Expanding our perspective: Building a sustainable metadata culture

Diana L. Magnuson[1] and Wendy L. Thomas[2]

Abstract

The Institute for Social Research and Data Innovation (ISRDI) at the University of Minnesota submitted an application for approval to the Core Trust Seal (CTS) in June 2022. In the course of the protracted process of preparing ISRDI materials for the application, we learned five lessons that expanded our perspective on the role of the archive within our organization and committed the Institute to building a sustainable metadata culture. After reviewing the specialized nature of ISRDI as it developed over time, clarifying and documenting the processes that developed as the Institute matured and expanded, and applying the standards and guidelines supported by the CTS, ISDRI staff are now better positioned to identify areas of future process development and to address outstanding needs for documenting and preserving the Institute's work. These lessons are applicable to research organizations responsible for preserving a record of their work in the mid- and long-term.

Keywords

Core Trust Seal, archive, metadata, business process model, preservation

Introduction

In June 2022, the Institute for Social Research and Data Innovation (ISRDI) submitted an application to the Core Trust Seal (CTS) for its IPUMS projects.[3] Application to the CTS for its professional approval culminates years of effort within the Institute for ensuring access to our signature collection of harmonized census and survey data from around the world. In the course of this protracted work, ISRDI learned five valuable lessons for data archivists organizationally positioned in a larger institutional context:

- **Institutional History** - Situating our institutional history in a larger social science context sharpened our understanding of our unique contribution to social science infrastructure and to data archiving.
- **Building the CTS Application** - Building our CTS application clarified our institutional strengths and illuminated areas to refine.
- **Business Process Model** - Developing a business process model documented roles and responsibilities of organizational components (project, administration, and archive) and highlighted metadata production and curation points.
- **Leveraging Documentation** - Leveraging documentation produced for the CTS application will support future funding applications, enumerate data archive responsibilities, identify cross-project technical systems, educate current staff, and facilitate onboarding new employees.
- **Preservation** - Preserving our data products and unique intellectual property relating to the processing and methodology that contributed to the development of our data products is an essential contribution to social science infrastructure.

These lessons are shaping the way we internally conduct our data archival work, externally relate to our funders and data collaborators, and prepare for future data harmonization projects. We believe our experience can be a guide for other organizations aiming to build a sustainable metadata culture. This paper presents the value of the CTS review and submission process in helping a non-traditional

 Magnuson, Diana L; Thomas, Wendy L. (2023) Expanding our perspective: Building a sustainable metadata culture, IASSIST Quarterly 47(2), pp. 1-11. DOI: https://doi.org/10.29173/iq1046

archive define its place within a research organization and clarifies the archive's role in supporting the standing of its parent organization with funders, data providers, and the research community.

Institutional history

While producing an application to submit to the CTS we came to appreciate the value of reflecting on our institutional history and situating that history in the larger context of data archiving and social science infrastructure. This intellectual exercise sharpened our understanding of our unique story and contribution to social science. This is particularly important within non-traditional archives where the focus of the organization may be research or providing a specialized data product. The review process forces the archive to clearly explain its role and archival activities within the larger organization.

Over the last thirty years, IPUMS has created the world's largest accessible database of census microdata. The Institute for Social Research and Data Innovation and its flagship data project, IPUMS, has its roots in the 1880 Historical Census Project, a Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) funded project to create a 1-in-100 public use microdata sample of the 1880 U.S. census of population. Housed in the History Department at the University of Minnesota, historical demographers and co-principal investigators Steven Ruggles and Russell Menard conceived of extending back the series of public-use microdata samples already in existence (1900, 1910, 1940, 1950, 1960, 1970) (Magnuson and Ruggles, 2022). Once the completed 1880 PUMS was disseminated, researcher feedback was overwhelmingly enthusiastic.[4] Funding to complete the decennial population series--1850-1870 and 1920-1930 and updates to 1900 and 1910--would come to the University of Minnesota between 1992 and 2002 (Magnuson and Ruggles, 2022).[5] By 1991 ten machine readable public use microdata samples covering the decennial censuses of population from 1880 and 1990 were publicly available or under development (for 1850, 1880, 1900, 1910, 1940, 1950, 1960, 1970, 1980, and 1990). The nagging problem facing the research community was the difficulty of using the data as a time series because the various datasets were created at different times, by different investigators, employing different formats, record layouts, coding schemes, and producing different documentation.

Between 1985 and 1991, Steven Ruggles "developed a set of FORTRAN programs that recoded selected variables into a common format across the available census samples, created subsets of the samples that were of manageable size, and pooled multiple censuses into a single file" (Magnuson and Ruggles, 2022). Initially Ruggles used a "lowest common denominator" approach for variable codes when combining samples, which naturally resulted in significant loss of information. Despite these limitations, the program could be customized to meet the requirements of any research question. Demand for customized data sets steadily increased in-house at the University of Minnesota, as well as coming from a few researchers at other universities. Clearly there was a user base for time series microdata if the compatibility issues could be resolved.

In 1991, Steven Ruggles was awarded a National Science Foundation grant to create a single integrated series that would "maximize comparability and minimize information loss" (Ruggles 1992-95; 1991a, 1991b). He proposed to name the finished product the "Integrated Public Use Microdata Series," and thus IPUMS was born.[6] Key technical innovations emerging from IPUMS included the first structured metadata system for data integration and the first interactive web-based system for user-customized data extraction. In 1993, IPUMS data were disseminated through an anonymous file transfer protocol (FTP) site and two years later the IPUMS website launched its own data extract system. Hypertext variable-level documentation became available in 1997 (Magnuson and Ruggles, 2022).

In 1999, the University of Minnesota Graduate School issued a call for competitive applications to receive funding for interdisciplinary centers. Collaborators from units representing geography,

history, public affairs, industrial relations, and health services successfully made their case to the University for establishing an interdisciplinary population center. Two smaller population centers merged to become one, and the Minnesota Population Center (MPC) thus emerged from a "strategic positioning process" that sought to prioritize and foster highly collaborative and interdisciplinary activities at the University (Magnuson 2015, Lawrenz and Paller 2006). Beginning in 2000, the MPC was a university-wide interdisciplinary cooperative for demographic research at the University of Minnesota. The Center had three main goals: "to foster connections among population researchers across disciplines, to develop large-scale collaborative research projects, and to provide infrastructure for demographic research" (Ruggles 2011).

In 2016, the MPC was reorganized in recognition of the diverse development of population research infrastructure at the University of Minnesota. The ISRDI became the parent organization of four centers: the MPC, IPUMS, the Life Course Center, and the Minnesota Research Data Center.[7] IPUMS separated from the MPC to become a co-equal center within the newly constituted Institute.

Over the course of our thirty-year history, the institutional entities that produce and disseminate ground-breaking IPUMS data products and technological innovations have formed an integral part of social science infrastructure as we know it today. At present, the IPUMS suite of products contain nine harmonized data collections.[8] Data comes from the United Nations Statistical Division, the United States Census Bureau International Division, and over 100 national and regional statistical organizations.

Building the CTS application

The time-consuming undertaking of building IPUMS policy documentation to complete the CTS application clarified our institutional strengths and identified areas to refine. The CTS process is intended to be on-going, including continuing to support and implement policies and processes, refine activities as needed, and to respond to changes in the environment over time. To do this the archive needs to obtain the initial buy-in of the parent organization as well as maintain on-going support for archival work. Recognizing our institutional strengths assure the parent organization that the goal of the archive is to improve and support the organization through its work.

2016 was a watershed year for the MPC as it reorganized into the ISRDI. As a co-equal entity within ISRDI, IPUMS clarified its mission in terms of data harmonization, access, curation, and preservation. This included expanding the use of metadata standards such as Dublin Core, DDI, and ISO-19115 in describing our data for archival purposes. At the same time, external funding organizations were increasing requirements for adherence to standard archival practice using the open archival information system (OAIS) model and digital object identifiers (DOI).[9] To address these external concerns, we began an internal assessment of our data products, metadata, and archival practices with respect to those standards. For microdata projects, variable definitions, source data for harmonization, data collection forms, collection instructions, and sampling information were relevant. For aggregate data, the table and dimension descriptions, data source, universe, geographic definitions, and imputation information were important to capture and preserve.[10] Our internal assessment revealed that we captured an extensive amount of metadata, but we did not capture changes to the metadata over time in a structured way. Developing clear guidelines regarding why and how we would be assigning DOI's, requirements for a versioning policy for each project, and capturing preservation copies for the archive that met OAIS standards, were the initial points of discussion. The needs of each project were reviewed and commonalities were documented. Communicating these issues and concerns across projects and administrative units was a challenging

but important part of the assessment process. Ultimately, this process began to nurture a sustainable metadata culture within our organization.

After a roughly three-year internal assessment, the decision to adopt the practice of using digital object identifiers (DOI) was made in 2016. DataCite (datacite.org) was selected as the University of Minnesota was already a member. Registering DOIs with DataCite required decision points around the following tasks: determining at what level to assign a DOI; capturing data and metadata for specific versions of our data products; providing persistent access to each identified version of our data products; and maintaining and providing access to those versions over time. The discussion of these issues was done in an iterative fashion and involved input from all of the IPUMS project groups. Our goal was to establish clear versioning rules around our data products while allowing each project the flexibility to decide when in their project workflow a version was triggered. Once guidelines were established, adhering to these requirements had a number of important internal payoffs. First, the digital object identifiers were persistent and unique. Unique, persistent identifiers provided support for our users to be able to accurately reference the data obtained from the IPUMS system. Second, references and related publications became trackable for our internal processes. Use of those DOIs by researchers made it much easier for IPUMS to track the research based on our products and to use this information in applying for continued funding for IPUMS projects. Third and most obviously, our data and metadata were captured and preserved, making our preservation work more accurate and complete. Perhaps the greatest pay off was the recognition by IPUMS project managers of the need to capture and retain the content of each product version in a format that could be preserved and accessed for the purpose of research replication. Finally, these developments motivated our organization to apply for the Data Seal of Approval (now Core Trust Seal).[11] While initially IPUMS focused on the value of the first two points, the preservation and access requirements of obtaining a DOI brought the role of the archive into clearer focus within the organization.

As we dug into the CTS application process, we quickly recognized that our policy documentation was scattered and incomplete, a byproduct of rapid institutional growth from 1991 to 2016. Pulling existing materials together, assessing policy documentation that needed to be updated, and crafting new documentation to reflect practices already in place, was time consuming but necessary to document our workflow.

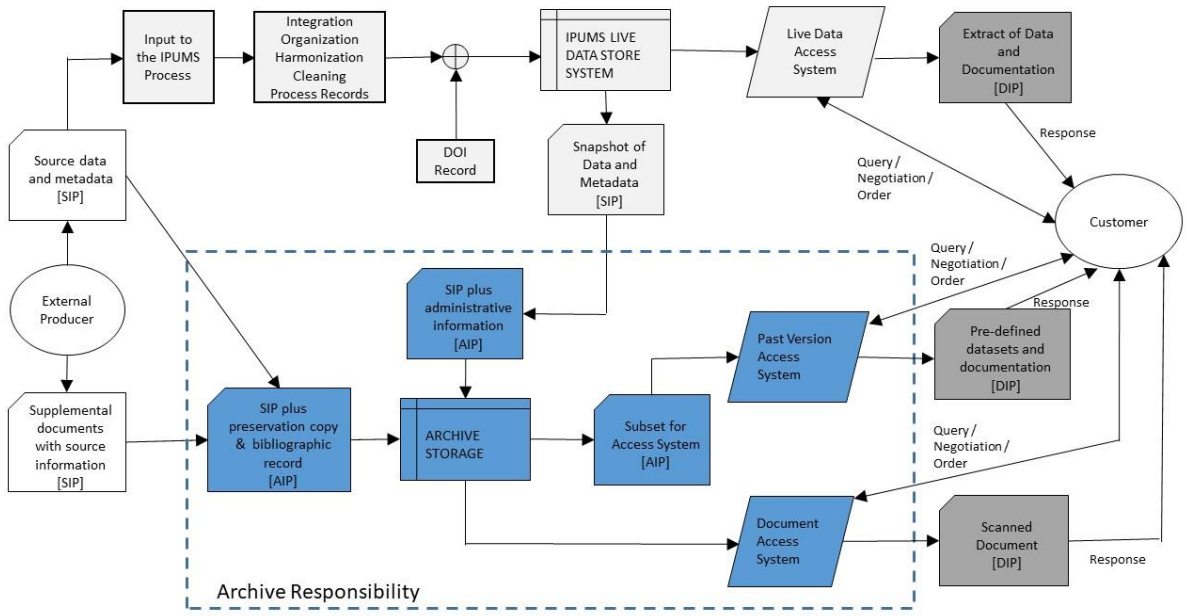
Business process model

Developing an organizational model that clearly and accurately reflected the workflow of our data projects and archival processes was a crucial step in developing our CTS application materials. The process clarified the role of the archive within the organization and provided us with the means to clearly present that role and identify specific touchpoints to IPUMS project workflows. It also encouraged IPUMS project managers to look at the commonalities of their work across IPUMS projects, rather than just the uniqueness of their specific project.

The CTS application process requires applicants to describe their archival responsibilities within their organization using an OAIS model.[12] Using the OAIS model, we worked to identify where our archive obtained submissions (both external and internal), what actions we took once we obtained those submissions, and how we delivered the products to users. These information packages were integrated into the IPUMS business process model (Figure 1) to clarify where the OAIS information packages originated and how they moved through the process model from external and project sources, to the archive for management and user access. The OAIS model helped us to establish an expanded workflow model of the collection, harmonization, and publication work done within the various IPUMS projects, and importantly, align that workflow with the role of the archive. The new workflow model made clear how and where the archive interacted with the projects in terms of

submitting data to the archive, packaging that data for persistent access, and delivering archived data to users once a dataset is replaced in the IPUMS live data access system by a new version.

Figure 1. IPUMS implementation of the OAIS Model



The OAIS model provided only a general model, however, and we soon determined that it was not flexible enough for our detailed, project-specific workflows; IPUMS is not a standard archive and thus the OAIS model could not reflect the full range of our activities. "The primary activities of IPUMS focus on acquiring data from an external producer, processing the data and related metadata to integrate it for the purposes of comparative research, providing a means of access to facilitate that research, and then delivering customized packages of data and metadata to the consumer."[13]

To identify the commonalities between the processes of individual IPUMS projects while allowing for differences in the selection and ordering of tasks within each project over time, Data Curator Wendy Thomas drew on two business process models, the Generic Statistical Business Process Model (GSBPM) and the Generic Longitudinal Business Process Model (GLBPM), to serve as templates in the creation of the IPUMS Business Process Model (IPUMS BPM). The GSBPM was designed to model a standard framework for statistical organizations that could be used modernize statistical products through the use of harmonized language and shared methodologies.[14] The GLBPM is a modification of the GSBPM, developed to focus "on the longitudinal survey process as employed in longitudinal data gathering by academic, governmental, and private research organizations."[15] The IPUMS Business Process Model is a customization of the GSBPM and the GLBPM, "reflect[ing] the use of secondary data sources and the work of harmonization and integration to create a data infrastructure that supports research across time and space.[16] Internally, use of the IPUMS BPM provides a clear visualization of our workflow from external submission of data, harmonization process, extraction systems, and archival preservation of metadata.[17] The upper levels of the IPUMS BPM also proved useful in identifying points where metadata is being produced by the projects (shown in green). (Figure 2)

Figure 2. IPUMS business process model

Evaluate / Specify Needs	Design / Redesign	Build / Rebuild	Collect	Process / Analyze	Archive / Preserve / Curate	Data / Dissemination / Discovery	Research / Publish	Retrospective Evaluation
1.1 Goal, research question, concepts, universe, conceptual variable	2.1 Identify sources	3.1 Develop data capture processes	4.1 Select sources	5.1 Validate data against metadata	6.1 Ingest data & metadata	7.1 Deploy release infrastructure	8.1 Obtain listing of publications based on the data product	9.1 Actors, when, inputs, methodology, list of criteria
1.2 Evaluation criteria, source list, evaluation results	2.2 Design sampling methods	3.2 Create or enhance infrastructure components	4.2 Negotiate access and distribution rights	5.2 Select and restructure data	6.2 Enhance metadata	7.2 Preserve dissemination products	8.2 Maintain publication database	9.2 Instrument, Actors, timeframe, resulting data
1.3 system requirements, estimation of development time, sub-projects/steps	2.3 Design capture process	3.3 Validate processes and tools	4.3 Capture data	5.3 Clean and anonymize data	6.3 Capture process/provenance metadata	7.3 Deploy access control system / policies	8.3 Manage versioning	9.3 Evaluation Form, Actors, Time, Data, Report
1.4 Criteria, concept list, representations, represented variables	2.4 Specify data elements and related metadata	3.4 Test production systems	4.4 Obtain metadata	5.4 Impute missing data	6.4 Preserve data & metadata	7.4 Promote dissemination products	8.4 Deposit metadata in related systems	9.4 Evaluation results, plan of action
1.5 Plan, create timetable, & identify needed infrastructure	2.5 Specify processing / data cleaning methods	3.5 Finalize production systems	4.5 Create sample	5.5 Harmonize selected data	6.5 Undertake ongoing curation	7.5 Provide data citation support	8.5 Manage disclosure risk	
1.6 Identify partners	2.6 Specify evaluation plan			5.6 Calculate weights		7.6 Enhance data discovery		
1.7 Prepare proposal and get funding	2.7 Organize research team			5.7 Calculate aggregates		7.7 Manager user support		
	2.8 Design infrastructure			5.8 Validate processed data				
				5.9 Finalize data outputs				

We are instituting a workflow mapping strategy to further identify IPUMS process and metadata capture points for the data archive. Currently our activity map has nine activity areas with sub-activities within each area. We have added depth in several of these activity paths to provide detail on specific activities. These activity paths will be expanded as we work with the individual projects to ensure that they each see their set of activities and process paths through the model. Our nine data projects have individualized project processes imposed by the "needs and constraints of their data sources and goals."[18]

The mapping approach has several advantages for the projects, the administrative team, the IT team, and the data archive. First, a common vocabulary is used across projects, administration, and IT. Because research staff sometimes move between projects, a common vocabulary streamlines those transitions. Second, the technical team can readily identify tools that can be developed and used across projects, developing efficiencies and economies of scale around data/metadata management, preservation, and delivery. Third, by establishing which products perform similar activities, IPUMS administration can identify process and tool developments that could benefit all projects. Further, the activity mapping approach allows each project to identify their own path through the process activities, thus preserving their individualized workflow, while maintaining our institutional standard. Finally, activity mapping identifies the areas of metadata production that require the attention of the archive for provenance and preservation purposes. (Figure 3)

Figure 3. SIP, AIP and DIP activity areas

Evaluate / Specify Needs	Design / Redesign	Build / Rebuild	Collect	Process / Analyze	Archive / Preserve / Curate	Data / Dissemination / Discovery	Research / Publish	Retrospective Evaluation
1.1 Define research needs, coverage & high-level concepts	2.1 Identify sources	3.1 Develop data capture processes	4.1 Select sources	5.1 Validate data against metadata	6.1 Ingest data & metadata	7.1 Deploy release infrastructure	8.1 Obtain listing of publications based on the data product	9.1 Establish evaluation criteria
1.2 Evaluate existing data & publications	2.2 Design sampling methods	3.2 Create or enhance infrastructure components	4.2 Negotiate access and distribution rights	5.2 Select and restructure data	6.2 Enhance metadata	7.2 Preserve dissemination products	8.2 Maintain publication database	9.2 Gather evaluation inputs
1.3 Establish outputs & needed infrastructure	2.3 Design capture process	3.3 Validate processes and tools	4.3 Capture data	5.3 Clean and anonymize data	6.3 Capture process/provenance metadata	7.3 Deploy access control system / policies	8.3 Manage versioning	9.3 Conduct evaluation
1.4 Identify specific concepts to be harmonized	2.4 Specify data elements and related metadata	3.4 Test production systems	4.4 Obtain metadata	5.4 Impute missing data	6.4 Preserve data & metadata	7.4 Promote dissemination products	8.4 Deposit metadata in related systems	9.4 Determine future actions
1.5 Plan, create timetable, & identify needed infrastructure	2.5 Specify processing / data cleaning methods	3.5 Finalize production systems	4.5 Create sample	5.5 Harmonize selected data	6.5 Undertake ongoing curation	7.5 Provide data citation support	8.5 Manage disclosure risk	
1.6 Identify partners	2.6 Specify evaluation plan			5.6 Calculate weights		7.6 Enhance data discovery		
1.7 Prepare proposal and get funding	2.7 Organize research team			5.7 Calculate aggregates		7.7 Manager user support		
	2.8 Design infrastructure			5.8 Validate processed data				
				5.9 Finalize data outputs				

SIP activity area
AIP activity area
DIP activity area

The combined OAIS and IPUMS BPM makes it clear at what point a new version of data is deposited in the archive as a submission information package (SIP). The model also accommodates any steps needed to meet the needs of individual projects--for example, handling the difference in creating snapshots for microdata and aggregate data products. Significantly, the model clarifies the point at which the content of the SIP becomes the custody of the archive and is no longer actively managed by the individual project. The content is then organized by the archive in an archive information package (AIP) for the purposes of management and future dissemination as a distribution information package (DIP) through a system separate from the IPUMS live data access system.

Leveraging documentation

The utility of leveraging the documentation collected, refined, and/or produced for the CTS application for various Institute purposes became evident as we organized our materials. For example, some of the documentation we produced for the CTS application will be used to support future funding application efforts. Our organization can efficiently demonstrate to potential funders the preservation policy practices constructed and maintained for our data production and metadata capture. It is particularly important that IPUMS can certify that it follows international standards; our work involves harmonization of official statistical data and contributing organizations need to know that their data are being responsibly handled. Further, because we took the time to thoughtfully think through and enumerate our data archive responsibilities, we will use this information and the visualizations we created to educate current staff, to inform stakeholders, and to onboard new employees. The focus of the CTS on documenting and providing access to the documentation of their methods and processes makes the role of the archive transparent to the parent organization, the user, funding agencies, and future archive staff. Staffing in non-traditional archives is often limited with low turnover. The danger of losing institutional knowledge of the reasoning, methods, and processes of the archive is very real as staff age out of their positions. Transparency becomes a means of retaining institutional coherence.

Lastly, our documentation provides valuable institutional and procedural history, both of which are often overlooked or attempted piecemeal long after processes have changed or been discontinued. Some of this documentation is publicly available, and other pieces are posted for internal use.[19] The CTS certification process provided justification for providing access to documentation on our processes in a more consistent and organized manner. It resulted in the addition of a working paper series focusing on IPUMS methodologies and development work. It has also highlighted the need to provide a process for routinely capturing the metadata currently provided on web pages and associating the content with the archived data files for past versions.

Preservation

Preserving our data products and resultant metadata is an obvious role of the IPUMS data archive. It became clearer to us as we worked to construct our CTS application that an important additional activity of our data archive is preserving the enormous intellectual investment that went into collecting, integrating, organizing, cleaning, documenting, and distributing our unique data products.

Our project managers have historically been primarily concerned with preserving the end product that is disseminated to users, and less attentive to preserving the pieces of intellectual activity that contributed to the data harmonization process. While the projects all operate within the nine activity areas identified on the activity map, as noted, each project follows its own unique path that can be preserved by the data archive. The major importance of this is in obtaining buy-in from the projects themselves. The purpose of the model in identifying common processes does not require individual process steps to occur in the same order for each iteration of a project or between two or more projects. Recognizing the individuality of each project reduces the concern of the project managers that they are being forced into a model as opposed to a model being developed based on the work that they do. Preserving the intellectual property relating to the processing and methodology that contributed to the development of our data products is a key and significant contribution to social science infrastructure.

While general archives have a relatively clear definition of what they need to preserve, the non-traditional archive often has a broader mandate. They are responsible for ensuring that the work done in the development of research and/or product development is also preserved. This information provides context for the traditional data and metadata and is a resource for future researchers in

terms of methodology, design, and decision making in the creation of data products or areas of research.

Guide for moving forward

It is vital that a non-traditional archive define its place within a research organization and clarify its role in supporting the standing of its parent organization with funders, data providers, and the research community. For the IPUMS archive, the CTS certification process was the opportunity to clearly articulate its role within the organization. As noted, questions from our funders provided the impetus by our administration to consider CTS certification, which in turn facilitated our organization considering the work of IPUMS more broadly, recognizing the value of preservation and future access to the unique products and by-products of IPUMS projects. The obvious significance of this to the archive was acknowledging its functions as an integral activity within IPUMS.

Drawing on earlier work conducted within the archive that explored process models from government statistical agencies and longitudinal survey projects (organizations that produced one or more statistical products on a regular basis and followed similar processes for creating each consecutive iteration), archive staff sought to adapt these process models to the IPUMS context. The archives' interest in these processes centered on capturing metadata generated by these processes and ensuring that metadata was being captured and preserved in a consistent format. The CTS certification application provided a framework for identifying gaps in our documentation, processes that needed clarification, and justification for ensuring that we had a complete and coherent workflow for ensuring access to our products far into the future.

As a first step, the CTS submission process required a review of the current IPUMS processes and encouraged us to clearly document all our activities. We modeled our workflows according to commonly used approaches (OAIS and GSBPM). The adaptation of these models helped us to present the role of the archive in a way that was understandable to the IPUMS organization and product groups. Specifically, we were able to:

- Clarify the role of the archive in IPUMS as a means of preserving the input to, and work of, each project within IPUMS
- Designate the touchpoints between the archive and the project teams to ensure that information was systematically passed into the archive as part of the project's production flow
- Identify the value-added information provided by the archive to the content deposited to it in terms of content, organization, and adherence to international standards
- Identify gaps in the archive process that required attention to meet the requirements of CTS certification

Overall, this process has been instrumental in presenting the importance of the role of the archive in ensuring that the intellectual work of the IPUMS projects is not lost over time and that the lessons of this important work in data harmonization and integration are not lost to future researchers. Data archives often serve a secondary role in research organizations, performing important work that is difficult to articulate and identify within the primary workflow of the research projects themselves. It is a bit like plumbing, its importance is only recognized when it fails. The CTS certification process is a valuable means of specifying and justifying the work of the archive in a research organization and brings that work to the attention of the organization in terms that they can understand. For the research organization, the CTS certification can be used to reassure data contributors that the organization understands and abides by international standards in the preservation of their data. Finally, in preparing for CTS certification, the organization goes through the process of updating,

completing, and providing access to information on its processes and policies. These documents can be referenced by applications for funding, showing the overall organization and care for the data being created, and the commitment of the organization to both data quality, as well as long term care and access.

Conclusion

The lessons we learned as part of the CTS application process are applicable in other data archive contexts as well, especially those in which preservation activities are not viewed as the primary function of the institution. In the IPUMS context, creation and distribution of harmonized datasets from census and survey data has always been the main focus of the projects, as required by our funders. IPUMS evolved from a single project (1880 PUMS) to a suite of products intended to be supported over time with the capacity to add new harmonized data products. Our institutional context has led to repositioning from a unit based in a small (history) department to an interdisciplinary research institute within the University of Minnesota.

Healthy institutions change over time, responding to a myriad of contingencies, both internal and external. Documenting this change provides a clear history of intent within the organization and offers a possible roadmap for other organizations experiencing similar growth, change, and development. The maturation of the role of the data archive within IPUMS reflects this dynamic growth and touches on the common issues of describing new functions, specifying the role of the archive in developing a sustainable metadata culture within the organization, and clarifying areas of management as specialization occurs within each contributing project. The multi-year process of preparing the IPUMS' application for the CTS encouraged us to provide models of our archival practices and clarify the details of our processes. Discussions with the project groups continue to clarify the role of the archive within IPUMS and to pinpoint where the project workflow intersects with the archive, improving overall communication. The disruption of the ongoing pandemic also reinforces the importance of preserving institutional history that is both clear and accessible to support the inevitable transition in personnel that occurs over the life course of an institution.

References

Lawrenz F. and Paller, M.S. (2006) "Transforming the University: Recommendations of the Task Force on Collaborative Research, University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/567.

Magnuson, D.L. (2014) Steven Ruggles interview, University of Minnesota, January 9, 2014.

Magnuson, D.L. (2015a) Wendy Thomas interview, University of Minnesota, March 24, 2015.

Magnuson, D.L. (2015b) "Curating Our Social Science Infrastructure: the MPC/IPUMS Institutional History as a Case Study," presented at the Social Science History Association, Baltimore, USA, November 12-15, 2015.

Magnuson, D.L. and Ruggles, S. (2022) "Challenges of Large-Scale Data Processing in the 1990s: The IPUMS Experience," IEEE Annals of the History of Computing, pp. 71-83.

Ruggles, S. (1991a) “Integration of the Public Use Samples of the U.S. Census,” 1991 Proceedings of the American Statistical Association, Social Statistics Section. Alexandria, VA, ASA, pp. 265-370.

Ruggles, S. (Summer 1991b) “The U.S. Public Use Census Microdata Files as a Source for the Study of Long-Term Social Change,” IASSIST Quarterly. https://doi.org/10.29173/iq703.

Ruggles, S. (1992-1995) “Integrated Public Use Microdata Series,” SES-9118299, NSF.

Ruggles, S. (2011) ”Minnesota Population Center: Self Study Report,” University of Minnesota, June 7, 2011.

Van Hook, J.L., Bleakley, C.H. and Hummer, R.A. (2016) ”External Review of the Minnesota Population Center,” June 14-16, 2016.

Endnotes

[1] Diana L. Magnuson is Curator and Historian at the Institute for Social Research and Data Innovation, University of Minnesota (magn0031@umn.edu).
[2] Wendy L. Thomas is retired Curator at the Institute for Social Research and Data Innovation, University of Minnesota (wlt@umn.edu).
[3] https://isrdi.umn.edu/. https://www.coretrustseal.org/. https://www.ipums.org/mission-purpose.
[4] Steven Ruggles, interviewed by Diana L. Magnuson, University of Minnesota, January 9, 2014.
[5] Steven Ruggles, interviewed by Diana L. Magnuson, University of Minnesota, January 9, 2014.
[6] Steven Ruggles, interviewed by Diana L. Magnuson, University of Minnesota, January 9, 2014.
[7] Steven Ruggles, ”MPC Strategic Plan,” May 2, 2016 (email in possession of author). Steven Ruggles, Institute Name,” August 19, 2016 (email in possession of author). Van Hook, J.L., Bleakley, C.H. and Hummer, R.A. (2016) ”External Review of the Minnesota Population Center,” ISRDI Institutional Archive, June 14-16, 2016. In 2016, all data projects took on the IPUMS prefix as part of their project name. Since not all projects are microdata and some have access conditions that limit their usage, it is inaccurate to describe IPUMS as a “public use” microdata series. Thus, since 2016 IPUMS is a brand, not an acronym. https://www.ipums.org/mission-purpose.
[8] https://www.ipums.org/.
[9] Wendy Thomas, interviewed by Diana L. Magnuson, University of Minnesota, March 24, 2015. https://assets.ipums.org/_files/ipums/workflows/IPUMS_Archive_Workflow_Nov2021.pdf.
[10] https://assets.ipums.org/_files/ipums/workflows/IPUMS_Archive_Workflow_Nov2021.pdf, p. 5.
[11] https://www.coretrustseal.org/about/history/data-seal-of-approval-synopsis-2008-2018/.

[12] https://assets.ipums.org/_files/ipums/workflows/IPUMS_Archive_Workflow_Nov2021.pdf, p. 12.
[13] https://assets.ipums.org/_files/ipums/workflows/IPUMS_Archive_Workflow_Nov2021.pdf, p. 11.
[14] https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1.
[15] https://ddialliance.org/sites/default/files/GenericLongitudinalBusinessProcessModel.pdf.
[16] https://assets.ipums.org/_files/ipums/workflows/IPUMS_Archive_Workflow_Nov2021.pdf, p. 16.
[17] https://www.ipums.org/workflows.
[18] https://assets.ipums.org/_files/ipums/workflows/IPUMS_Archive_Workflow_Nov2021.pdf, p. 17.
[19] https://www.ipums.org/about/more.