

Samples

Question 2: How will the data be used and managed in your research project?

SAMPLE 1:

Data originally recorded on paper will be transferred into spreadsheets using .csv formats). Data will be checked for outliers in the R statistical program, and any outliers will be checked for transcription errors. DGVM simulation runs will be performed on a high performance parallel computing platform, a 96-node Linux cluster, maintained jointly by USFS Pacific Northwest Research Station and Oregon State University. DGVM output will be analysed and displayed with the ESRI ArcGIS software suite.

As the data will be generated, processed and analysed by different project team members, I will recommend the project team members to name the data file by using their name initials, date and version, e.g. LGH_20150801_v1.

Adapted from:

Cleland, E., Lipson, D., & Kim, J. The influence of plant functional types on ecosystem responses to altered rainfall. Retrieved Nov 24, 2015, from UC San Diego Sample NSF Data Management Plans website:

<https://libraries.ucsd.edu/services/data-curation/data-management/dmpsample/DMP%20Example%20Cleland.pdf>

SAMPLE 2:

The audio and video recordings will be saved in common standardized formats, e.g. WMV and the photographs will be saved in JPEG format. The interview recordings will be transcribed into Luganda and then translated into English. Both translations will be saved in word documents. The English translated interview will be coded by using the ethnographic software.

The raw data will all be stored in a folder titled "Raw data_YYYYMMDD"; the processed or analysed data will be kept at different folders by data type, e.g. all audio recordings will be saved in the same folder and video recordings will be stored at another folder. We will be using the following file-naming convention for each data file and folder:

- data file name: Subject_v1 (e.g. interview_v1)
- folder name: datatype_v1_YYYYMMDD (e.g. audiorecordings_v1_20151120)

Adapted from:

Sapp Nelson, Megan and Beavis, Katherine (2013) "History / Sustainable Development - Purdue University," *Data Curation Profiles Directory*: Vol. 5, Article 1. <http://dx.doi.org/10.7771/2326-6651.1032>

SAMPLE 3:

All acquired original data from government sources in different format, such as .xls, .csv, .html, .shp, etc will be loaded and appended to existing time series in the MS SQL database. After that, all data will be aggregated by economic regions. Mapping will be carried out and created by using ArcGIS. Eventually, the data will be made available for searching and retrieving (querying and visualising) in html on a website. Analytical reports are also made available for downloading on the website. SAS, Excel, ColdFusion (query, retrieve and display data from the MS SQL), Google Chart API (to visualized data in various forms, e.g.

population pyramids and trendlines) and Google Map API (to display dynamic maps of the data) will be used in the process of making the data for public to access on the website.

JISC has provided a guide on choosing a file name. We will name our data files based on the recommendations available in this website: <http://www.jiscdigitalmedia.ac.uk/guide/choosing-a-file-name>. All data files will be stored in different folders organised by researchers' initials and date.

Adapted from:

Jenkins, Keith (2012) "Sociology / Demographics - Cornell University," *Data Curation Profiles Directory*: Vol. 4, Article 6. <http://dx.doi.org/10.5703/1288284315013>

SAMPLE 4:

The audio files in wav format will be transcribed by using a transcription software ELAN and saved in Microsoft Word document. Both version of data will be cleaned up and normalised. Metadata will be created using morphological glosses so that the textual files can be searched for either the native language or English or the gloss for specific linguistic features and then call up the audio files. Since data will be added each year, the data will be organised by folder with year indicated. Ultimately, the data will be ingested in a publicly accessible searchable database or repository. The audio and the transcription are synchronised.

The data will be organised and stored in different folders with the following file-naming convention: Subjectkeyword_V2_YYYYMMDD; Subjectkeyword_V2_YYYYMMDD...

Adapted from:

Tancheva, Kornelia (2012) "Linguistics - Cornell University," *Data Curation Profiles Directory*: Vol. 4, Article 7. <http://dx.doi.org/10.5703/1288284315007>

SAMPLE 5:

The data will be analysed which involves generating proprietary files for processing software and convenient printable formats for manually examining the data, for example Excel spreadsheets or PDF files. The pressure trace graphs and chromatographs are the focus of analysis. Hence, pressure trace data will be analysed using Matlab, Excel and Adobe Portable File Format. Chromatograms will be interpreted for Clarity software. Some graphs on Arrhenius plots and concentration plots will be generated using Origin software. The video from the experiment will be used primarily for verification that the experiment ran correctly. Video stills will be generated from the video files and will be merged with some graphs using Photoshop.

We will store all the data in a shared drive and will name each file by the following file-naming convention:

- 20140603_MAEProject_DesignDocument_Tan_v2-01.docx
- 20140809_MAEProject_MasterData_Daniel_v1-00.xlsx
- 20140825_MAEProject_Ex1Test1_Data_Jason_v3-03.xlsx
- 20141023_MAEProject_ProjectMeetingNotes_Kumar_v1-00.docx

Adapted from:

1. Kashyap, Nabil (2011) "Aerospace Engineering / Chemical Kinetics - University of Michigan," *Data Curation Profiles Directory*: Vol. 3, Article 1. <http://dx.doi.org/10.5703/1288284314989>

2. Brandt, S. (29 July 2015). Data Management for Undergraduate Researchers: File Naming Conventions, from <http://guides.lib.purdue.edu/c.php?g=353013&p=2378293>

SAMPLE 6:

The Gas Chromatography- Mass Spectrophotometry (GC-MS) data will be analysed using the instrument specific proprietary software to measure the area underneath the peaks for specific known Volatile Organic Compounds (VOCs). The peak area will then be entered into an Excel spreadsheet along with the field survey data. Statistical analysis of the data will be performed using StatView to prepare the tables and graphs for the research.

We have not decided on how the data files will be organised yet. However, we will follow the file naming conventions recommended by the Stanford University Libraries (<https://lib.stanford.edu/data-management-services/file-naming>) to name our data files.

Adapted from:

Wright, Sarah J. (2012) "Environmental Science / Herbivory - Cornell University," *Data Curation Profiles Directory*: Vol. 4, Article 3. <http://dx.doi.org/10.5703/1288284315002>

SAMPLE 7:

The data will be moved to Excel for automated and filtering to removing errors and noise which occur due to the system being sensitive to light (e.g. reflections) and motion marker occlusion. More automatic threshold- based filtering will be carried out after that, along with visual review of the data and manual cleaning. This process will take place in Matlab and the data will eventually be converted to represent several variables (e.g. angle data, displacement velocity, or acceleration of joint segments). The data will then be aggregated across subjects and will be stored in an Excel spreadsheet. The data will be organised through a file folder system where each trial will be documented in a single spreadsheet, and all the files from particular study will be stored in the same folder structure.

Adapted from:

Cragin, Melissa; Kogan, Marina; and Collie, Aaron (2011) "Bio-Mechanics Motion Studies - University of Illinois Urbana-Champaign," *Data Curation Profiles Directory*: Vol. 3, Article 6. <http://dx.doi.org/10.5703/1288284314998>