



Research Data Services | DMP Example

Prepared for	University of Maryland, College Park
Prepared by	UMD Libraries Karl Nilsen
Contact	knilsen@umd.edu 301-405-9204
Date	2015-09-08
Link	<a href="http://lib.umd.edu/data">http://lib.umd.edu/data</a>
License	This work is licensed under a <a href="#">Creative Commons Attribution 3.0 Unported License</a>
Purpose of this document	<ul style="list-style-type: none"><li>• This is an example of a data management plan. It is not intended to provide boilerplate language, but to illustrate typical content and writing style. Your plan should reflect the unique characteristics of your research and the technology infrastructure available to you.</li><li>• The example plan is not designed to meet the requirements of any particular funding agency or organization, but it contains information that is often found in successful data management plans.</li><li>• The example plan is not based on a real research project, though it contains elements that are common to many research projects.</li></ul>

## **Data Management Plan**

### **I. Roles and responsibilities**

The PI will be primarily responsible for data management, quality, and dissemination. Students supervised by the PI will receive instruction in data management and guidance on technical and ethical practices. The PI will periodically review student work for quality control. In the event that the PI leaves the institution, responsibility for data management will remain with the PI.

### **II. Expected data**

This project will generate a number of research products:

1. Journal papers, technical reports, and presentation slides in PDF format. PDF is widely used in our field and facilitates efficient re-use and replication.
2. Movement data will be collected from participants' smartphones and combined with GIS and demographic data from publicly available municipal sources. De-identified data will be processed and analyzed in MATLAB and ArcGIS data formats, but disseminated to researchers and the general public in platform-independent format such as CSV (comma-separated values) and KML. CSV and KML are widely used and facilitate long-term access to these data.
3. Fully-commented software code will be written in the MATLAB script format (\*.m).

All data management activities will be consistent with the terms of our IRB. While the study is being conducted, the team will store data, code, and other materials on password-protected servers maintained by the university's IT division. Movement data that contain direct or indirect identifiers will be encrypted for storage and only decrypted for the purposes of de-identification through perturbation and/or aggregation. Backup copies will be produced nightly by the university's IT division. De-identified data will be analyzed on the PI's and students' password-protected laptop computers.

### **III. Documentation and metadata**

Data and software code will be accompanied by plain-text 'readme' files containing any information necessary for re-use, replication, and validation that is not automatically embedded in the data files themselves. Data documentation will include source information, processing and transformation steps, definition of variables, codes, and flags, and other information. Code documentation will include required libraries and packages, system and configuration instructions, and version information.

### **IV. Public access to research products**

Publication of data and other research products shall occur during the project, if appropriate, or at the end of the project, consistent with practices in our field. De-identified data that document, support, and validate research findings will be made available when the main findings have been accepted for publication. However, data from publicly available municipal sources will not be re-distributed. The documentation will describe those sources.

Since the analyses will be conducted on de-identified data, none of the data disseminated by the research team will contain direct or indirect identifiers of human subjects. In the event that reviewers or researchers wish to examine higher-granularity data, they will be required to conform to the terms of a data-use agreement mandating security, distribution, retention, and destruction practices.

Research products from this project will be archived at the Digital Repository at the University of Maryland (DRUM) unless a more appropriate facility can be identified. DRUM is a long-term, open-access repository managed and maintained by the University of Maryland Libraries. Researchers and the general public can download data and code files, associated metadata and documentation, and any guidelines for re-use. All records in DRUM are assigned a persistent DOI to support consistent discovery and citation. The project description will be automatically indexed in Google and Google Scholar to support global discovery. Whenever possible, data curation specialists in the University Libraries work with researchers to document and format materials for long-term access.

## **V. Conditions on re-use and re-distribution**

For the research products available to the public, there are no conditions on re-use or re-distribution beyond the attribution and credit norms of the scientific community. As indicated above, in the event that reviewers or researchers wish to examine data containing direct or indirect identifiers, they will be required to conform to the terms of a data-use agreement. The agreement will prohibit re-distribution of sensitive data.

## **VI. Long-term archiving and preservation**

The research products archived in DRUM will be available indefinitely. The University of Maryland Libraries' DRUM repository is built on DSpace software, a widely used, reliable digital repository platform. DRUM performs nightly bit-level integrity tests on all files, and all contents are regularly copied to back-up storage. DRUM conforms to the digital preservation principles outlined in the University of Maryland Libraries' Digital Preservation Policy.

The PI will retain encrypted copies of the raw smartphone data for seven years after the completion of this project. This retention period satisfies the University of Maryland's records management policy and satisfies the Office of Human Research Protections' requirements for retention of IRB documents (45 CFR 46.115(b)). After seven years, the raw data will be destroyed.