

Holocaust Ghettos Project Data Management Plan

Roles and responsibilities: The Advanced Computing Group at the University of Maine (ACG@UMaine) will provide computer and data services, including backup and long term archiving. The ACG@UMaine's primary data repository is the Maine Dataverse Network (MDVN). Data produced by the individual researchers on this project will be stored on this resource. ACG@UMaine services include training and support for all technical aspects of the data produced during the course of this project necessary to maintain security, dissemination, and preservation. The PI will have decision-making authority over all data management. Those who generate data as a result of this project are responsible for adding it to the MDVN repository immediately after publication. It is the responsibility of the individual PI, Co-PIs, and Consultants to NOT add data to the repository that violates privacy, confidentiality, security or intellectual property concerns. The archival life cycle and retention policy for archived data will be managed by ACG@UMaine. Data will be retained indefinitely after the end of the project until or unless doing otherwise is necessary to meet legal obligations or adhere to an established data management policy. The PI will check adherence to this plan at least 90 days prior to the expiration of the award. Adherence checks will include review of the MDVN content, number of studies released, availability for each study of subtable/preservation friendly data formats (possibly embargoed, but listed); availability of documentation (public); and correctness of data citation, including an integrity check.

Expected data and period of data retention: The data generated by this project will be in a variety of formats, including program code, databases, images, shapefiles, text, and binary data. Once cleaned and finalized, all data will be shared with the scholarly community via the MDVN. All associated metadata will also be shared, including sources, analytical methods, and rules for data entry. Several database management systems are available on the data server. We plan to use MySQL for this project, although alternative DBS solutions will provide flexibility if needed. All data are stored native in journal-based file systems ext4 with a configuration of RAID-5 underneath. For program codes, version control systems, such as git and cvs, are available to record the history of changes. There are no known legal or ethical restrictions on access to non-aggregated data generated in this project. During the project, all data files will be stored, maintained, and regularly backed up daily on Seafile, UMaine's cloud storage system, which is maintained by ACG@UMaine and will be accessible to all team members.

The primary tool for providing data retention on this project is the Maine Dataverse Network (MDVN). The MDVN is a public repository, which runs on secure servers hosted and maintained by the ACG@UMaine. Data will be deposited into the MDVN as soon as it has been cleaned and checked by the PI and the Co-PI or Consultant responsible for its creation, in every case within six months of the completion of data analysis. Data will be retained indefinitely, never destroyed (unless required by law).

Data formats and dissemination: The primary tool for providing access to shared data produced by the individual researchers on this project is the MDVN. The MDVN facilitates access to data through descriptive and variable/question-level search; topical browsing; data extraction and re-formatting; and on-line analysis. All data will be deposited within six months of the completion of data analysis. Such data may be embargoed until the publication of research based on the data or until three years after the expiration of the award, whichever is sooner. Users will be required to agree to click-through terms at the discretion of the PI that prohibit unlawful uses and intentional violations of privacy, and require attribution. Use of the data will be otherwise unrestricted and free of charge.

Data generated by this project can be released without privacy restrictions. The data extracted from the U.S. Holocaust Memorial Museum's *Encyclopedia of Camps and Ghettos* does not constitute private information about identified human subjects. The data extracted from transcripts of Holocaust survivor video testimony comes from sources for which the collecting organizations (USHMM and the USC Shoah Foundation's Visual History Archive) have permission to distribute the oral histories to the

public. The data will not be encumbered with intellectual property rights (including copyright, database rights, license restrictions, trade secret, patent or trademark) by any party (including the investigators, investigators' institutions, and data providers); nor are the data subject to any additional legal requirements. Depositing with the MDVN does not require a transfer of copyright, but instead grants permission for the Maine Dataverse Network to re-disseminate the data and to transform the data as necessary for preservation and access. Access to restricted data at any point in the data life cycle may be granted through MDVN request for access function. At the discretion of the PI, data will be shared on the MDVN and an email sent to the requestor with the access information. Requestors are required to agree to click-through terms and conditions of reuse and re-distribution (e.g., authorship, acknowledgment, citation use). Terms and conditions are included in associated data files on MDVN and determined by the PI on a case-by-case basis. The PI, Co-PI, or Consultant responsible for each part of the project will create documentation detailing the sources, coding, and editing of all data, in sufficient detail to enable another researcher to replicate them from original sources; and descriptive metadata for each study including a title, author, abstract, descriptive keywords, and file descriptions. The PI will coordinate metadata creation to ensure consistency. The project will include bibliographic information for any publication by the project based on that data. Sound, videos, and image files will be maintained in WAV, JPEG2000, and TIFF formats. Other documentation will be deposited in PDF/a, or plain-text formats, to ensure long-term accessibility, including a readme file in each directory to describes detailed metadata information, such as data formats, access methods, usage, authors, notes, and revision history. Quantitative data will be converted to CSV and ESRI shapefile (SHP) format. These formats are fully supported by the Maine Dataverse Network (MDVN), which performs archival format migration; metadata extraction; and validity checks. Deposit in these formats will also enable on-line analysis; variable-level search; data extraction and re-formatting; and other enhanced access capabilities. The MDVN repository system's "templating" feature will be used for consistency of information across studies. The MDVN system automatically generates persistent identifiers, and Universal Numeric Fingerprints (UNF) for studies; extracts and indexes variable descriptions, missing-value codes and labels; creates variable-level summary statistics; and facilitates open distribution of metadata with a variety of standard formats (DDI v 2.0, Dublin Core, and MARC) and protocols (OAI-PMH and Z39.50).

Data storage and preservation of access: The archiving and preservation of all data will be managed by ACG@UMaine, whose primary data repository is the Maine Dataverse Network (MDVN). The MDVN is hosted by ACG@UMaine on a secured shared storage array acquired under the funding support of a previous NSF MRI grant. The storage array is fast and reliable, with a total of 120 TB of raw storage interconnected with multiple 4 Gb/s Fiber Channel paths and offsite backups. ACG@UMaine commits to good archival practice including 24x7 machine operation support versioning and deaccession compliant, regular off-site backup, and regular content migration to ensure that data are available for access consistently and kept secure. All data will be stored, during and post project, within the UMaine network infrastructure, which employs firewalls and secure authentication and authorization methods for login and access. Data deposited into the MDVN will be retained indefinitely, never destroyed (unless required by law), backed up on a daily basis and replicated across multiple locations for long-term access. Long-term access to data will be provided in accordance with the permissions, terms of use, policies and procedures established by the PI in consultation with the research team at the time of archiving. ACG@UMaine services include personnel training and support for all technical aspects of the data produced during the course of this project necessary to maintain security, dissemination, and preservation. Should the archiving entity be unable to perform, transfer agreements with the University of Maine System are in place to easily migrate to another entity within the university system.