

Research Data Management Framework: Capability Maturity Guide

Introduction

The outline set out below shows five levels of attainment or maturity which institutions may achieve in managing their research data. Organisations can use it as a guide to assessing their current level of attainment and identifying areas where they may wish to concentrate in the future. In this way the model can serve as a form of gap analysis.

This is an adaptation of the well-known Capability Maturity Model (CMM) which was originally applied to software development. This has subsequently been developed into the Capability Maturity Model Integration¹ (CMMI) and applied to other areas of business activity.

Given the breadth of activities and processes involved in the Framework, it is unlikely that any organisation will demonstrate consistent levels of attainment across the four major areas of Organisational policies and procedures, IT infrastructure, Support services and Managing metadata. Furthermore, it is up to each organisation to decide where on the model it wants to be: not all organisations will seek to attain Level 5 (Optimised).

Who should read this

This guide is intended for practitioners within research organisations who have responsibility for some aspects of research data infrastructure provision. Future ANDS activity related to the provision of top-down, coordinated research data infrastructure is planned.

Level 1 - The Initial Level

At Level 1, the organisation does not provide a stable environment to support research data management. Expertise is likely to be concentrated within a few individuals, and successes are likely to be due to their efforts rather than to proven processes. It may prove difficult, therefore, to repeat successes. Co-ordination and cohesion across the various groups (e.g. research office, IT, library, records office, research areas) associated with research data management is patchy, if non-existent.

Policies and procedures may not exist at all, or if they do, may be out of date, not address data-related issues, be inconsistently applied or not include compliance measures.

IT infrastructure is disorganized and inadequate to support research data storage and management. Expertise is held by only a few individuals who are unable to cope with demand. (*cont. p3*)

¹ CMMI Product Team, CMMI for Services, Version 1.3, November 2010. Software Engineering Process Management Program. <http://www.sei.cmu.edu> (downloaded 8 December, 2010).



Research Data Management Framework: Capability Maturity Guide

	Level 1 Initial	Level 2 Development	Level 3 Defined	Level 4 Managed	Level 5 Optimised
	Process is disorganised & <i>ad hoc</i>	Process is under development	Process is standardised, communicated	Process is managed, measured	Focus is on continuous improvement
Institutional policies & procedures	Policies & procedures may be undeveloped, not up to date, and/or inconsistent.	Policies & procedures are developed & harmonised.	Policies & procedures are promulgated & absorbed into behaviours.	Policies & procedures accepted as part of the culture & subject to audit.	Policies & procedures are subject to review & improvement.
IT infrastructure	IT infrastructure provision is patchy, disorganised & poorly publicised.	Funds are invested in technology & skills. Responsibilities are defined. Processes are established, defined & documented.	Management shows active support. Facilities are well-defined & communicated, standardised & integrated.	Funding adapted to need. Management actively engaged. Documentation kept up to date.	Concerted efforts to maintain, update & publicise infrastructure. Metrics & feedback used to optimise services.
Support services	Training is ad hoc, curation & preservation services are disorganised, data management planning is unsupported & other services inconsistent & poorly publicised	Investment in skills. Services identified & staffed. Responsibilities are defined. Documentation & training developed.	Active participation in training & widespread availability of services.	Widespread take up of services. Curation & preservation acknowledged as critical to the institutional mission.	Customer feedback used extensively to update & improve services.
Managing metadata	Metadata management is chaotic & understood by only a few.	Responsibilities are defined & skills developed. Processes are established, defined & documented. Metadata applied to key datasets & shared externally.	Processes are standardised & integrated. Metadata applied to new datasets & shared externally.	Metadata quality metrics collected. All datasets described & metadata shared.	Continuous improvement applied to processes & capabilities.



Support services are *ad hoc*, with training largely unavailable, data management planning is unsupported, partly because the issue has not been dealt with through the development of policies and procedures, standards are undefined, curation and preservation services are disorganised, and other services are inconsistent (if not nonexistent) and largely unknown.

Metadata management is chaotic & understood only by unsupported individuals. It is regarded as low priority by the organisation, so that institutional and area responsibilities for it are undefined and resources to support it are inadequate. Few, if any, attempts are made to share metadata externally and its role in potentially supporting institutional administrative activities is unexplored. Those directly involved in research projects, however, may be adept at meeting the metadata needs of their own research, but they are unsupported by institution-wide services which may help them to provide better metadata services to support data sharing and re-use.

Level 2 - The Development Level

At Level 2, the organisation has moved on from *ad hoc* processes, as the role of research data management is increasingly recognized as an important aspect of the organisation's business. There is a more systematic approach to research data management, with some attempts at coordination across the various areas involved (e.g. research office, IT, library, records office, research areas). All aspects of research data management are being developed, with assigned responsibilities, trained staff and improved resourcing. The institution is moving from being reactive to being proactive, but still has some way to go.

Policies and procedures: Key to moving from Level 1 to Level 2 is development of a research data management policy. This should define responsibilities, provide definitions and set out expected behaviours around the creation, storage and management of data, while reflecting the requirements of the *Australian Code for the Responsible Conduct of Research*. Other related policies should also have been developed, updated and/or harmonized with the research data management policy. Relevant procedures should be in place for each policy. While the policy and procedures exist at Level 2, they have still to be institutionalized. Instead, there is reliance on a central person or group to understand the issues and implement procedures reliably and consistently.

IT infrastructure: Reaching Level 2 means that the organisation has recognized the need for IT infrastructure provisioning and IT equipment. Processes are in place, responsibilities are defined, training has been delivered to staff immediately engaged with the IT infrastructure, funding needs have been assessed and budgets readjusted. Institutional needs for storage of research data, and its accompanying metadata, have been recognized and procedures put in place to ensure that all researchers have access to persistent storage for their data, whether it has been created in-house or imported from elsewhere. Data storage may be in-house, shared with another institution or outsourced. Processes and tools have been put into place to manage identities, authentication and access, and to ensure that licence conditions for shared data are met. Tools are available to meet researcher needs for analysis, visualization and collaboration, even if these have not yet been widely publicized. There is periodic monitoring and evaluation of processes and the organisation can respond to periods of stress in a systematic rather than *ad hoc* fashion.

Support services. At Level 2, the organisation has a wide range of support services in place. These include services which publicise the need for and importance of research data management and provide training to data creators, curators and specialists working in data centres. Training activities are starting to be supported by web pages linking to internal policies, procedures, templates and facilities, as well as external sources of information. Data curation and preservation services are well under way, and data is starting to be assessed for longer-term retention.



Metadata management. The organisation recognizes the need for comprehensive metadata to underpin good research data management. It recognises the benefits that this can bring to both research and the administrative requirements of the organisation, in addition to meeting any external requirements (as from funders, for example). Standards are identified for both object-level (technical) and collection-level metadata records. Tools are developed to support the automatic collection of metadata from various instruments. Processes are in place to identify new datasets for administrative purposes as well as sharing with other entities, such as Research Data Australia. Some metadata records are created and shared with Research Data Australia and/or other portals.

Level 3 - The Defined Level

Level 3 organizations typically understand the business meaning of research data (e.g. as an asset) and have a stated policy that 'data is treated as an organisational asset.' This is reflected in an emphasis on skills development, improved communications, the establishment of standard processes and an increase in consistency across the organisation. Processes are typically described more rigorously and more widely adopted.

Policies and procedures: Moving to Level 3 shows that the organisation is taking responsibility not just for having policies and procedures in place, but for implementing change processes to ensure that these are widely understood and put into practice. Responsibility for the implementation of appropriate procedures is devolved outwards from a central few as more individuals become engaged in implementation of procedures and understand their benefits. Policies are enforced and testing is done to ensure that requirements are being met.

IT infrastructure: Management actively supports data IT Infrastructure and ensures it is coordinated and publicised across the organisation. IT infrastructure needs are coordinated with capital expenditure planning and with technology transition planning. IT support roles are more specialised than in Level 2, although some staff may enact multiple roles. Some use is made of data support tools, but not consistently across the organisation. IT facilities are more standardised and some formal testing of facilities and services is done.

Support services: All services are now fully supported financially. They are well developed, documented and publicized. Training is regarded as a normal part of induction, for graduate students as well as new staff in relevant areas. Checks are made to verify that the services are meeting the requirements of the user community.

Metadata management: While data is recognized as an organizational asset, having high quality metadata is recognized as being critical to good data management. Metadata records are created for all newly identified datasets and shared with Research Data Australia and/or other portals. Automatic collection of metadata from instrumentation is improved in terms of both accuracy and content relevance, and processes are developed to align object-level and collection-level metadata. Linkages are in place between the various systems from which metadata can be gathered so that information is only collected once.

Level 4 - The Managed Level

At Level 4, senior administrators recognize research data as an organizational asset. This implies that it is treated in the same way as other assets (personnel, finance, buildings, equipment, etc) and subject to audit. Audit measures are both quantitative and qualitative, ensuring that there is a focus on both efficiency and effectiveness. There is consistent coordination between those responsible for policies and procedures, IT infrastructure provision, support services and metadata management.



Policies and procedures: Policies which have an impact on research data management have now been absorbed into the organizational culture and are consistent across all areas. Procedures are fully understood and implemented. Compliance is audited.

IT infrastructure: Management views data as a key organisational asset and supports IT infrastructure through targeted support and value-adding of data assets. IT infrastructure costs and planning are integrated with other asset-based planning systems across the organisation. There is a clear separation of specialised IT support roles. Data support tools are largely standardised across the organisation and regular measures of their effectiveness and value are captured. Formal, automated testing of facilities, services and data quality is performed.

Support services: There has been widespread take up of services across the organisation. Curation and preservation are acknowledged as critical to the institutional mission, and supported accordingly. Some auditing is undertaken.

Metadata management: Metadata is accepted as a key aspect of effective research data management and advanced tools are in place to manage it. The institution is now in a position to know what data exists and where it is located, and to use this information for internal administrative processes, such as research reporting or researcher profiles. All datasets are fully described, allowing this information to be available across disciplines. Metadata is routinely contributed to Research Data Australia and/or other portals. Metadata is audited to assess quality and availability.

Level 5 - The Optimizing Level

The Level 5 organization uses the policies, procedures, practices, services and facilities already developed as the basis for continual improvement.

Policies and procedures: policies subject to review and improvement in order to ensure that external requirements and internal needs are being met.

IT infrastructure: Management optimises the IT infrastructure evolved in previous Levels through continuous focus on management and improvement of data assets. Automatically captured metrics and audit findings are used to maintain service quality and drive change in rapidly changing technology support areas. Individual IT support roles enact well publicised technology transition and process improvement plans. Use of data support tools is driven by measures of effectiveness. All aspects of data IT infrastructure facilities and services are monitored through highly visible automated testing and feedback systems.

Support services: Customer feedback forms the basis of updating and continuous improvement of services.

Metadata management: Metadata provides the basis for rapid discovery of data in all disciplines, both within the institution and beyond. Continuous improvement is applied to processes & capabilities.

Further information

ANDS Guide to Creating a Data Management Framework

<http://ands.org.au/guides/dmframework/data-management-framework.html> and linked pages.

