Plan Overview

A Data Management Plan created using DMPTool

Title: Digital Curation Innovation Center Data Management Plan

Creator: Mary Kendig

Affiliation: University of Maryland, College Park (umd.edu)

Principal Investigator: Mary Kendig

Data Manager: Mary Kendig

Funder: National Science Foundation (nsf.gov)

Funding opportunity number: 21243

Template: NSF-CISE: Computer and Information Science and Engineering

Last modified: 05-06-2016

Copyright information:

Digital Curation Innovation Center Data Management Plan

Data management responsibilities are under the jurisdiction of the two graduate assistants employed by the Digital Curation Innovation Center (DCIC). The two graduate assistants will check the adherence to the data management plan on a monthly basis. For Fall 2016 and Spring 2017, the data management plan will be monitored and implemented by Myeong Lee and Mary Kendig.

In addition to monitoring and implementation, the graduate students will facilitate the short and long term preservation of digital assets. Short term preservation consists of coordinating with project members and faculty to confirm data uploads into UMD Box. The graduate assistants will monitor the organization and activity within UMD Box. Graduate assistants are expected to dedicate at least an hour per week to ensure that data is uploaded and organized into UMD Box. For long term preservation, the graduate students will facilitate the ingestion of data from UMD Box into the DCIC's long term archival storage repository Indigo at the end of every semester. This will involve coordination with the Center Directors and the Research Software Architect. At the creation of this data management plan, the Center Directors are Dr. Richard Marciano, Dr. Michael Kurtz, Dr. Kenneth Heger, and Dr. Ricky Punzalan. The Research Software Architect is Greg Jensen.

To ensure transferability of responsibility, the graduate assistants will be required to train the next set of graduate assistants. Prior to this training, the graduate assistants and Center Directors must meet to discuss the viability the data management plan. If the plan continues to support the center's activity, then the graduate assistants can train the students. If not, the plan must be updated and then training can occur. Decisions about data will fall under the DCIC's current Center Directors, regardless of original personnel's availability.

Data for the DCIC varies by projects. Common data types include archival images, spreadsheets, software code files, reports, and website links to final products. Captured data includes JPEG or PDF images of the archival documents. Research planning documents, spreadsheets, software code, reports, and websites links will be generated from project member creation.

To supplement captured data, existing data may be collected. Due to potential unstructured and uncurated data, the DCIC will be implementing the extensible data transformation service Brown Dog. This service will not only be utilized to access current legacy data or unreadable data formats, but to ensure future access to the data types presently recognized. To assure quality control, original data copies will be maintained in a UMD Box cloud folder accessible only to Center Directors and graduate assistants. A duplicate folder will be available for project member, faculty, and student use.

As an information institution, the DCIC is expected to grant public access to all collected data. The data will be accessible through three different public data repositories and one information medium. Data will be made available publically through GitHub and Figshare. This data will be organized by project. In addition to public access, the data will be made accessible to university students and faculty through the Digital Repository for University of Maryland Research (DRUM). Finally, every project is required to maintain a project blog/website. At the end of the semester, a compressed folder will be embedded into the project website for user download.

Before sharing the data, the DCIC must consult with the Maryland Intellectual Property Legal Resource Center to understand the legal, ethical, and policy issues associated with DCIC technology and intellectual property. In addition to consulting with the Maryland Intellectual Property Legal Resource Center, the graduate students will be required to skim through the data to remove PII or potentially damaging information (emails, addresses).

Short term strategy for storing the data is the University of Maryland's free cloud base service UMD Box. As

project members capture and generate data, they will upload their data onto the project's cloud folder.

Long term data storage will be facilitated through the DCIC's own data repository software called Indigo. Indigo is a record manager platform based on Apache Cassandra and software developed by Archive Analytic Solutions. Partnering with NetApp Data Management/Cloud Storage Solutions, the DCIC has built a Cyberinfrastructure Center with 4 Dell Servers and 3 NetApp E-Series storage arrays for a total of 720 terabytes of raw storage. The Indigo/Cassandra software works in conjunction with this facility and is designed for long-term archival storage and preservation. As noted in roles and responsibilities, data will be ingested into Indigo at the conclusion of the semester.

The data selected for long term preservation includes the original archival images, the extracted data from images, software code, planning documents, archival metadata, and final reports.

(leave open for Dr. Marciano/Jensen comments)