IPUMS qualitative resources: Supporting robust social science research

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 contains nine harmonized census and survey microdata and aggregate geographic data collections. In addition to archiving these data collections, the IPUMS archival staff is also responsible for curating, preserving, and making discoverable three key pieces of qualitative IPUMS content: ancillary census and survey materials acquired primarily by IPUMS International for their data harmonization work; DDI-Codebook metadata and documentation produced during the IPUMS data extract system and collection preservation; and working papers authored by IPUMS staff documenting technical innovation around IPUMS data collections. This paper describes the development of IPUMS infrastructure to manage, preserve, annotate, and disseminate its qualitative collection, tightly connecting these materials to the IPUMS quantitative data collections produced for the purpose of supporting robust social science research. The IPUMS archival staff has developed processes to receive, organize, tag, and distribute a large and diverse body of qualitative material. Attention to the range of processing and research uses of this qualitative collection has been instrumental in identifying useful tags to provide targeted access to support these uses. Operationalizing this work has informed archival organizational knowledge control and efforts to raise awareness of archival work and research possibilities for those who provide qualitative content in all its forms. These efforts support strong social science research and offer a path forward for archivists preserving qualitative metadata, ancillary materials, and working papers, in an organizational setting in which archival curation, preservation, discoverability, and dissemination activities are essential, although often considered secondary to the main product.

Keywords

IPUMS harmonized data, qualitative data, DDI-Codebook, ancillary materials, IPUMS working paper

Introduction

For over thirty years, IPUMS at the University of Minnesota has created the world's largest accessible database of census and survey microdata (Magnuson & Ruggles, 2022). As of this writing, the IPUMS suite contains nine harmonized census and survey microdata and aggregate geographic data collections. IPUMS launched in 1993, but its qualitative data and ancillary material curation infrastructure only began to take shape in 1999. The impetus for this development was the IPUMS International (IPUMS-I) data project. With a social science infrastructure grant from the National

Science Foundation, IPUMS-I data project goals were ambitious: to preserve the world's microdata resources and democratize access to those rich resources (Ruggles & McCaa, et al., 1999-2004).

At the core of the IPUMS archival approach is a commitment to building and maintaining infrastructure that provides FAIR (findable, accessible, interoperable, reusable) access to our quantitative and qualitative data to support vigorous social science research. Regarding qualitative data and associated ancillary materials, what began as an internal effort to support IPUMS project teams performing data harmonization has expanded to include access by external researchers using IPUMS harmonized data collections. While structured metadata is generally seen as a data source regarding the development and relationship of concepts and the means of measuring them, discussions with survey statisticians around the world saw this collections as a unique resource for: research on survey development, including structured questionnaires; semi-structured enumerator instructions, and documents covering the development of certain question sets over time; and approaches to harmonization between countries to support international comparison of results (Magnuson, 2015). Significantly, materials documenting the process of data collection including the identification of difficult to enumerate populations and the use of public relations campaigns to address these issues exist in these collections within the context of national approaches over time and international support for census procedures. Visibility and access to IPUMS qualitative documentation provides the context in which canvassers historically collected the quantitative data; the secondary IPUMS harmonization of the data; and the diverse research uses of the data. Further, IPUMS rich qualitative data sources can inform researchers' use of quantitative data and enhance the stories quantitative research seeks to tell.

This article describes the development of IPUMS infrastructure around its qualitative collection in three areas: ancillary census and survey materials acquired primarily by IPUMS-I for their data harmonization work and collateral materials received from other donors; DDI-Codebook metadata and documentation produced during IPUMS quantitative data harmonization collection preservation; and working papers authored by IPUMS staff documenting technical innovation around IPUMS data collections. The IPUMS archival staff have developed processes to receive, organize, tag, and distribute a large and diverse body of qualitative data and associated materials. Attention to the range of processing and research uses of this qualitative collection has been instrumental in identifying useful tags to provide targeted access to support these uses. Operationalizing this work informs archival organizational knowledge control and efforts to raise awareness of archival work and research possibilities for those who provide qualitative content in all its forms. These efforts support robust social science research and offer a path forward for archivists preserving a diverse body of qualitative data in an organizational setting in which archival curation, preservation, discoverability, and dissemination activities are essential, although often considered secondary to the main product.

Defining the IPUMS qualitative collection

The IPUMS qualitative collection consists of three main types of documentation: ancillary census and survey materials, DDI-Codebook metadata and documentation, and working papers.

The first component of the IPUMS qualitative collection is composed of ancillary census and survey materials acquired by IPUMS-I for their direct data harmonization work and collateral materials from other archives and donors not specifically intended for use in the harmonization process. Over time,

IPUMS-I amassed tens of thousands of ancillary materials 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 (Magnuson, 2024b; Ruggles et al., 2015). Examples of this qualitative material include but are not limited to: correspondence, maps, enumerator instructions, supervisor instructions, training materials, codebooks, publicity, reports, newspaper clippings, unpublished papers, census timetables, data processing materials, and technical manuals. A portion of IPUMS-I grant money was targeted for the curation and preservation of these qualitative materials (Ruggles et al., 2003).

The second component of the IPUMS qualitative collection is DDI-Codebook metadata. The IPUMS Extract System creates an XML codebook that is compliant with the DDI-Codebook standard as output in the preservation of each updated version of IPUMS quantitative harmonized data collection.[2] The use of DDI-Codebook in supporting social science qualitative research is significant because DDI-Codebook structures the metadata, allowing researchers to examine the development and change of variable, value domains, and questions across time and space in a way unstructured documents cannot. The use of DDI-Codebook to structure, manipulate, discover, and preserve IPUMS harmonized data collections dates back to 2003 (Block & Thomas, 2003).

IPUMS harmonized data collections use the DDI-Codebook metadata standard in two ways. First, the IPUMS quantitative data extract system allows researchers to curate their own data file tailored to their specific research needs. Every researcher-initiated data extract produces a customized DDI-Codebook in XML format for that particular data request. Second, an IPUMS project version 'snapshot' of data and metadata produces a DDI-Codebook compliant document in XML format for the complete dataset as well as for each subset. The 'snapshot' data is created from the IPUMS Live Data Storage System through an extract tool and is downloaded and deposited in permanent archive storage (Figure 1). Thus DDI-Codebook is a powerful tool for structuring metadata with extensive detail on variables, source questions, and harmonization, providing a qualitative data resource for comparative survey statisticians and those interested in the comparability and change in census taking over time and space. IPUMS DDI-Codebook metadata file, documentation, and data are available through the specific project web site either through the download system for the current version or through the past version landing pages.

The third component of the IPUMS qualitative collection working papers authored by IPUMS staff. The IPUMS Working Paper Series is a valuable resource for social science researchers using IPUMS quantitative data collections. First, the working papers provide a means for publishing documents describing the IPUMS data harmonization processes and capturing the development of decision making within the process. Second, the working paper format is flexible for presenting a range of documentation from discussion of specific topics addressed in the harmonization process to detailed descriptions of how and why particular processes were employed. Third, the IPUMS Working Papers Series provides a venue for documenting a dynamic and growing body of work around IPUMS methodology, features, applications, and analysis.

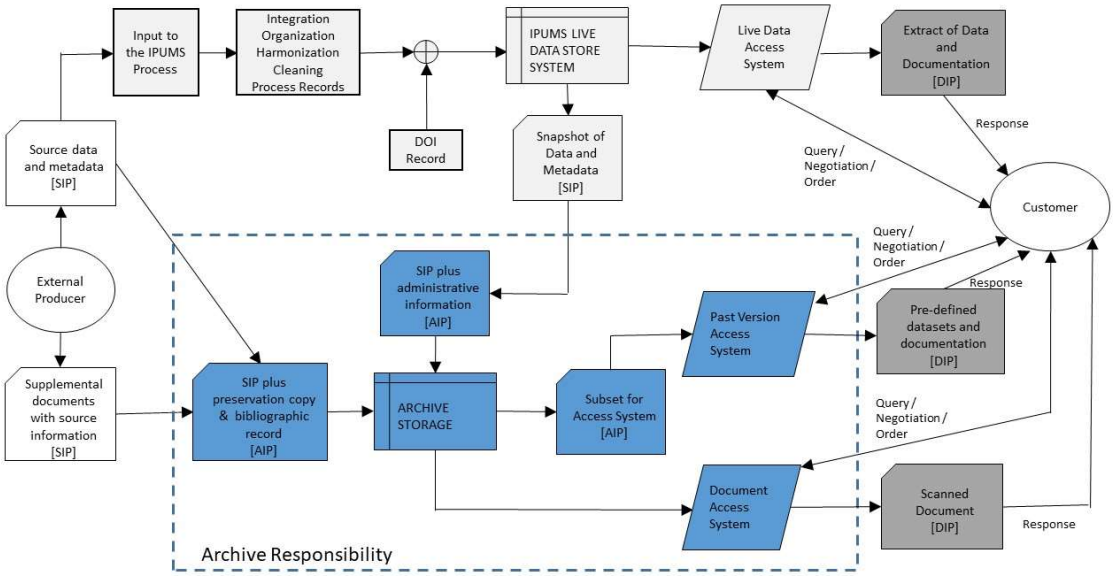
Developing the IPUMS qualitative collection infrastructure

The hire of Wendy Thomas in 2001 as data curator of the Minnesota Population Center (MPC, precursor to the Institute for Social Research and Data Innovation (ISRDI), organizational home of IPUMS), led to two important infrastructure initiatives around the IPUMS qualitative materials collection. The first initiative was formalizing curation efforts. In particular, Thomas operationalized accessioning of the diverse body of qualitative materials accumulated by IPUMS-I. The second initiative focused on implementing the Data Document Initiative (DDI) DDI-Codebook standard to structure metadata produced by IPUMS harmonized data collections.[3]

With respect to manuscript archival materials, Thomas drew on her past technical expertise curating data and metadata for the Machine Readable Data Center (MRDC) at the University of Minnesota to develop a qualitative material curation workflow and infrastructure (University of Minnesota: Office of Information Technology, 2000; The Minnesota Daily, 1991). Archival materials did not arrive at IPUMS on a fixed schedule, in uniform structure, or with significant metadata; but rather, as cooperative agreements were made witstatistical entities (Ruggles et al., 2015; Ruggles et al., 2003; McCaa & Ruggles, 2000). Further, some archival elements were collateral material bundled with documents being shipped to IPUMS; that is, not directly related to IPUMS data harmonization work but adjacent to censuses and surveys being harmonized by IPUMS. Still more qualitative materials arrived pertaining to censuses and surveys not currently part of the IPUMS data harmonization workflow. It was thus necessary to create an intake system that addressed these issues and at the same time anticipated future requirements around access, use, and discoverability, for both internal and external researcher needs (Magnuson, 2015).

Using an Open Archival Information System model (OAIS), archival staff visualized the broad IPUMS archival workflow: where submissions were obtained (submission information package, SIP), what actions staff took with submissions (archival information package, AIP), and how staff delivered the collections to users (dissemination information package, DIP). With respect to qualitative data curation and ancillary materials workflow and infrastructure, the focus of this article, the archive responsibility is highlighted in blue (Figure 1).

Figure 1. IPUMS Workflow (Magnuson and Thomas 2023)



Taking archival control of the qualitative ancillary materials arriving as part of the IPUMS-I harmonization work followed an approach that simultaneously addressed immediate project needs for access to the materials and the longer term preservation, curation, and dissemination requirements of the archive (Table 1).

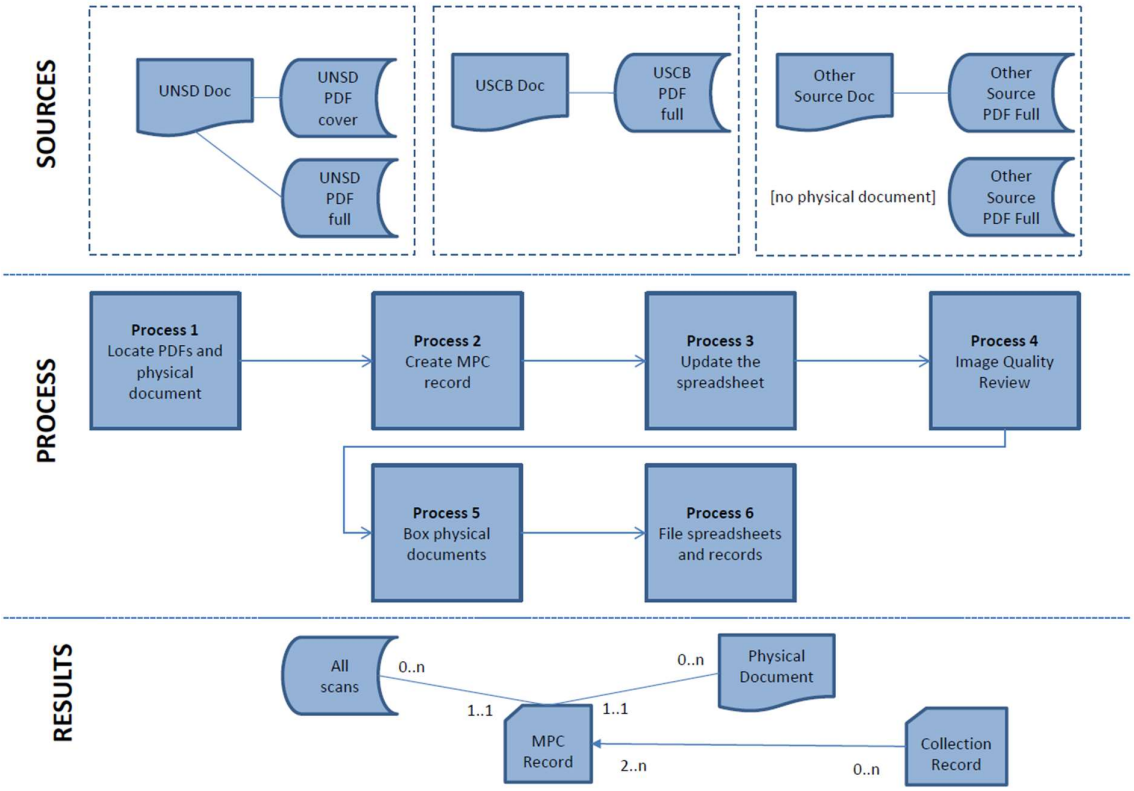
Table 1. Ancillary material intake and workflow (Magnuson 2024b)

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	
Scan materials	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 (Magnuson 2024b, pp. 5-7)

Operationalizing the IPUMS ancillary qualitative materials infrastructure

The volume of ancillary materials received by IPUMS-I over time dictated operationalizing the curation workflow to address issues of volume, scale, component parts, fragility of the materials, undergraduate student worker training and turnover, and metadata creation pinch points (Figure 2 and Table 2). As noted, the archive was receiving a variety of original documents and scans with different provenance and different use cases for IPUMS-I data harmonization work; thus, the curation workflow created a structured environment to process qualititative materials in a consistent manner, described and linked by a biblographic record (denoted by an 'MPCacquisition number', Tables 1 and 2).

Figure 2. Ancillary material curation workflow (overview)



A key task in the IPUMS archival workflow is the essential but time consuming work of creating bibliographic metadata for all materials in the ancillary collection. As noted, after Thomas' arrival in 2001, mapping out and documenting this workflow was particularly important. The newly assembled archival team needed to address a large backload of existing materials obtained from other organiztions, address new materials arriving through on-going projects, and address the high turn-over of staff to do this work. The archive is staffed with up to 1.5 FTE of undergraduate student staff. Undergraduate employees are transient by nature, and thus their training needs to be efficient and consistent to ensure a standard quality to their work. In short, archival processes needed to be clear, address a number of common exceptions, and provide staff with sufficient written resources to

address most issues as they were encountered. Undergraduate student staff are trained in basic archival managment and assisted the archivist in scanning, cleaning, creating bibliographic metadata, and preserving the ancillary materials in the IPUMS document collection (Table 2). The detailed workflows for processing census related documentation was addressed in a previous *IASSIST Quarterly* issue (Magnuson, 2024b) and internally the archive maintains process documentation containing instructions and reasoning, as well as multiple examples and resources to ensure the consistent use of coding and content creation.

The decision by archival staff to use Dublin Core as the metadata standard was threefold: its practicality as an internationally recognized basic standard for creating bibliographic records; the ease of extending the vocabulary as new material types or issues of scale arose; and the ability to map Dublin Core to other standards. Of particular importance in chosing Dublin Core was the ability to create specialized types with extensions containing local controlled vocabularies. Our internal rule is to allow expansion of the controlled vocabulary as needed, but to retain all past valuues (Thomas, 2023). In practice, this allows for archival staff to designate association of a document with a particular census or survey, specify the census or survey type (e.g., population, housing, agriculture), and specify document type (e.g. enumeration form, supervisor instructions, code list, public relations, training materials). While initially added to provide support to IPUMS-I research staff in identifying materials to harmonize data, it also allowed researchers to identify structured and sesmi-structured document that may be used as input to qualitative researcher methods (Magnuson, 2024b).

Table 2. Ancillary material curation workflow (detail)

Component	Goals
Sources	
UNSD, USCB, CELADE, EASTWEST, CEPED, Other	Acquire, organize, and scan
Processes	
P1: Collect sources	Collect and organize source materials: spreadsheet, physical documents, scans
P2: Create records	Utilize hand-tailored IPUMS-I controlled vocabulary to create a bibliographic record in XML (Magnuson, 2024b, pp. 5-7)
P3: Update spreadsheet	Populate spreadsheet for each record: MPCacquisition number, path to directory/folder for pdfs, file name, region, post-scan cleaning directions, notes
P4: Review image quality	Confirm scans are clean, straight, and legible; if not address issues.
P5: Box physical documents	Box physical documents; use MPC numbering taxonomy, e.g., [ISO region]-[ISO country]-[box number]
P6: File spreadsheets, digital records, and boxes	Move completed spreadsheets and digital files to verification area; move new scans to appropriate folder; file boxes to archival storage
Results	

Bibliographic record, collection record, physical document, scans	All components ready for verification to move into documents.ipums.org workflow (Magnuson, 2024b, p. 9)

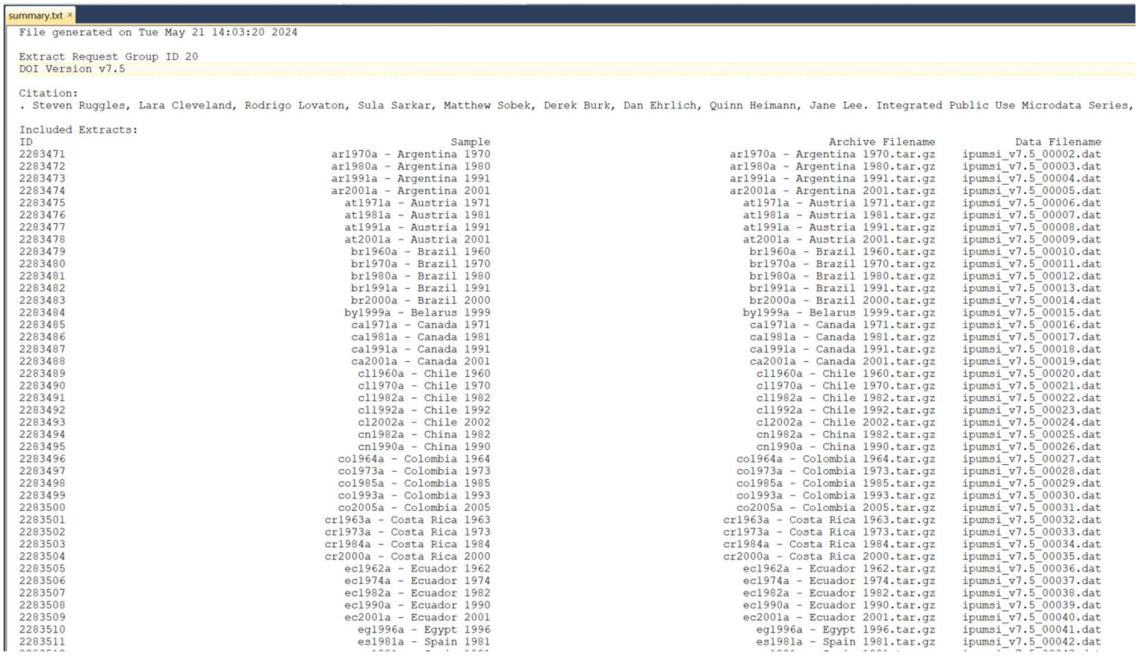
Operationalizing archival intake and workflow (Table 2) laid the groundwork for the creation of the IPUMS Document Collection, a publicly accessible document access system launched in 2024 (Magnuson, 2024b, Magnuson, 2025).[4] Figure 1 identifies where the document access system resides in the IPUMS workflow (lower right). The ongoing work of scanning, cleaning, creating bibliographic metadata, and preserving the ancillary materials used for harmonization work and 'bonus' materials received by IPUMS-I continues to expand the online holdings of the IPUMS Document Collection. Public visibility and access to IPUMS qualitative documentation provides the context in which canvassers historically collected the quantitative data, deepens FAIR access to our ancillary collection, enriches use of IPUMS quantitative data collections, and supports strong social science research.

Implementing DDI-Codebook standard to IPUMS harmonized data collections

IPUMS use of DDI-Codebook to structure, manipulate, discover, and preserve IPUMS harmonized data collections dates back to 2003 (Block & Thomas, 2003). As noted above, data archivist Thomas was well acquainted with structured metadata and she used this expertise to strongly encourage adoption of DDI-Codebook for IPUMS quantitative data collections, a metadata standard first released in 2000 (subsequent version were released in 2003 and 2005).[5]

DDI-Codebook metadata standards are used in two ways with regard to IPUMS harmonized data collections. First, the IPUMS data extract system allows researchers to curate their own data file tailored to their specific research needs. Every researcher-initiated data extract produces a DDI-Codebook in XML format for that particular data request (Figure 1). Second, an IPUMS project version 'snapshot' of data and metadata produces a DDI-Codebook in XML format for the complete archived dataset (Figures 1 and 3). Creation and preservation of the DDI-Codebook metadata file for new versions ('snapshots') of IPUMS quantatitive data collections is carried out using a line-command software tool developed by the IPUMS Information Technology team. The output created in the snapshot process is a set of files including: a folder of DDI-Codebook metadata files for each sample in the dataset and one for the complete dataset; a summary sheet listing the samples created; and data extract folders for subsets of the project (data extract in a fixed format single record type file; simple codebook file in text format; SPSS setup File; SAS setup file; R setup file; XML DDI-Codebook). The output is created from the IPUMS Live Data Store System and moved into permanent archive storage (Figure 1). Individual IPUMS data harmonization project versioning utilizes the IPUMS Business Process Model (IPUMS BPM) for identifying archival activity areas (Magnuson, 2024a).

Figure 3. Screenshot example from the summary.txt of samples created during a "snapshot"



summary.txt

```
File generated on Tue May 21 14:03:20 2024

Extract Request Group ID 20
DOI Version v7.5

Citation:
. Steven Ruggles, Lara Cleveland, Rodrigo Lovaton, Sula Sarkar, Matthew Sobek, Derek Burk, Dan Ehrlich, Quinn Heimann, Jane Lee. Integrated Public Use Microdata Series,

Included Extracts:
```

ID	Sample	Archive Filename	Data Filename
2283471	ar1970a - Argentina 1970	ar1970a - Argentina 1970.tar.gz	ipumsi_v7.5_00002.dat
2283472	ar1980a - Argentina 1980	ar1980a - Argentina 1980.tar.gz	ipumsi_v7.5_00003.dat
2283473	ar1991a - Argentina 1991	ar1991a - Argentina 1991.tar.gz	ipumsi_v7.5_00004.dat
2283474	ar2001a - Argentina 2001	ar2001a - Argentina 2001.tar.gz	ipumsi_v7.5_00005.dat
2283475	at1971a - Austria 1971	at1971a - Austria 1971.tar.gz	ipumsi_v7.5_00006.dat
2283476	at1981a - Austria 1981	at1981a - Austria 1981.tar.gz	ipumsi_v7.5_00007.dat
2283477	at1991a - Austria 1991	at1991a - Austria 1991.tar.gz	ipumsi_v7.5_00008.dat
2283478	at2001a - Austria 2001	at2001a - Austria 2001.tar.gz	ipumsi_v7.5_00009.dat
2283479	br1960a - Brazil 1960	br1960a - Brazil 1960.tar.gz	ipumsi_v7.5_00010.dat
2283480	br1970a - Brazil 1970	br1970a - Brazil 1970.tar.gz	ipumsi_v7.5_00011.dat
2283481	br1980a - Brazil 1980	br1980a - Brazil 1980.tar.gz	ipumsi_v7.5_00012.dat
2283482	br1991a - Brazil 1991	br1991a - Brazil 1991.tar.gz	ipumsi_v7.5_00013.dat
2283483	br2000a - Brazil 2000	br2000a - Brazil 2000.tar.gz	ipumsi_v7.5_00014.dat
2283484	by1999a - Belarus 1999	by1999a - Belarus 1999.tar.gz	ipumsi_v7.5_00015.dat
2283485	ca1971a - Canada 1971	ca1971a - Canada 1971.tar.gz	ipumsi_v7.5_00016.dat
2283486	ca1981a - Canada 1981	ca1981a - Canada 1981.tar.gz	ipumsi_v7.5_00017.dat
2283487	ca1991a - Canada 1991	ca1991a - Canada 1991.tar.gz	ipumsi_v7.5_00018.dat
2283488	ca2001a - Canada 2001	ca2001a - Canada 2001.tar.gz	ipumsi_v7.5_00019.dat
2283489	cl1960a - Chile 1960	cl1960a - Chile 1960.tar.gz	ipumsi_v7.5_00020.dat
2283490	cl1970a - Chile 1970	cl1970a - Chile 1970.tar.gz	ipumsi_v7.5_00021.dat
2283491	cl1982a - Chile 1982	cl1982a - Chile 1982.tar.gz	ipumsi_v7.5_00022.dat
2283492	cl1992a - Chile 1992	cl1992a - Chile 1992.tar.gz	ipumsi_v7.5_00023.dat
2283493	cl2002a - Chile 2002	cl2002a - Chile 2002.tar.gz	ipumsi_v7.5_00024.dat
2283494	cn1982a - China 1982	cn1982a - China 1982.tar.gz	ipumsi_v7.5_00025.dat
2283495	cn1990a - China 1990	cn1990a - China 1990.tar.gz	ipumsi_v7.5_00026.dat
2283496	co1964a - Colombia 1964	co1964a - Colombia 1964.tar.gz	ipumsi_v7.5_00027.dat
2283497	co1973a - Colombia 1973	co1973a - Colombia 1973.tar.gz	ipumsi_v7.5_00028.dat
2283498	co1985a - Colombia 1985	co1985a - Colombia 1985.tar.gz	ipumsi_v7.5_00029.dat
2283499	co1993a - Colombia 1993	co1993a - Colombia 1993.tar.gz	ipumsi_v7.5_00030.dat
2283500	co2005a - Colombia 2005	co2005a - Colombia 2005.tar.gz	ipumsi_v7.5_00031.dat
2283501	cr1963a - Costa Rica 1963	cr1963a - Costa Rica 1963.tar.gz	ipumsi_v7.5_00032.dat
2283502	cr1973a - Costa Rica 1973	cr1973a - Costa Rica 1973.tar.gz	ipumsi_v7.5_00033.dat
2283503	cr1984a - Costa Rica 1984	cr1984a - Costa Rica 1984.tar.gz	ipumsi_v7.5_00034.dat
2283504	cr2000a - Costa Rica 2000	cr2000a - Costa Rica 2000.tar.gz	ipumsi_v7.5_00035.dat
2283505	ec1962a - Ecuador 1962	ec1962a - Ecuador 1962.tar.gz	ipumsi_v7.5_00036.dat
2283506	ec1974a - Ecuador 1974	ec1974a - Ecuador 1974.tar.gz	ipumsi_v7.5_00037.dat
2283507	ec1982a - Ecuador 1982	ec1982a - Ecuador 1982.tar.gz	ipumsi_v7.5_00038.dat
2283508	ec1990a - Ecuador 1990	ec1990a - Ecuador 1990.tar.gz	ipumsi_v7.5_00039.dat
2283509	ec2001a - Ecuador 2001	ec2001a - Ecuador 2001.tar.gz	ipumsi_v7.5_00040.dat
2283510	eg1996a - Egypt 1996	eg1996a - Egypt 1996.tar.gz	ipumsi_v7.5_00041.dat
2283511	es1981a - Spain 1981	es1981a - Spain 1981.tar.gz	ipumsi_v7.5_00042.dat

Table 3. Example of (sample).tar file

File name	File content	File content description
ipumsi_v7.5_00315.xml	XML File	DDI-codebook metadata
ipumsi_v7.5_00315.cbk	CBK File	Simple codebook listing of variables
ipumsi_v7.5_0035.dat	Compressed Archive Folder	Compressed data file
ipumsi_v7.5_0035.do	Stata File	Stata setup file
ipumsi_v7.5_0035.R	R File	R setup file
ipumsi_v7.5_0035.sas	SAS File	SAS setup file
ipumsi_v7.5_0035.sps	SPS File	SPS setup file

Publishing IPUMS Working Papers Series

Creating and maintaining an infrastructure for the working paper series generates internal and external benefits to IPUMS. As noted above, the working papers describe the IPUMS data harmonization process and capture the development of decision making within the process. The working paper format is flexible for presenting a range of documentation from discussion of specific topics addressed in the harmonization process to detailed descriptions of how and why particular processes were employed. Lastly, the IPUMS Working Papers Series provides a venue for documenting a dynamic and growing body of work around IPUMS methodology, features and applications.

The workflow for publishing the working paper series is relatively straightforward (Table 4). The working papers all have registered DOIs and are publicly discoverable and available for download in pdf format (Magnuson & Thomas, 2023, p. 7). IPUMS technical documentation can thus be used and properly cited by internal IPUMS research staff and social science researchers more broadly. Further, the working paper series curates material that can be documented and used to support the grant proposal work of ISRDI.

Table 4. IPUMS Working Paper Series workflow

Component	Source
Obtain information	'External Producer' (Figure 1)
Title, creators, publication year, abstract, funding, pdf	PDF is archived at an internal IPUMS drive
Create DOI record	DataCite Kernal
XML bibliographic record that validates against DataCite Kernal	<?xml version="1.0" encoding="UTF-8"?> <resource xmlns="http://datacite.org/schema/kernel-4" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://datacite.org/schema/kernel-4 http://schema.datacite.org/meta/kernel 4.2/metadata.xsd"> <identifier identifierType="DOI"></identifier> <creators> <creator> <creatorName></creatorName> <givenName></givenName> <familyName></familyName> <affiliation></affiliation> </creator> </creators> <titles> <title xml:lang="en"></title> </titles> <publisher></publisher> <publicationYear></publicationYear> <resourceType resourceTypeGeneral="Text"></resourceType> <language>en</language> <relatedIdentifiers> <relatedIdentifier relatedIdentifierType="DOI" relationType="IsPreviousVersionOf"></relatedIdentifier> <relatedIdentifier relatedIdentifierType="DOI" relationType="IsPartOf"></relatedIdentifier> </relatedIdentifiers> <formats> <format></format> </formats> <descriptions> <description descriptionType="Abstract" xml:lang="en"></description> </descriptions> </resource>
Update DOI collection record	DataCite Kernal
Update collection record to include relationship identifier	<?xml version="1.0" encoding="UTF-8"?> <resource xmlns="http://datacite.org/schema/kernel-4" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://datacite.org/schema/kernel-4 https://schema.datacite.org/meta/kernel-4.2/metadata.xsd">

	<identifier identifierType="DOI">10.18128/CP-IPUMS</identifier> <creators> <creator> <creatorName nameType="Organizational"></creatorName> </creator> </creators> <titles> <title xml:lang="en"></title> </titles> <publisher></publisher> <publicationYear></publicationYear> <language>en</language> <resourceType resourceTypeGeneral="Collection">Collection</resourceType> <relatedIdentifiers> <relatedIdentifier relatedIdentifierType="DOI" relationType="HasPart"></relatedIdentifier> </relatedIdentifiers> <version></version> <descriptions> <description descriptionType="Abstract" xml:lang="en"></description> </descriptions> </resource>
File DOI	DataCite (doi.datacite.org)
Assign a new DOI, provide URL, upload XML record	Use Fabrica (DataCite DOI management tool)
Landing page	Working Paper Series landing page[6]
Update working paper series website landing page to reflect new citation	Use Drupal (open source content management system) and populate with relevant bibliographic metadata

The IPUMS Working Paper Series began in 2020.[7] A separate MPC Working Papers Series began earlier, in 2005. It primarily publishes research papers by Minnesota Population Center (MPC) members, often using IPUMS data, but it also contains papers addressing IPUMS technical documentation.[8] ISRDI now publishes three working paper series: MPC, IPUMS and the Life Cource Center (LCC).[9] All working papers are internally reviewed before publication.

Conclusion

Quantitative data is the 'skeleton' of social science research, providing form and structure for analysis. The best social science research also utilizes qualitative sources. The 'skeleton' needs 'muscle' and 'skin' and it is this use of descriptive metadata that supports intelligent analysis by research data users. In the IPUMS context, metadata includes related qualitative documents created both as part of the data and metadata creation process, and as a reflection of the environment in which IPUMS staff created harmonized data.

The purpose of developing the IPUMS qualitative archival infrastructure is thus to manage, preserve, annotate, and make discoverable, its qualitative collection, tightly connecting these materials to IPUMS harmonized quantitative data produced for supporting robust social science research. IPUMS archive staff consciously designed quantitative and qualitative archival infrastructure to capture IPUMS data harmonization process information. IPUMS qualitative archival infrastructure is operationalized in a way that is flexible and addresses the variety of ancillary materials and artifacts, the irregularity of when items are received from partner organizations, and the rapid undergraduate staff turnover found in academic settings, thus maintaining quality control. While managing the vast quantity of documentation related to IPUMS projects is important in practical terms, what is most unique about this approach lies in the development of an annotation process that highlights both the unique characteristics of the collection as well as providing insight and access to materials as they contribute to the IPUMS harmonization process. The use of controlled vocabularies arising from the needs and work of the project researchers reflects those work processes and provides detailed access to future researchers interested in the development of census taking across time and space.

The development of IPUMS infrastructure around its qualitative collection supports researcher access to these important materials. Collectively, access to ancillary census and survey materials, DDI-Codebook metadata and other documentation produced during IPUMS quantitative data harmonization, and working papers authored by IPUMS staff that document technical innovation and analysis around IPUMS data collections, add substance to quantitative evidence. Visibility and access to IPUMS qualitative documentation illuminate the context in which canvassers historically collected the quantitative data; the secondary IPUMS harmonization of the data; and the diverse research uses of the data. The IPUMS archive model offers a path forward for archivists preserving qualitative data in an organizational setting in which archival curation, preservation, discoverability, and dissemination activities are essential although often considered secondary to the main product. IPUMS rich qualitative materials in all its forms can inform researchers' use of quantitative data and enhance the stories quantitative researchers seek to tell.

References

Block, W., & Thomas, W. (2003). Implementing the Data Documentation Initiative at the Minnesota Population Center. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *32*(2). https://doi.org/10.1080/01615440309601219

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

Magnuson, D. L. (2024a). The IPUMS business process model: Instituting a workflow mapping strategy to support archival processes. *IASSIST Quarterly*, *48*(4). https://doi.org/10.29173/iq1130

Magnuson, D. L. (2024b). Stewarding our resources: Building a sustainable IPUMS archival document system. *IASSIST Quarterly*, *48*(1). https://doi.org/10.29173/iq1095

Magnuson, D. L. (2025). The IPUMS document collection: Progress & prospects [Conference poster]. IASSIST Annual Meeting, Bristol, United Kingdom. https://doi.org/10.5281/zenodo.15473847

Magnuson, D. L., & Ruggles, S. (2022). Challenges of large-scale data processing in the 1990s: The IPUMS experience. IEEE Annals of the History of Computing, 71–83. https://ieeexplore.ieee.org/abstract/document/9972862

Magnuson, D. L., & Thomas, W. L. (2023). Expanding our perspective: Building a sustainable metadata culture. *IASSIST Quarterly*, *47*(2). https://iassistquarterly.com/index.php/iassist/article/view/1046

McCaa, R., & Ruggles, S. (2000). IPUMS-International: A global project to preserve machine-readable census microdata and make them useable. In P. K. Hall, R. McCaa, & G. Thorvaldsen (Eds.), Handbook of international historical microdata for population research (pp. 335–346). https://international.ipums.org/international/microdata_handbook.shtml

Ruggles, S., King, M. L., Levison, D., McCaa, R., & Sobek, M. (2003). IPUMS International. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *32*(2). https://doi.org/10.1080/01615440309601215

Ruggles, S., McCaa, R., Levison, D., Gardner, T., & Sobek, M. (1999–2004). International Integrated Microdata Access System (SBR9908380). Methodology, Measurement, and Statistics Program, National Science Foundation.

Ruggles, S., McCaa, R., Sobek, M., & Cleveland, L. (2015). The IPUMS collaboration: Integrating and disseminating the world's population microdata. *Journal of Demographic Economics*, 81. https://doi.org/10.1017/dem.2014.6

The Minnesota Daily. (1991, February 7). The Minnesota Daily. University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/246061

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

University of Minnesota: Office of Information Technology. (2000). Information Technology Newsletter, 5(8). University Digital Conservancy. https://hdl.handle.net/11299/60522

Endnotes

[1] Diana L. Magnuson is curator and historian at the Institute of Social Research and Data Innovation (ISRDI) at the University of Minnesota. Magnuson can be reached by email at magn0031@umn.edu.
[2] For a descriptive history of the creation and use of the IPUMS Extract System, see Magnuson and Ruggles 2022.
[3] This work was initially funded through a cooperative project with the World Bank and funded by the Organization for Economic Cooperation and Development, Development Cooperation Directorate (OECD/DCD-PARIS21, JADE#:60525;MEHLB(2010)12).
[4] https://documents.ipums.org/

[5] https://ddialliance.org/ddi-codebook#v1.x
[6] https://www.ipums.org/impact/working-papers and https://pop.umn.edu/research/working-papers
[7] https://www.ipums.org/impact/working-papers
[8] https://pop.umn.edu/research/working-papers
[9] https://lcc.umn.edu/research/working-papers