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.

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’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 procedures 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 Services (ANDS) long-term sustainability scoring model, the system compares the information collected during the data interview process with these data management best practices. 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 to the Crowston and Qin Capability Maturity Model (CMM). Using this model as a base, 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 “manage” the situation).
- Level 4: Optimizing (the researcher is continually improving the 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 eradicate the recommendation report process by providing valuable actionable feedback that the researcher can immediately apply to improve the sustainability of their data.

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.

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:

- 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.

Final Assessment (Given to Researcher):

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

References:


References:


A Capability Maturity Model indicating a researcher’s “grade” on the Data Management practice continuum.

Aggregated Results: Cumulative Results for Researchers at UVA.

- 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.