Census Data Archiving and Preservation

Select Topics in International Censuses[1]

Released July 2019

INTRODUCTION

Individual census records are considered a special kind of record within national statistical offices (NSOs) because of their reach and sensitive nature. These records include virtually every person living in a country at a given time and are comprised of confidential personal information with great potential for misuse. For these reasons, their transfer and archiving procedures and their storage security protocols should be considered the "gold standard" of recordkeeping at NSOs. Ideally, other less sensitive materials would be held to the same standards as the procedures and regulations used with individual census records. However, in practice this varies from country to country. The ultimate goal of implementing good archiving practices in a NSO is to keep individual census records—whether as digital or paper testimonies—safe and available for future use.

This Select Topics in International Censuses technical note provides NSOs with information on internationally recognized standards on special procedures and protocols for the archiving and preservation of individual census records.

Purpose of Census Archives

The essential purpose of archiving individual records, such as census questionnaires, is to keep them safe for future use by the government, researchers, and other interested parties.

When released, public historical records can be aggregated to produce inferential statistics; or analyzed individually, such as in the case of historical figures. Archived individual records have been used for diverse and sensitive purposes including identifying membership in aboriginal communities and for claiming pensions and social services (United Nations, 2016). To protect respondents' confidentiality, the public release of individual census records is subject to statutory timetables. The waiting period for public release usually takes several decades.

Figure 1 presents a copy of a population schedule (enumeration district 82-12 in Dearborn, Michigan) from the 1940 Census available at National Archives and Records Administration's (NARA) 1940 Census Web site.[2] The first person listed in this schedule is Henry Ford, founder of the Ford Motor Company. At the time the census was taken, Ford was 76 years old and was married to Clara Ford. They owned the house in which they lived, which had a reported value of $200,000 at the time, or roughly $3.7 million in today's U.S. dollars. Box 1 provides yet another example of how census records can be utilized to provide a public service.

[1] This technical note is part of a series on *Select Topics in International Censuses (STIC),* exploring matters of interest to the international statistical community. The U.S. Census Bureau helps countries improve their national statistical systems by engaging in capacity building to enhance statistical competencies in sustainable ways.

[2] https://1940census.archives.gov/.

United States® Census Bureau

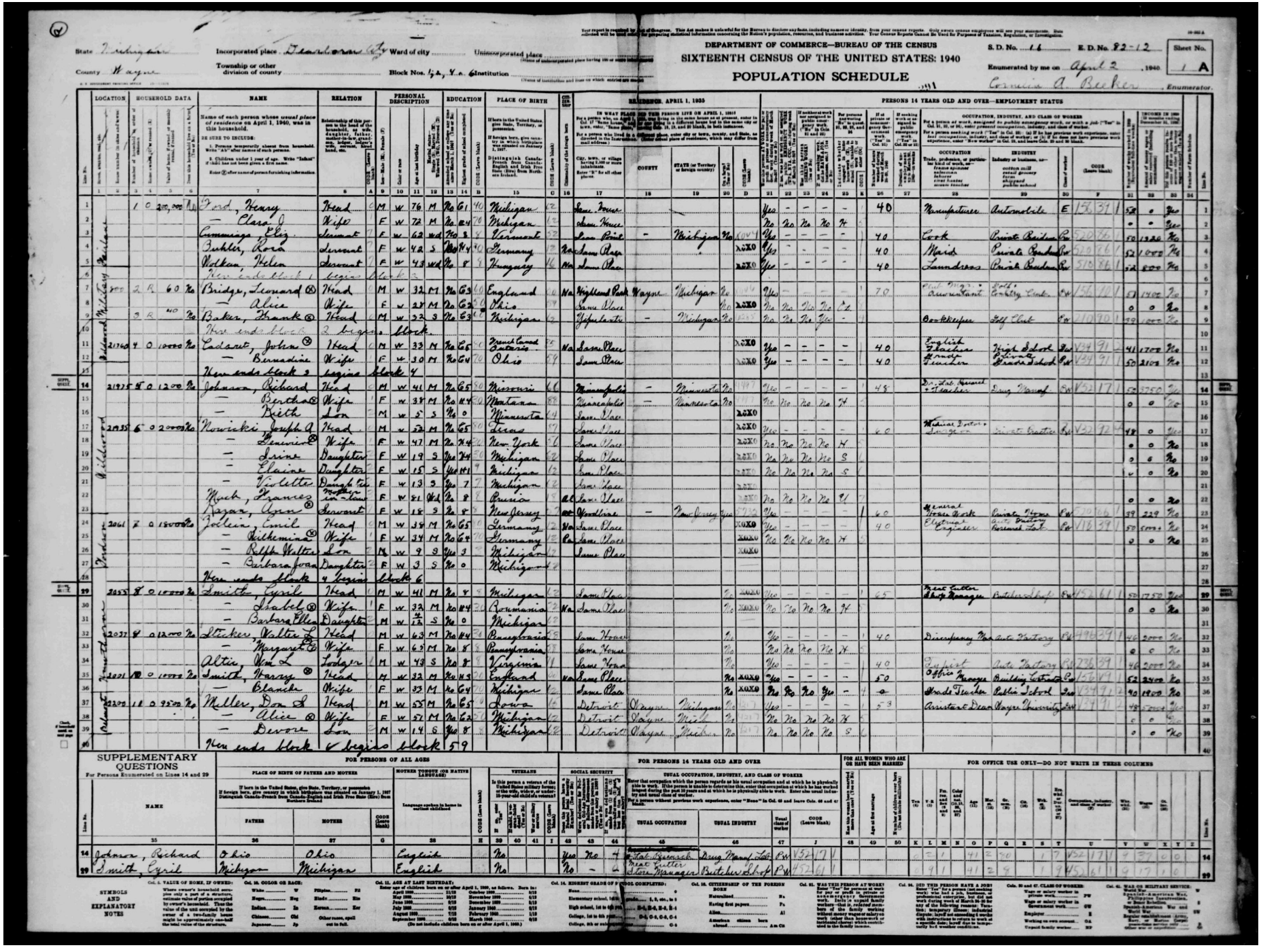
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

UNFPA

USAID
FROM THE AMERICAN PEOPLE

Figure 1.

Population Schedule for Michigan's Enumeration District 82-12: 1940 Census



Source: National Archives, 2018.

Box 1.

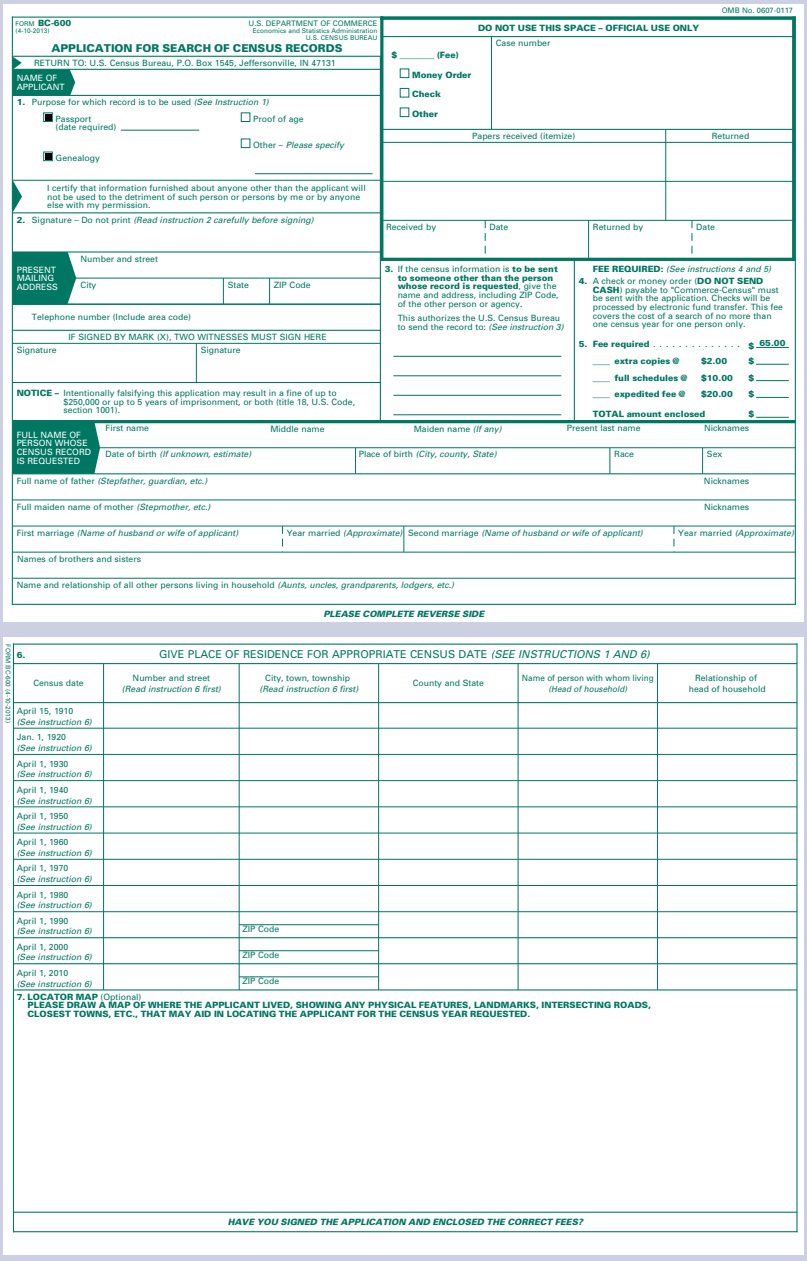
Search of Census Records

In the United States, copies of census records are often accepted as evidence to qualify for social security and other retirement benefits, in making passport applications, to prove relationship in inheritance claims, or to satisfy other situations where a birth or other certificate may be needed but is not available. Decennial census records are confidential for 72 years after "Census Day." Thus, the most recent individual census records that can be accessed publicly are those of the 1940 Census; records from the 1950 Census will be released on April 1, 2022.

The Census Bureau's National Processing Center in Jeffersonville, Indiana, maintains copies of the 1910 to 2010 census records. After the 72-year period has passed, NARA makes the records publicly available for viewing or purchase. Individual census records from 1910 through 2010 can be released upon request of the person named in the record or their heirs or guardians by submitting a form and paying a fee. Figure 2 represents a copy of Form BC-600—Application for Search of Census Records, used for this purpose.

Figure 2.

Form BC-600—Application for Search of Census Records



FORM **BC-600** (4-10-2013)

U.S. DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. CENSUS BUREAU

OMB No. 0607-0117

APPLICATION FOR SEARCH OF CENSUS RECORDS

RETURN TO: U.S. Census Bureau, P.O. Box 1545, Jeffersonville, IN 47131

DO NOT USE THIS SPACE – OFFICIAL USE ONLY

$ _____ (Fee) ☐ Money Order ☐ Check ☐ Other

Case number

Papers received (itemize) | Returned

Received by | Date | Returned by | Date

NAME OF APPLICANT

1. Purpose for which record is to be used *(See Instruction 1)*

■ Passport (date required) _____ ☐ Proof of age

■ Genealogy ☐ Other – *Please specify* _____

I certify that information furnished about anyone other than the applicant will not be used to the detriment of such person or persons by me or by anyone else with my permission.

2. Signature – Do not print *(Read instruction 2 carefully before signing)*

PRESENT MAILING ADDRESS — Number and street | City | State | ZIP Code

Telephone number (Include area code)

IF SIGNED BY MARK (X), TWO WITNESSES MUST SIGN HERE

Signature | Signature

NOTICE – Intentionally falsifying this application may result in a fine of up to $250,000 or up to 5 years of imprisonment, or both (title 18, U.S. Code, section 1001).

3. If the census information is **to be sent to someone other than the person whose record is requested**, give the name and address, including ZIP Code, of the other person or agency.

This authorizes the U.S. Census Bureau to send the record to: *(See instruction 3)*

FEE REQUIRED: *(See instructions 4 and 5)*

4. A check or money order (**DO NOT SEND CASH**) payable to "Commerce-Census" must be sent with the application. Checks will be processed by electronic fund transfer. This fee covers the cost of a search of no more than one census year for one person only.

5. **Fee required** . . . . . $ 65.00
___ **extra copies @** **$2.00** $_____
___ **full schedules @** **$10.00** $_____
___ **expedited fee @** **$20.00** $_____
TOTAL amount enclosed $_____

FULL NAME OF PERSON WHOSE CENSUS RECORD IS REQUESTED — First name | Middle name | Maiden name *(If any)* | Present last name | Nicknames

Date of birth *(If unknown, estimate)* | Place of birth *(City, county, State)* | Race | Sex

Full name of father *(Stepfather, guardian, etc.)* | Nicknames

Full maiden name of mother *(Stepmother, etc.)* | Nicknames

First marriage *(Name of husband or wife of applicant)* | Year married *(Approximate)* | Second marriage *(Name of husband or wife of applicant)* | Year married *(Approximate)*

Names of brothers and sisters

Name and relationship of all other persons living in household *(Aunts, uncles, grandparents, lodgers, etc.)*

PLEASE COMPLETE REVERSE SIDE

FORM BC-600 (4-10-2013)

6. GIVE PLACE OF RESIDENCE FOR APPROPRIATE CENSUS DATE *(SEE INSTRUCTIONS 1 AND 6)*

Census date	Number and street *(Read instruction 6 first)*	City, town, township *(Read instruction 6 first)*	County and State	Name of person with whom living *(Head of household)*	Relationship of head of household
April 15, 1910 *(See instruction 6)*					
Jan. 1, 1920 *(See instruction 6)*					
April 1, 1930 *(See instruction 6)*					
April 1, 1940 *(See instruction 6)*					
April 1, 1950 *(See instruction 6)*					
April 1, 1960 *(See instruction 6)*					
April 1, 1970 *(See instruction 6)*					
April 1, 1980 *(See instruction 6)*					
April 1, 1990 *(See instruction 6)*		ZIP Code			
April 1, 2000 *(See instruction 6)*		ZIP Code			
April 1, 2010 *(See instruction 6)*		ZIP Code			

7. **LOCATOR MAP** (Optional)
PLEASE DRAW A MAP OF WHERE THE APPLICANT LIVED, SHOWING ANY PHYSICAL FEATURES, LANDMARKS, INTERSECTING ROADS, CLOSEST TOWNS, ETC., THAT MAY AID IN LOCATING THE APPLICANT FOR THE CENSUS YEAR REQUESTED.

HAVE YOU SIGNED THE APPLICATION AND ENCLOSED THE CORRECT FEES?

Source: U.S. Census Bureau.

Census Archive Policy Considerations

Areas for policy development necessary to properly manage a census archive should include:

- Storage management policies for both paper and digital records.
- Policies for managing access control and maintaining confidentiality through format and documentation standards.
- Emergency preparedness policies, to limit damage from natural disasters or other emergencies.
- Disaster recovery policies, to provide mechanisms for duplicating digital content and storing it in separate physical facilities (with regular back-ups).
- Security policies to prevent accidental or malicious loss.

Archiving Individual Census Records

Individual census records refer to the original completed paper questionnaires or digital records on each enumerated person and household in a census, including the respondents' personal unique identifiers such as name, address, and date of birth.

A country's national strategy for archiving should include planning for three components: organizational infrastructure, technological infrastructure, and the human/economic resources required.

Organizational infrastructure refers to the policies within the NSO to ensure the efficiency of records' archival and retrieval processes. In most cases, these duties will be handled by a specific department within the NSO that is put exclusively in charge of archiving, storage, and eventual release of individual records. Typically, once a mandated time period for public release has passed, NSOs transfer census records to national libraries or a national archives institution—in the case of the United States, the NARA.

Technological infrastructure refers to the actual technology used for archiving. In the case of paper questionnaires, a physically secure structure is required, with regulated temperature and humidity, and a host of other requirements (see Box 2), including protection from fire hazards and extreme weather events. However, in most cases, questionnaires are scanned and stored electronically.

Technological infrastructure also refers to the development of an archiving and cross-referencing scheme that ensures the efficient storage and retrieval of records (with accompanying metadata and documentation). Technological infrastructure considerations include provision for production, archival, and dissemination of microdata data sets (census records with personal identifiers removed).

Finally, human and economic resources for archiving and retrieval of records should be planned in the context of the technological and organizational infrastructure of the NSO. In budgeting resources for archiving and retrieval of records, a long-term strategy should be adopted. There will always be a need to prepare for the next round of records release or for archiving newly collected records.

Archival Facilities

NSOs should ensure that appropriate arrangements for storage of paper and electronic records have been made. Storage should ensure security, maintenance, and accessibility of the records, including provisions for structural/environmental control and fire safety for storage facilities. Facilities should comply with national and local regulations. Ideally, a storage facility should meet with the minimum standards contained in Box 2.

These directives focus on physical records as the requirements for digital records depend on the technologies used. In addition to the measures in Box 2, long-term electronic records archival requires dedicated, stable storage space with environmental controls separate from those for the facility. Consideration should be given to server rack accessibility for maintenance, and to the susceptibility of digital media to damage from temperature, dust, power fluctuations, and magnets. In the United States, the Code of Federal Regulations (CFR) (U.S. Publishing Office, 2018) dictates the way in which records centers should be established, maintained, and operated. These regulations apply to all records storage facilities storing official federal records, whether federally owned and operated by NARA or another federal agency; federally owned and contractor operated; or privately owned commercial storage facilities that store government records. See Appendix A for more detail.

Box 2.
Minimum Standards for Storage Facilities

- General structural standards: facility must be designed in accordance with building codes to prevent building collapse or failure and constructed with noncombustible materials and building elements.
- Protection against water damage: facility must be sited away from flood zones; the roof membrane must not permit water to penetrate; piping must avoid leakage into the storage area.
- Heating, ventilation, and air conditioning (HVAC): system must provide sufficient air exchanges to maintain temperature, relative humidity, and pollutant control requirements.
- Finishing materials: use only permitted paint on walls, ceilings, and shelves.
- Lighting: ultraviolet (UV) radiation must be avoided as much as possible (sunlight is the greatest source of UV radiation).
- General fire safety: facilities must comply with local requirements. Archival facility must have an approved, supervised automatic smoke detection and automatic sprinklers system.
- Security (physical, network, computer systems): facility must comply to minimum security specifications to prevent unauthorized data access, modification, disclosure, or destruction.

Source: United Nations, 2016, pp. 375–376.

PRESERVATION

Physical Record Preservation

Paper records can deteriorate physically and chemically. While paper records naturally deteriorate, poor handling, storage, or display conditions accelerate this process. Ink or pencil on paper also deteriorate with time. While deterioration of paper media cannot be stopped, there are measures that can minimize deterioration. See Box 3 for some measures for preserving paper documents.

Digitization of Paper Documents for Electronic Access

Maintaining a paper-based retrieval and storage system for a national census is a monumental task with high monetary and nonmonetary costs. Such systems are both labor- and time-intensive, regardless of country size. By transforming paper documents into digital records, handling and storage costs can be greatly reduced. An additional advantage of digitizing paper documents is that original records are protected from deterioration or destruction. A better user experience combined with declining prices of digital

Box 3.
Preservation of Paper Documents

- Careful handling is the essential basic strategy for the preservation of paper files.
- Archival enclosures, such as boxes or folders, protect documents against mechanical damage, light, dust, and temperature and humidity fluctuations (detrimental to paper documents).
- Before placing files in protective packaging, ensure that they are free of dust, mold, insects, or active deterioration.
- Shelving and cabinets should be designed to minimize damage to stored items. Use powder coat painted metal shelves for paper records and plan cabinets for flat storage of maps or other large paper documents.
- Drawers should be clearly labelled so that items may be retrieved with a minimum of handling.
- Keep the storage area clean to reduce the chance that pests will be attracted to these areas.
- Photocopying and digital scanning can be used to preserve a high-resolution copy of the information on a fragile or deteriorating record, as an access copy to preserve a heavily used original, or to exhibit a copy and preserve the original record in storage.

Source: United Nations, 2016, pp. 374–375.

storage have made the use of this type of technology increasingly attractive for NSOs (United Nations, 2016).

When digitizing paper records, it is crucial to also digitize all accompanying documentation such as code books and technical documents. As a number of variables are stored as code, data are useless if technical documentation is not available. Along with the digitized copies of paper records themselves, digital preservation should include storage of metadata (including those forms listed below), records of institutional knowledge from the census including successes and areas of challenge, and all data sets including original, semi-edited, and fully edited records.

Metadata is generally categorized into four or five groupings based on the information the metadata captures, as seen in the list below.

- Descriptive metadata: information on the intellectual content of a resource.
- Administrative metadata: information about management including ownership and rights management.
- Structural metadata: information for displaying and navigating resources, focused on relationships between multiple files.
- Technical metadata: information on the technical features of a digital file.
- Preservation metadata: information facilitating management and access over time.

When considering digitization—at the managerial level—administrators should take into account the cost and benefits of digitizing records; disposition of original paper documents; and planning for technical issues such as the selection of a systems architecture.

Digital Record Preservation

In the United States, typical standards for electronic records management—such as digitized census records—are covered by the U.S. Department of Defense 5015.2 Standard (U.S. Department of Defense, 2007) and the U.S. Code of Federal Regulations (U.S. Publishing Office, 2013). These guides provide detailed and actionable guidance for anyone designing organization electronic records management systems.

While there is some overlap between these terms, the primary approaches to long-term digital record preservation are migration, encapsulation, emulation, normalization, and conversion.

- Migration or "media refreshing": a shift in computer platform, storage medium, or physical format.
- Encapsulation: preservation of the original files including the full set of technologies needed to access those records.
- Emulation: digitally mimicking the functionality of older or obsolete technologies to enable modern computers to read older, obsolete file formats.
- Normalization: limiting formats for preservation to a small set of common formats that are as software-independent as possible. By avoiding proprietary software formats and standardizing to selected well-maintained, generally usable formats, risks of data becoming inaccessible are reduced.
- Conversion: associated with normalization, conversion is a shift in file format, usually to an open/standard format to avoid loss of backward compatibility.

A combination of these approaches is usually adopted, however strategies most often consist of a combination of migration and conversion, with records being transferred to modern platforms, media, and file formats purposively as part of a rigorously planned digital retention strategy.

In planning preservation strategies, an NSO will make decisions based on the form of storage media to use (magnetic, optical, solid state drives, or dedicated repository software) and on the file formats to use. It is recommended to create multiple copies of each record using multiple media types. The current general consensus in digital preservation is a move towards using dedicated records management systems and repository software with embedded filing and search/retrieval functionality. For sensitive census records, local-area-network and online-based strategies are risky, as they provide an avenue for illicit access leading to data breach risks.

Comprehensive planning should include evaluation of archive contents and recommendations for treating current archive holdings, developing archive standards and policies, providing periodic risk analysis reports, and monitoring technological changes, as well as changes in the user community's service requirements (United Nations, 2016). Monitoring technology is a crucial part of planning. Emerging digital technologies, information standards, and computing platforms should be tracked to identify technologies, which could cause obsolescence in the archive's computing environment.

Clearly defined retention and disposal schedules should be set which establish standards for each of the following:

- When and how records will be assessed for possible updates, conversion, or migration.
- Annual sampling of 3 percent of working media, back-ups, and indexes to check data integrity.
- Establish internal and external naming and indexing conventions.
- Establish data integrity processes with clearly defined tolerances for every time data are transferred or copied.

CONCLUSION

Safely storing census records and setting methods for their retrieval are among the most important functions that NSOs are tasked with. Doing so in a secure and cost-effective manner can be challenging. Extensive planning must be done on a periodic basis in order to be successful. Planning should account for budgetary, technological, and legal constraints. This note summarizes the practices and recommendations of the U.S. Government and the Census Bureau on archiving and preserving census data.

REFERENCES

Inter-university Consortium for Political and Social Research, *Principles and Good Practice for Preserving Data*, IHSN Working Paper No. 003, International Household Survey Network, 2009.

Minnesota Historical Society, *Electronic Records Management Guidelines,* Version 5, Saint Paul, 2002.

Royal Statistical Society & the UK Data Archive, *Preserving & Sharing Statistical Material, Working Group Report on the Preservation and Sharing of Statistical Material: Information for data producers,* Essex, 2002.

United Nations, *Handbook on the Management of Population and Housing Censuses*, Revision 2, New York, 2016.

United States Census Bureau, "Age Search Service," <www.census.gov/topics/population/genealogy/agesearch.html>, accessed on October 31, 2018.

_____, "Decennial Census Records," <www.census.gov/history/www/genealogy/decennial_census_records/>, accessed on October 31, 2018.

_____, "The 72-Year Rule," <www.census.gov/history/www/genealogy/decennial_census_records/the_72_year_rule_1.html>, accessed on October 31, 2018.

United States Department of Defense, *Electronic Records Management Software Applications Design Criteria Standard*, DoD 5015.02-STD, Version 3, Washington, DC, 2007.

United States National Archives, "Records Storage Standards Toolkit," <www.archives.gov/records-mgmt/storage-standards-toolkit>, accessed October 31, 2018.

_____, "Search the 1940 Census," <https://1940census.archives.gov/search/?search.state=MI&search.enumeration_district=82-12#filename=m-t0627-01825-00587.tif&name=82-12&type=image&state=MI&index=1&pages=40&bm_all_text=Bookmark>, accessed on October 31, 2018.

United States Publishing Office, "Records Management and Preservation Considerations for Designing and Implementing Electronic Information Systems Facility Standards for Records Storage Facilities," Code of Federal Regulations, Title 35, Vol. 3, part 1236.10, Washington, DC, 2013.

_____, "Facility Standards for Records Storage Facilities," Code of Federal Regulations, Title 36, Vol. 3, part 1234, Washington, DC, 2018.

Appendix A.

Facility Standards for Records Storage Facilities

According to the U.S. Code of Federal Regulations (USPO, 2018), records storage facilities should comply with a minimum of safety standards. These are:

- The facility must be constructed with noncombustible materials and building elements, including walls, columns, and floors. Roof elements may be exempted from this requirement if they are protected by a wet-pipe automatic sprinkler system.
- A facility with two or more stories must be designed or reviewed by a licensed fire protection engineer and civil/structural engineer to avoid catastrophic failure of the structure due to an uncontrolled fire on one of the intermediate floor levels.
- The building must be sited a minimum of 5 feet above and 100 feet from any 100-year floodplain areas or be protected by an appropriate flood wall.[3]
- The facility must be meet applicable national, regional, state, or local building codes (whichever is most stringent) to provide protection from building collapse or failure of essential equipment from earthquake hazards, tornadoes, hurricanes, and other potential natural disasters.
- Roads, fire lanes, and parking areas must permit unrestricted access for emergency vehicles.
- A floor load limit must be established for the records storage area by a structural engineer.
- The facility must ensure that the roof membrane does not permit water to penetrate the roof. NARA strongly recommends that this requirement be met by not mounting equipment on the roof and placing nothing else on the roof that may cause damage to the roof membrane.
- Piping (with the exception of fire sprinkler and roof drainage piping) must not be run through records storage areas.
- The records storage facility must be equipped with an anti-intrusion alarm system to protect against unlawful entry and to monitor interior storage spaces at all times.
- Records infiltrated by insects or exhibiting active mold growth must be stored in separate areas having separate air handling systems from other records.
- To eliminate damage to records or loss of information due to insects, rodents, mold, and other pests; the facility must have an integrated pest management program. This program should emphasize prevention, least-toxic methods, and should be based on a systems approach.[4]
- Do not install mechanical equipment, excluding material handling and conveyance equipment that has operating thermal breakers on a motor rated in excess of 1 HP within records storage areas.
- Do not install high-voltage electrical distribution equipment (i.e., 13.2kv or higher switchgear and transformers) within records storage areas.
- A redundant source of primary electric service, such as a second primary service feeder, should be provided to ensure continuous, dependable service to the facility especially to the HVAC systems, fire alarm, and fire protection systems.
- A facility storing permanent records must be kept under positive air pressure, especially in the area of the loading dock. In addition, to prevent fumes from vehicle exhausts from entering the facility, air intake louvers must not be located in the area of the loading dock, adjacent to parking areas, or in any location where a vehicle engine may be running for any period of time. Loading docks must have an air supply and exhaust system that is separate from the remainder of the facility.

In regards to records storage shelving and racking systems, the U.S. Code of Federal Regulations dictates that:

- All storage shelving and racking systems must be designed and installed to provide seismic bracing that meets state, regional, or local building code (whichever is most stringent).
- Racking systems, steel shelving, or other open-shelf records storage equipment must be braced to prevent collapse under full load. Each racking system or shelving unit must be industrial style shelving rated at least 50 pounds per cubic foot supported by the shelf.
- Compact mobile shelving systems (if used) must permit proper air circulation and fire protection.

[3] A 100-year floodplain is the land that is predicted to flood during a "100-year storm," which is a hypothetical storm that has a 1 percent chance of occurring every given year.

[4] The program must be effectively coordinated with all other relevant programs that operate in and around the building.

Appendix B.

Records Management and Preservation Considerations for Designing and Implementing Electronic Information Systems Facility Standards for Records Storage Facilities

The U.S. Code of Federal Regulations (USPO, 2013) dictates that electronic information systems that host federal records should include the following types of records management controls:

- Reliability: controls to ensure a full and accurate representation of the transactions, activities, or facts and that can be depended upon in the course of subsequent transactions or activities.
- Authenticity: controls to protect against unauthorized addition, deletion, alteration, use, and concealment.
- Integrity: controls, such as audit trails, to ensure records are complete and unaltered.
- Usability: mechanisms to ensure records can be located, retrieved, presented, and interpreted.
- Content: mechanisms to preserve the information contained within the record itself that was produced by the creator of the record.
- Context: mechanisms to implement cross-references to related records that show the organizational, functional, and operational circumstances about the record.
- Structure: controls to ensure the maintenance of the physical and logical format of the records and the relationships between the data elements.