

## Building Capabilities for Research Data Infrastructure and Services

Jian Qin <sup>(1)</sup>, Kevin Crowston <sup>(1)</sup>, Liz Lyon <sup>(2)</sup>, Angus Whyte <sup>(3)</sup>, Jonathan Rans <sup>(3)</sup>

1) School of Information Studies, Syracuse University 2) School of Information Sciences, University of Pittsburgh 3) Digital Curation Centre, University of Edinburgh

### SHORT DESCRIPTION

Capability modeling is an established business process management support tool, with proven value for planning products and service delivery. This tutorial and workshop is aimed at service managers, data stewards and others interested in applying them to help evaluate research data management (RDM) services and data stewardship practices. Participants will learn how to apply a model suited to their contexts, and gain awareness of how such tools may be further developed to represent evolving RDM community norms and guidelines, help services prepare for Trustworthy Digital Repository assessment, and inform skills and career development processes.

### KEYWORDS

Capability Maturity Model, Research data management, Community Capability Model, Best practices

### AGENDA (HALF DAY)

- |               |                                                                                                                                  |
|---------------|----------------------------------------------------------------------------------------------------------------------------------|
| 9.00-9.30     | Introduction: goals of the workshop and an overview of capability models                                                         |
| 9.30-10.30    | CMM for RDM                                                                                                                      |
|               | Walkthrough of key practices covered, how and where to apply the model (15 minutes)                                              |
|               | Exercise – in groups work with a facilitator to apply the model to a scenario describing examples of good and bad practice (30m) |
|               | Group feedback– how did this model work with your examples? (15m)                                                                |
| 10.30- 10.45  | Coffee break                                                                                                                     |
| 10.45 – 11.45 | Community Capability Model – walkthrough, exercise and discussion as above                                                       |
| 11.45- 12.45  | RISE/ ReCap– walkthrough, exercise and discussion as above                                                                       |
| 12.45 – 13.00 | Discussion to share experiences of service development activities these models could and should support                          |

### LONG DESCRIPTION

Capability models are tools offering different perspectives from which to systematically review key RDM process areas and activities (capabilities), and assess their level on a scale such as ‘maturity’ or ‘organizational readiness’. The models to be presented are works in progress, and the workshop organizers aim to further link their models to guidelines for data policy making, personnel training, and performance assessment. The organizers have each organized similar workshops previously to test and validate their approaches. This workshop is the first to offer a comparison across a number of approaches.

Institutes, Universities and Research Infrastructures need to offer researchers high-quality Research Data Management (RDM) support services. Capability models can provide a benchmark for tracking improvements to services and the practices they support, and for external comparisons. By participating in this workshop you will gain awareness of different approaches to process management and assessment for RDM, and learn how to apply these models. Doing so can help sustain a service in various ways, for example by providing a framework for:

- Service managers to appraise the support a research data service offers its users, and outline business case options to improve the service

- Data stewards to support researchers in data-intensive research communities to self-assess their collective RDM practices, and describe shared support requirements

The workshop will focus on the following models:

- CMM for RDM (Capability Maturity Model for Research Data Management) - Jian Qin and Kevin Crowston, Syracuse University
- CCMF (Community Capability Model Framework) - Liz Lyon, University of Pittsburgh
- RISE (Research Infrastructure Self Assessment), and ReCap (Repository Capability) - Jonathan Rans and Angus Whyte, Digital Curation Centre

A common reference point is the Capability Maturity Model (CMM) developed at the Software Engineering Institute (SEI) at Carnegie Mellon University (Paulk et al., 1993). The CMM addressed a need to transform software development from a cottage industry to one that could deliver quality software on time and within budget. CMM quickly proved successful as a tool for evaluating the processes needed for that, and has since been generalized across sectors as the CMMI framework (CMMI Product Team, 2006).

A common feature of CMM and other capability models is a list of process areas, used to construct a rubric, in tabular form, whose rows list the relevant capabilities. A number of columns (e.g. 3 to 5) are labelled to describe a scale of attainment, and each cell provides a statement describing a level for each capability. In the CMM approach the scale represents maturity, i.e. the level of organizational capability to reliably perform the process. This maturation reflects the extent to which each process is institutionalized and managed, ideally with quantified measures enabling continuous process improvement.

The *CMM for RDM* draws parallels between current challenges RDM services face and those addressed in software engineering, indicating CMM should also help structure efforts to improve RDM practices. The model has been developed since 2011 by gathering evidence to identify and cluster key RDM practices. The current draft offers rubrics for assessing where the project or institution stands in terms of the maturity level of five key areas: 1) Data management in general; 2) Data acquisition, processing and quality assurance; 3) Data description and representation; 4) Data dissemination; and 5) Repository services and preservation.

The *Community Capability Model Framework (CCMF)* for data-intensive science has been developed in parallel by UKOLN, University of Bath UK in partnership with Microsoft Research, as a self-assessment tool for disciplinary researchers (Lyon, et al., 2011). The *CCMF* tool has eight dimensions located within three broad areas: environmental, human and technical, and is framed as a MS Excel Profile template which can be adapted in terms of language and exemplars to suit different disciplinary audiences. The *CCMF* has been tested in diverse domains including agronomy and the social sciences.

The *RISE* and *ReCap* models have been developed more recently by the Digital Curation Centre, drawing on their previous models for institutions to assess progress in developing RDM services to meet UK Research Council expectations (Rans & Whyte, 2016; Whyte & Macdonald, 2016). *RISE* is based on ten RDM service components. *ReCap* focuses on repository functions associated with three of these; preservation, discovery, access and publishing. *RISE/ReCap* are influenced by the *Digital Preservation Readiness* model developed at Cornell University (Kenney and McGovern 2003), and describing five stages of ‘organizational response’. *RISE/ReCap* adopts a similar three-point scale: ‘complying’ with external policy, ‘institutionalising’ a service meeting local requirements, and ‘externalising’ sector-leading practice.

**Room layout and style:** Cabaret **Minimum/Maximum delegate numbers:** 25/30 **Number of speakers:** 4~5

**Funding model:** Delegate fees **Equipment requirements:** projector, sound, flipcharts **Length of workshop:** half day

## REFERENCES

CMMI Product Team. (2006). *CMMI® for Development, Version 1.2: Improving processes for better products*. CMU/SEI-2006-TR-008 ESC-TRo2006-008. Pittsburgh, PA: Carnegie Mellon University Software Engineering Institute.

Crowston, K. & J. Qin. (2011). A capability maturity model for scientific data management: Evidence from the literature. In: *Proceedings of the American Society for Information Science and Technology*, 48: 1–9. doi: 10.1002/meet.2011.14504801036

Lyon, L., Ball, A., Duke, M., & Day, M. (2011). Community Capability Model Framework. <https://communitymodel.sharepoint.com/Documents/CCMDIRWhitePaper-24042012.pdf>

## IDCC 2017 Workshop Proposal

Kenney, A. & McGovern, N. (2003) "The Five Organizational Stages of Digital Preservation," in Digital Libraries: A Vision for the Twenty-first Century,. Available from University of Michigan Scholarly Monograph Series:  
<http://dx.doi.org/10.3998/spobooks.bbv9812.0001.001>

Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. (1993). Capability maturity model, Version 1.1. IEEE Software, 10(4): 18–27

Rans, J. & Whyte, A. (2016). Introduction to RISE, a Research Infrastructure Self-Evaluation Checklist'. Edinburgh: Digital Curation Centre (to appear) <http://www.dcc.ac.uk/resources/how-guides>

Whyte, A., Macdonald, S. et al. (2016). How to Evaluate Research Data Repository Platforms. Edinburgh: Digital Curation Centre (to appear) <http://www.dcc.ac.uk/resources/how-guides>