “conflicting views” of experts, the law affords an agency the “discretion to rely on the reasonable opinions of its own qualified experts.” *See Marsh v. Or. Nat. Res. Council*, 490 U.S. 360, 378 (1989); *Mississippi v. EPA*, 744 F.3d 1334, 1348 (D.C. Cir. 2013) (stating “it is not our job to referee battles among experts”).

3. *In this highly technical area that calls for an agency’s predictive judgment, deference is owed to the agency.*

The deference that would be afforded to Defendants in an APA challenge is the sort of deference that should be given to Defendants in this case, in the same way it is afforded to executive branch affidavits in FOIA cases in the national security context, where “courts lack the expertise necessary to second-guess . . . agency opinions [assessing the harm that could be caused by disclosure].” *Am. Civil Liberties Union v. U.S. Dep’t of Defense*, 628 F.3d 612, 619 (D.C. Cir. 2011). Fair Lines notes that this deference has not been extended “to an agency’s invocation of Title 13 (pursuant to Exemption 3) for protecting the confidentiality of data,” Opp’n at 33, but fails to otherwise explain why the two contexts are not analogous. Further, this sort of deference has been extended to agencies in other FOIA contexts. For example, prior to the Supreme Court’s decision in *Food Marketing Institute v. Argus Leader Media*, 139 S. Ct. 2356 (2019), in order to

justify withholding information pursuant to Exemption 4,⁶ the government had to show that disclosure of the information would cause a government contractor substantial competitive harm. “In reviewing an agency’s determination as to substantial competitive harm,” the courts “recognize[d] that predictive judgments are not capable of exact proof, and [they] generally defer[red] to the agency’s predictive judgments as to the repercussions of disclosure.” *United Techs. Corp. v. U.S. Dep’t of Defense*, 601 F.3d 557, 563 (D.C. Cir. 2010) (cleaned up). Thus, such deference has been extended outside of the national security context in FOIA cases and there is no reason not to extend it in this context as well.

To avoid this deferential standard, Fair Lines calls upon its own expert to explain that this is not a highly technical issue, although that same expert spends six pages of his expert report explaining in technical statistical terms the errors he believes the Census Bureau made in its simulated reconstruction experiment. *See* Ruggles Decl., App’x A at 10-16. Dr. Ruggles’s criticism of the Census Bureau’s simulated reconstruction experiment is, in any event, not credible because he admittedly lacks the full context to analyze it. As he notes, he has not had access to “a full description of their experiment,” “some details remain obscure,” and “[t]here are no peer-reviewed publications explaining their methodology.” *Id.* at 9. While he states that the Census Bureau has released “a more detailed description of the experiment,” he still does not claim to have the full details necessary to make an accurate assessment of the experiment. Rather, he simply deduces that the experiment did not demonstrate a risk of reconstruction by comparing the results to what he terms a “null model of random guessing.” *Id.* at 11. However, his null model is flawed.

⁶ Exemption 4 prohibits the disclosure of “trade secrets and commercial or financial information obtained from a person and privileged or confidential.” 5 U.S.C. § 552(b)(4).

Dr. Abowd details the methodological flaws in Dr. Ruggles's null model in his Third Declaration. *See* Third Abowd Decl. ¶¶ 23-31. Dr. Ruggles's experiment's primary flaw is that it compares apples to oranges. . In Dr. Ruggles's experiment he essentially tests whether he can randomly guess an age/sex combination in a census block. *Id.* ¶ 25. But that is not how reconstruction works. A malicious attacker is not going to randomly guess. A valid reconstruction experiment tests whether one can find each person in the reconstructed data, not whether one can find a person with random characteristics in a given census block. *Id.* ¶ 26. The meaninglessness of Dr. Ruggles's experiment is shown by replacing his random guesser with a one-note guesser who always guesses the same age/sex combination. *Id.* ¶ 27. One would expect that one-noted guesser to be a very poor guesser of population characteristics, but as Dr. Abowd shows, under Dr. Ruggles's methodology, the one-note guesser does even better than the random guesser. *Id.* It has a matching rate of 57% even though the most common age/sex combination is shared by less than 1% of the U.S. population. *Id.* That disconnect reflects the significant flaws in Dr. Ruggles's methodology, not any flaws in the Census Bureau's simulated reconstruction experiment.

But, to reiterate, the Court's job in this FOIA action is not to referee a battle between Dr. Abowd and Dr. Ruggles about the merits or flaws of the Census Bureau's simulated reconstruction experiment. Nor does Fair Lines suggest that such a deep dive into statistical theory is necessary.⁷

⁷ Although Fair Lines states that "there is no genuine issue of material fact in this case," Pl.'s Statement of Material Facts Not in Genuine Dispute at 9, ECF No. 14-2, it says in a final footnote that discovery might be merited to enable it to challenge the Census Bureau's determination of the risks disclosure poses to the DAS. *See* Opp'n at 36 n.22. The cases Fair Lines cites are inapposite. In neither *CREW v. U.S. Dep't of Justice*, No. Civ. 05-2078, 2006 WL 1518964 (D.D.C. June 1, 2006), nor *Local 3, Int'l Bhd. Of Elec. Workers v. NLRB*, 845 F.2d 1177 (2d Cir. 1988), did the

See Opp'n at 33 (characterizing the dispute in this case "as simple both as a matter of law and of logic"). At bottom, this is a legal dispute about the breadth of Title 13's confidentiality provisions. If the Court concludes, as it should, that Title 13's confidentiality provisions apply to imputed data, it should rule in Defendants' favor. There is no legitimate factual dispute about the risks concomitant with the disclosure of such data. Defendants have shown that a destabilized DAS poses a significant risk to the confidentiality of census data, and Fair Lines has failed to rebut that common sense conclusion.

CONCLUSION

For the reasons stated above, Defendant respectfully request that this Court grant judgment in Defendants' favor and deny Fair Lines' cross-motion for summary judgment.

court permit discovery into an agency's predictive judgment in the absence of bad faith (or something close to it). Indeed, FOIA actions are typically resolved without discovery. *See Wheeler v. CIA*, 271 F.Supp.2d 132, 139 (D.D.C.2003) ("Discovery is generally unavailable in FOIA actions."). "When allowed, the scope of discovery is usually limited to the adequacy of the agency's search and similar matters." *Voinche v. FBI*, 412 F. Supp. 2d 60, 71 (D.D.C. 2006). This is not such a case.

November 2, 2021

Respectfully submitted,

BRIAN M. BOYNTON
Acting Assistant Attorney General

MARCY BERMAN
Assistant Branch Director, Federal Programs

/s/ Jonathan D. Kossak

JONATHAN D. KOSSAK
Trial Attorney (DC Bar # 991478)
U.S. Dep't. of Justice, Federal Programs Branch
1100 L Street, NW
Washington, D.C. 20005
Tel. (202) 305-0612; Fax. (202) 616-8460
Email: jonathan.kossak@usdoj.gov

Counsel for Defendants

CERTIFICATE OF SERVICE

I hereby certify that I filed the foregoing Reply in Support of Defendants' Motion for Summary Judgment and in Opposition to Plaintiff's Cross-Motion for Summary Judgment with the Clerk of the Court through the ECF system on November 2, 2021. This system provided a copy to and effected service of this document on all parties.

/s/ Jonathan D. Kossak
JONATHAN D. KOSSAK
Trial Attorney (DC Bar # 991478)
United States Department of Justice
Civil Div., Federal Programs Branch

Exhibit A

I, John M. Abowd, make the following Declaration pursuant to 28 U.S.C. § 1746, and declare that under penalty of perjury the following is true and correct to the best of my knowledge:

This is my third Declaration in this lawsuit. My first Declaration supported the Defendants' opposition to Plaintiff's motion for preliminary injunction, and the second supported Defendants' motion for summary judgment. I incorporate those first two Declarations herein by reference. In this third Declaration I do not repeat the material in prior declarations, rather I respond to certain points identified in Plaintiff's cross-motion for summary judgment and to the opinion of Plaintiff's expert, Dr. Steven Ruggles. Specifically, I:

- Explain the decision-making process at the U.S. Census Bureau regarding confidentiality determinations;
- Explain why the Census Bureau protects the confidentiality not only of respondents' identities, but also their characteristics and response status;
- Describe how the Census Bureau's imputation process worked for the 2020 Census;
- Explain why imputed count data is considered information furnished by, or on behalf, of census respondents under Title 13's confidentiality provisions;
- Document how the Census Bureau has been transparent about the use of imputation in the 2020 Census;
- Clarify what data the Census Bureau has previously published about the use of imputation in past censuses;
- Explain the disclosure risk of releasing Group Quarters Count Imputation (GQCI) data without proper disclosure avoidance;
- Identify significant methodological flaws in Dr. Ruggles's expert declaration.

DECISION-MAKING ABOUT CONFIDENTIALITY AT THE CENSUS BUREAU

1. As described in my second declaration, confidentiality practice at the Census Bureau is governed by the Data Stewardship Executive Policy Committee (DSEP). DSEP serves as the Census Bureau's focal point for policy issues related to respondent privacy, security, data confidentiality, data management, record linkage, and administrative data. DSEP is chaired by the Deputy Director/Chief Operating Officer of the Census Bureau and is composed of career senior executives with expertise in confidentiality practice, the uses of Census Bureau data, and policy. Specifically, the Standing Members of DSEP other than the Chair are: the Associate Director for Decennial Census Programs; the Associate Director for Demographic Programs; the Associate Director for Economic Programs; the Associate Director for Field Operations; the Associate Director for Research and Methodology and Chief Scientist (me); the Chief Information Officer; the Chief of the Office of Program, Performance, and Stakeholder Integrations; the Chief of Staff; the Assistant Director for Communications; the Assistant Director for Research and Methodology; and the Chief of the Policy Coordination Office/Chief Privacy Officer. In addition to the information about DSEP in my second declaration (see especially footnotes 34 and 46 and paragraph 38), general information about this longstanding body can be found on the Census Bureau's website at: https://www.census.gov/about/policies/privacy/data_stewardship/dsep_committee.html.
2. DSEP has five supporting committees, one of which is the Disclosure Review Board (DRB), whose mission is to ensure that the Census Bureau protects Title 13 respondent confidentiality. It serves as the focal point for issue identification, research coordination, and policy development on issues related to disclosure review of all data products. The DRB also reviews and clears all Census Bureau microdata, tabulation, and research result table releases under its purview for confidentiality protection. For additional information, see footnotes 34 and 46 and paragraph 69 of my second declaration.

TITLE 13 CONFIDENTIALITY PROTECTS RESPONDENTS' IDENTITIES, CHARACTERISTICS, AND RESPONSE STATUS

3. In his expert report, Dr. Ruggles contends that "Census disclosure control requires the protection of identities, not concealment of characteristics." (Ruggles Expert Report, p. 23). However, as Dr. Ruggles notes, Title 13 U.S.C. § 9(a)(2) requires that the Census Bureau "shall not make any publication whereby the data furnished by any particular establishment or individual ... can be identified..." Protecting against re-identification of respondents requires applying disclosure avoidance to respondents' characteristics (e.g., location, sex, age, number of persons living in the household or group quarters, relationship among members of the same housing unit, racial identity, racial composition of the household or group quarters, ethnic composition of the household or group quarters, whether the housing unit is rental or owner occupied, imputation status) because those characteristics are what an attacker can use to link published census data to third-party data sources, which is how re-identification attacks work. In the Census Bureau's simulated reconstruction-abetted re-identification attack (see paragraphs 40-41 and Appendix B of my second declaration), for example, respondent characteristics allowed linkage of individuals' census responses to commercially available data in order to attach names and addresses to those records. Thus, insufficient protection of individuals' characteristics is what undermines the protection of their identities. The necessity of protecting individuals' characteristics has been understood since Ivan Fellegi's seminal publication in 1972 on statistical confidentiality where he noted that "... it is not enough to withhold the publication of the names or other identification of respondents; it is not enough to insure the security of confidential data; it is not enough to publish data for populations of particular interest (rather than for individuals); it is also incumbent on the statistical offices to maintain

a continuous scrutiny of their own publications to insure that from their publications no information can be deduced concerning particular respondents.”¹ Fellegi’s understanding of statistical confidentiality has been affirmed by the Federal Committee on Statistical Methodology (organized by the Chief Statistician of the United States under the auspices of the Interagency Council on Statistical Policy), which has stressed the importance of “minimizing the risk of disclosure (public identification) of the identity of individual reporting units *and information about them*” (emphasis added).²

4. Historically, Census Bureau confidentiality protections have also extended to whether and how an individual responded to the census. Protection of individuals’ response status is common practice across federal agencies, for both statistical and administrative information collections, because it is widely recognized that knowing the response status of an individual not only reveals information about that individual, but also reveals information about their neighbors (see my example in paragraph 7 below). The Internal Revenue Service, for example, recognizes that protection of confidential tax information requires protecting the “fact of filing,” that is, whether an individual or business has filed a tax return.³

¹ Fellegi, I.P. 1972. “On the Question of Statistical Confidentiality.” *Journal of the American Statistical Association*, Vol. 67, No. 337 (March), pp. 8-9.

² Federal Committee on Statistical Methodology. 2005. Report on Statistical Disclosure Limitation Methodology. Statistical Working Paper 22 (second revision), page 2, available at <https://nces.ed.gov/fcsm/pdf/spwp22.pdf> (cited October 22, 2021).

³ IRS Publication 1075 Tax Information Security Guidelines For Federal, State and Local Agencies: Safeguards for Protecting Federal Tax Returns and Return Information. “Status of whether a return was filed, under examination, or subject to other investigation or processing, including collection activities.” (page 5)

THE CENSUS BUREAU'S COUNT IMPUTATION PROCESS

5. Here is a brief overview of how count imputation worked in the 2020 Census. For a housing unit address, the Census Bureau first tried to obtain a self-response from the housing unit directly. If unsuccessful, the Bureau inserted the case into the nonresponse follow-up (NRFU) universe. That is, the Bureau tried to obtain an interview with a person in that housing unit (a householder). If (a) the Bureau was still unsuccessful after one attempt and (b) the administrative records for that housing unit were of sufficient quality, the Bureau drew the count from those records. If the administrative records were not of sufficient quality, the Bureau gave the housing unit the full NRFU treatment: up to 6 visits, sometimes even more, until the Bureau obtained an interview. If still unsuccessful, the Bureau relaxed slightly the administrative record quality criteria, to see if it could still use their administrative record-derived count. If all this failed, the Bureau inserted the case into the housing unit count imputation operation, which used information from a similar nearby housing unit to obtain a count. If the count was greater than 0, the Bureau then used information from a similar nearby housing unit (not necessarily the one that supplied the count) to insert the characteristics (race, Hispanic origin, age, sex, etc.).
6. For group quarters (GQs), if the Census Bureau couldn't obtain the count in the data collection but had sufficient evidence the GQ was occupied, it inserted the GQ into the GQCI operation. The Census Bureau often had relevant information about the expected or maximum person count in the GQ from the GQ Advance Contact (an earlier 2020 Census operation) or prior Census Bureau surveys. (This is quite analogous to the use of administrative records for housing unit count imputation.) If so, the Bureau used this information--combined with other information from GQs of the same type and in the same state that responded--to impute a count for the GQ. If not (that is, if such information on the nonresponding GQ was not available), the Bureau used other information from other federal records (for some college GQs) or from the

data collected successfully in the GQ operation from other GQs to impute a count. Then, after imputing the count, the Bureau inserted characteristics for each imputed person in the GQ based on characteristics from respondents in other GQs of the same type, often located nearby.

IMPUTED COUNT DATA IS CONSIDERED INFORMATION REPORTED BY, OR ON BEHALF OF, INDIVIDUALS RESPONDING TO THE CENSUS AND THEREFORE IS COVERED BY TITLE 13'S CONFIDENTIALITY PROVISIONS

7. Plaintiff contends that “[t]he Census Bureau imputed (i.e., made up) the requested group quarters data to fill in gaps caused by the absence of responses from individuals or establishments to the Census.” (Motion, p. 3-4). But Plaintiff fails to recognize that imputed data are considered to be reported “by, or on behalf of” census respondents, and therefore are covered by Title 13’s confidentiality provisions. See 13 U.S.C. § 8(b) (prohibiting the Census Bureau from disclosing “information reported by, or on behalf of, any particular respondent” (emphasis added)). As I described in the previous section, when an individual does not respond to a census inquiry, the Census Bureau uses methods, including proxy interviewing and imputation, for obtaining information about that individual. The Bureau does not create these records out of thin air; the imputed records are derived from actual census responses collected from similarly placed individuals. For example, if a particular area has five GQ facilities, and four of those GQs did not properly respond to the Census, then the Census Bureau’s statistical methods for imputation use the existing records from the fifth GQ to infer the individual census responses for the residents of the other four. Revealing which of those facilities housed imputed persons exposes the personal information of the residents of the GQ that contributed the donor records for the imputations.

THE CENSUS BUREAU WAS TRANSPARENT ABOUT ITS USE OF GROUP QUARTERS COUNT IMPUTATION

8. The Census Bureau has been transparent about both the data collection challenges it faced and the methods it used to mitigate any harm to data quality that could have resulted from those difficulties. As early as January 2021, while the Bureau's imputation procedures were being developed and refined, the Bureau made information available both to oversight agencies and scientific experts about the use of imputation for "group quarters that submitted a zero-population count."⁴
9. On February 2, 2021 Acting Director Jarmin wrote a Director's Blog alerting the public to the difficulties the Census Bureau faced in Group Quarters enumeration, stating, "Enumerating Group Quarters (GQs) facilities is a challenge in every decennial census, but we are seeing additional complications brought on by the COVID-19 pandemic. GQs are facilities such as college dormitories, prisons and nursing homes. We delayed this and other field operations due to the pandemic. This delay, and the fact that some facilities emptied in the spring due to the pandemic, has caused issues with our GQs enumeration."⁵
10. In March 2021, GAO released a report that included information that the Census Bureau shared with GAO regarding the Group Quarters count, explaining, "The pandemic made it difficult to count group quarters... Bureau officials told us that in

⁴ GAO, "Decennial Census: Bureau Should Address Significant Data Collection Challenges as It Undertakes Planning for 2030," March 2021. Available at: <https://www.gao.gov/assets/gao-21-365.pdf>; American Statistical Association "2020 Census State Population Totals: A Report from the American Statistical Association Task Force on 2020 Census Quality Indicators" September 2021. Available at: <https://www.amstat.org/asa/files/pdfs/POL-CQI-Task-Force-final-report.pdf>; JASON report JSR-20-2N (WS'21) "Assessment of 2020 Census Data Quality Processes" February 8, 2021. Available at: <https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/plan/planning-docs/2020-census-data-quality-processes.html>.

⁵ Dr. Ron Jarmin, "2020 Census Processing Updates," February 2, 2021. Available at: [2020 Census Processing Updates](#).

December 2020 they decided to re-contact more than 24,000 out of approximately 272,000 group quarter facilities to collect data, and that imputation would be used to count individuals at the remaining facilities still reporting a zero-population count.” That report provides additional detail about the data collection issues the Bureau faced and how the Bureau attempted to mitigate them through additional follow-up operations and, as a last resort, count imputation.⁶

11. Later in March, the Census Bureau published additional information about the 2020 Census Group Quarters enumeration, including how the COVID-19 pandemic had impacted the count, how post-data collection processing was conducted, the Census Bureau’s efforts to conduct additional outreach to Group Quarters facilities, and its use of count imputation to “fill in the gaps” where information was still missing.⁷
12. To ensure that the public better understood how imputation was performed, in April 2021 the Census Bureau published the blog referenced in Plaintiff’s brief on “How We Complete the Census When Households or Group Quarters Don’t Respond.”⁸ This blog explains how the Census Bureau conducts imputation for households, for characteristics, and for group quarters.
13. Additional information about the group quarters imputation process was provided to the American Statistical Association’s (ASA) Task Force on 2020 Census Quality Indicators during their examination of the collection and post-data collection operations

⁶ GAO Report. op. cit. front matter.

⁷ Deborah Stempowski and James Christy, “2020 Census Group Quarters,” March 16, 2021. Available at: [2020 Census Group Quarters](#).

⁸ Pat Cantwell, “How We Complete the Census When Households or Group Quarters Don’t Respond,” April 16, 2021. Available at: [How We Complete the Census When Households or Group Quarters Don’t Respond](#).

of the decennial census. In addition to the narrative discussion of group quarters imputation, the ASA included process statistics in its report on the percent of group quarters with an imputed count.⁹

PAST PUBLICATION OF COUNT IMPUTATION DATA

14. Dr. Ruggles incorrectly claims that “[i]n the past, the Census Bureau has released detailed information—down to the block level—on the number of imputed persons in each locality” (Ruggles Expert Report, p. 10). In the 2010 Census, the data in Summary File 1 (SF1) contained the block-level Table P-44, which tabulated “substituted persons” in the universe of the total population. Table P-44 was subjected to the disclosure avoidance procedure used for SF1. The complete definition of substituted persons is derived from Decennial Statistical Studies Division Memorandum #2010-G-01 in Table 1, which defines “Whole-Person Census Imputation Categories” as:

“1. Status Imputation - No information about the housing unit; housing unit imputed as occupied, vacant, or non-existent. Those imputed as non-existent were removed from the census files.”

“2. Occupancy Imputation - Existence of housing unit confirmed, but no information as to occupancy status; imputed as occupied or vacant.”

“3. Household Size Imputation - Occupied status confirmed, but no information as to household count; the household population count was imputed.”

“Population Count Already Known for the Housing Unit.”

“4. Whole Household - Population count known; all characteristics imputed for the entire household.”

⁹ American Statistical Association, “2020 Census State Population Totals,” September 2021.

“5. Partial Household - Population count known; all characteristics imputed for some, but not all, persons in the household.”

15. Substituted persons are categories 1 to 4 in this definition. Only housing units in categories 1, 2 and 3 of the whole-person census imputations created individuals who were added to the enumeration via count imputation. Therefore, the Census Bureau did not release any tabulations in the 2010 Census data publications of “imputed persons” as the Plaintiff’s expert has defined them. Even so, the households tabulated in Summary File 1 Table P-44 were still subjected to the disclosure avoidance procedures in effect for the 2010 Census, contrary to Dr. Ruggles’s assertion.
16. Following the 2010 Census a few tables of household count imputation data were released at the state level. DRB-approved disclosure avoidance methods in place in 2011 were applied to these tables. Those methods did permit state-level tabulation without rounding, suppression, or noise injection. Current DRB-approved methods for unweighted tabulations, which pre-date the adoption of differential privacy for 2020 Census publications, require rounding as described in footnote 60 of my second declaration. The rules were strengthened precisely because of the increased threat from attacks that combine information from multiple publications.

THE SPECIFIC DISCLOSURE RISK OF GROUP QUARTERS COUNTS

17. Dr. Ruggles states that “[t]here is no possible means by which the number of imputed cases could be used in combination with other statistics to allow for identification of an individual.” (Motion, p. 4-5). This statement is mathematically false, and the incorrectness of his statement has been known since 1972. In fact, whole chapters of

traditional disclosure avoidance textbooks are devoted to explaining exactly why the statement is false.¹⁰

18. The disclosure risk to individual respondents is created by a failure to protect all statistical tabulations, whether of imputed data or otherwise. The protection of all non-invariant statistical tabulations is an acknowledged component of prior disclosure limitation procedures carefully documented in scientific articles that build on the work of Fellegi. This is the principle of complementary disclosure that I discussed in my two previous declarations. For example, the equations that defeat suppression-based disclosure avoidance systems when complementary suppressions are not used are shown in Section 4.1 of a 2011 textbook on statistical confidentiality. The chapter illustrates the principle “[i]mportantly, a table satisfies a linear system of equations. All tables in practice (including multi-dimensional, linked, hierarchical, and others) fit this structure” (emphasis in original).¹¹ If you modify one cell in a table it affects many of these equations, and those connections are the reason that all data contributing to the table must be considered when designing any disclosure avoidance system, whether differentially private or otherwise. Put simply, every statistic you publish potentially interacts with every other statistic in data products derived from the same confidential source. Any disclosure risk analysis must account for this interaction. Failure to do so compromises the entire disclosure avoidance system, whether differentially private or otherwise. The Foundations for Evidence-based Policymaking Act of 2018, which updated the Confidential Information Protection and Statistical Effi-

¹⁰ See, e.g., Duncan, G.T., Elliot, M. and Salazar-Gonzalez, JJ. 2011. *Statistical Confidentiality: Principles and Practice* (New York: Springer), Chapter 2 pages 33-35 and Federal Committee on Statistical Methodology op. cit. pages 66-70.

¹¹ Duncan et al. op. cit., page 68.

ciency Act in Title III, explicitly acknowledges the necessity of this kind of comprehensive assessment of disclosure risk, when it instructs agencies to consider “risks and restrictions related to the disclosure of personally identifiable information, *including the risk that an individual data asset in isolation does not pose a privacy or confidentiality risk but when combined with other available information may pose such a risk*” (emphasis added).¹²

19. This is why Dr. Ruggles’s criticism of my Delaware ship example misses the mark. He argues that “[i]f we knew, for example, that the GQCI added 42 persons to a vessel in Delaware, that information could not possibly reveal any particular person’s identity or individual census responses” (Plaintiff’s motion p. 30). As explained previously, one must think of the data point of 42 persons to a vessel in Delaware as just one cell of a table that is related to many other cells in the millions of tables produced by the 2020 Census. It is the Census Bureau’s duty to safeguard the confidentiality of the data it collects by accounting for the interactions between all such cells.
20. Unobscured information about imputed persons directly compromises the data from direct respondents because the Census Bureau’s imputation system usually substitutes data from another member of the group quarters facility, household or a nearby household (known as “donor records”) when doing characteristic imputation. The donor records are direct census responses (or responses from confidential administrative records). Providing unobscured imputed information derived from donor records puts those nearby individual donor’s data at risk. Differential privacy systems automatically provide various levels of protection for these data in either their direct use or their use as “donor records.”
21. Dr. Ruggles’s claim that releasing the requested information does not create a new invariant is, by definition, false: an invariant is any statistic published exactly from

¹² 44 U.S. Code §3504(b)(6)(A)

the Census Edited File (CEF) without differentially private noise added. The GQCI information is part of the CEF, and so releasing it without differentially private noise involves the release of an invariant statistic. And, because the GQCI statistics cannot be derived from the current approved set of invariants (and not, in particular, from the invariant state populations and the invariant block-level counts of occupied group quarters facilities), they would constitute 52 new invariants. Because of this, the GQCI statistics would leak a poorly controlled amount of information about the CEF and could be combined with other 2020 Census releases to improve inferences about the CEF records (recall the analogy to linear systems of equations previously given: releasing the GQCI information without differentially private noise is like releasing 52 additional equations that can be used to help solve for the underlying CEF microdata).

22. It is true that, even with the release of the GQCI information as an invariant, differentially private noise would still remain in the total GQ population counts. But this does not mitigate the practical consequences of the release of the GQCI information as an invariant: differentially private noise is carefully structured to guarantee that additional leaks of information from each new published statistic are small and controlled. This means that the kind of uncertainty Dr. Ruggles asserts would remain in the total GQ population counts, if the GQCI information were released without differentially private noise, is not comparable to differentially private uncertainty. Some uncertainty may remain, but it would not be easily quantified, and unlike with carefully structured differentially private noise, it would not be possible to promise that the additional disclosure risk from the published statistics remained small and controlled. Finally, depending upon the magnitudes of the GQCI values, their release could completely defeat the noise in the GQ population. This is why the Bureau would still consider the forced release of the requested information as “unobscured.”

REBUTTAL OF RUGGLES EXPERT REPORT

23. In his critique of the Census Bureau's simulated reconstruction-abetted re-identification attack (Ruggles Expert Report, pages 15-22), Dr. Ruggles claims that "[t]he Census Bureau's detailed description of database reconstruction provides overwhelming evidence that the database reconstruction experiment failed to demonstrate a realistic disclosure risk" (p. 16). The risk of disclosure is amply demonstrated by the reconstruction attack.¹³ One of the principal sources of disclosure risk is the inherent uniqueness of an individual within their community. Elamir and Skinner (2004) affirmed this, stating "...a measure of disclosure risk is the proportion of individuals in the microdata sample which have a unique combination of values of the key variables (assumed categorical) in the population...*Such individuals, referred to as population unique, may be judged to be particularly 'at risk of disclosure'*" (emphasis added).¹⁴ As I showed in Table 5 of Appendix B of my first declaration, 44% of the entire U.S. population has a unique combination of sex and age at the census block level. In sparsely populated blocks, those with populations of between 1 and 9 residents, over 95% of individuals are population uniques at the census block level. This prevalence of population uniques suggests that the underlying disclosure risk of census data published down to the census block level is alarmingly high, and the results of the Census Bureau's reconstruction-abetted reidentification attack (as reflected in the precision rates of those reidentifications) confirm this vulnerability. The precision rates of the Census

¹³ The Federal Committee on Statistical Methodology has also cited this research and explicitly recognized the risk to confidentiality from database reconstruction in their recently released Data Protection Toolkit. Available at: <https://nces.ed.gov/fcsm/dpt/content/26> (cited on October 28, 2021).

¹⁴ Elamir, E. A., and Skinner, C. J. (2004). "Record-level measures of disclosure risk for survey microdata" page 2. Available at: <https://eprints.soton.ac.uk/8175/1/8175-01.pdf> (cited on October 29, 2021).

Bureau’s simulated attack, shown in Table 1, demonstrate that one disclosure risk inherent to the reconstructed data is particularly high for respondents with unique combinations of block, age, and sex. If an attacker knew a respondent’s census block, age, and sex and used the reconstructed data to infer their race and ethnicity, then the attacker would have a very high success rate in blocks with populations between 1-9 people and between 10-49 people (96.98% and 91.68%, respectively). For these individuals, if the attacker finds a match in the reconstructed data, they can be highly confident that their re-identification of that individual is accurate. In blocks of these very small sizes, the data provided by these respondents will very often be information that is unique to their response.

Block Population Bin	Putative Re-identifications (Source: Commercial Data)	Confirmed Re-identifications (Source: Commercial Data)	Precision (Source: Commercial Data)	Putative Re-identifications (Source: CEF)	Confirmed Re-identifications (Source: CEF)	Precision (Source: CEF)
TOTAL	137,709,807	52,038,366	37.79%	238,175,305	178,958,726	75.14%
0						
1-9	1,921,418	1,387,962	72.24%	4,220,571	4,093,151	96.98%
10-49	25,148,298	13,481,700	53.61%	47,352,910	43,415,168	91.68%
50-99	30,567,157	12,781,790	41.82%	51,846,547	42,515,756	82.00%
100-249	38,306,957	13,225,998	34.53%	63,258,561	45,807,270	72.41%
250-499	21,789,931	6,408,814	29.41%	35,454,412	22,902,054	64.60%
500-999	13,803,283	3,460,118	25.07%	23,280,718	13,514,134	58.05%
1000+	6,172,763	1,291,984	20.93%	12,761,586	6,711,193	52.59%

DRB clearance number CBDRB-FY21-DSEP-003.

24. Dr. Ruggles criticizes the Census Bureau’s simulated attack for lacking a “null model” (p. 17). He compares the Census Bureau’s database reconstruction simulation to the results of a random-guessing strategy he devised as a supposedly valid null model (p.

- 18). He argues that the results of the Census Bureau's simulation are no better than a random guesser would do.
25. But Dr. Ruggles makes substantial methodological errors in his random-guessing strategy comparison. For example, oversimplifying things for instructive purposes, I would describe Dr. Ruggles procedure as, essentially:
- i. Choosing a random census block, favoring densely populated blocks much more than sparsely populated blocks; and
 - ii. Randomly guessing one combination of age and sex (e.g., 50-year-old female).
26. Dr. Ruggles developed his proposed "null model for comparison" experiment to determine the probability that such a random guesser would match anyone in the block. Considering that a block of 1,000 people will almost certainly contain a 50-year-old female, it is very easy to succeed in such an experiment. However, this experiment has nothing do with actual reconstruction—proposing 1,000 records and checking how well each and every one of those proposed records match the actual characteristics of the population. In other words, actual database reconstruction must go beyond establishing the existence of a single 50-year-old-female. It must estimate, for each age and sex combination, the number of individuals with those characteristics in a particular census block. The question is not "can one find a person with random characteristics" but rather "can one find each person in the reconstructed data". I explained this very carefully in Appendix B to my original declaration, but Dr. Ruggles ignores this distinction, rendering his conclusions meaningless in the context of the Census Bureau's research.
27. To further explain the flaws in Dr. Ruggles's methodology, consider what would happen if we replaced the random guessing mechanism ("random guesser") in his experiment with a one-note guesser who always chooses one age/sex combination for

everyone (such as the most frequently occurring age/sex combination in the U.S. population). The premise underlying Dr. Ruggles's baseline, what he calls the "null model," is that the random guesser, working from a good estimate of the population distribution by sex and age, will replicate that distribution over the 10,000 guesses in his experiment. The one-note guesser will produce a population distribution that has a single value in the 10,000 simulations. This modification can be done in Dr. Ruggles's published code by taking the censim3.f file and adding hyp=146 after line 65 to make the guesser always choose the age/sex combination that is encoded as the number 146 in the code (50-year-old female). One would expect such a one-note guesser to produce data that would not match population characteristics at all because even the most common age/sex combination is shared by less than 1% of the U.S. population. Hence, if Dr. Ruggles's proposed baseline methodology were a reasonable "null model," one would expect the simulation to produce a very low expected match probability. However, when used in Dr. Ruggles's simulation methodology (a one-line change to his code), the match rate for this one-note guesser (57%) is even better than the match rate for his random guesser, (52.6%) (Ruggles Expert Report, p. 18). Now clearly, 57% of the U.S. population do not share the same exact sex and age, so this kind of one-note guesser should not achieve such a high score when used in a properly designed baseline simulation methodology. And if Dr. Ruggles's simulation had replicated a real record linkage model, 1% is the baseline it would have produced from the one-note guesser because each time it found a 50-year-old female in the target population, that record would be removed from the target population before the next guess simulation.¹⁵ As soon as all the 50-year-old females in the target population were guessed, the remaining 50-year-old female guesses would fail. Having the one-

¹⁵ Removing records once they have been matched is the way the Census Bureau's reconstruction statistics were generated. (See Appendix B of my second declaration.)

note guesser always guess the most common value in the population produces the largest baseline probability his methodology can generate. The striking 57% baseline probability from the one-note guesser demonstrates that the simulation methodology that Dr. Ruggles uses to critique the Census Bureau is deeply flawed.

28. Later, in Dr. Ruggles's critique of recent research that the Census Bureau performed to evaluate the impact of higher swapping rates on disclosure risk (p. 22), he misinterprets the baseline used for the comparison. The re-identification rates and precision for what he refers to as the "unswapped data" were in fact the rates for analysis of the 2010 Census Hundred Percent Detail (HDF) file—the swapped microdata used for tabulation of the published tables. The swapping rate used for the HDF is not specified because that parameter is confidential. Furthermore, Dr. Ruggles fails to note that while data swapping was primarily concentrated in low-population blocks in 2010, the recent data swapping experiments at the Census Bureau did not concentrate selected records for swapping in this way, and they revealed the same vulnerability—even at swap rates of 50%—as the 2010 swapping system.¹⁶ Therefore, the conclusions he draws from this recent research are not supported by the evidence he presents.
29. A more appropriate measure of disclosiveness of the reconstructed data is how well an attacker can predict a person's characteristics relative to how well the attacker could predict those characteristics if the individual's response were removed from the

¹⁶ Hawes, Michael and Rolando Rodriguez "Determining the Privacy-loss Budget: Research into Alternatives to Differential Privacy," Presentation to the Census Scientific Advisory Committee. May 25, 2021. Available at: <https://www2.census.gov/about/partners/cac/sac/meetings/2021-05/presentation-research-on-alternatives-to-differential-privacy.pdf>; Hawes, Michael and Rolando Rodriguez "Determining the Privacy-loss Budget: Research into Alternatives to Differential Privacy" Recorded webinar. June 4, 2021. Available at: <https://www.census.gov/data/academy/webinars/2021/disclosure-avoidance-series/research-into-alternatives-to-differential-privacy.html>.

Census before tabulation. Dr. Ruggles's expert report provides no analysis of this "inference with the person" vs. "inference without the person" comparison that could account for the high success rate demonstrated in the Census Bureau's simulated attack. This type of comparison is critical to evaluating disclosure risk because it helps differentiate between what you can learn about an individual from improperly protecting their confidential data (what I'll call unauthorized disclosure or confidentiality-violating inference) on the one hand, versus what you can infer about an individual from broader societal trends reflected in the data (what I'll call statistical inferences based on aggregates) without using any of that individual's confidential information (scientific learning or legitimate statistical use of the data) on the other.¹⁷ Being able to quantify this difference is what distinguishes formally private disclosure avoidance methods, like differential privacy, from the approaches used in the past.

30. Here is a clarifying example of the difference between statistical inferences based on aggregates and confidentiality-violating inferences based on using the respondent's data. Some inferences about confidential information can be achieved purely by relying on aggregate statistical information about groups and do not rely on any individual's confidential census responses. For example, suppose Alicia is trying to learn how Roberto responded to the race question in 2010, and she already knows Roberto lived in Montana at the time of the 2010 Census enumeration. Alicia could then review the 2010 Census tables, and, because she can find that 89.4 percent of respondents reported "White Alone" in Montana, Alicia can guess with high confidence that Roberto's census response was "White Alone." This is an example of an inference based on aggregate statistical information about groups, rather than knowledge of Roberto's

¹⁷ Dwork, C., & Naor, M. (2010). On the difficulties of disclosure prevention in statistical databases or the case for differential privacy. *Journal of Privacy and Confidentiality*, 2(1).

confidential census response. Roberto's census response was one of 989,415 in Montana in 2010, and so, even if Roberto had never participated in the census, it would still be easy for Alicia to guess that Roberto's race is probably "White Alone," just by reviewing the responses of the other participants and guessing that Roberto's response would match the most common response. Now suppose that on his 2010 Census form Roberto reported being "Some Other Race Alone," that Roberto was the only resident of his census block, and that Alicia knows Roberto's address (and consequently his census block). Alicia could then easily review the published tables for Roberto's block, find that a single person who reported "Some Other Race Alone," and, if disclosure avoidance techniques were not used, guess with complete confidence that Roberto reported "Some Other Race Alone." This is an example of a confidentiality-violating inference—if Roberto had not participated in the census, Alicia would not be able to infer Roberto's race in this way as his block would have a reported count of zero. I emphasize that the key issue here is that Alicia's inference could not have been made without Roberto's data being present in the census and could only be made with his data present. This is what makes the inference unique to Roberto's census response. That Roberto is the only resident of his block and the inference is 100 percent certain, rather than just highly confident, both help to make the example simple. Privacy-violating inference can still take place in blocks with large populations (even if it is more common in small populations) and when an attacker can be confident but not certain.

31. In his overview of disclosure avoidance methods used for prior censuses, Dr. Ruggles claims that these methods "have worked extremely well" (p. 14); however, he fails to note the historical context. The disclosure avoidance rules in place for the 2010 Census reflected the understanding that releasing all the swapped microdata directly (at the chosen parameter settings) would constitute a violation of Title 13's confidentiality

provisions. But, at the time, the Census Bureau did not believe that the published aggregated data tables based on all the microdata were equivalent to microdata. Thus, at the time, the Census Bureau believed that such aggregated data tables could be protected using less stringent disclosure avoidance methods. The Census Bureau's simulated reconstruction attack demonstrated that the differentiation between aggregated data vs. microdata is no longer tenable in the context of disclosure avoidance and that additional research on the strengths and weaknesses of the techniques used in 2010 is necessary. This research is currently ongoing. DSEP decided to replace the swapping method in 2018 on the basis of the reconstruction alone because the reconstructed data violated the 2010 Census microdata disclosure avoidance requirements, as I have noted in my previous declarations.¹⁸ Specifically, the reconstructed microdata (1) contain all census person records, not a sample, (2) contain geography identifiers for areas with populations less than 100,000 persons, (3) contain demographic variables for which the national population is less than 10,000. All three are prohibited by the 2010 Census microdata disclosure avoidance procedures.¹⁹ To be completely clear, when the microdata file contains a record for every person who was enumerated, the presence of a population unique—a person who is the only one in the country enumerated with a particular combination of characteristics—is a well-documented, known disclosure risk requiring active disclosure avoidance measures.²⁰ For this reason, I have repeatedly stressed that the salient disclosure risk

¹⁸ Abowd second declaration Appendix B.

¹⁹ McKenna, Laura. 2019. "Disclosure Avoidance Techniques Used for the 1960 Through 2010 Census" available at <https://www.census.gov/library/working-papers/2019/adrm/six-decennial-censuses-da.html> (cited October 25, 2021).

²⁰ Duncan et al. op. cit. Chapter 2.

statistic from the 2010 Census tabulations in Summary File 1 is that 44% of the population of the U.S. is unique using just their block location, sex and age (Abowd supplemental declaration, Table 1). That is, 44% of the population was at known disclosure risk because the disclosure avoidance techniques used in the 2010 Census proved to be insufficient. DSEP and the Census Bureau's Operating Committee have understood this disclosure avoidance issue since 2018 and have consistently instructed the 2020 Census Disclosure Avoidance System to actively mitigate this risk.²¹

I declare under penalty of perjury that the foregoing is true and correct.

DATED and SIGNED:

JOHN ABOWD

 Digitally signed by JOHN ABOWD
Date: 2021.11.01 11:01:55 -04'00'

John M. Abowd

Chief Scientist and Associate Director for Research and Methodology

United States Bureau of the Census

²¹ Data Stewardship Executive Policy Committee Meeting Record, February 15, 2018 (Appendix A).

Appendix A

DSEP Meeting Record

Topics:



Database Reconstruction Issue Mitigation (John Abowd, ADRM)

Meeting Date: February 15, 2018

<i>Position</i>	<i>Attending for Position</i>
COO (Chair)	Enrique Lamas
ADDC	Albert Fontenot
ADDP	Karen Battle
ADEP	Nick Orsini and Ron Jarmin
ADFO	Tim Olson
ADITCIO	Nitin Naik
ADRM	John Abowd
ADCOM	Stephen Buckner
CAO	David Ziaya
CFO	Joanne Crane
Asst. DRM	John Eltinge
Chief PCO/ Chief Privacy Officer	Robin Bachman
S.A. Director's Office	Douglas Clift
CISO	<i>Absent</i>
At-Large	Howard Hogan
At-Large	Frank Vitrano
Also Attending:	William Samples, David Waddington, Burton Reist, Victoria Velkoff, Robert Sienkiewicz, Jim Treat, Cynthia Hollingsworth, Clifford Jordan, Julia Naum, Jim Dinwiddie, Simson Garfinkel, Melissa Creech, Pat Cantwell, Byron Crenshaw, Hampton Wilson, Ashley Landreth, Mike Castro, Julie Atwell, Michael Snow

[REDACTED]

- [REDACTED]
- [REDACTED]

Database Reconstruction Issue Mitigation

Background

The Census Bureau's Operating Committee (OPCOM), serving as the Enterprise Risk Review Board, elevated the enterprise risk of database reconstruction to an enterprise issue based on the results of a database reconstruction attack research effort the Census Bureau launched to understand that risk better. When an enterprise risk is elevated to an enterprise issue, the risk owner must implement an active mitigation plan to mitigate the risk. To that end, the Research and Methodology Directorate presented six recommendations to help manage the Census Bureau's publication strategy in ways that will protect its databases from reconstruction attacks.

NOTE: presenters and DSEP recognized that implementing several of the recommendations will require decisions on budget and staffing resources and that those decisions would need to be handled by other bodies at the Census Bureau. DSEP confined its discussion to establishing policy in response to the recommendations.

The following 6 recommendations were presented to DSEP:

- 1. Suspension until September 30, 2019 of ad hoc releases of sub-state geography from any confidential source unless vetted differential privacy tools, or a DRB-approved noise-infusion alternative, have been used to produce the publication. This applies to all research projects whether they are external or internal. It does not apply to scheduled publications from sponsored survey clients for whom there is already an approved DRB protocol. Those clients should be put on notice for subsequent contracts. The complete list of approved exceptions, including sponsored survey products, is provided in 20180215b-External_Internal_Substate_Geography.xlsx. The suspension will be reviewed prior to September 30, 2019.**

NOTE: This suspension does not apply to state and national publications. It also does not apply to already scheduled publications from regular production activities. Program areas provided ADRM a list of those scheduled publications that should be exempted from the suspension. ADRM proposed ending those exemptions by September 30, 2019 even for those publications if they were not being produced using formally private systems by that point.

Discussion: DSEP recognized the need to modernize the Census Bureau's disclosure avoidance systems. DSEP acknowledged that by approving a list of exemptions they are agreeing to hold elevated levels of risk of database reconstruction associated with all of these data products. However, DSEP acknowledged the Census Bureau is obligated to provide the data the public needs for decision making and some of the release dates are required by law.

DSEP also acknowledged the need to set a target date for making these changes. While the ultimate goal is to make the publications of all of our programs formally private, that likely will not happen by September, 2019. However, in the meantime significantly improved noise infusion methods will be put in place to mitigate reconstruction risk.

DSEP members expressed concern that the list of already scheduled publications presented might be incomplete and asked for additional time for program areas to review the list and submit updates. DSEP agreed that the Center for Disclosure Avoidance Research (CDAR) should continue to accept submissions and finalize the list in advance of the next DSEP meeting. DSEP will formally approve the list at that point.

Decision: DSEP will finalize their approval of this recommendation at the March 15 DSEP meeting once the list of excepted publications has been finalized.

Action Items: Program areas will send updates on the table of exempted data releases to the Chief of CDAR by February 23. The Chief of CDAR will redistribute the combined list to all contributors by February 28. CDAR will finalize the list of approved exceptions for distribution before DSEP's meeting on March 15.

- 2. Suspension of all proposed tables in Summary File 1 and Summary File 2 for the 2020 Census at the block, block-group, tract, and county level except for the PL94-171 tables, as announced in Federal Register Notice 170824806–7806–01 (November 8, 2017, pp. 51805-6). To add a summary file table at any level of geography, racial/ethnic subpopulation other than OMB aggregate categories as specified in the 1997 standard (Federal Register October 30, 1997, pp. 58782-90), or group quarters type below the 2010 P42 seven categories, an affirmative case must be made for that table, use cases identified, and suitability for use standards developed. In addition, we recommend that the voting-age invariant in PL94-171 be removed, so that voting-age would be protected. DSEP will be asked to approve the SF1 and SF2 table specifications once they have cleared 2020 governance.**

NOTE: The PL94-17 tables from the 2018 End-to-End Census Test have been designed with a formally private system already and will be published, with the voting-age invariant, as planned.

Discussion: DSEP recognized that the SF1 and SF2 involved a very detailed set of tables that had been created to suit a wide set of data users. These tables were created, as a rule, to produce as much highly accurate data as possible within the existing disclosure avoidance framework. However, DSEP acknowledged that these data in many cases were accurate to a level that was not supported by the actual uses of those data, and such an approach is simply untenable in a formally private system.

DSEP acknowledged a fundamental need to take stock of what data the Census Bureau is required to publish, both by statute and the needs of our data users, and at what level of accuracy. This is not an activity that should be done by our Disclosure Review Board. Program areas have to make the case of what the data will be used for, and the actual minimum level of accuracy needed for those uses, so that CDAR and the DRB can build the system to allocate the privacy-loss budget according to those use cases.

A redesign of SF1 and SF2 based on formally articulated use cases will take a tremendous amount of effort but cannot be done in a vacuum. Program areas will have to reach out to data-user communities on developing the use cases for the needed data accuracy and levels of geography.

NOTE: DSEP discussed but tabled until later any decision on changing the voting-age invariant for the PL94-171 table produced as part of the 2020 Census.

Decision: DSEP approved this recommendation. For the 2020 Census, SF1 and SF2 will be rebuilt based on use cases.

Action Items: DCMD, POP, and ADDC divisions will work with the relevant program management governing board (PMGB) to establish a plan to execute this redesign.

3. Immediate review of all sub-state geography scheduled publications from the American Community Survey (ACS) to determine which ones can be delayed until there is a formally private publishing system for ACS.

Discussion: DSEP acknowledged that many of the ACS tables are already in production and that production needs to move forward. DSEP acknowledged that there are likely no publications currently suitable for delay, however they emphasized that ACSO needs to ensure that all exceptions are added to the list.

Decision: DSEP approved this recommendation.

Action Items: ACSO will verify that they have included all of the necessary publications on the list of exempted data releases.

4. Consideration of postponing ACS PUMS releases indefinitely.

NOTE: DSEP recognized that all of the publication systems and methods for the Census of Island Areas are identical to the ACS. DSEP emphasized that any changes made to the ACS should also reflect consideration of the needs of the Island Areas.

Discussion: DSEP acknowledged that while the threat of database reconstruction and reidentification attacks applies to all of the Census Bureau's data products, should the ACS data be subject to a reidentification attack, from a public perception standpoint, our continued publication of the ACS PUMS files would appear to be an egregious mistake.

However, DSEP also acknowledged that the ACS PUMS is a heavily used dataset for research and recognized that discontinuing this publication could generate a great deal of traffic for the FSRDCs. DSEP acknowledged that, before the Census Bureau restricts use the ACS PUMS to the FSRDCs, it needs to verify that they can handle the increased workload. Additionally, at present there are no FSRDCs that are readily accessible from the Island Areas.

DSEP recognized that immediate suspension of the ACS PUMS would cause a great deal of concern among data users and others. DSEP discussed the need to work on messaging around

any suspension and to brief the Department of Commerce before the Census Bureau implements the suspension.

Decision: DSEP deferred for one month any decisions to suspend release of the ACS PUMS pending further consideration of the ability of the FSRDC network to support increased demand, the impact on the data needs of the Island Areas, and development of a messaging plan.

Action Items: ADRM will prepare an assessment of the potential increased demand on the FSRDC network, and Decennial will prepare an assessment of the impact of suspending this publication on the Island Areas. ADCOM will work on a messaging plan.

5. Mandate for the 2022 Economic Censuses to use formally private publication systems for all tables.

Discussion: DSEP recognized that it is too late to begin creating a formally private system for data releases from the 2017 Economic Census. DSEP additionally discussed how modernizing disclosure avoidance systems will involve much more than just budgeting extra funds. It also will require having the adequate number of people with the right skills to do the work.

DSEP recognized that program areas will have to involve their PMGB in setting resources, budgets, and timelines and that it should be feasible to put formally private systems in place in time for the 2022 Economic Census.

Decision: DSEP approved this recommendation. The Census Bureau will move forward with designing and implementing formally private systems for the 2022 Economic Census.

6. Mandate to the Demographics Directorate to begin negotiations with survey clients for increased use of restricted-access microdata protocols and formally private table publication systems.

POST MEETING NOTE: a member in attendance recommended that there should also be outreach to reimbursable clients for the Economic Directorate.

Discussion: DSEP recognized the need to begin discussions with sponsors of Census Bureau surveys but determined that the Census Bureau should have a communications plan in place before mandating that the Demographic Directorate speak to sponsors.

Decision: DSEP will reconsider in one month whether to mandate conversations with survey and report sponsors.

Consolidated Action items:

- Program areas will send updates on the table of exempted data releases to the Chief of CDAR by February 23.
- The Chief of CDAR will redistribute the combined list to all contributors by February 28.
- DCMD, POP, and the ADDC will work with the relevant PMGBs to establish a plan to execute the redesign of SF1 and SF2 based on use cases.
- ACSO will work to determine that all ACS data releases in production are listed on the spreadsheet of exceptions to the suspension.
- ADRM will prepare an assessment of the potential increased demand on the FSRDC network from suspension of the ACS PUMS.
- ADCOM will work on a messaging plan related to the suspension of the ACS PUMS.
- Decennial will prepare an assessment of the impact of suspending publication of the ACS PUMS on the Island Areas.