

IPUMS Digital Preservation Policy Framework

May 2022

1. Purpose

IPUMS democratizes access to the world's social and economic data for current and future generations. We enable transformative scholarship, teaching and policy making by integrating data and documentation across time and space and by streamlining data access through innovative dissemination technology.

The original IPUMS project began disseminating U.S. census data in 1993, with the goal of making existing U.S. census microdata interoperable across time. IPUMS is currently the home of several data collections, described on [the IPUMS website](#). This framework draws upon relevant IPUMS documentation.

The Digital Preservation Policy Framework is the highest level digital preservation policy document for IPUMS. The audience for the framework includes IPUMS data users, primarily academic researchers, policy analysts, and students, as well as IPUMS staff and partners, data producers, and funders.

2. Standards Compliance

This framework aligns with the de facto digital preservation community standard, *Trusted Digital Repositories: Attributes and Responsibilities*. The following sections correspond to the attributes of a trusted digital repository. The *Open Archival Information System (OAIS) Reference Model* (ISO 14721: 2012) is the second core digital preservation standard. IPUMS aligns its digital preservation program with the OAIS Reference Model, providing a sound technological foundation for the program. [IPUMS Workflows](#) documents our conformance to these standards.

3. Administrative Responsibility

IPUMS is part of the [Institute for Social Research and Data Innovation](#) (ISRDI), a University-wide research center that reports to the [Office of the Vice President for Research](#) (OVPR) at the University of Minnesota. This framework makes explicit IPUMS' commitment to preserve its digital assets through a sustainable digital preservation program that is conformant with the prevailing standards and practice of the digital preservation community.

3.1. Mandate

IPUMS is responsible to our funders and to current and future scholars to preserve access to the data and metadata associated with these data collections.

- *Grants and contractual obligation*: As part of the grants and contracts that fund these projects, we have committed to retaining access, usually for 5 to 10 years past the life of the grant. In some cases, we also commit to longer-term preservation of both the IPUMS data and the source materials.
- *Scholarly commitment*: Our commitment goes beyond our grant promises. We understand that the resources IPUMS has created need to be accessible long into the future. We have established this framework to support preservation and to allow ongoing access to the data for new research, for scholarly replication, and for unanticipated uses of the data.

3.2 Objectives

Our goal is to preserve the data and metadata from each IPUMS collection, encompassing the following objectives:

- Meet the needs of the user community by providing access to IPUMS collections, including previous versions of the data.
- Serve IPUMS internal needs by retaining data and metadata for the purpose of provenance and quality assurance.
- Meet archival requirements of funding agencies.
- Demonstrate compliance with data, metadata, and archival standards, including disciplinary standards of data users and the digital preservation community.
- Maintain a digital preservation program that is responsive to the changing technological environment.

4. Organizational Viability

IPUMS has an established track record for responsible and sustained data management. The first IPUMS data (now known as IPUMS USA) were released in 1993. IPUMS data, now more than ten data collections, have been available continuously since that time.

4.1 Scope

IPUMS builds its collection by securing funding for projects to integrate and expand existing data that can contribute to the IPUMS mission. Each project identifies data appropriate for inclusion in the infrastructure. IPUMS is committed to building, managing, preserving, and making available its data collections.

4.2 Operating Principles

IPUMS has adopted and addressed the [Ten Principles for Digital Preservation Repositories](#). We are committed to complying with OAIS and other digital preservation standards and practice, and to developing a sustainable and auditable digital preservation repository. See also 8.1 Audit and Transparency.

4.3 Roles and Responsibilities

IPUMS manages digital preservation as a shared responsibility, including contributions from IPUMS staff throughout the life cycle of the data, from identification through acquisition, preservation, and long-term discovery and access. [IPUMS Organization](#) outlines how IPUMS projects coordinate with each other and with related organizations.

4.4 Selection and Acquisition

All IPUMS data collection and infrastructure development is initiated by grants and contracts. The grant proposal and any subsequent negotiations with the funding agency outline the specifics of each project to develop or enhance an IPUMS data collection. Many of these projects have an advisory board to provide advice on priorities, including data selection and acquisition. The [IPUMS website](#) describes the data collections for existing IPUMS infrastructure.

4.5 Access and Use

IPUMS provides usable, integrated, and comprehensive documentation for all data collections. IPUMS investigates and documents comparability issues and provides guides on the use of complex variables and advice for avoiding analytic pitfalls. Most users download data directly in the format of their statistical software or in a format that is easily converted by their software. Online tabulation software is opening IPUMS resources to researchers and students who do not have access to or training in statistical software, broadening the potential for IPUMS resources for users with a diverse range of backgrounds and circumstances. IPUMS provides dissemination and outreach services for all data collections, offering tutorials, workshops, forums, and email support to help users download, analyze and understand the data.

IPUMS resources are most frequently used by researchers in economics, sociology, demography, epidemiology, and geography, but also we also have users in a diverse range of fields, including anthropology, computer science, nursing, political science, psychology, and neuroscience.

4.6 Challenges and Risks

IPUMS actively manages its data collections and programs to identify challenges and to anticipate, detect, and mitigate risks. As an organization that is engaged in digital preservation, IPUMS is responsive to continually changing technology, resulting in the emergence of new digital content types, the need for new skills, and potential preservation risks to new and existing digital content. Therefore, IPUMS is committed to providing appropriate training for staff who manage data and to supporting the requirements of data users, as needs and requirements evolve.

5. Financial Sustainability

IPUMS is supported by externally sponsored research grants and contracts and by University of Minnesota funds provided by the Office of the Vice President for Research (OVPR). University funds are governed by University of Minnesota policies and all sponsored funds must adhere to agency guidelines. Most IPUMS grants are from the federal government and therefore follow regulations stipulated by OMB Uniform Guidance, which replaced requirements of OMB Circular A-21, A-133 and A-110.

5.1 Institutional Commitment

Sponsored revenue has been consistently growing since IPUMS was established in 2000. As part of the institutional commitment to IPUMS, the Vice President for Research and ISRDI established a \$1 million endowment fund. In the unlikely event that funding for IPUMS ends without any additional revenue sources, the University of Minnesota is ultimately responsible for the orderly dissolution of the center and appropriate archiving of the data collections and associated materials.

5.2 Cooperation and Collaboration

IPUMS has two types of collaborating partners: data producers that provide data for IPUMS infrastructure and organizations who collaborate with IPUMS on other aspects of data infrastructure development. Data producers include national statistical agencies, genealogical companies, data archives, and academic researchers (see more at [IPUMS Data Producers](#)). Some data producers prefer to restrict usage to scholarly research and educational purposes only, but other producers have no restrictions. Several IPUMS collections are the result of collaborations with other organizations, such as IPUMS Time Use (University of Maryland), IPUMS DHS (the Demographic and Health Surveys Program), IPUMS MICS (Multiple Indicator Cluster Surveys, UNICEF), and IPUMS PMA (Performance Monitoring for Action, the Bill & Melinda Gates Foundation and Johns Hopkins University).

6. Technological and Procedural Suitability

The ISRDI Information Technology (IT) Team provides cutting-edge software development and computing support for IPUMS data projects, including for the digital preservation program. The large scale of IPUMS data collections requires sophisticated infrastructure for software development and data processing. IPUMS software has well-documented code, supplemental documentation on a dedicated wiki, and intentionally pervasive institutional knowledge about how the systems work. Information about servers is captured and accessible. The Operations Team partners with other IT organizations on campus to maintain the hardware and software to support IPUMS members and staff, and to enable the digital preservation program. IPUMS has

effective disaster planning in place, including data archiving protocols and copies in multiple geographic locations to ensure that archived data are kept permanently.

IPUMS infrastructure projects share many common workflows and processes that are well-defined and well-documented. Processes for validation of incoming data and cleaning, transformation, and creation of the final data product are managed by each project. Metadata are explicitly and intentionally managed in accordance with a full range of standards and practices for data management and for digital preservation. See [IPUMS Workflows](#).

7. System Security

The technical environment maintained by IPUMS ensures controlled and appropriate access to its data collections and services (see [IPUMS IT Security Plan](#)). IPUMS data are freely available to registered users (see [Terms of Use](#)). IPUMS policy is to keep data access as open as possible, and all data collections are available with the least restrictions possible. Some datasets have restrictions that are imposed by data providers or partner organizations. User registration is unique to each dataset. Unrestricted data are distributed free of charge with the conditions that users cite the data appropriately and that any redistribution of the data is limited to non-commercial applications and is approved by IPUMS staff. Permission for commercial redistribution is required. User registration for unrestricted data is automatic, and users gain access to restricted data files once their application is approved. Metadata for all data collections are open to all users.

8. Procedural Accountability

IPUMS is committed to building and maintaining the documentation that is required to demonstrate good digital preservation practice. This documentation will serve as a baseline against which future evaluation is measured.

8.1 Audit and Transparency

IPUMS commits to periodic self-assessments and audits to evaluate, measure, and adjust the procedures, preservation approaches, and practices of the digital curation and preservation program. The self-assessment and audit program for IPUMS is undertaken in accordance with the *CoreTrustSeal* and the *Ten Principles*, which build on the *Attributes of a Trusted Digital Repository: Roles and Responsibilities* and the *Open Archival Information System (OAIS) Reference Model*.

8.2 Framework Administration

This document, Version 1 of the IPUMS Digital Preservation Framework, was developed by Catherine Fitch, Tracy Kugler and Wendy Thomas. Preliminary versions were developed in

consultation with Nancy McGovern. IPUMS will maintain the document, reviewing it every three years as part of the CoreTrustSeal review, and adjusting it as needed.

8.3 Definitions

DDI Glossary: IPUMS adopted the terminology used by the DDI Alliance available at: <https://ddialliance.org/resources/ddi-glossary>

OAIS Reference Model definitions: IPUMS adopted the June 2012 version of OAIS definitions, pages 1-8-1-16. <http://public.ccsds.org/publications/archive/650x0m2.pdf>.

8.4 References

Audit and Certification of Trustworthy Digital Repositories, September 2011 version available at: <http://public.ccsds.org/publications/archive/652x0m1.pdf>.

CoreTrustSeal, information on requirements and certification process available at: <https://www.coretrustseal.org/>.

Data Documentation Alliance (DDI), current versions of codebook and specifications available at: <https://ddialliance.org/> (see Products).

Open Archival Information System (OAIS) Reference Model, the June 2012 version is available at: <http://public.ccsds.org/publications/archive/650x0m2.pdf>.

Ten Principles, 2007 version of core criteria for digital preservation repositories available at: <https://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/core-re>

Trusted Digital Repositories: Attributes and Responsibilities, May 2002, available at: <https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>.