DM Vitals: A Data Management Assessment Recommendations Tool

Sherry Lake, Senior Scientific Data Consultant, shlake@virginia.edu
Scientific Data Consulting Group, University of Virginia, Library,
University of Virginia, Charlottesville, Virginia, U.S.A.

Susan Borda, Graduate Student Summer Intern, shborda@syr.edu
School of Information Studies,
Syracuse University, Syracuse, New York, U.S.A

SCIENTIFIC DATA CONSULTING GROUP

Abstract:

In an effort to develop an understanding of how science and engineering researchers at the University of Virginia (UVa) manage their research data, UVa Library's Scientific Data Consulting (UVa SciDaC) group began a series of research data interviews. The goals of the data interview process include identifying common research data problems, identifying research support needs, and providing recommendations on improving data management. In practice, however, providing objective suggestions for data management practices proved to be troublesome. It was difficult to make reliable customization recommendations; be objective in a timely fashion.

In response to these challenges, the UVa SciDaC group developed a system (DM Vitals) that would easily and objectively rate the current state of the researcher's data management practices. Using best practice statements from UVa sources (Information Technology Services' Risk Management Program and SciDaC guidelines) and the Australian National Data Service's (ANDS) long-term sustainability scoring model, the system compares the information collected during the data interview process with these data management best practice statements. The model then further correlates the researcher's data management practices with the eight data management practice components developed by the SciDaC group: File Formats and Data Types, Organizing Files, Security/Storage/Backups, Funding Guidelines, Copyright & Privacy/Confidentiality, Data Documentation & Metadata, Archiving & Sharing and Citing Data.

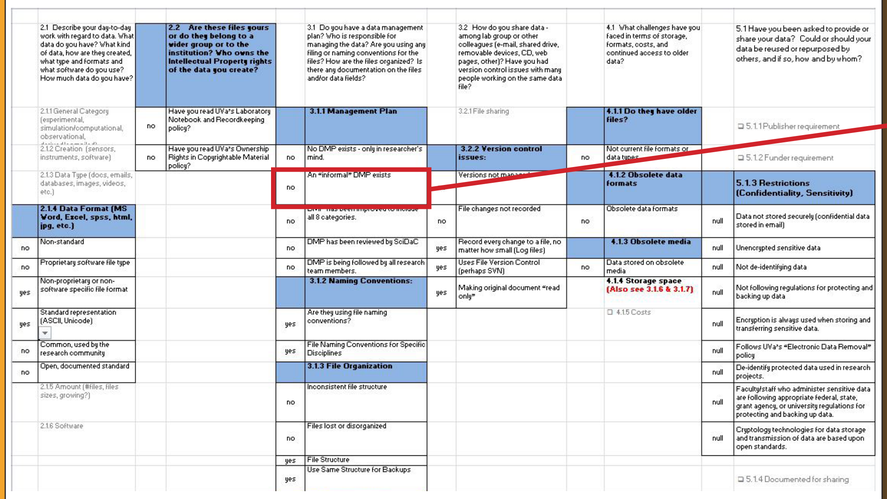
To provide a framework for comparing and improving departmental data management practices, we took the value resulting from the average of the data management best practice statements and compared them the Crowston and Qin Capability Maturity Model (CMM). Using this model as a basis, the data management maturity levels are defined as: Level 0: Initial (this includes current practices that can be seen as counter-productive or even "risky" from a security standpoint), Level 1: Managed (the researcher begins to uniformly apply some of the lower level/easier best practices, really starting to "manage" the situation), Level 2: Defined (the researcher is further "defining" their DM practices), Level 3: Quantitatively Manage (the researcher begins to use central and outside services to manage their data), and Level 4: Optimizing (the researcher are continually improving their data management practices).

The strength of the DM Vitals tool is in generating tasks customized to each researcher. These tasks can then be easily grouped into phases, creating a data management implementation plan for each researcher based on their personal data interview and subsequent information gathering. The DM Vitals assessment tool differs from the Digital Curation Center's (DCC) CARDIO in that its focus is not on consensus and collaboration by individuals responsible for the research data (PI, IT, Data manager, etc.).

Once this tool is fully integrated into the existing UVa SciDaC Data Interview and Data Management Plan process, it will expedite the recommendation report process by providing valuable actionable feedback that the researcher can use immediately to improve the sustainability of their data.

UNIVERSITY of VIRGINIA LIBRARY

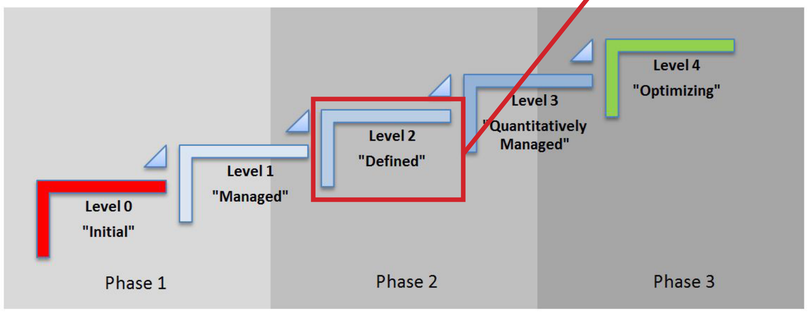
Interview Questions:



- **UVa's "Data Interview Question Template Sections" are split into discrete questions.**
- **Data management practices are added to these questions, as sub-statements with three possible answers: no, yes, or null.**

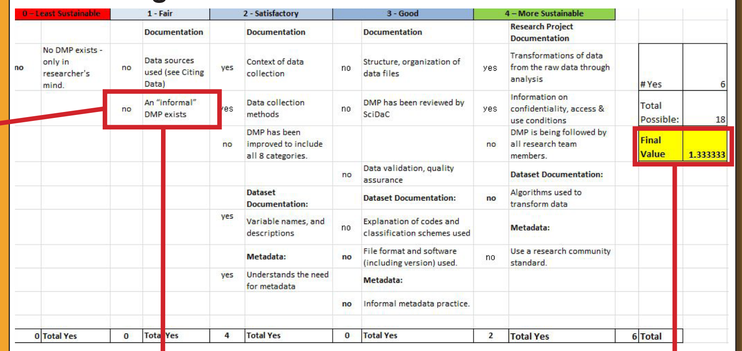
Final Assessment (Given to Researcher):

	Average	
File Formats Data Types	2	Satisfactory
Organization of Files	4	More Sustainable
Security Storage Backups	3.5	Good
Copyright Privacy Confidentiality	0	Least Sustainable
Data Documentation Metadata	1.333333333	Fair
Cumulative	2.166666667	Satisfactory



- **Final DMP component values and an overall sustainability grade;**
- **A Capability Maturity Model indicating a researcher's "grade" on the Data Management practice continuum.**

Data Management Practice Statement Valuation:



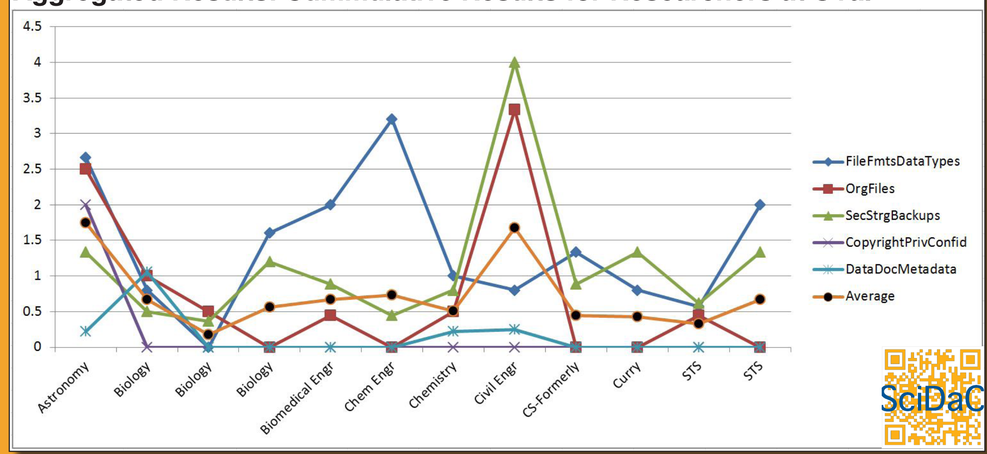
Each sub-statement is correlated with a "sustainability level" that rates the quality of existing DM practices from worst to best and assigns the point values to each level.

Implementation Phases and Action Statements:

Applicable?	Interview Topic	Sustainability	Phase	Action Statement
X	Data sources used	1	1	Properly document data sources used.
X	An "informal" DMP exists	1	1	Have an "informal" Data Management Plan (DMP): a DMP is the basis of all data management, and is a critical tool in protecting the continuity of your research process. Once in place, it can continually be updated, provided to new members of the lab as guidelines, and easily be applied to future grant proposals. We will work with you to develop an appropriate plan for managing your data. This is a fundamental first step in improving process.
X	Have you read UVa's Laboratory Notebook and Recordkeeping policy?	1	1	Read UVa's "Laboratory Notebook and Recordkeeping" Policy: https://policy.itc.virginia.edu/policy/policydisplay?id=RES-002
X	Have you read UVa's Ownership Rights in Copyrightable Material policy?	1	1	Read UVa's "Ownership Rights in Copyrightable Material" Policy: https://policy.itc.virginia.edu/policy/policydisplay?id=RES-001
X	DMP has been improved to include all 8 categories.	2	2	Informal DMP has been improved to include these 8 categories: File Formats and Data Types, Organizing Files, Security/Storage/Backups, Funding Guidelines, Copyright & Privacy/Confidentiality, Data Documentation & Metadata, Archiving & Sharing Data, and Citing Data.
X	Common, used by the research community	3	2	Use file types and data formats that are commonly used by your research community
X	Structure, organization of data files	3	2	Document structure and organization of data files
X	Data validation, quality assurance	3	2	Document data validation and quality assurance processes
X	Explanation of codes and classification schemes used	3	2	Document and explain any codes and classification schemes used
X	File format and software (including version) used.	3	2	Document any file formats and software (including version) used
X	Informal metadata practice	3	2	Begin applying appropriate metadata standard to data and other research materials
X	Adhering to ITC recommended backup cycles with off-site storage	3	2	Work with ITC to be sure that all servers and computers containing research data are being backed up to their specifications (with regular cycles and off-site storage)
X	DMP has been reviewed by SciDaC	3	3	Have SciDaC review DMP
X	Open, documented standard	4	3	Use an open and documented standard for file formats
X	Algorithms used to transform data	4	3	Document any algorithms used to transform data
X	In physically secure environment	4	3	Store data storage media and servers in a physically secure environment
X	Should be formally administered	4	3	Have data backups formally administered by ITC or some other system administrator
X	DMP is being followed by all research team members.	4	3	DMP is being followed by all members of the research team
X	Use a research community metadata standard.	4	3	Fully apply the appropriate research community metadata standard to research data and materials.

- **Data management "action statements" are created from corresponding sub-statements with values of "no."**
- **Implementation phases corresponding to the complexity of implementing each action are added to each "action statement."**
- **Action statements are sorted according to implementation phase.**

Aggregated Results: Cummulative Results for Researchers at UVa.



References:

Australian National Data Service. (2011) ANDS and Data Storage. Available: http://ands.org.au/guides/storage.html. Last accessed August 30, 2011.

Crowston, K., & Qin J. (2010). A capability maturity model for scientific data management. American Society for Information Science and Technology Annual Meeting. Pittsburg, PA. Working Paper available: http://crowston.syr.edu/content/capability-maturity-model-scientific-data-management-0. Last accessed August 23, 2011.

Digital Curation Center. (2011). CARDIO. Available: http://cardio.dcc.ac.uk/. Last accessed August 30, 2011.

Information Technology Security (2010). University of Virginia Information Technology Security Risk Management (ITS-RM) Program. Available: http://its.virginia.edu/security/riskmanagement/docs/ITS-RM_3-0.pdf. Last accessed August 23, 2011

University of Virginia Library (2011). Scientific Data Consulting Data Management Home. Web Site available: http://www2.lib.virginia.edu/brown/data/. Last access August 30, 2011.