

Member Node Description: Minnesota Population Center

Version 1.0 11/19/14 Catherine Fitch

General

Name of resource:	Minnesota Population Center (MPC)
URL(s):	www.popdata.org
Institutional affiliation(s):	University of Minnesota
Primary geographic location:	Minneapolis, Minnesota, United States
Project Director & contact info:	Steve Ruggles Director, Minnesota Population Center ruggles@umn.edu
	Cathy Fitch Associate Director, Minnesota Population Center fitch@umn.edu
Technical Contact & contact info:	Wendy Thomas wlt@umn.edu
Age of resource:	MPC was funded in 2000, IPUMS has disseminated data since 1995
Funding support:	Federal grants from the National Science Foundation (NSF) and National Institutes of Health (NIH)
Proposed Unique Identifier:	urn:node:US_MPC

Content

Content description/collection policy (1 paragraph, domain and spatial/temporal coverage, uniqueness of content, exclusions, as applicable):

MPC disseminates integrated census and survey data from the U.S. and around the world. These resources describe the populations of more than 70 countries and are designed to be interoperable for spatiotemporal comparisons. We current disseminate census microdata--data describing individuals and households--from 72 countries for the period 1960 to the present; for nine countries, we have census microdata from the nineteenth century. We disseminate integrated versions of four high-value demographic and health surveys from the United States. We also disseminate small-area summary data from the United States for places such as tracts and counties, and boundary files describing these places within the U.S. for the period 1790 to present. MPC also disseminates additional small-area summary data as well as environmental data, such as land cover, land use and climate information, for the whole globe. MPC data resources are described at www.popdata.org.

Types of data (complex objects, text, image, video, audio, other):

MPC disseminates 4 kinds of data: textual metadata, microdata (describing persons and households), aggregate summary data (describing places), and raster data. We plan to disseminate metadata and selected boundary and raster data through DataONE. For other data, users will be directed to our websites.

Data and metadata availability (rights, licensing, restrictions):

MPC policy is to keep data access as open as possible. Metadata for all of our projects is freely available. All projects require user registration to download data. For some projects registration is automatic

once agreeing to common terms and conditions. Some projects must restrict usage to approved research projects and teaching applications.

Option for embargo (yes/no, duration):

Data do not have an embargo period.

Size of holdings (number and size of datasets, mean and median granules (files) per dataset):

Holdings for each of the MPC projects are described on the following websites; size of each dataset is also noted. We do not plan to include these data in DataONE.

IPUMS Interational (670 GB): <https://international.ipums.org/international/samples.shtml>
IPUMS-USA (287 GB): <https://usa.ipums.org/usa/sampdesc.shtml>
IPUMS-CPS (146 GB): https://cps.ipums.org/cps/sample_sizes_all.shtml
North Atlantic Population Project (NAPP) (108GB): <https://www.nappdata.org/napp/samples.shtml>
Integrated Health Interview Series (IHIS) (96GB): <https://www.ihis.us/ihis/surveys.shtml>
American Time Use Survey Extract Builder (ATUS-X) (99GB):
<https://www.atusdata.org/samples.shtml>
National Historical Geographic Information System (NHGIS) (2.4TB):
<https://www.nhgis.org/documentation>

We plan to include selected data from TerraPop in DataONE. The full TerraPop collection currently comprises 455 GB. Complete holdings are outlined on the Datasets tab at: <https://beta.terrapop.org>

Please describe recent usage statistics, if known, including information on annual data product downloads, annual number of users, annual number of data products used in publications:

Project: Vists / Unique Visitors/ Page Views/ Extract Downloads*
IPUMS-USA: 200,545 / 99,114 / 1,845,635 / 31,056
IPUMS-CPS: 56,807 / 24,517 / 580,998 / 9,307
IPUMS Interational: 55,653 / 30,510 / 580,748 / 7,717
NAPP: 8,291 / 5,529 / 58,120 / 437
IHIS: 23,794 / 11,568 / 248,552 / 2,514
ATUS: 8,459 / 4,249 / 58,767 / N/A
NHGIS: 98,033 / 64,680 / 490,117 / 20,072

*Everything measured over April 2012 - March 2013 except downloads, reported for 2012 calendar year

User interactions

How does a user contribute data? (what can be deposited, how are data prepared, are specific software required, documentation/support available)

Users cannot deposit data. Data is sought that supports specific projects along with supporting documentation. Data are validated and integrated into the specific project. Level of integration is project dependent. Documentation on processing within a project is available from site.

How does a user acquire / access data?

Users register to use project website, and then create customized data extracts for download. Our data extraction systems create datasets based on selections the user makes (e.g., country, years, variables,

and case selections are available). Data are offered in several formats: fixed-width text data file, comma delimited (.csv), Stata (.dta), SAS (.sas7bdat) or SPSS (.sav). Metadata available in DDI codebook or ISO-19115. All websites have comprehensive documentation.

What user support services are available (both for depositing and accessing/using data)?

User support is available for all projects. We encourage email questions at ipums@umn.edu; email is answered during business hours. An answerbase website contains many common questions and allows users to pose questions to other data users.

How does the resource curate data at the time of deposit?

Data are reviewed for internal consistency, reformatted for MPC processing and codes are harmonized for interoperability with other datasets.

Technical characteristics and policies

Software platform description, incl. data search and access API(s):

Our webapps are written in Ruby 1.9 and utilize the Rails 3 web application framework and run on the Ubuntu linux distribution. Our websites offer a search function to find variables and documentation, but we do not offer search across our websites. Right now there is no API to the data or metadata but we are working on several

Service reliability (including recent uptime statistics, frequency of hardware refresh, if known):

Uptime on MPC websites is 99 - 99.5%. We have an annual datacenter shut down with advance warning that requires a 24-hour downtime. Our aim is to refresh hardware every 3 years, but with changes in technology our infrastructure will be virtualized and hardware refresh will be less relevant.

Preservation reliability (including replication/backup, integrity checks, format migration, disaster planning):

MPC will be improving our preservation reliability in the next two years. We have current replication/backup and disaster plans, but we will extend and improve them.

User authentication technology (incl. level of create/modify/delete access by users):

We have implemented custom password-based authentication within our application code. We store the passwords in a one-way hashed form within the database. As mentioned above, authorization is managed within each project; a user has to register for each project separately before getting access to that project. We are working on a single-sign-on system that will consolidate access for all projects that a user has registered for. Each user has the ability to create/modify his or her own customized data extracts; user's data extracts are eventually deleted from our website, but request details are saved and users can re-create the extract. Users cannot modify or delete the underlying data.

Data identifier system and data citation policy, if available:

Users agree to properly cite the data when they register. Data citation is available from each project website and included in email notifications when data extracts are ready for download. We do not yet use a DOI system but will be implementing it in the next year.

Metadata standards (including provenance):

Descriptive content is consistent with the Data Documentation Initiative (DDI) and available as a DDI XML document. Spatial metadata is consistent with ISO/IEC 19115. Bibliographic records of data, metadata, and related documents is expressed in Dublin Core (qualified DC). Provenance and preservation specific metadata is currently modeled on PREMIS (an implementation of OAIS). We are exploring a cooperative provenance/preservation management system with the University of Minnesota Libraries and may change to comply with that system (will be OAIS based).

Capacity/services to DataONE

At what functional tier will you initially be operating? (see <http://bit.ly/MNFactSheet> for definitions)

- Tier 1: Read only, public content
- Tier 2: Read only with access control
- Tier 3: Read/write using client tools
- Tier 4: Able to operate as a replication target

If you can host data from other member nodes, what storage capacity is available?

Can you provide computing capacity to the broader network? If so, please describe.

Other Services

What other services or resources (such as expertise, software development capacity, educational/training resources, or software tools) can be provided of benefit to the broader network?

MPC experts could possibly provide expertise and training related to the Data Documentation Initiative (DDI) and to the use of census and survey data.