



ANDS Guide



Creating a data management framework

Level: Awareness

Last updated: 23 March 2018

Web link: www.ands.org.au/guides/creating-a-data-management-framework

This Guide is for research institutions which are intending to assess the capability maturity of their current infrastructure supporting the management of institutional research data assets.

Creating a data management framework	1
Introduction.....	2
Developing a Data Management Framework using the Capability Maturity Model (CMM)	2
CMM overview	3
CMM viewed through the 5 Levels	4
Level 1 - the initial level.....	4
Level 2 - the development level	5
Level 3 - the defined level	6
Level 4 - the managed level.....	7
Level 5 - the optimising level.....	8
CMM viewed through the 5 Elements	9
Policies and procedures	9
Support services	9

IT infrastructure.....	10
Metadata management	12
Research Data management	13

Introduction

A Data Management Framework will provide an integrated and effective institutional data management approach to the 5 essential framework elements:

1. **Policies and procedures:** establishing and promulgating policies and procedures in accordance with the Australian Code for the Responsible Conduct of Research and other external legal and regulatory frameworks.
2. **Support services:** providing the requisite advice and support services such as online toolkits, information, training.
3. **IT infrastructure:** providing the requisite IT infrastructure including the hardware, software and other facilities which underpin data and metadata management.
4. **Metadata:** metadata is managed findable, accessible, interoperable and reusable (FAIR), for both internal and external purposes.
5. **Research Data:** research data is stored and managed in accordance with the FAIR principles.

Applying to each of these five elements is the expectation that there will be:

- Appropriately defined roles and responsibilities
- Defined and promulgated standards
- Adequate resources – financial, staffing and equipment
- Appropriate skills across all staffing groups

Effective data management is best achieved through teamwork and collaboration between researchers, research offices, information specialists and technical support staff. Integrating both the approach and infrastructure will enable researchers to meet their obligations to funders, improve the efficiency of research, and make richly described data available for sharing, validation and reuse.

Developing a Data Management Framework using the Capability Maturity Model (CMM)

Organisations can use the Data Management Framework Capability Maturity Model (CMM) set out below as a form of gap analysis:

- assessing current levels of data management capability and
- identifying areas to enhance and extend data management capability

This Data Management Framework CMM is an adaptation of the well-known Capability Maturity Model (CMM) which was originally applied to software development.

Given the breadth of activities and processes involved developing an effective institutional data management framework, it is unlikely that any organisation will demonstrate consistent levels of attainment across the 4 major elements. Each organisation should determine where on the model it wants to be: not all organisations will seek to attain Level 5 (Optimised) in any, or all, of the 4 elements.

CMM overview

This CMM uses:

- **Five elements of data management capability:** Policies and procedures; support services, IT infrastructure; managing metadata, managing research data
- **Assessed across five levels of maturity:** Initial; Development; Defined; Managed; Optimised

	Level 1 Initial	Level 2 Development	Level 3 Defined	Level 4 Managed	Level 5 Optimised
	Process is disorganised & ad hoc	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 is 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 with widespread availability of support 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 ad-hoc, chaotic & understood by only a few and without established standards.	Responsibilities are defined & skills developed. Processes are established & documented. Metadata applied	Processes are standardised & integrated. Metadata created for new datasets & shared externally, to ensure data is findable and	Metadata quality metrics are collected. All datasets described in machine-readable format & metadata shared. Metadata aligns	Continuous improvement applied to processes & capabilities.

		to key datasets & shared externally.	accessible.	with the FAIR data principles.	
Managing Research Data	Data is stored in ad-hoc facilities. Only data custodians know where the data is stored and accessible only to the researcher or small group of researchers, Standard formats are not applied and the potential for reusability is limited	Institutional data storage facilities are being developed. Data standards established	Well defined data storage facilities are established. Data managed using standard open formats widely used. Some connectivity between systems permitting limited reuse and sharing.	Data routinely stored in established repositories in machine-readable formats using open standards in established facilities according to FAIR principles.	Continual improvements to maintain update & publicise infrastructure. Metrics & feedback used to optimise services.

CMM viewed through the 5 Levels

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 to repeat successes. Coordination and cohesion across the various groups (i.e. research office, IT, library, records office, research areas) associated with research data management is patchy or non-existent.

1. **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.
2. **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 (or non-existent) and largely unknown.
3. **IT infrastructure** is disorganised and inadequate to support research data storage and management. Expertise is held by only a few individuals who may be unable to cope with the demand. The capability and processes to mint persistent identifiers are non-existent or limited.
4. **Metadata management** is chaotic and understood only by unsupported individuals. It is regarded as a 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 may be adept at meeting the metadata needs of their own research, but they are unsupported by institution-wide services which would help them to share data outside of their group.
5. **Research Data** storage facilities may be rudimentary such as shared drives or free web services. The data has little-associated metadata with the data stored in a range of formats including proprietary formats requiring the use of proprietary software to access and use the data. The ability and inclination to share or reuse the data outside of the purposes for which it was collected is very limited. Tools to analyse the

data are often proprietary or limited and the ability to share data between systems is cumbersome, time-consuming and requires specialist skills.

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 recognised 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 (i.e. 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.

1. **Policies and procedures:** The key to moving from Level 1 to Level 2 is the 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 harmonised with the research data management policy. Relevant procedures should be in place for each policy. While policy and procedures exist at Level 2, they have still to be institutionalised. Instead, there is reliance on a central person or group to understand the issues and implement procedures reliably and consistently.
2. **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 underway, and data is starting to be assessed for longer-term retention.
3. **IT infrastructure:** Reaching Level 2 means that the organisation has recognised 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 recognised 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 in place to manage identities, authentication and access with the ability to mint persistent identifiers. There are periodic monitoring and evaluation of processes and the organisation can respond to periods of stress in a systematic rather than ad hoc fashion
4. **Metadata management:** The organisation recognises the need for comprehensive metadata to underpin good research data management. It recognises the benefits that this can bring to research, and the administrative requirements of the organisation and/or external funder requirements. 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 and/or sharing purposes. A data licensing regime preferably using an existing standard is adopted. Limited numbers of metadata records with persistent identifiers (PIDs) not just for collections but also for related people, activities, services and institutions and are created and shared with Research Data Australia and/or other portals.

5. **Research Data:** The organisation recognises data as an institutional asset and persistent data storage facilities and standards are required to underpin and support good research practice. Suitable data storage facilities internal and external are identified with some being developed in-house. Appropriate open data standards, formats, protocols, ontologies and vocabularies are identified and begin to be applied to the data. Tools are developed to support the automatic storage of data with the metadata from various systems and instruments. The data incorporates persistent identifiers. Tools are available to meet researcher needs for analysis, visualisation and collaboration, even if these have not yet been widely publicised and may still be rudimentary with a limited ability to interoperate other tools and systems.

Level 3 - the defined level

Level 3 organisations typically understand the business meaning of research data (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.

1. **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 the implementation of procedures and understand their benefits. Policies are enforced and testing is done to ensure that requirements are being met.
2. **Support services** are now fully supported financially. They are well developed, documented and publicised. 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.
3. **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. IT facilities are more standardised and some formal testing of facilities and services is done.
4. **Metadata management:** Data is recognised as an organisational asset, and having high-quality metadata is recognised as being essential. Domain-relevant metadata standards are used to create rich object-level metadata which begins to be applied to instrument-generated metadata. Processes exist to link this to collection-level metadata. Metadata records are created for all newly identified datasets, people, activities, services and institutions and shared with Research Data Australia and/or other portals. All new datasets and related objects are assigned a PID (e.g. a DOI or ORCID) and are fully described using standard metadata schemas and vocabularies. Standard data licences are clearly specified. Systems in which metadata can be entered and/or gathered from are connected, so information is only collected once, in a formal structure that can be crosswalked to other schema.
5. **Research Data:** Well-defined internal and external data storage facilities are available and are in the process of being integrated with other data management services. Data services of the individual domains within the institution are still being fully developed. Provenance information and standards

begin to be routinely collected but may not be standardised. Data begins to be managed using standard open formats, using machine-readable vocabularies and ontologies enabling it to be linked and shared with other systems. Systems begin to be connected and the data shared and reused. APIs are developed enabling data retrieval using standardised communications protocols. Development of support and analysis tools to release the benefits of reusing and sharing the data is undertaken.

Level 4 - the managed level

At Level 4, senior administrators recognise research data as an organisational 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.

1. **Policies and procedures** which have an impact on research data management have now been absorbed into the organisational culture and are consistent across all areas. Procedures are fully understood and implemented. Compliance is audited.
2. **Support services:** There has been widespread uptake of services across the organisation. Curation and preservation are acknowledged as critical to the institutional mission and supported accordingly. Some level of auditing is undertaken.
3. **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. Formal, automated testing of facilities, services and data quality is performed.
4. **Metadata management** 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 along with related entities are findable and accessible. They are assigned a DOI and are fully described using standard metadata schema, identifiers and vocabularies in a machine-readable format and are available across disciplines. Standard data licences are clearly specified, and provenance metadata is available to enhance reusability. Metadata is routinely contributed to Research Data Australia and/or other portals. Metadata is audited to assess quality and availability.
5. **Research Data:** Routine data management is an established practice with well-defined internal and external data storage facilities capable of handling the unique data requirements of the entire range of domains within the institution. The data has associated metadata including provenance information automatically collected and is stored in recognised standard open machine-readable formats with standard variable names, units and vocabularies. Data are hosted and managed in formats and structures that enable retrieval using standardised communications protocols and tools and shared across domains. Tools have been developed to analyse and link data automatically and quickly between differing systems turning data into knowledge not only for the domain the data was created for but also for other domains. Regular measures of their effectiveness and value are undertaken

Level 5 - the optimising level

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

1. **Policies and procedures** are subject to review and improvement in order to ensure that external requirements and internal needs are being met.
2. **Support services:** Customer feedback forms the basis of updating and continuous improvement of services.
3. **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.
4. **Metadata management** provides the basis for rapid discovery of data in all disciplines, both within the institution and beyond. Continuous improvement is applied to processes and capabilities.
5. **Research Data:** Data management is fully automated and machine accessible throughout the organisation and provides the basis for rapid discovery of data in all disciplines, both within the institution and beyond. Continuous improvement is applied to processes and capabilities

CMM viewed through the 5 Elements

Policies and procedures

Level 1 - the initial level

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.

Level 2 - the development level

Policies and procedures: Key to moving from Level 1 to Level 2 is the 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 harmonised 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 institutionalised. Instead, there is reliance on a central person or group to understand the issues and implement procedures reliably and consistently.

Level 3 - the defined level

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 the implementation of procedures and understand their benefits. Policies are enforced and testing is done to ensure that requirements are being met.

Level 4 - the managed level

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

Level 5 - the optimising level

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

Checklist for assessing Policy and Procedure capability

[Outline of a Research Data Management Policy](#) for Australian Universities/Research Institutions is a basic starting point for institutions that are intending to write, or update, their research data management policy.

Support services

Level 1 - the initial level

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.

Level 2 - the development level

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 underway, and data is starting to be assessed for longer-term retention.

Level 3 - the defined level

Support services are now fully supported financially. They are well developed, documented and publicised.

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.

Level 4 - the managed level

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

Level 5 - the optimising level

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

Checklist for assessing Research data support capability

The following list is a starting point for assessing the capability maturity for providing requisite advice and support services: such as online toolkits, information, training.

- Do support staff (eg Librarians, data managers, research office staff etc) have the skills needed to ensure data is managed as an institutional asset?
 - Conducting [data interviews](#)
 - [Librarians, IT, and skills for data management Projects](#)
 - [23 \(research data\) Things](#)
- Is [institutionally contextualised research data management training](#) accessible to all students and staff?

IT infrastructure

Level 1 - the initial level

IT infrastructure is disorganised and inadequate to support research data storage and management. Expertise is held by only a few individuals who may be unable to cope with the demand. The capability and processes to mint persistent identifiers are non-existent or limited.

Level 2 - the development level

IT infrastructure: Reaching Level 2 means that the organisation has recognised 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 recognised 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 in place to manage identities, authentication and access with the ability to mint persistent identifiers.

There are periodic monitoring and evaluation of processes and the organisation can respond to periods of stress in a systematic rather than ad hoc fashion

Level 3 - the defined level

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. IT facilities are more standardised and some formal testing of facilities and services is done.

Level 4 - the managed level

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. Formal, automated testing of facilities, services and data quality is performed.

Level 5 - the optimising level

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.

Checklist for assessing IT Infrastructure capability

[Checklist for assessing IT Infrastructure capability for Data Management](#) is a basic starting point for institutions who wish to assess their IT infrastructure capability maturity in terms of research data management.

This checklist asks questions a number of questions around various IT infrastructure considerations for research data which will help institutions gather a good sense of the level of maturity of their IT infrastructure

e.g. If datasets and their metadata are separate, are they stored together or could they become separated? How closely are the data and metadata linked?

- Storage for data and metadata

- Identity management, authentication & access
- Internal & external network connectivity
- Access to discipline-specific tools to support analysis
- Software development
- Visualisation
- Collaborative environments
- High-performance computing (HPC)

Metadata management

Level 1 - the initial level

Metadata management is chaotic and understood only by unsupported individuals. It is regarded as a 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 may be adept at meeting the metadata needs of their own research, but they are unsupported by institution-wide services which would help them to share data outside of their group...

Level 2 - the development level

Metadata management: The organisation recognises the need for comprehensive metadata to underpin good research data management. It recognises the benefits that this can bring to research, and the administrative requirements of the organisation and/or external funder requirements. 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 and/or sharing purposes. A data licensing regime preferably using an existing standard is adopted. Limited numbers of metadata records with persistent identifiers (PIDs) not just for collections but also for related people, activities, services and institutions and are created and shared with Research Data Australia and/or other portals.

Level 3 - the defined level

Metadata management: Data is recognised as an organisational asset, and having high-quality metadata is recognised as being essential. Domain-relevant metadata standards are used to create rich object-level metadata which begins to be applied to instrument-generated metadata. Processes exist to link this to collection-level metadata. Metadata records are created for all newly identified datasets, people, activities, services and institutions and shared with Research Data Australia and/or other portals. All new datasets and related objects are assigned a PID (e.g. a DOI or ORCID) and are fully described using standard metadata schemas and vocabularies. Standard data licences are clearly specified. Systems in which metadata can be entered and/or gathered from are connected, so information is only collected once, in a formal structure that can be crosswalked to other schema.

Level 4 - the managed level

Metadata management 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 along with related entities are findable and accessible. They are assigned a DOI and are fully described using standard metadata schema, identifiers and vocabularies in a machine-readable format and are available across disciplines. Standard data licences are clearly specified, and provenance metadata is available to enhance reusability. Metadata is routinely contributed to Research Data Australia and/or other portals. Metadata is audited to assess quality and availability.

Level 5 - the optimising level

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

Checklist for assessing Metadata management capability

The following list is a basic starting point for assessing the capability maturity for providing the requisite Metadata infrastructure so data records can be used for both internal and external purposes.

- Is there provision for [storage of metadata](#)?
- Are research [data collection descriptions](#) able to contain all elements necessary for discovery and reuse?
- Are appropriate [metadata schemas](#) used for collection descriptions?
- Are persistent identifiers available and interconnected to enable assessment of research data impact?
 - [data citation](#) including DOIs (Digital Object Identifiers)
 - [Identifiers for researchers](#) including ORCID

Research Data management

Level 1 - the initial level

Data is managed in an ad-hoc manner: data sharing and reuse is inconsistent and undertaken in an unsupported manner. Some groups or individuals may utilise good practices but are unsupported by institution-wide services and data management may be costly, time-consuming with considerable duplication.

Research Data storage facilities may be rudimentary such as shared drives or free web services. The data has little-associated metadata with the data stored in a range of formats including proprietary formats requiring the use of proprietary software to access and use the data. The ability and inclination to share or reuse the data outside of the purposes for which it was collected is very limited. Tools to analyse the data are often proprietary or limited and the ability to share data between systems is cumbersome, time-consuming and requires specialist skills.

Level 2 - the development level

The organisation recognises data as an institutional asset and persistent data storage facilities and standards are required to underpin and support good research practice. Suitable data storage facilities internal and external are identified with some being developed in-house. Appropriate open data standards, formats, protocols, ontologies and vocabularies are identified and begin to be applied to the data. Tools are developed to support the automatic storage of data with the metadata from various systems and instruments. The data incorporates persistent identifiers. Tools are available to meet researcher needs for analysis, visualisation and collaboration, even if these have not yet been widely publicised and may still be rudimentary with a limited ability to interoperate other tools and systems.

Level 3 - the defined level

Well-defined internal and external data storage facilities are available and are in the process of being integrated with other data management services. Data services of the individual domains within the institution are still being fully developed. Provenance information and standards begin to be routinely collected but may not be standardised. Data begins to be managed using standard open formats, using machine-readable vocabularies and ontologies enabling it to be linked and shared with other systems. Systems begin to be connected and the data shared and reused. APIs are developed enabling data retrieval using standardised communications protocols. Development of support and analysis tools to release the benefits of reusing and sharing the data is undertaken.

Level 4 - the managed level

Routine data management is an established practice with well-defined internal and external data storage facilities capable of handling the unique data requirements of the entire range of domains within the institution. The data has associated metadata including provenance information automatically collected and is stored in recognised standard open machine-readable formats with standard variable names, units and vocabularies. Data are hosted and managed in formats and structures that enable retrieval using standardised communications protocols and tools and shared across domains. Tools have been developed to analyse and link data automatically and quickly between differing systems turning data into knowledge not only for the domain the data was created for but also for other domains. Regular measures of their effectiveness and value are undertaken

Level 5 - the optimising level

Data management is fully automated and machine accessible throughout the organisation and provides the basis for rapid discovery of data in all disciplines, both within the institution and beyond. Continuous improvement is applied to processes and capabilities.

Checklist for assessing Data Management capability

The following list is a basic starting point for assessing the maturity for providing the requisite Data Management infrastructure so data is Findable, Accessible, Interoperable and Reusable.

- [The FAIR data principles](#)

Feedback?

We welcome your feedback on this guide. Please email contact@ands.org.au with any comments or questions.

About ANDS

The Australian National Data Service (ANDS) makes Australia's research data assets more valuable for researchers, research institutions and the nation.

ANDS is a partnership led by Monash University in collaboration with the Australian National University (ANU) and the Commonwealth Scientific and Industrial Research Organisation (CSIRO). It is funded by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).

This work is licensed under a [Creative Commons Attribution 4.0 International License](#). You are free to reuse and republish this work, or any part of it, with attribution to the Australian National Data Service (ANDS).