Stewarding our resources: Building a sustainable IPUMS archival document access system

Diana L. Magnuson[1]

Abstract

IPUMS at the University of Minnesota has created the world's largest accessible database of census and survey microdata. The IPUMS suite of products contains nine harmonized data products. The largest of these products, IPUMS International (IPUMS-I), has supported the curation and preservation of ancillary materials received during data acquisition efforts. Archival staff have preserved thousands of unique pieces of census and survey documentation, creating bibliographic records using an extended Dublin Core profile that supports the use of controlled vocabularies to enhance findability for the project staff and outside users. The goal of this curation work was to create a findable, searchable, and downloadable document access system for our internal use and to support IPUMS researchers. This paper describes our experience constructing a web interface that supports exploration and dissemination of these archived materials. During this development, we gained valuable insight about stewarding our resources that are applicable to research organizations responsible for curating, preserving, and disseminating archival materials.

Keywords

Archive, preservation, metadata, IPUMS

Introduction

Over the last thirty years, IPUMS at the University of Minnesota has created the world's largest accessible database of census and survey microdata (Magnuson and Ruggles 2022). The primary work of IPUMS is data harmonization—making census and survey data compatible across time and space. IPUMS integration and documentation makes it easy for researchers to study change, conduct comparative research, merge information across data types, and analyze individuals within family and community contexts. As of this writing, the IPUMS suite of products contains nine harmonized data collections. International data comes from over one hundred national and regional statistical organizations. All data are freely available to the global public.

The context of this paper is specific to the IPUMS International (IPUMS-I) data project, the largest of the nine harmonized data collections. Beginning in 1999, with a social science infrastructure grant from the National Science Foundation (NSF), IPUMS-I had a simple yet audaciously ambitious goal: preserve the world's microdata resources and democratize access to those resources. Twenty-five years later, the project goals continue to be: collecting and preserving census and survey data and documentation; harmonizing those data; and disseminating the harmonized data free of charge

(Ruggles and McCaa et al. 1999-2004; McCaa and Ruggles 2000; Ruggles et al. 2003). IPUMS-I data are coded and documented consistently across countries and over time to facilitate robust comparative research.

IPUMS-I has amassed tens of thousands of ancillary materials in support of its data harmonization work. These materials came from United States Census Bureau (USCB), United Nations Statistical Division (UNSD), Latin American and Caribbean Demographic Center (CELADE), The East-West Center, Centre Population et Dévelopement (CEPED), and over one hundred national statistical agencies. Examples of this material include correspondence, maps, enumerator instructions, supervisor instructions, training materials, codebooks, publicity, reports, newspaper clippings, unpublished papers, census timetables, data processing materials, and technical manuals. The ancillary materials in our collection attest to the varied technical, business, social, and economic aspects of conducting censuses and surveys across time and space.

Acquisition, preservation, and dissemination of data products is central to the full IPUMS workflow (Figure 1). For the purposes of this paper, archive-specific workflows are highlighted in blue. We used an Open Archival Information System model (OAIS) to identify where our archive obtained submissions both external and internal (submission information package, SIP), what actions we took once we obtained those submissions (archival information package, AIP), and how we delivered the products to users (dissemination information package, DIP) (Magnuson and Thomas 2023). Figure 2 highlights the archive responsibility within the IPUMS workflow. This paper will focus on the development of the document access system within the archive workflow (Figure 2, circled in red).

Figure 1. Full IPUMS workflow (Magnuson and Thomas 2023)

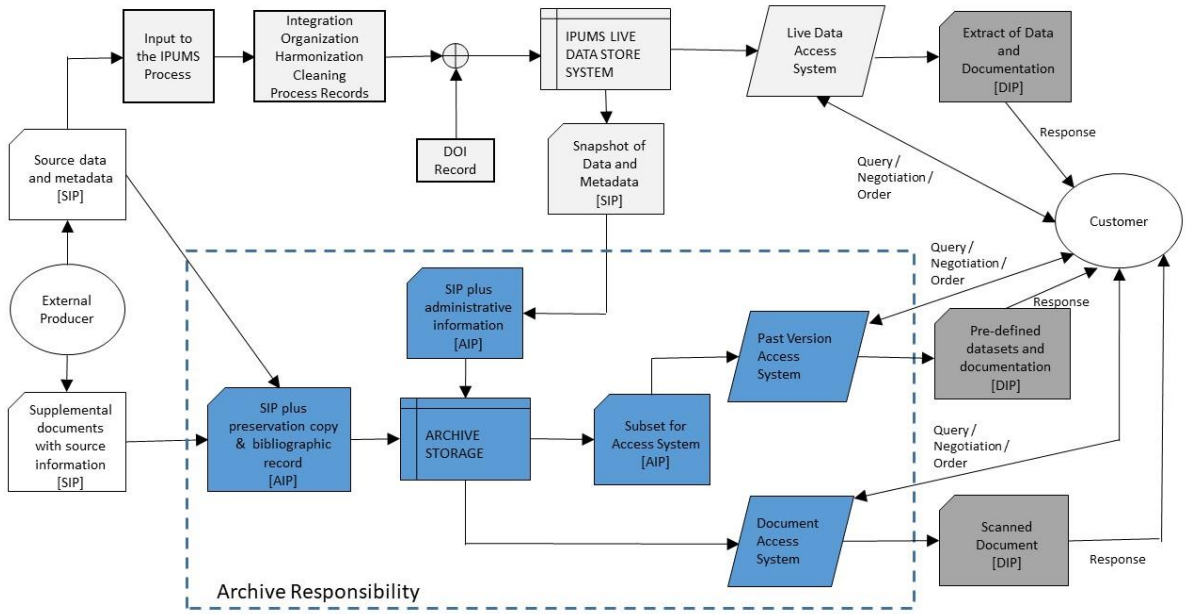
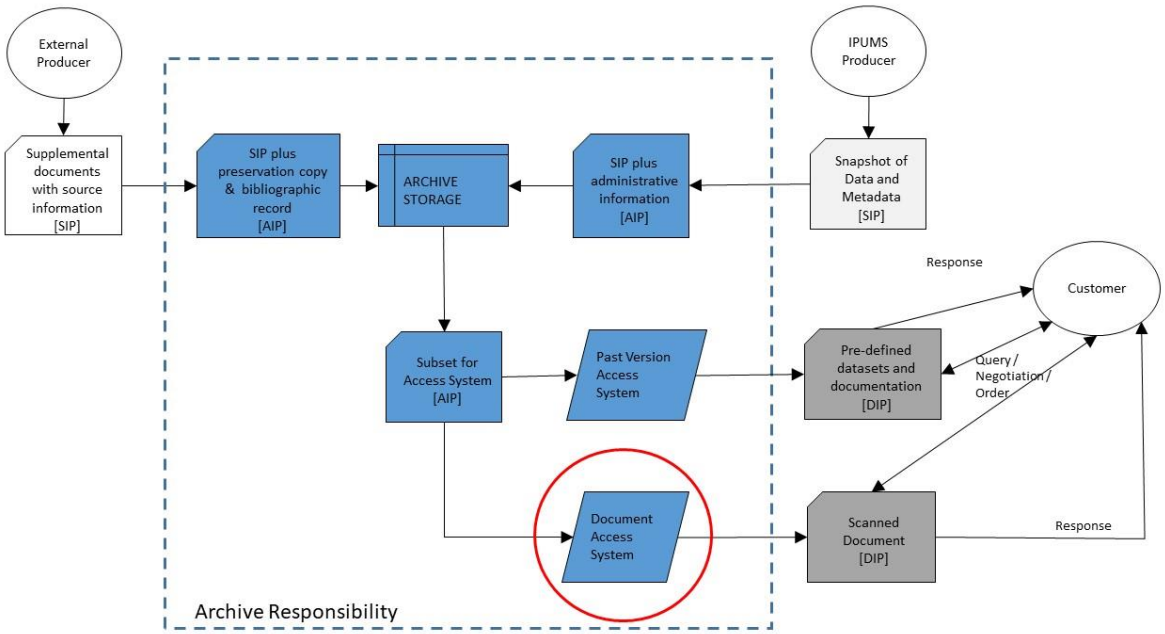


Figure 2. IPUMS archive workflow and document access system



A portion of IPUMS-I grant money has funded the curation and preservation of the ancillary materials acquired by the project. For over two decades, archival staff have been preserving thousands of unique pieces of census and survey documentation, creating bibliographic records using an extended Dublin Core profile that supports the use of controlled vocabularies to enhance findability for the project staff and outside users. The goal of this work was the creation of a simple, findable, searchable, and downloadable document access system.

An early iteration of the effort to make available basic international census documentation was the creation of a World Population Census Forms webpage. The aim of the page was primarily to demonstrate to potential international data partners the scope and capacity of IPUMS-I. It provided a tabular list of census forms organized by country and year, with minimal functionality offered by hyperlinks to download the documents (Ruggles et al. 2003).[2] While useful for organizing, identifying, and accessing standard international census forms, this static utilitarian webpage challenged us to consider how we could provide access to the thousands of additional materials entrusted to us. We needed a simple but more sophisticated tool for stewarding our archival resources.

Building the IPUMS document access system

Building the IPUMS document access system required identifying short- and long-term goals of the archive, building a warehouse of flexible metadata, navigating the priorities and schedule of the Institute for Social Research and Data Innovation (ISRDI is IPUMS' parent organization) and working

with the ISRDI Information Technology (IT) Product Team to develop a web application that was simple, sustainable, discoverable, and capable of disseminating basic metadata and PDFs of digitized archival materials.

Vision

Beginning in 1999, ancillary documents began streaming into the archive as a product of IPUMS-I acquisition efforts. It is likely that IPUMS-I uniquely holds some of these materials. A portion of the materials arrived with the express understanding that they would become available through some means after their lifecycle in the IPUMS-I microdata harmonization work had concluded. Other acquired materials were broader than the IPUMS-I data collection and covered countries, censuses, and surveys for which IPUMS-I does not disseminate microdata. The fragility of some documents created a race against time to carefully preserve them for future use. For example, the high acid content of some paper documents meant they were brittle and at considerable risk of eventual disintegration. Other materials were produced using Thermofax technology (heat sensitive copy paper), which darkens over time, rendering the documents illegible. Taken together, this collection reflects an exciting diversity of archival materials. We cannot predict the uses to which future researchers might put these materials, but it is our obligation to preserve them and make their innovative future use possible.

Data curator Wendy Thomas recognized the significance of all these materials and anticipated the day when an archival document access system would be a reality. Thomas' experience providing research support for social science data users and her technical expertise curating data and metadata positioned her well for envisioning an online IPUMS document access system. Thomas immediately took steps to assemble the scaffolding necessary to build a flexible metadata warehouse. In this context, "flexible" refers to creating metadata in an open-source format that allowed for its mapping to a range of standards. From the beginning, Thomas advocated for a simple online tool that would provide basic metadata and discoverable, accessible, searchable, and downloadable PDFs of digitized archival materials. The tool was envisioned to be sustainable long term by the archivist, with minimal intervention of IT staff beyond the development of the web application (Magnuson 2015).

Building a metadata warehouse

Building a metadata warehouse for the IPUMS archival document access system has been years in the making. First, Thomas built up the scaffolding for logical intake and workflow for digital and manuscript materials (Table 1). This was no small task. Documents did not arrive on a fixed schedule or in uniform packaging, but rather, as agreements were made with supporting statistical entities (McCaa and Ruggles 2000; Ruggles et al. 2003). Thousands of physical and digital documents needed to come under archival control before they were subject to the metadata creation stage.

During this initial stage, Thomas was balancing the creation and implementation of intake and workflow with a data harmonization project that was already underway. Prior to Thomas' hire, IPUMS-I researchers had pulled a number of documents from their individual collections and scanned them for use in the IPUMS-I project. After Thomas' hire, archival staff were fielding requests from researchers to locate specific types of materials from within the newly acquired materials. Thomas determined that the most expedient way to know the contents of the multilingual collection and to efficiently access the documents, was through scanning the cover or first page of all documents.

Original boxes were sorted and labeled by region and country (using ISO numeric codes) and then transferred into archival safe boxes. This approach facilitated an understanding of the scope of the collection by country, the percentage of documents needing special attention due to their physical condition and determining the extent of documents requiring language translation. Thus, the short-term needs of the IPUMS-I project followed a systematic, country-by-country approach to creating bibliographic records for this extensive collection.

Table 1. Archival intake and workflow

Intake	Workflow
Acquisition	
Acquire physical and digital documents from data partners	Accessioning e.g., UNSD, USCB, CELADE, EASTWEST, CEPED, Other
Organization	
Organize by source/region in file cabinets or boxes	e.g., UNSD, USCB, CELADE, EASTWEST, CEPED, Other
Label	Label box with standard [ISO3166 region]-[ISO3166-country]-[box ###]
Create country spreadsheet	Populate spreadsheet with item MPCacquisition number, folder, source, file name, region, country, notes
Scan	
Follow scanning decision tree	Determine whether full, cover, or partial scan is warranted based on type, size, uniqueness, and/or fragility
Label scans	Follow MPC numbering taxonomy, e.g., 142-050-CB-09 or 142-048-001-09-full
Bibliographic record creation	
Extended Dublin Core profile	Utilize hand-tailored IPUMS-I controlled vocabulary

Next, Thomas developed a process using a text editor for creating structured bibliographic records for each piece of archival material. Using an extended Dublin Core profile, Thomas hand-tailored a controlled vocabulary to the archival dimensions of IPUMS-I ancillary materials (Table 2).[3] The decision to use Dublin Core was a practical one. Dublin Core is recognized worldwide as a basic standard for creating bibliographic records. In addition, Dublin Core vocabulary is easily extended as needed to address new material types or issues of scale. Cataloging systems that do not use Dublin Core invariably provide information on mapping to Dublin Core (Thomas 2023).[4] Prior to the creation of the web application, archive staff used Thomas' controlled vocabulary to search for and organize materials, using text string searches employing simple programming scripts.

Table 2. IPUMS-International bibliographic record extended Dublin Core controlled vocabulary (excerpt)

Line No.	Text	Adds Content
00	Add Document [includes lines: 01=title, 03=creator 11=MPCacq 12=FullURI 17=MPCloc 23=format[pdf]	<mpcrecord> <dc:title xml:lang="en"></dc:title> <dc:creator></dc:creator> <dc:identifier xsi:type="MPCacq">MPC</dc:identifier> <dc:identifier xsi:type="FullURI"></dc:identifier> <dc:identifier xsi:type="MPCloc">Box </dc:identifier> <dc:format>application/pdf</dc:format>

	34=spatial 38=temporal 42=language 47=type 49=provenance]	<dcterms:spatial xsi:type="ISO3166_N"></dcterms:spatial> <dcterms:temporal xsi:type="census"></dcterms:temporal> <dc:language xsi:type="ISO639-1">en</dc:language> <dc:type xsi:type="dcterms:DCMIType">PhysicalObject</dc:type> <dcterms:provenance>USCB</dcterms:provenance> </mpcrecord>
32	MPC subject	<dc:subject>Agricultural Survey</dc:subject> <dc:subject>Agricultural Census</dc:subject> <dc:subject>Economic Census</dc:subject> <dc:subject>Economic Survey</dc:subject> <dc:subject>Health Survey</dc:subject> <dc:subject>Housing Census</dc:subject> <dc:subject>Housing Survey</dc:subject> <dc:subject>Labor Force Survey</dc:subject> <dc:subject>Population Census</dc:subject> <dc:subject>Population Survey</dc:subject> <dc:subject>Economic Classifications</dc:subject> <dc:subject>Educational Classifications</dc:subject> <dc:subject>Geographic Classifications</dc:subject> <dc:subject>Temporal Classifications</dc:subject>
44	Description	<dc:description>data collection form</dc:description> <dc:description>non-census questionnaire</dc:description> <dc:description>process management form</dc:description> <dc:description>UN questionnaire form</dc:description> <dc:description>data collection instructions</dc:description> <dc:description>supervisor instructions</dc:description> <dc:description>IPUMS-I inventory form</dc:description> <dc:description>geographic list</dc:description> <dc:description>From: ; To: </dc:description> <dc:description>[provider] Documents inventory</dc:description> <dc:description>agenda</dc:description> <dc:description>budget</dc:description> <dc:description>census guide</dc:description> <dc:description>census methods</dc:description> <dc:description>census report</dc:description> <dc:description>census topics</dc:description> <dc:description>data processing</dc:description> <dc:description>evaluation</dc:description> <dc:description>hiring test</dc:description> <dc:description>information for schools</dc:description> <dc:description>meeting notes</dc:description> <dc:description>progress report</dc:description> <dc:description>UN mission reports</dc:description> <dc:description></dc:description>
46	MPCLimited Types	<dc:type xsi:type="MPCLimited">code book</dc:type> <dc:type xsi:type="MPCLimited">data</dc:type> <dc:type xsi:type="MPCLimited">data collection form</dc:type> <dc:type xsi:type="MPCLimited">data collection instructions</dc:type> <dc:type xsi:type="MPCLimited">inventory</dc:type> <dc:type xsi:type="MPCLimited">extract</dc:type> <dc:type xsi:type="MPCLimited">form</dc:type>

		<dc:type xsi:type="MPCLimited">gov doc</dc:type> <dc:type xsi:type="MPCLimited">image</dc:type> <dc:type xsi:type="MPCLimited">instructions</dc:type> <dc:type xsi:type="MPCLimited">correspondence</dc:type> <dc:type xsi:type="MPCLimited">map</dc:type> <dc:type xsi:type="MPCLimited">article</dc:type> <dc:type xsi:type="MPCLimited">book</dc:type> <dc:type xsi:type="MPCLimited">ephemera</dc:type> <dc:type xsi:type="MPCLimited">publicity</dc:type> <dc:type xsi:type="MPCLimited">regulations</dc:type> <dc:type xsi:type="MPCLimited">serial</dc:type> <dc:type xsi:type="MPCLimited">table</dc:type> <dc:type xsi:type="MPCLimited">technical manual</dc:type> <dc:type xsi:type="MPCLimited">training material</dc:type> <dc:type xsi:type="MPCLimited">unpublished</dc:type>

Accessing archive metadata has consistently relied on use of controlled vocabularies to enhance findability and flexibility both for immediate use by project staff and for future external researchers. Along the way, archivists developed and refined training and reference materials to support undergraduate student workers who created bibliographic records for each unique piece of archival material. An unintended but happy result of the development and refinement of these instructional and reference materials over time was the curation of institutional history documenting the work, growth, and development of the IPUMS archival processes.

Institutional priorities

As noted above, census and survey data harmonization and dissemination are the central activities of IPUMS within ISRDI at the University of Minnesota. In this environment, it has been incumbent upon archival staff to nurture understanding and investment in expanding our institutional archival curation activities. IPUMS administrators, principal investigators, project managers, and research data scientists all recognize the importance and utility of an archival data access system, but quite naturally their priorities have historically focused on grant writing, data acquisition, data harmonization, and data dissemination.

The key for archival planning and productivity, then, has been identifying IPUMS priorities that naturally connect with, and enhance, the short- and long-term goals of the archive. First, the bold mission of IPUMS is to democratize access to the world's social and economic data for current and future generations.[5] Second, a "central goal" of IPUMS-I from the beginning was "to create an inventory of surviving census microdata and documentation," including "enumerator instructions, census forms, codebooks, studies of data quality, and any other ancillary documentation we can locate for all countries that will allow us access to this documentation" (Ruggles et al. 2003). Third, the decision to begin assigning DOIs to IPUMS data products was triggered by the increasing requirements of external funding organizations to conform to "standard archival practice using the open archival information system (OAIS) model and digital object identifiers (DOI)" (Magnuson and Thomas 2023). Assigning DOIs forced broader internal discussions around access, curation, and preservation—all concerns of the archive. Fourth, documenting these internal discussions became a steppingstone to building a successful Core Trust Seal application, which in turn highlighted the preservation work of the archive (Magnuson and Thomas 2023).[6] These priorities all have strong ties to IPUMS archive

concerns. Intentionally communicating and nurturing those connections to internal IPUMS stakeholders has been essential to moving archival goals forward.

The steady growth of the IPUMS-I data project, increasing expectations of external funding organizations regarding IPUMS preservation practices, and a transition in IPUMS archive leadership due to a retirement, eventually led to concrete steps toward building an online document access system. The IPUMS IT Product Team, the group of developers responsible for the IPUMS web dissemination system and related components, added that task to their quarterly project calendar in late 2022 (Ruggles et al. 2015; Magnuson and Thomas 2023). The persistent commitment to building a warehouse of flexible metadata positioned the archive well to take advantage of this opportunity to fulfill a long-standing grant deliverable. In addition, the IT Product Team was looking for a discrete project to train a new hire and introduce them to project workflows. Exploration of existing document access product solutions had occurred periodically, and the conclusion reached each time was that too much modification was necessary to leverage the use of our controlled vocabularies. Now the evolutionary moment had finally arrived for IPUMS to undertake building an archival document access system.

The IT Product Team invited archival staff to propose a small defined project, and the timing was ideal. Archive personnel had already submitted a presentation proposal to the 2023 IASSIST Conference promising to document the effort to build a document access system. The acceptance in February 2023 for an IASSIST presentation in early June 2023 fit the IPUMS IT Product Team timeline and meant the goal of the presentation would not merely be documentation but an actual demonstration of the new user interfaces.

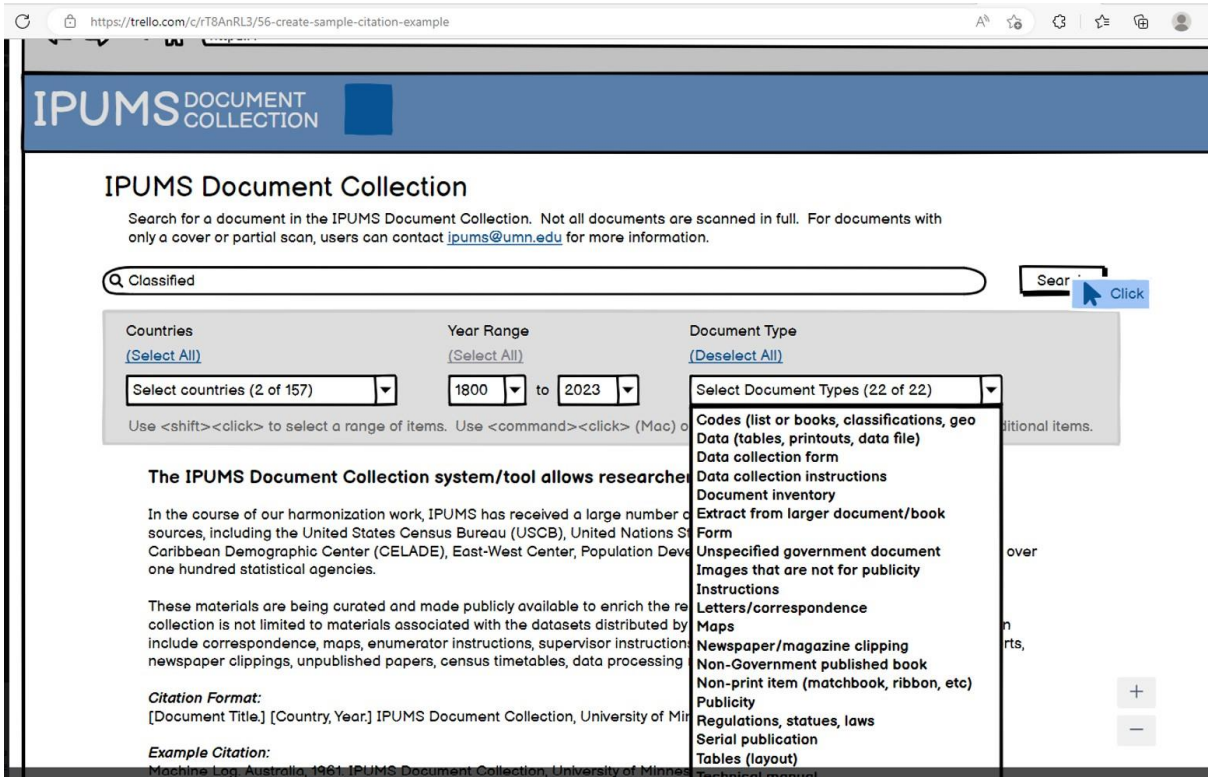
IPUMS IT Product Team

The process of working with the IPUMS IT Product Team to develop a web interface was collaborative. The first formal step in the collaboration was completing an IT Project Data Sheet (PDS) to record information from the archive covering four areas: 1) objective statement, project motivation, project context, primary stakeholder; 2) major milestones and deadlines, project scope; 3) Minimum Viable Product (MVP), most important quality attributes, trade-off matrix; 4) known business issues and risks, and how success will be measured. Working through the PDS to its completion was a worthwhile intellectual exercise for archival stakeholders. It forced us to problematize the project in productive directions, identifying “must have,” “valuable to have,” and “aspirational” (future) developments to the system functionality and user interfaces.

Once the project was approved by the IT Product Team and underway, there was regular written (email) and oral (Zoom) dialogue clarifying project goals and developing a workflow. Tasks were tracked via the web-based management tool Trello. For five months leading up to the project “sprint” (final two weeks of the project), periodic check-ins to track progress and clarify action points were conducted over Zoom every few weeks. Archive stakeholders were first presented by the IT Product Team software developers with proof of concept and shortly thereafter a prototype of the user interface. Two weeks before the project deadline the IT Product Team launched their sprint and communication increased to daily brief “standup” meetings. During the standup, IT and archive contributors assessed and assigned tasks, asked questions, and held everyone accountable to the specifications of the PDS. During the sprint period, the web developer created “wireframes”

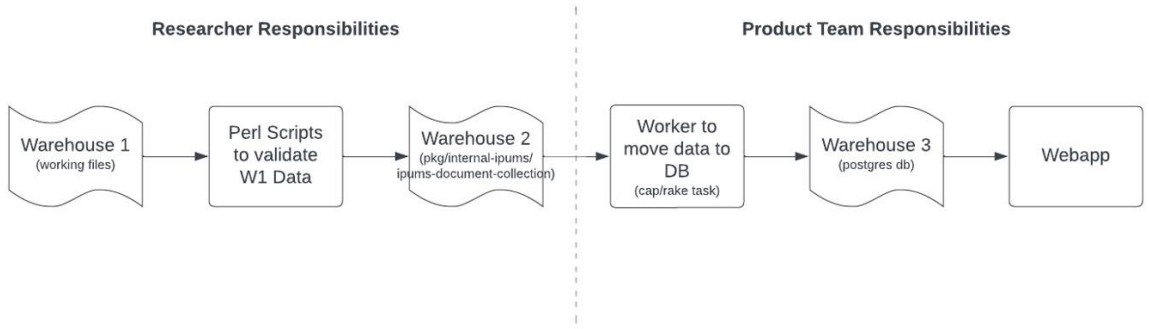
(illustrations) of the proposed user interface content, functionalities, and intended behaviors for archival review and feedback (Figure3).

Figure 3. Example of IPUMS Document Collection wireframe



The Product Team visualized a product workflow containing three dynamic data "warehouses" (Figure 4). Warehouse 1 contains PDFs of the scanned documents and the XML metadata working files created by undergraduate student workers. Files in Warehouse 1 are subject to archivist ("Researcher") validation against internal archive standards and cleaning protocols using Perl Scripts. After completing validation and cleaning the archivist moves the metadata to Warehouse 2, where it is ready to be processed by the web application. The IT Product Team then takes the PDF and XML metadata files from Warehouse 2, subjecting these files to IT computational validation and moves the files to Warehouse 3, to be disseminated through the web interface.

Figure 4. Metadata workflow responsibilities for researcher and product team



The collaboration between the IT Product Team and archive personnel produced valuable learning in both directions. Archive staff gained an appreciation and understanding of the elements necessary to build a scalable user interface: including consistency of content, reliable search filters, and clear expectations regarding the specific elements of the tool. In turn, the role and concerns of the archive are now in the IT Product Team's line of sight. This is especially important for the archive, as its work is often viewed as being conducted on the periphery of IPUMS product workflows.

IPUMS Document Collection

The IPUMS Document Collection web interface was the minimum viable product (MVP) resulting from the collaboration between archive staff and the IPUMS IT Product Team. The launch of the IPUMS Document Collection (documents.ipums.org) from the development site to the live website in early June 2023 triggered a new phase in product development.

PDF documents and metadata from the region of Oceania served as a test dataset to represent the quality and content of the metadata records and to evaluate the ability of these records to support the functionality of the web interface. The Oceania dataset is small (9 countries, 626 document records, 29 collection records, and 653 PDF files), but it is representative of the overall IPUMS-I collection. Thus, the Oceania dataset served as an essential diagnostic in the development of the web interface.

First, the Oceania dataset made clear for IPUMS IT staff the nature of the files to be provided to them by the archive. Second, archive staff divided responsibility for validation based on where in the workflow was most effective to assess the metadata. Third, communication during the refinement of the MVP clarified record content in terms of immediate and future requirements that will support additional search options. Going forward, the IPUMS Document Collection will be built out by archival staff as IPUMS-I ancillary materials from new regions are processed through Warehouses 1, 2 and 3.

Conclusion

Adopting a long-range strategy of acquisition, establishing archival control, and employing a process of flexible metadata creation, were essential to building and successfully launching the online IPUMS Document Collection. This effort took place over time, in expectation of an eventual funding and/or institutional opportunity. When the institutional opportunity presented itself, archival staff were ready. A strategy of anticipatory and flexible preparation can be adopted by archive staff who operate in an institutional setting in which their role supports the main product.

As noted, the primary work of IPUMS is data harmonization, making census and survey data compatible across time and space. IPUMS integration and documentation makes it easy to study change, conduct comparative research, merge information across data types, and analyze individuals within family and community contexts. All IPUMS resources are free of charge. In this organizational context, the archive has a supporting but vital role. Stewarding and expanding access to our archival resources in support of IPUMS involved:

- Articulating what areas of the organization our archival functions are related to. It was incumbent upon archive staff to clearly and consistently identify where the work of the archive touched the main product of the organization and how archival stewardship enhanced the main product.
- Knowing the priorities of our organization. Framing archival goals within the context of IPUMS deliverables required archival staff to plan metadata creation in terms of how it would be efficiently conveyed to users of IPUMS data products.
- Investing in building flexible and scalable metadata processes and infrastructure.
- Leveraging work with the IPUMS IT Product team. Through constructive and deliberate communication, archival staff developed short- and long-term goals, identified strengths and areas for refinement of archival data management, and educated IPUMS IT staff on the vital role of the archive within the organization.
- Strategizing scalable deliverables. In the IPUMS-I context, we focused on developing the scaffolding for intake of digital and manuscript materials, tailoring a controlled vocabulary using an extended Dublin Core profile, and building a simple web interface.

The ancillary materials in the IPUMS Document Collection attest to the technical, business, social, and economic aspects of conducting censuses and surveys across time and space. In addition, ancillary materials contextualize contemporary uses of, and responses to, historical censuses and surveys. Expanding and deepening our search and delivery system will provide findability and accessibility to a rich set of supporting archival documentation that will illuminate census development and implementation processes across time and space. For example, access to materials documenting the development of enumeration forms and procedures over time supports researchers' understanding of how statistical entities responded to the challenges of collecting demographic data on difficult to enumerate populations. Creating a sustainable, discoverable, and searchable access system for a broad range of archival census and survey materials will support the IPUMS mission of democratizing access to the world's social and economic data and support transformative scholarship. We believe curation and public availability of these materials enrich IPUMS products but also the innovative research of IPUMS data users.

While the IPUMS Document Collection and history is unique, it reflects a common situation for any archive that operates in a supporting role to a large research project. Typically, research project managers focus on the dissemination of their primary product; the organization and management of ancillary materials is secondary to that effort. The role of the archivist is to ensure future researcher access to the rich documentary evidence amassed during any research project. This access may be a locally supported access system or deposited into an external system. At the beginning of the IPUMS-I project, we did not know who would be making the materials accessible, how they might be accessed, or what standards would be employed by this future system. For these reasons, we focused on organizing and creating a highly flexible structure for our metadata. The core of our system and that of any system operating in a similar environment is the need to accurately:

- Preserve the content in a manner that will support future research use.
- Track provenance of the materials, including distribution rights.
- Retain the logical relationship between the materials (based on archival needs).
- Select metadata and preservation standards that will ensure the greatest ease in depositing the content into one or more access systems.
- Provide as much descriptive information as possible in a manner that can be mapped to an existing or emerging access system and expanded as needed by the growth of the collection.

Developing short- and long-term strategies to organize and manage a collection within the constraints of current project funding is essential. In the IPUMS context, short-term efforts focused on: scanning cover pages so documents could be searched on a country-by-country basis; creating bibliographic records that formalized terms used by the project (MPCLimited controlled vocabulary); and promoting the interests of funders by providing access to full documentation and provenance in some manner. The long-term IPUMS archival strategy included providing subsets of scans and bibliography records (in our context geographical regions) so that when the opportunity for developing an online access system arose, we were ready to populate it with a significant collection of materials that represented the range of materials in our collection. The payoff was an online collection with immediate usefulness to the global research community. These lessons are applicable across research organizations with unexploited or underexploited collections of valuable ancillary materials.

References

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

Magnuson, D.L. and Ruggles, S. (2022) "Challenges of Large-Scale Data Processing in the 1990s: The IPUMS Experience," IEEE Annals of History and Computing, pp. 71-83. https://ieeexplore.ieee.org/abstract/document/9972862

Magnuson, D.L. and Thomas, W.L. (2023) "Expanding our perspective: Building a sustainable metadata culture," IASSIST Quarterly, Volume 47, No. 2. https://iassistquarterly.com/index.php/iassist/article/view/1046

McCaa, R. and Ruggles, S. (2000) "IPUMS-International: A Global Project to Preserve Machine-Readable Census Microdata and Make Them Useable," Handbook of International Historical Microdata for Population Research, edited by Patricia Kelly Hall, Robert McCaa and Gunnar Thorvaldsen, pp. 335-346. https://international.ipums.org/international/microdata_handbook.shtml

Ruggles, S., King, M.L., Levison, D., McCaa, R and Sobek, M. (2003) "IPUMS International," Historical Methods: A Journal of Quantitative and Interdisciplinary History, Volume 32, No. 2. https://www.tandfonline.com/doi/abs/10.1080/01615440309601215

Ruggles, S., McCaa, R., Sobek, M. and Cleveland, L. (2015) "The IPUMS Collaboration: Integrating and Disseminating the World's Population Microdata," Journal of Demographic Economics, Volume 81. https://doi.org/10.1017/dem.2014.6

Ruggles, S., McCaa, R., Levison, D., Gardner, T. and Sobek, M. (1999-2004) "International Integrated Microdata Access System." SBR9908380, Methodology, Measurement, and Statistics Program, NSF.

Thomas, W. (2023) "Why Dublin Core," General Information, ISRDI Archive.

Endnotes

[1] Diana L. Magnuson is Curator and Historian at the Institute for Social Research and Data Innovation, University of Minnesota. Magnuson can be reached at magn0031@umn.edu.

[2] https://international.ipums.org/international/census_forms_shtml

[3] The use of "MPC" refers to the Minnesota Population Center, which predated the Institute for Social Research and Data Innovation (ISRDI).

[4] https://www.dublincore.org/specifications/dublin-core/profile-guidelines/

[5] https://www.ipums.org/mission-purpose

[6] https://www.coretrustseal.org/