

Values in a National Information Infrastructure: A Case Study of the U.S. Census¹

Timothy M. Weber

Doctoral Student, Media Ecology

Department of Culture and Communication

New York University

In July of this year, the Census Bureau complied with a request from the Department of Homeland Security and released to the Customs and Border Protection division a specialized compilation of demographic tabulations on the Arab-American population (specifically tabulations outlining urban areas with more than 10,000 inhabitants reporting Arab descent and zip-code specific tabulations subdivided by country of origin)². While the sharing of information between the two agencies was declared "common practice", the public outcry which resulted prompted the Census Bureau to revise its information³ sharing policies - requiring all special data requests from law enforcement and intelligence agencies to undergo review by an appropriate Associate Director. In light of this controversy, this talk seeks to flesh out the relationship between the American census, public concerns over privacy, and technology. By tracing this triangulated relationship throughout U.S. census history, I will argue for the importance of keeping census practices both reflexive and transparent.

From its inception in 1790 to its twenty-second iteration in 2000, the U.S. Census as a practice and institution has undergone tremendous change. In the course of this transformation, what one might call the surrounding "rationale" for the census - its

¹ Part of this research was made possible by the National Science Foundation for the ITR collaborative grant "Sensitive Information in a Wired World" (PI: Helen Nissenbaum).

² Clemetson, Lynette. "Homeland Security Given Data on Arab-Americans". NY Times, July 30, 2004. Available at: <http://www.nytimes.com/2004/07/30/politics/30census.html>.

purpose per se - has steadily shifted from an emphasis on the "mere" enumeration of the population for use in Congressional apportionment to a commitment toward an ever expanding collection of socio-economic data aimed at intelligent policy making. To compare the explicitly constitutional beginnings of the census with the current strategic goals laid out by the Census Bureau is a step toward grasping this change. In its original Constitutional form, the census clause reads: "Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers, which shall be determined by adding to the whole Number of free Persons, including those bound to Service for a Term of Years, and excluding Indians not taxed, three fifths of all other Persons. The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct."⁴ The contemporary census is represented in its own literature as primarily geared to "meet the needs of policymakers, businesses and nonprofit organizations, and the public for current measures of the U.S. population, economy, and governments"⁵. In short, the current census is about much more than determining seats in Congress.

At the core of this historical development, one might posit a triangulated relationship between the census, certain technological means for statistical processing, and extra-governmental understandings of national data (including both private sector desires for a source of in-depth national data profiling as well as popular concerns regarding the intrusion of the government into everyday life). With each iteration, one

³ "U.S. Census Bureau Announces Policy Regarding Sensitive Data". Press Release August 30, 2004. Available at: http://www.census.gov/Press-Release/www/releases/archives/directors_corner/00249-1.html

⁴ U.S. Constitution, Article I, Section 2, Clause 3

⁵ U.S. Census Bureau Strategic Plan, FY 2004-2008 Issued September 2003 by the U.S. Dept. of Commerce

finds the census growing in scope - from the addition of question "types" and the expansion of social fields surveyed to a steadily increasing timeline of operations and a growing quantity of information made publicly available. Some noteworthy dates here include: 1810 (the first collection of data pertaining to economic and religious institutions), 1830 (the first collection of data on health and disability), 1850 (a shift to the individual instead of the household as the basic unit of enumeration), 1902 (the establishment of a full-time Census Bureau), 1940 (the beginnings of statistical sampling and the birth of the "long-form") and 1970 (the publishing of 100 percent city-block data for any town with at least 10,000 inhabitants across the country)⁶.

This historical expansion echoed the statistical sentiments of one of the census' primary founders - James Madison. Madison's initial proposal for the census included its extension "so as to embrace some other objects besides the bare enumeration of the inhabitants"; in this way, Congress' possession of this "most useful information" would "enable them to adapt the public measures to the particular standards of the community"⁷. What is crucial here is that Madison's vision of the census as a vehicle for statistical inquiry beyond mere "headcounting" was simultaneously a surpassing of what was specifically outlined in the Constitution and an undertaking that was realized with the first census in 1790. In other words, from "day one" the census has been an institution which has pushed the limits of its Constitutionally enumerated powers.

This initial surpassing of its Constitutionally enumerated mandate set an important precedent, because (unsurprisingly) the expansion of the census was not

⁶ Anderson, Margo J. *The American Census: A Social History*. Yale University Press: New Haven. 1988

⁷ *Annals of Congress*, 1. P. 1077. See also The Federalist Papers for Madison's perspective on Government Intelligence.

without its detractors. Initial resistance came from segments of the economic sector who reacted strongly to government inquiries as early as 1820. Resistance culminated in the first direct legal challenge to the census in the 1901 court case *U.S. v. Moriarity*; the ruling of the court, however, reflected the Madisonian logic in which the census was birthed: (It's worth quoting at length here the verdict of District Judge Edward B. Thomas) "The functions vested in the national government authorize the obtainment of information in order to enact laws adapted to the needs of the vast and varied interests of the people, after acquiring detailed knowledge thereof. The government has the right to make the researches in order to meet its ever-widening obligations to the welfare of its citizens and to the world. For the national government to know something, if not everything, beyond the fact that the population of each state reaches a certain limit, is apparent, when it is considered what is the dependence of this population upon the intelligent actions of the general government."⁸

The staunchness with which the government approaches its "right" to data collection is apparent in U.S. Code Title 13, Chapter 7 which makes refusal to answer census inquiries or the providing of false information to census officials a punishable offense⁹; that is to say, filling out your census form is the law. Nevertheless, resistance to the census has not disappeared. The most recent decennial census in 2000 was preceded by public statements from such figures as Sen. Trent Lott and then-presidential-candidate George W. Bush who encouraged citizens to "leave-blank" those questions which they

⁸ *U.S. v. Moriarity*, 106 Fed. 886, 691, 692 (C.C.S.D.N.Y. 1901).

⁹ U.S. Code Title 13, Chapter 7, Subchapter II, Sec. 221. Available at: <http://www4.law.cornell.edu/uscode/13/221.html>

found intrusive¹⁰; and the 2000 census was followed by another lawsuit - *Morales v. Evans* - challenging the constitutionality of compulsory inquiries beyond the purposes of apportionment. The court ruling, however, once again upheld the Madisonian rationale of the necessity of statistics for intelligent governance¹¹.

In the midst of such concerns over governmental intrusion, the position of the Census Bureau on the American landscape is perhaps best refracted through the following event: "Census 2000 featured the first-ever paid advertising campaign. So as to reach all adults living in the United States (including Puerto Rico and the Island areas), the Census Bureau awarded a contract to Young & Rubicam, totaling \$167 million, for print, television, and radio advertising for its national, regional, and local advertising campaign. The advertising campaign consisted of more than 250 TV, radio, print, outdoor, and Internet advertisements - in 17 languages - reaching 99 percent of all U.S. residents. By the end of the campaign, the census message - 'This is your future. Don't leave it blank.' - had been heard or seen an average of 50 times per person.¹²"

It is here, perhaps, that the third term of the triangle - technology - can best be brought to bear on this historical analysis. In talking about technology with respect to the census, one opens up a complicated labyrinth of analytical pathways. In a broad sense - contiguous with something like the notion of technique - the census itself can be construed as a technological type. Alternatively, technology as a social category might

¹⁰ "Everyone counts: It's Census Day in the U.S." April 1, 2000. From CNN.com. Available at: <http://www.cnn.com/2000/US/04/01/census.day02/> and "Take my Privacy, Please". *Time*. 155, no 15, 22, April 17, 2000.

¹¹ Petition for Writ of Certiorari – Public Advocate of the United States. Edgar Morales, Laique Rehman, Nouhad K. Bassila, George Breckendridge, and William Jeffrey Van Fleet v. Don Evans, U.S. Secretary of Commerce, and Kenneth Prewitt, Director of the United States Bureau of the Census. U.S. Supreme Court, January 8, 2002

¹² "Measuring America" U.S. Census Bureau. p.140.

be understood as an inter-census referent of sorts - that is to say that much of the information elicited via the census pertains to technology in a more traditional sense (such as questions on housing facilities [plumbing, electricity, etc.], transportation methods, etc.). For the purposes of this talk, however, I find it useful to encounter the census through technologies of information processing.

In many respects, the census as it currently stands (housed in a "full-time" government bureau) was made possible only through a technological breakthrough in data processing. Due to dynamic population growth, the 1880 census was not completely tabulated until 1888, rendering much of the data obsolete before the census had officially ended. Such circumstances prompted the government to hold a competition for more efficient statistical processing methods; the winner of this competition was none other than census employee Herman Hollerith, whose early system of punch-cards allowed a more timely tabulation for the eleventh census in 1890. Hollerith went on to father IBM, and the census became its own bureau in 1902.

While with Hollerith's system, the historical stage might seem set for an evolutionary line to be drawn from punch-cards to contemporary statistical techniques, in the case of the census such a view seems misconstrued. In what one might term a rare case of technological "restraint", the Census Bureau is required (by the Constitution and recently via a lawsuit brought against the Bureau by Congress)¹³ to conduct an "actual" enumeration of the population - such a statute translates technologically as follows: the Census Bureau is prohibited from using sampling techniques (a staple of contemporary statistics) with respect to any information gathered for apportionment; thus, every ten years the census must be conducted via the canvassing of the country and the (for-lack-

of-a-better-word) physical enumeration of the population. Trivial as this might seem, the inability for the Census Bureau to conduct its collection phase via sampling seems to have important consequences for worries about the expansion of governmental information surveillance; namely, because a "successful" census hinges on a cooperative populace (that is, refusal to answer or providing false information stands to jeopardize the very "accuracy" such person-to-person enumeration is supposed to ensure). As such, the Census Bureau must be in a position of sensitivity toward the worries of those who comprise its data set.

Such concern on the part of the census is reflected not only through such phenomena as the massive 2000 advertising campaign and the rapid response to concerns regarding the recent Homeland Security requests, but most significantly through the bureau's policy toward what it terms "confidentiality". Here, one finds the historical implementation of statutes and techniques aimed at "protecting" the individual responses garnered via enumeration. As part of the census' statutory obligations, results from each collection phase must be published; as such, the census bureau has developed an architecture of information handling policy designed to obscure the identity of all information at the individual level - in other words, before any information can "leave" the bureau, it must be stripped of any identifiers which might indicate the actual individual respondent. Included in this policy are such statutes as oath-taking and punitive measures which make breaches of confidentiality by census employees a felony offense, as well as a variety of statistical processing techniques (including disclosure

¹³ Dept. of Commerce v. U.S. House of Representatives. 525 U.S. 316 (1999). Docket Number: 98-404

limitation, data suppression, data swapping, and the introduction of statistical "noise")¹⁴ aimed at ensuring that any published information cannot be used for purposes of identifying individuals.

To move this diachronic analysis into the present moment, it becomes important to consider both the current socio-political relevance of census data and contemporary challenges to the bureau's commitment toward confidentiality. In their work on the consequences of classification systems, Bowker and Star discuss the tendency of information infrastructures toward "embeddedness" - being "sunk" into other social structures and arrangements, as well as into technologies¹⁵. In the case of the U.S. census, one finds a prime example of this notion; census data (and the typologies contained therein) diffuse throughout a wide range of American sites - from realty to academic research, from Starbuck's and Wal-Mart's¹⁶ planning decisions to the data mining of "private sociology" at large. Indeed it is far from a stretch to argue that the entire industry of third-party information providers in large part owes its very existence to information freely provided via the census. While in the terminology of economics, this might be construed as an instance of a "significant positive externality", such diffusion of information into a variety of private sector sites is not entirely inconsistent with the current governmental understanding of the census' role (one must keep in mind that the Census Bureau is housed within the Department of Commerce).

The tension arises, however, when technological development exceeds the capacity of statutory protections. In short, while the Census Bureau might face statutory

¹⁴ "Confidentiality" U.S. Census Bureau. American FxxtFinder. Available at: http://www.factfinder.census.gov/jsp/staff/SAFFInfo.jsp?_lang=en&_sse=on&_content=su5_confidentiality.html.

¹⁵Geoffrey C. Bowker and Susan Leigh Star, Sorting Things Out: Classification and its Consequences. MIT Press, Cambridge Massachusetts: 2000.

limitations on the sampling techniques it may employ, other institutional sites exhibit much less restraint with regards to the techniques of statistical analysis they might utilize. While it is no longer "news", increases in computer processing power have allowed for more sophisticated information processing algorithms; the diffusion of such technical capabilities across a variety of sites has engendered an environment in which publicly available census data might be used for a variety of purposes. When one considers the current sophistication of geodemographic analysis, a growing danger is emerging that when cross-referenced with statistical information from other sources, the data published by the census can be used to "reidentify" (and locate) individual respondents. A short search on the world wide web will garner one a host of private workshops to attend that "teach" methods for "reading" and "putting to use" census data.¹⁷ When coupled with the historical logic of typological expansion which has seemingly governed the census from its inception, the threat to privacy seems imminent.

To understand this threat, it is helpful perhaps to consider the notion of privacy as a contextually derived value. In her theory of privacy as contextual integrity, Helen Nissenbaum argues that conflicts around privacy are often the result of an upsetting or challenge to the informational norms inherent to a social context¹⁸. Contexts here include both norms of appropriateness and norms of transmission. In the case of the census, then, one finds a historically robust context wherein the former norms (appropriateness) would bear for example on the questions appearing on census forms and the latter (transmission) are brought to bear on determining who has access to individually identifiable census

¹⁶ Kysar, Douglas A., "Kids & Cul-de-Sacs: Census 2000 and the Reproduction of Consumer Culture" . <http://ssrn.com/abstract=280778>

¹⁷ See for instance www.sgtinfo.com

¹⁸ Nissenbaum, Helen. "Privacy as Contextual Integrity". *Washington Law Review*. 2004.

data. The root of contemporary worries regarding privacy and the census, therefore, might be grounded in something akin to a violation of contextual informational norms. In other words, a controversy such as the sharing of census information between the bureau and the Department of Homeland Security might be understood as a disruption of both norms of appropriateness and norms of transmission - i.e. those Arab-American respondents whose data comprised the aggregate set had proffered their information to the Census Bureau under terms quite distinct from the norms governing the context of Homeland Security. To identify oneself as "of Arab descent" on a census form is an informational act governed by a different understanding of appropriateness than such an identification carries in the context of law enforcement; the former context is imbued with rationales such as affirmative action legislation, while the latter is a context aligned with more suspect acts such as racial profiling. In sum, the very informational category of race changes its appropriateness across contexts - where it is acceptable for the purposes of a census, such a category is perhaps not so acceptable for purposes of policing the populace.

To conclude, then, it becomes important to revisit the rationale and role of the census in the American informational infrastructure and to recognize its centrality therein. The challenges posed by new technologies involve the ease with which census data might be exceeding the norms of its context - i.e. used for purposes inappropriate to its collection or transmitted to those with no ethical claim to the data. To mitigate this, it becomes absolutely crucial to keep census practices transparent to the public, and the bureau's operations autonomous from possibly detrimental influences (governmental or otherwise). The bureau has a history of reflexivity that should be applauded (engendered

primarily by its own mandate to secure the "trust" of the public) but also kept open for review. In the words of Bowker and Star, "We need to recognize that all information systems are necessarily suffused with ethical and political values, modulated by local administrative procedures. These systems are active creators of categories in the world as well as simulators of existing categories. Remembering this, we keep open and can explore spaces for change and flexibility that are otherwise lost forever" (321)¹⁹. It is hoped that this paper contributes to a relatively modest goal of keeping open such a space for "flexibility", and in doing so will encourage a more public discussion on the importance of census techniques.

¹⁹ Ibid n.15