A Data Management Planning Tool for Understanding and Reporting Data about Data

Joshua E. Massey and Dr. Kimberly A. Tryka
National Institute of Standards and Technology

Introduction

When conducting a research project, it is essential to have a method for collection, maintenance and preservation, and publication of all associated data. Many research scientists compile this information the form of a data management plan.

A data management plan (DMP) is a document that describes all aspects of a research project including a description of the project, methods used, how the project's data will be recorded, and how the results will be preserved and made available for future reference. Many funding sources and all federally-funded research projects require a DMP.

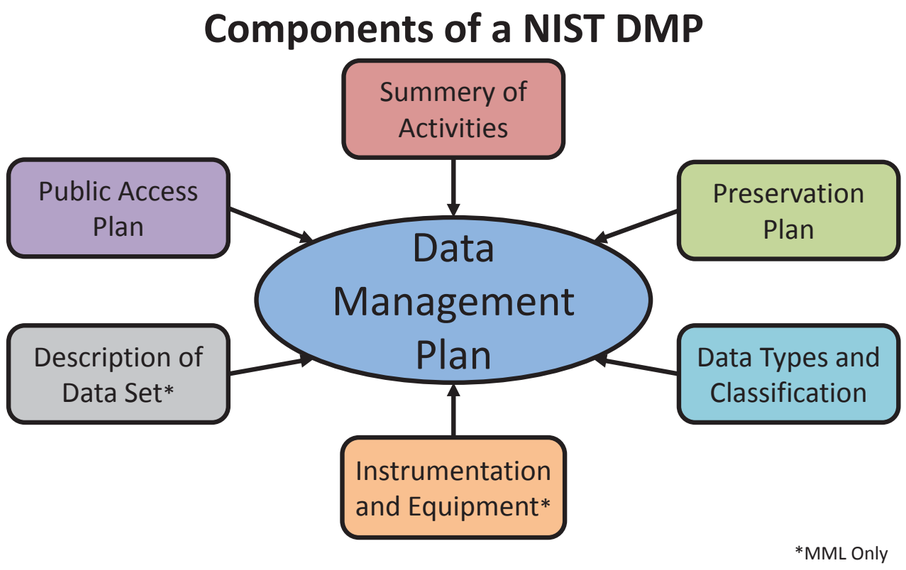

Figure 1. This figure depicts the four major sections of a NIST data management plan: 1) The Summary of Activities details the steps taken throughout the project; 2) The Preservation Plan explains how data will be collected and maintained for future use; 3) The Public Access Plan states how the data will be released to the public; 4) Data Types and Data Classification provide information about formats. Additionally, the MML collects information about Instrumentation and Equipment used and a specific Description of Data Set information (such as the creation date and title of the data set).

The MML Data Management Planning Tool

In Spring 2015, the Office of Data and Informatics (ODI) of the NIST Material Measurement Laboratory (MML) created a data management planning tool to be used for the laboratory. It was established as a centralized location where MML researchers could enter details about all of their projects and research activities. The tool also serves as a way for MML management to monitor the activity of the divisions and groups. In Fall 2015, the tool was given the permanent name *Minerva.*

Project Goals:

1. **Develop a reporting tool to better understand content of the MML Data Management Plans.**
2. **Analyze and interpret results to understand how researchers interact with the DMP tool so that it can be improved upon.**

Reports

The main aspect of the project was to create an algorithm that output query results in a format from which reports could easily be generated. Two of the reports that were generated include a divisional summary and an employee summary.

Employee	# Creator	# Contributor	# Contact
Employee 01	1	0	0
Employee 02	0	0	0
Employee 03	0	0	0
Employee 04	0	0	0
Employee 05	1	0	0
Employee 06	1	0	0
Employee 07	0	0	0
Employee 08	0	0	0
Employee 09	0	0	0
Employee 10	0	0	0

Employee 11

ID#	Title	Description	Keywords
133	EPMA WDS using JEOL JXA-8500F	Electron probe microanalysis (EPMA) wavelength dispersive spectroscopy (WDS) raw data, including peak and background intensities for unknown materials and for reference materials, peak intensity maps, and wavelength-scanned spectral intensities, are collected and stored in a proprietary format by Probe for EPMA software.	electron probe microanalysis, epma, wavelength dispersive spectroscopy, wds
134	EPMA EDS using JEOL JXA-8500F	Electron probe microanalysis (EPMA) energy dispersive spectroscopy (EDS) raw data, including spectral intensities as a	electron probe microanalysis, epma, energy dispersive spectroscopy, eds

Figure 2. These are samples of reports generated from the output of the algorithm. The report sample on the left shows the employees from an entire division/group and the number of DMPs on which they are listed as a contact. The report sample on the right shows more detailed information with DMP details broken down by employee.

Visualizations

Visualizations were found to be a more appropriate reporting form when trying to get an idea of the relationships across the entire data set and amongst different fields in the data such as projects and people.

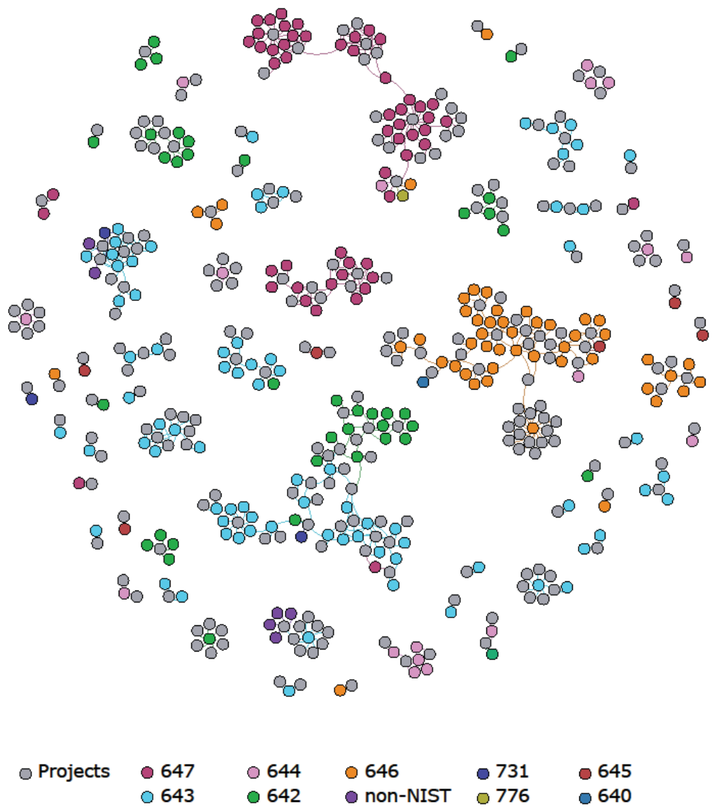

Figure 3. This visualization depicts the employees of different divisions (colored dots), the projects they are associated with (gray dots), and the relationships between people and projects. It shows many collaborative projects (some projects are connected to many employees) but it also many cases where there is a one-to-one (or one-to-many) employee-to-project relationship.

Summary

1. **Developed a Python algorithm that generates metadata that reports and visualizations can easily be generated from.**
2. **Demonstrated how the DMP tool is being used by the laboratory for further refinement and expansion of the tool.**

Acknowledgements: Supported by the NIST Research Library through the NIST Summer Undergraduate Research Fellowship directed by Dr. Brandi Toliver.