**DMVitals Tool**

The DMVitals is a tool designed to take qualitative interview information and use it to systematically assess a researcher’s data management practices in direct comparison to institutional and domain standards. Using the DMVitals, a data management consultant matches a list of evaluated data management practices with responses from a research data interview and ranks the researcher’s current practices by their level of data management “sustainability.” The tool then generates customized and actionable recommendations, which a consultant then provides to the researcher as guidance to improve his or her data management practices. By design, the recommendations are far more objective, repeatable, and can be generated rapidly. The use of the DMVitals tool has helped our team expedite and standardize the data management consulting process.

Using best practice statements from UVa sources (Information Technology Services’ Risk Management Program and SciDaC Group 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 data management best practice statements. The model then further correlates the researcher’s data management practices with the eight data management practice categories developed by the SciDaC Group: file formats and data types; organizing files; security/storage/backups; funding guidelines; copyright and privacy/confidentiality; data documentation and metadata; archiving and sharing; and citing data.

A key part of the tool is the data management (DM) sustainability ratios, which are created for each of the previously mentioned categories using the best practice statements. To provide a framework for defining and improving researchers’ data management practices, the DM sustainability ratios are averaged to define a data management maturity level. These levels of maturity are a synthesis of the levels described by Crowston and Qin (2010) and the Australian National Data Service (ANDS, 2011), which are based on the Capability Maturity Model (CMM), typically used in software development (Paulk, Curtis, Chrissis, & Weber, 1993).

The data management maturity level is compared to the maturity levels as defined in the ANDS’s *Research Data Management Framework: Capability Maturity Guide* (2011):

* Level 1: Initial (environment is not stable enough to support DM; few individuals have expertise; and infrastructure is disorganized)
* Level 2: Development (the researcher increasingly recognizes the lower level/easier best practices; DM process is under development)
* Level 3: Defined (the researcher is further defining his or her DM practices)
* Level 4: Managed (DM is seen as important at an organizational level and has more coordination between researcher and other organizational IT units )
* Level 5: Optimizing (the researchers are continually improving their data management practices; DM practices are not static)

The strength of the DMVitals tool is the creation of the DM report, which generates tasks customized to each researcher. These tasks can then easily be grouped into phases, creating a DM implementation plan for each researcher based on his or her personal data interview and subsequent information gathering. Combining this tool with assessment and planning methods helps to expedite the recommendation report process and provide valuable actionable feedback that the researcher can use immediately to improve the sustainability of his or her data.

*UVa’s Data Interview Initiative*

The DMVitals tool evolved out of a need to systematically assess, score, and deliver mostly objective recommendations to a researcher following a start-up consultation interview. When we first began developing our DM services, we recognized that a data interview structure could help us develop a deeper understanding of how UVa’s science and en­gineering researchers manage their research data while also initiating a discussion about how to simplify processes and improve practices. These interviews were constrained to 60 minutes and included the scientific data consultants, the subject librarian, and the researcher. With the data interviews we hoped to:

a. Identify common research data problems and needs,

b. Identify the types of digital data that are being created,

c. Identify communities and individuals who are under the most pressure from emerging grant regulations,

d. Identify potential partnerships for institutional repository data deposit, and

e. Develop opportunities to provide data management recommendations and training.

In creating our interview model we consulted the following models: the Data Asset Framework (DAF) for background reading (Jones, Ball, & Ekmekcioglu, 2008), the University of Oregon for consultation and information on implementation and buy-in , the University of Glasgow for interview questions and format (Ward, Freiman, Molloy, Jones, & Snow, 2010), and Purdue’s Distributed Data Curation Center (D2C2) Data Curation Profile framework for question refinement . Our interview protocol was based on the protocol from Wisconsin’s Summary Report of the Research Data Management Study Group (Wolf et al., 2009).

Over the first two years of our services, we conducted about 25 data interviews. Through the process, we learned about research data practices at UVa, identified service needs and opportunities, and opened the door to consulting opportunities with researchers. Additionally, we encountered the dilemma of how to manage “unique” conditions of each research environment against common characteristics of DM within domains and institutional frameworks. In terms of support, we were having trouble customizing data recommendation requirements for each researcher.

We recognized a need to reduce the subjectivity and increase the speed at which we produced a report with recommendations. Additionally, we wanted to weigh all assessment factors from our interview, create actionable and repeatable recommendations, and address current DM conditions while showing paths for improvements. These needs are what led to the development of the DMVitals tool.

*Development of the DMVitals Tool*

The DMVitals tool is built using Microsoft Excel and consists of three types of worksheets: interview questions sheet, data management category sheets, and the data management report sheet. The interview questions sheet (Interview) contains the questions from the data interview. Each of the data management categories, as defined on the SciDaC Group website, are a sheet. In version (1.0) of DMVitals, we use five of the eight categories as sheets: file formats data types (FileFmtsDataTypes), organization of files (OrgFiles), security storage backups (SecStrgBackups), copyright privacy confidentiality (CopyrightPrivConfid), and data documentation metadata (DataDocMetadata). Each category sheet is populated with DM best practice statements for each category from UVa sources (Information Security, Policy, and Records Office (ISPRO), SciDaC Group guidelines) and the ANDS long-term sustainability scoring model. The category sheets also contain the calculated ratio of best practices statements. The third type of sheet, the data management report sheet (Report), is where the DM sustainability index ratios are displayed from the data management category sheets’ ratio of best practice statements. This sheet also displays the data management maturity level (the average of the DM sustainability index scores) and the action statements for DM improvement (corresponding to best practices).

Interview Questions Sheet

The questions from the SciDaC Group’s Data Interview Protocol are entered on the Interview sheet, one question per column. The current version maps questions from sections 2–5 (see Figure 1). Each question is then associated with one or more DM best practice statements. These DM best practices are listed under each question (or sub-question). Using the answers from the interview, each best practice was coded “yes,” “no,” or “null.” “Yes” meant that the researcher was already doing that action; “no” meant that the researcher was not doing that action and “null” meant that best practice did not apply (for example, the best practice of “data is de-identified” in cases where data need not to be de-identified). “Null” is the default answer.

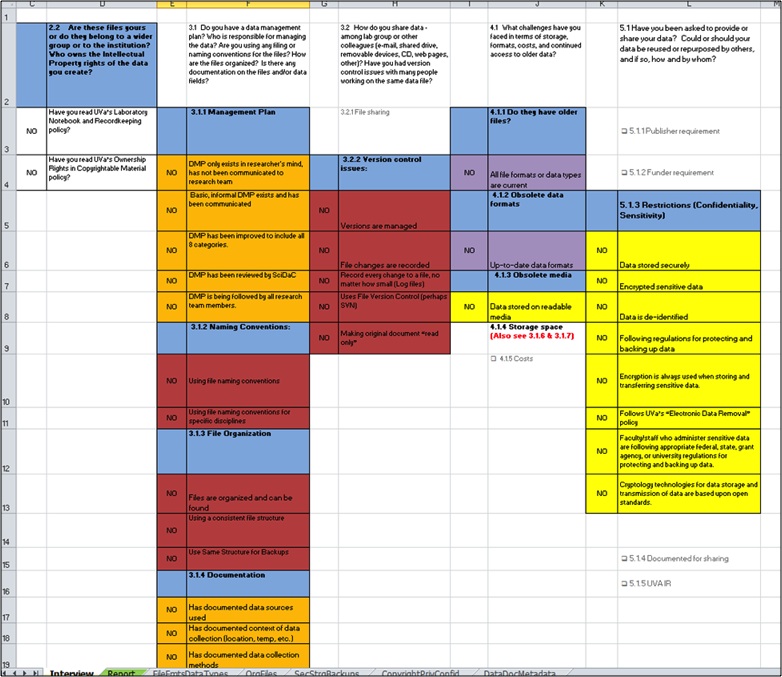
******

Figure 1: Interview Sheet

Data Management Category Sheets

Each best practice statement from the interview questions sheet is mapped to one of eight data management categories (file formats data types: FileFmtsDataTypes, organization of files: OrgFiles, security storage backups: SecStrgBackups, copyright privacy confidentiality: CopyrightPrivConfid, data documentation metadata: DataDocMetadata, funding guideline: FundingGuide, archiving and sharing data: ArchSharing, citing data: CitingData). Note that in this version, only five of the management categories are being used. The categories funding guidelines, archiving and sharing, and citing data are not used at this time. These will be added in a future version of DMVitals.

Each best practice statement from the interview questions sheet is categorized, given a sustainability level, and is then put in the corresponding column per one of the five sustainability levels (least sustainable, fair, satisfactory, good, and more sustainable). For ease of editing, each best practice statement is linked from its cell on the interview questions sheet to the sustainability level. The actual response to the best practice statement from the interview questions sheet—“yes,” “no,” or “null”—is also linked. The mapping and linking of each best practice is done for each of the categories on the corresponding category sheet.

To calculate the sustainability index ratio (current best practice to total-possible best practice), each sustainability level was given a “weight”: least sustainable (\*1), fair (\*2), satisfactory (\*3), good (\*4), and more sustainable (\*5). The total number of “yes” responses (current practice) for each level is multiplied by the sustainability weight. These levels are then totaled and divided by the total possible score (total number of best practices that apply, for that sustainability level, multiplied by the sustainability weight). The ratio for each category is then automatically recorded (via a link) on the data management report sheet as the sustainability ratio. See the screenshot in Figure 2 for the OrgFiles sheet for an example.

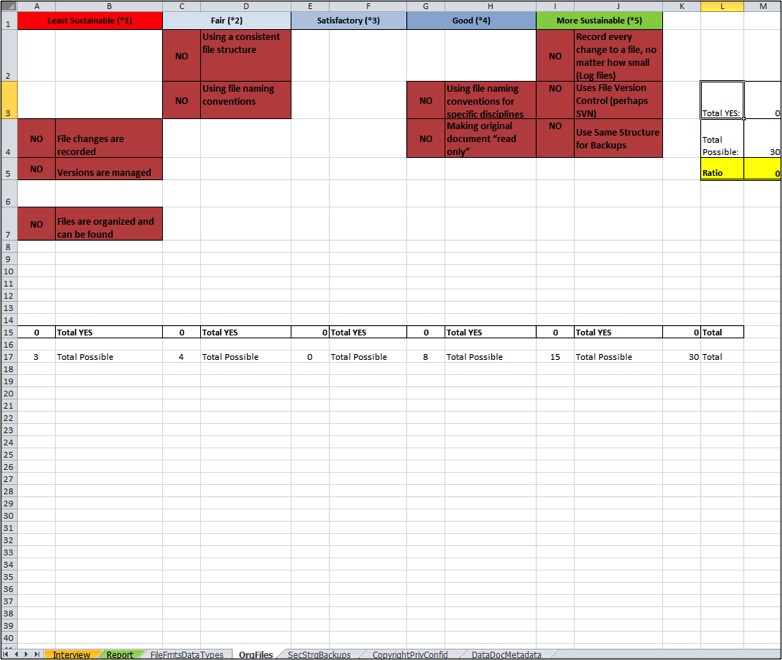
******

Figure 2: Data Management Category Sheet- Organization of Files (**OrgFiles**)

Data Management Report Sheet

The data management report sheet is comprised of three distinct sections: sustainability index, data management maturity level and action statements (see Figure 3). The top chart (sustainability index) shows the DM category and the resultant sustainability index (displayed as a percent—a ratio). The actual ratios are linked from the corresponding data management category sheets. With five levels of sustainability, we divided the ratio values into five groupings—0–20 percent (Level 1), 21–40 percent (Level 2), 41–60 percent (Level3), 61–80 percent (Level 4), and >81 percent (Level 5)—and color-code the values using the colors on the data management maturity level scale (see the colors on Figure 3). This gives a visual view of how the researchers’ current DM practices, per category, are ranked, according to the level of “sustainability.” The ratios of the categories are averaged for a sustainability index. The average is also color-coded using the above percent groupings.

The bottom of the sheet contains a chart of action statements. The chart includes actionable recommendations targeting improvements for DM. Each best practice statement from the interview questions sheet has a corresponding action statement. The best practice a researcher is not doing is marked with “X.” These are the basis for DM improvement recommendations. The phases are customizable and can be moved around as the consultant sees fit. In our consulting approach, we avoid placing too many actions in any phase, which in turn may put researchers more at ease with improving practices.

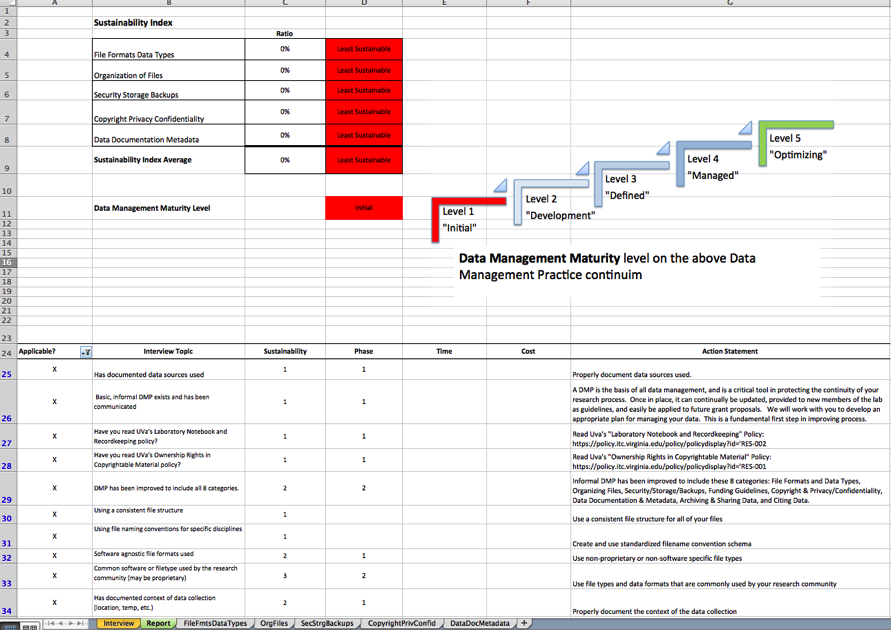
******

Figure 3: Report Sheet

*Use of DMVitals Tool*

Recommendations Report

The recommendations report is designed for distribution to researchers (see Figure 4 for an example). It begins with general information on DM and the goals of the report. The DMVitals report sheet provides the rest of the information that goes in the report. The sustainability index chart includes their data management maturity “grade.” The chart is copied and pasted from the report sheet. The action statements are grouped into implementation phases (Phase 1: short-term; Phase 2: long-term; and Phase 3: future).

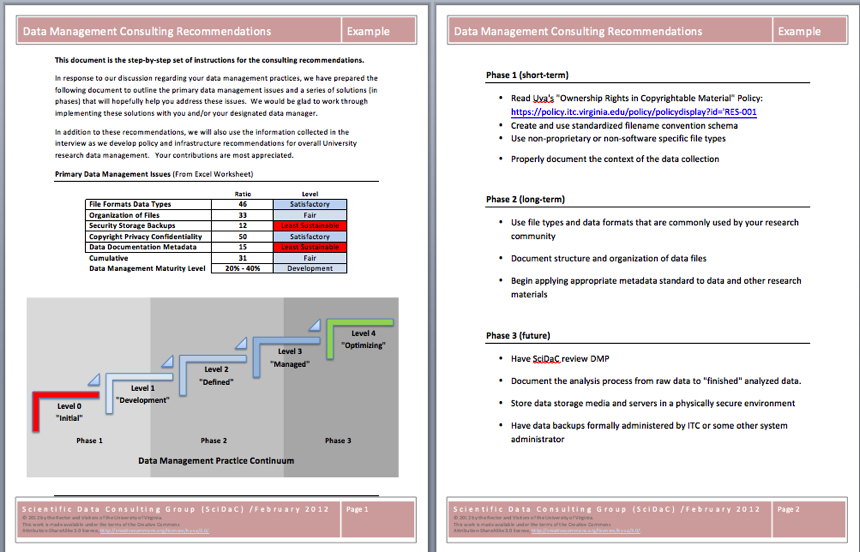
******

Figure 4: Data Management Consulting Recommendations Report

Data Management Implementation

The next step in UVa’s data interview workflow is to distribute the final report, with recommendations, and begin implementation with the researcher. We then sit down with the researcher go over the recommendations and make adjustments on what actions are done in each phase (see Figure 5). The DMVitals tool is to be used throughout the implementation. As researchers improves their practices, their DMVitals score is recalculated and reflected in their new sustainability index.

The goal of repeating these steps—going back to the DMVitals, updating the best practices, and reevaluating the sustainability index—is for the researcher to obtain a data management maturity score of five “Optimizing,” on the ANDS CMM for research DM. But as the model warns, level five is not the “final” level. Level five is labeled “Optimizing.” At this level, researchers should be focusing on continually improving their DM practices.

Implementation: DMVitals Workflow

*Applying the DMVitals Tool at Your Institution*

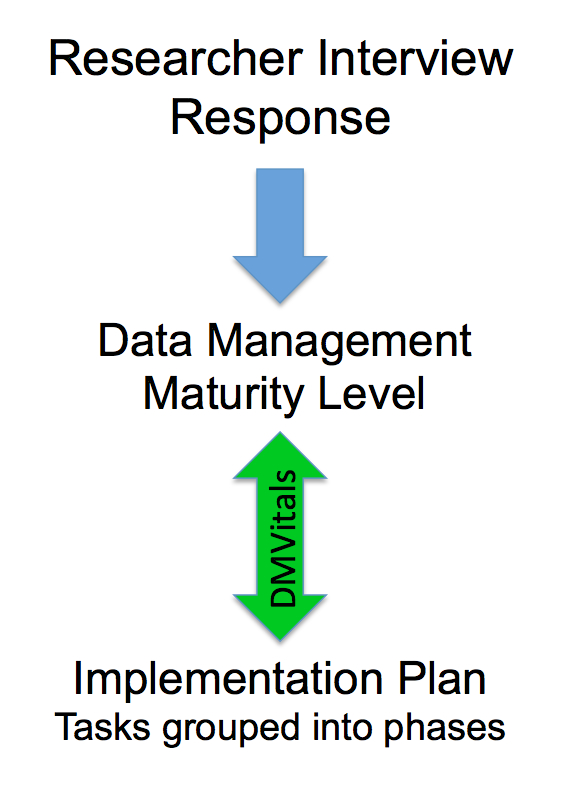
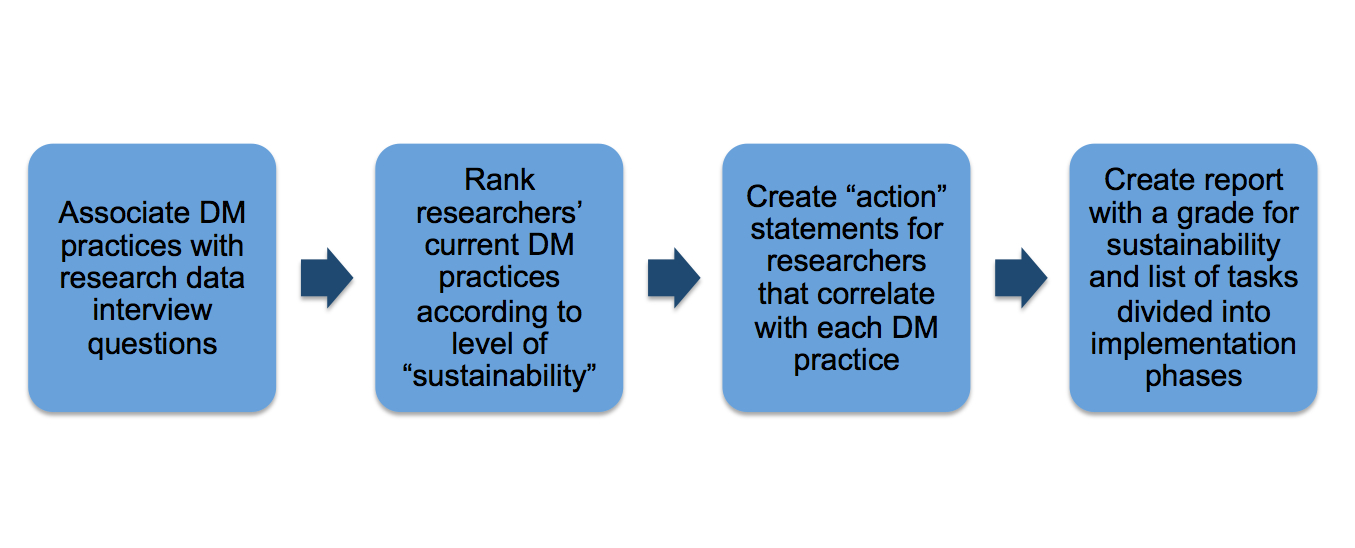
******The DMVitals tool easily can be configured for your institution. It can be used with any data assessment tool. The main interviewworksheet can be customized using your assessment questions, local institution policies, and best practices. Ranking of sustainability can be adjusted per discipline or institution. The action statements definitely will require local customizations. These are the actions that your researchers need to do for your institution. Actions might include contacting specific service providers for support. Figure 6 shows the steps in the DMVitals workflow.

Figure 5: Data Management

******Figure 6: DMVitals Modification Flow

The DMVitals tool will continue to undergo improvement to further evaluate whether the scoring accurately represents improved practices. We recognize that the selection of assessment criteria, the calculation and weighting of scores, and how performance is communicated to the researcher must be done carefully in order to be effective as a communication tool with researchers. We plan for continued refinement of all steps, and we aim to produce a tool that offers value to both DM support providers and the researchers who receive support.

At the date of writing this chapter, this is still a prototype tool, and we do not yet have metrics to assess the impact or benefit of its use upon effective delivery of services. We are in the early planning phases for an expanded version of the tool, and we likely will integrate it with other broadly used interview models.

**Acknowledgments**

We would like to acknowledge Susan Borda, a SciDaC Group intern in the summer of 2011 and presently the Digital Curation Librarian at University of California-Merced, for her help on the creation and development of the DMVitals tool.

**References**

Australian National Data Service (ANDS). (2011). *Research data management framework: Capability maturity guide*. Retrieved from http://ands.org.au/guides/dmframework/dmf-capability-maturity-guide.html

Carlson, J., & Witt, M. (2007, December). *Conducting a data interview*. Presented at the 3rd International Digital Curation Conference, Washington DC. Retrieved from http://docs.lib.purdue.edu/lib\_research/81

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

Jones, S., Ball, A., & Ekmekcioglu, C. (2008). The data audit framework: A first step in the data management challenge. *International Journal of Digital Curation*, *3*(2), 112–120. http://dx.doi.org/10.2218/ijdc.v3i2.62

Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. V. (1993). *Capability maturity model for software, version 1.1* (CMU/SEI-93-TR-024). Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University. Retrieved from http://www.sei.cmu.edu/reports/93tr024.pdf

Ward, C., Freiman, L., Molloy, L., Jones, S., & Snow, K. (2010). *Incremental scoping study and implementation plan.* Project Report. University of Cambridge, Online. Retrieved from http://eprints.gla.ac.uk/54623/

Westra, B. (2010). Data services for the sciences: A needs assessment. *Ariadne, 64.* Retrieved from http://www.ariadne.ac.uk/issue64/westra/

Wolf, A., Simpson, M., Salo, D., Flee, D., Cheetham, J., & Barton, B. (2009). *Summary report of the Research Data Management Study Group*. Madison: University of Wisconsin. Retrieved from http://digital.library.wisc.edu/1793/34859