

Metadata Stores Solutions Guide

Updated July 2011

The Metadata Stores Program in ANDS has been set up to support development of solutions for gathering, managing, and publishing metadata about research collections and other related entities, resulting in collection discovery and reuse. This guide primarily focuses on solutions which can be deployed at an institutional level, in line with the Metadata Stores Program focus; an appendix outlines considerations for deploying solutions with narrower scope. The guide describes the current functionality and feature base of mature institutional solutions, as well as the status of solutions currently under development. The guide will be updated periodically as solutions mature, and the Metadata Stores Program may fund other solutions in the future if necessary.

Table of Contents

1. Types of metadata store	2
2. ANDS Priorities	3
3. Choosing a solution	4
4. Solution Integration	4
5. Mature Institutional Solutions	5
5.1 VITRO (ANDS-funded project: EIF029, EIF 002)	5
Contact: Simon Porter, simon.porter@unimelb.edu.au ; Jo Morris, j.morris@griffith.edu.au	5
5.2 ReDBox: Research Data Box (ANDS-funded projects: EIF040, EIF048)	6
5.3 MyTARDIS (Squirrel: ANDS-funded project: EIF019; MeCAT: ANDS-funded project: EIF020, EIF037)	7
Contact: Anthony Beitz, anthony.beitz@monash.edu (Squirrel); Alistair Grant, alistair.grant@synchrotron.org.au (MeCAT)	7
5.4 ORCA	8
5.5 Geonetwork (ANDS-funded project: EIF023)	9
6. Institutional Solutions under development	9
6.1 University of Queensland	9
6.2 University of Sydney	10
6.3 Generic Object Store	10
7. Appendix: Local Metadata Stores	10



1. Types of metadata store

ANDS distinguishes between metadata stores based on their coverage; the granularity of data that they describe; and the specialisation of their descriptions.

Based on coverage,

- A *local* metadata store has coverage over data produced by a single instrument or research group.
- An *institutional* metadata store has coverage over data produced across the institution, typically by a variety of research groups and disciplines.
- A *national* metadata store has coverage over data produced across a country, by a variety of institutions. (*Research Data Australia* is an instance of a national store.)
- A *discipline-specific* metadata store has coverage over data produced within a discipline, across a variety of research groups, institutions, and (typically) countries.

Metadata about research collections should ideally be created and managed close to where the research data is gathered, in local metadata stores tightly integrated with research groups and projects. This metadata should be relevant to the researcher's needs, and be accessible within the researcher's immediate work context.

However, the metadata stores with broader coverage are essential if the collections are to be discovered and tracked outside that immediate context, across a discipline or an institution. Stores with broader scope have more users than local stores, and institutional and national stores use more generic formats, applicable to more domains. Stores with broader scope typically act as metadata aggregators, gathering metadata (or appropriate distillations of metadata) from local systems.

Based on granularity,

- A *collection-level* metadata store describes data collections (collections, datasets, etc).
- An *object-level* metadata store describes individual data objects (files, database rows, spreadsheets, physical objects).
- An *integrated* metadata store describes both individual data objects and the collections that they comprise, in the one system. They are typically coupled with data storage for the data being described.

Based on specialisation,

- A *specialist* metadata store captures metadata of interest to a discipline specialist.
- A *generic* metadata store only captures metadata which is of interest to a general audience. (For example, university administration, university research office, general public, researchers in other fields).

The specialisation of a metadata store depends on who will be using it. Both are necessary: specialist metadata may be what is generated first (especially if automated), and what other researchers need; but it usually cannot be repurposed automatically into generic metadata.

Institutional solutions tend to be generic, since their metadata descriptions cannot be discipline-specific. However an institutional solution can be configured to provide different solutions for different disciplines.

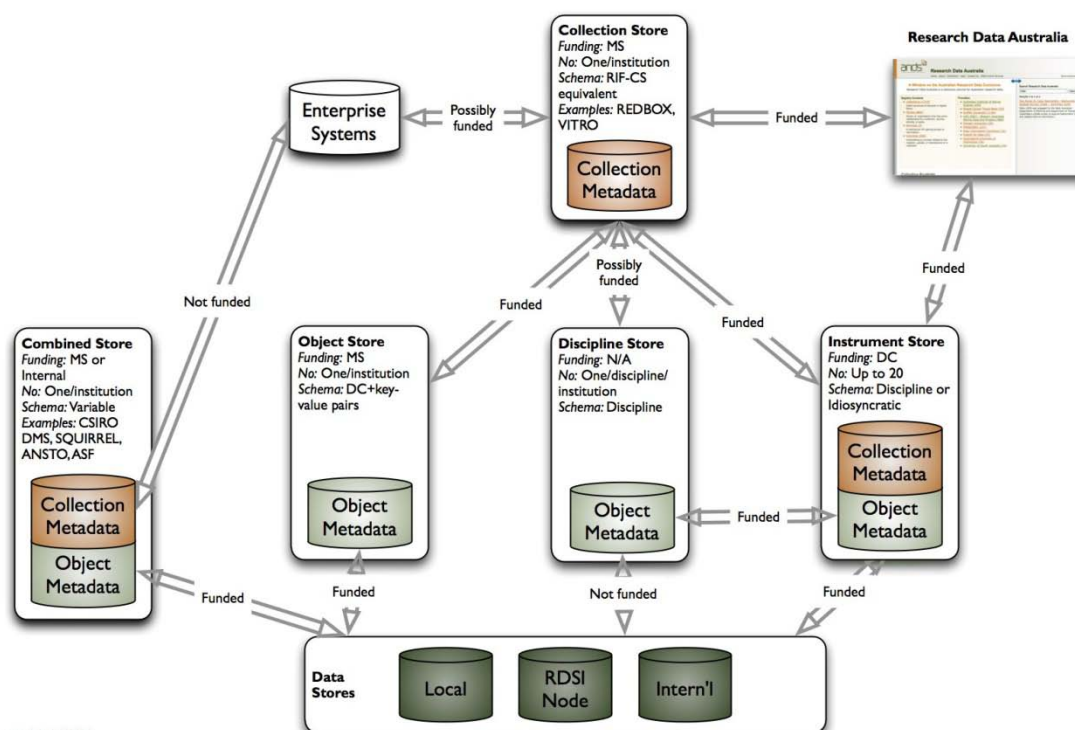
Object-level stores are typically specialist, because discipline knowledge is needed to make sense of individual data objects; data capture often produces specialist metadata automatically. If a specialist store is managing data objects, and the discipline needs to organise those objects into collection, it will usually do so as an integrated store, so that the management of objects and collections is co-located.



The solutions described in this guide can be classified as follows:

	Generic	Specialist
Collection-level	VITRO, ReDBox, ORCA, University of Queensland	Geonetwork
Object-level	(Institutional Repositories)	(Data Capture stores)
Integrated		Squirrel, MeCAT

The dependencies between the different types of metadata store and other systems in institutions are given in the following diagram:



Version 4.1, 16/3/11

The Metadata Stores program has been funding institutional solutions, which means they are generic. The integrated solutions funded by the program are configurable across multiple disciplines; so they are tantamount to being generic. The Metadata Stores program has not funded any object-level stores (though the Data Capture projects have); the program is considering funding a generic object-level metadata store, as we describe later in the guide.

2. ANDS Priorities

ANDS is interested in sound management of research data, so it encourages the use of local metadata stores. (See appendix for some criteria.) However, the priority for the Metadata Stores Program is to make metadata (and the collections it describes) more widely available. To make this happen, ANDS prioritises building broader-scope metadata stores, which will do aggregation of metadata. It is easier for users to discover collections if they need to search for collections in fewer places, and if the descriptions of collections in those places look similar to each other. It is also easier to deal with description and links between them and other entities, if consistency has been applied beforehand.



Discipline-based stores, such as portals and databanks, are where researchers go first to describe and discover collections within their field of expertise. ANDS wants to ensure that the metadata fed to existing international disciplinary repositories is of high standard. However the priority for the Metadata Stores Program is to fund institutional metadata stores, and to have them feed into a national metadata store (Research Data Australia). Few institutions already have this infrastructure in place; this guide describes options for institutional solutions.

Information on activities and parties is already being managed at an institution-wide level, and needs to be sourced from the institution to avoid duplication. Institutions have a stake in the research collections they host; metadata stores should support the institution's requirements in working with its data collections. Institutions are best placed to implement sustainable data management practices, which includes tracking, managing, and curating research collection metadata.

3. Choosing a solution

The ANDS Metadata Stores Program is funding development of several metadata stores solutions which are described below; other solutions are also being developed by ANDS partners outside the program framework. The solutions which ANDS is funding take different approaches; this is deliberate, to ensure that a range of environments can be supported. ANDS encourages its partners to consider deploying one of these solutions, rather than duplicating development effort internally.

In considering what metadata store solution to take up, organisations should bear the following questions in mind:

- What is the coverage of the metadata store?
 - A local metadata store, specific to a department or research lab, will describe a relatively homogeneous set of objects and can afford to provide more detailed, discipline-specific descriptions across its objects. An institutional store will need to use more generic, high-level descriptions of data objects.
- How will the metadata store provide information to other consumers?
 - Research Data Australia is not the only or even the most important consumer for the store. Organisations need to consider the following as other possible consumers: the institution itself (including university administration and research planning), discipline repositories, institutional repositories, government and funding agencies, research collaborators, individual researchers, and the general public.
- What will the granularity and specialisation of metadata be in the metadata store?
 - The broader the coverage of the metadata store, the more difficult it becomes to provide metadata about individual data objects, or detailed metadata specific to a discipline. Metadata stores with broad coverage, such as institutional stores, are normally built as aggregators, and should refer users to local stores for more detailed metadata.
- Will the metadata store act as an aggregation point for other metadata stores?

Several organisations may have a federal structure for managing data—as is typically the case for a university. The organisation may need to source collection descriptions from other stores, rather than assume that all descriptions will be pushed into the central store.

4. Solution Integration

Descriptions of data collections should not be seen in isolation: they need to be related to other kinds



of information, which may be stored and managed in different data stores. ANDS, for example, requires information about related parties and activities to accompany collections. The authoritative sources for such metadata are HR and Research Office systems. Metadata stores should be reusing that metadata, rather than creating its own records, with potentially inaccurate information. Metadata should, ideally, only ever be created once and then re-used as needed.

If the contextual information is common across different institutions, it is appropriate to have a common external authority for the information. A common description of a grant or researcher across institutions allows users to navigate between data collections held by different institutions, but involving the same research team members. This requirement is why ANDS is supporting the National Library of Australia, the Australian Research Council, and the National Health and Medical Research Council, in developing national infrastructure for researcher identity and research grant descriptions.

This means that deploying a metadata stores solution involves integrating multiple sources of information, possibly including external sources. If such data has already been aggregated or centralised in the institution (e.g. as a data warehouse), it can be exploited by institutional metadata stores. The form and purpose of any existing aggregation influences how it can be exploited; whether the metadata aggregation is being driven by the library or the research office affects what solution will best match the institution.

But providing context for research collections is a novel purpose for aggregating metadata in institutions. Most institutions will not have aggregated all the metadata needed for that purpose, and will not have integrated external sources such as the NLA and the ARC, in the forms ANDS is promoting. If institutions are not strongly centralised, they will need to work out how best to integrate those systems with their metadata stores. This includes working out whether to use manual or automatic feeds; how frequently to update their information based on external changes; and whether to include historical data. Because of the disparate systems involved, they will also need to do data modelling and analysis, to ensure that the data comes together coherently. This may require considerable effort in cleaning and deduplicating the metadata; but it has the payoff of ensuring that not only researchers, but the institution itself has a better understanding of the research it produces.

ANDS is endeavouring to ensure that the solutions it funds are redeployable elsewhere; but institutions will still need to do some work to get their instances up and running. The ANDS Metadata Stores Program may provide funding to help some institutions deploy metadata stores; but ANDS cannot guarantee that it will support all institutions. ANDS also cannot guarantee that it will be able to fund customisation of metadata stores solutions to meet local requirements.

5. Mature Institutional Solutions

5.1 VITRO (ANDS-funded project: EIF029, EIF 002)

Contact: Simon Porter, simon.porter@unimelb.edu.au; Jo Morris, j.morris@griffith.edu.au

This project is using VIVO, a semantic web, triplestore-based approach to gathering and sharing research data. VIVO has been developed by a consortium in the US (originally Cornell University): <http://www.vivoweb.org/> . The project is based on the code base, VITRO, and the VIVO ontology for describing research: <http://vitro.mannlib.cornell.edu/> . This ontology has been enhanced to support ANDS requirements, and the enhancements (called the ANDS VITRO ontology) are being built as a community initiative involving several Australian universities. The ANDS VITRO ontology is extensible and more detailed than RIF-CS, and can be applied to a wide variety of purposes. The ANDS VITRO ontology is available at <http://eresearch.griffith.edu.au/ANDS/vitro/ANDS-VITRO.owl> .

The VIVO approach provides an integrated University-wide view of research. VIVO came into being



because there was a need to present views of research identity that crosses organizational boundaries, needed in the absence of established whole-of-University reporting practices in the US. VITRO is well-suited to institutions in which the research takes the lead in implementing aggregation in collaboration with the library. VITRO can provide such institutions with detailed modelling of their research collections and researchers—e.g. in publishing researcher profiles across the institution.

As a semantic web-oriented product, Vitro is based on triplestore technology, which enables powerful SPARQL queries of metadata and benefits from inferencing capabilities. The VIVO approach also offers institutions the ability to create a whole of University Research Data Registry. (see <http://docs.lib.purdue.edu/iatul2010/conf/day3/3/>)

The ANDS VITRO metadata store solution enables Linked Data approaches to research data, being RDF-based, but it is currently oriented to collection-level descriptions of data, and not more fine-grained descriptions.

The VITRO platform can handle both automated and manual feeds of research data into the triple store, from single or multiple data sources. As with all University-wide metadata aggregator solutions, building those feeds is still the responsibility of the deployer, and may involve significant effort in cleaning up the data, and in modelling the connections to outside data. The effort is substantially less if the institution already has a data warehouse in place. In any case, data must be mapped from existing data stores to the ANDS-VITRO ontology to be ingested by the system.

Code is already in place for converting the RDF of VITRO to RIF-CS and for providing an OAI-PMH harvest point from VITRO: <http://code.google.com/p/ands-vitro-code/> . EIF 002 produced Kepler workflows to automate populating VITRO for their metadata hub, as well as providing a harvester; the code and documentation is available at <https://df.arcs.org.au/quickshare/b77f99d1cfea2ddc/Package.zip>

The current authoritative overview presentation for VITRO in Australia is at:

http://vivoweb.org/files/2010_Australian_Community.pdf

<http://www.vimeo.com/15252525>

VITRO is being taken up by the University of Melbourne (EIF029), Griffith University (EIF 002), the Queensland University of Technology (EIF 002), Victoria University, Curtin University, and the University of Western Australia. Of these deployments, most are using VITRO as an interface and export tool: the deployments are ingesting mapped research activity data that is stored and managed in more traditional formats, in institutional data silos (Oracle, Mediaflux, Research Master). EIF 002 is now complete. Project development for EIF 029 commenced in March 2010, and is scheduled to run until September 2011.

5.2 ReDBox: Research Data Box (ANDS-funded projects: EIF040, EIF048)

Contact: Duncan Dickinson, duncan@dickinson.name

This solution uses its own instance of the Fedora-commons data store to store and disseminate metadata on research collections. The store uses as its front-end The Fascinator faceted search software developed at the University of Southern Queensland.

The ReDBox solution takes an Institutional Repository approach to research metadata: metadata is collected through user forms as an interface to the repository, as well as automated integration of the repository with other campus systems. Metadata already collected in the repository is repurposed for disseminating research data. The solution is well-suited to institutions which already have a strong repository presence, with established work practices for repository management (so that the repository



is enhanced by deploying the solution), and in which the library takes the lead in implementing aggregation.

Metadata can be added to the system either manually or via automatic harvesting from other systems. The manual entry of research metadata is supplemented by alerts issued by interfacing systems—including grants databases and disk storage: these point repository maintainers to new instances of research data to be processed. The metadata is internally stored using the VITRO ontology, in which the ReDBox project are also stakeholders. This means that the solution offers semantic consistency with institutions using VITRO; but deploying and using the software does not require developing semantic web skills.

The ReDBox solution also includes the “Mint”, infrastructure supporting controlled vocabularies used in research metadata, and treating them as Linked Data. The Mint allows validation of data entered by users in the forms interface, and it is also how ReDBox deals with party and activity identifiers. The use of unique identifiers ensures data integrity for the records they identify.

Because the solution is repository-driven, it can support description of individual resources as repository objects. The repository can also store descriptions of data collections which themselves are stored remotely (e.g. in a large disk array) or can be used to house data collections themselves. Being Fedora-based, the solution has built-in support for OAI-PMH, and for multiple metadata schemas describing the same object.

The project is documented online in the ANDS partners project blog and the project wiki:

<http://www.ands-partners.org/blog/category/redbox/>

<https://sites.google.com/site/redboxmint/>

The code for ReDBox is available at <http://code.google.com/p/redbox-mint>

ReDBox has been taken up at the University of Newcastle, and will be taken up at Flinders University. Project development on ReDBox 1.0 is now complete. The Queensland Cyber Infrastructure Foundation is funded until June 2012 to help deploy ReDBox, and to develop fixes and enhancements to the solution (ReDBox 1.1).

5.3 MyTARDIS (Squirrel: ANDS-funded project: EIF019; MeCAT: ANDS-funded project: EIF020, EIF037)

Contact: Anthony Beitz, anthony.beitz@monash.edu (Squirrel); Alistair Grant, alistair.grant@synchrotron.org.au (MeCAT)

The Squirrel and MeCAT projects are both extending the MyTARDIS codebase for use as an institutional metadata store. MyTARDIS (<http://tardis.edu.au>) was initially developed for storing datasets and metadata in protein crystallography. The code is now being made more versatile to easily fit in with other discipline-specific and generic approaches to research data management and reuse. The system under development allows researchers to organise, describe, find, reuse and share their data, which is stored in a central data store.

The two projects are coordinating their work, to ensure that the codebase remains common between the two. We refer to the projects together as TARDIS systems in the following.

The Squirrel project involves Monash University. Squirrel includes a schema registry allowing users to define their own metadata schemas to describe their data. Squirrel aims to support the self-deposit, by researchers and research support intermediaries, of discipline-specific metadata and the descriptive and administrative metadata which needs to be provided to research offices, libraries, records and archives, and Research Data Australia. It will provide integration with externally stored information about parties (e.g. researchers) and activities (e.g. grant-funded projects) through web services. Web



services for Monash are being provided by another ANDS-funded project (EIF 038), and would require local development for any new deployment.

The MeCAT project is extending MyTARDIS for deployment on the Infrared and Small Angle X-Ray Spectroscopy beam lines at the Australian Synchrotron (EIF 020), and five beam lines at the Bragg Institute, which is part of the Australian Nuclear Science and Technology Organisation (ANSTO) (EIF 037). The enhancements being made to MyTARDIS will enable users to search and download data and metadata from the facilities and assist beam line scientists in managing data from their beam lines, supporting users and improving beam line operations. These enhancements include:

- Storing more detailed information on the equipment being used to conduct experiments and the samples being analysed at the facilities;
- Extending the authentication and authorisation capabilities of MyTARDIS to provide more fine grain control over who can access data;
- Extending the search capabilities to work with scientific data in multiple disciplines;
- Detailed logging and audit trails for tracking access and modification of metadata.

TARDIS systems will include an independent OAI-PMH provider, meaning several complementary dissemination models can be supported, including:

- direct harvest of RIF-CS for registering collections with Research Data Australia; and/or
- provision of discipline-specific metadata to a discipline-specific portal (e.g. TARDIS in the case of crystallography); and/or
- transfer of metadata to other repositories or aggregators that need information about research outputs, to meet institutional goals around research assessment/impact, and compliance with legislation and protocols for record-keeping and responsible research.

TARDIS systems are aimed at facilitating research data management and reuse, and are not intended to function as a metadata aggregator.

TARDIS systems are currently being taken up at: Monash University, the Australian Synchrotron, ANSTO, the Ian Wark Institute, and the Royal Melbourne Institute of Technology. Squirrel project development commenced in February 2010, and is scheduled to run until August 2011. MeCAT project development commenced in March 2010, and is scheduled to run until December 2011.

The MyTARDIS codebase is available at <http://code.google.com/p/mytardis/>. Further more information about the projects please visit the following websites: <http://www.monash.edu.au/ands> (Squirrel), <http://mecatproj.wordpress.com/> (MeCAT)

5.4 ORCA

Contact: ANDS, services@ands.org.au

ANDS has internally developed ORCA as a metadata store for managing the RIF-CS records that are collected in Research Data Australia. ORCA is set up to provide OAI-PMH feeds of the RIF-CS records it stores, and also has authoring support for RIF-CS. That means that ORCA adequately supports the narrow goal of authoring and disseminating RIF-CS records.

However ORCA does not provide the broader support of research data management or integration with external data that ANDS sees as desirable in metadata stores. For that reason, ANDS does not encourage using ORCA as a substitute for deploying a fully-fledged metadata store.

Currently ANU is planning to augment ORCA for use as an institutional metadata store.



5.5 Geonetwork (ANDS-funded project: EIF023)

Contact: <http://geonetwork-opensource.org/>

Geonetwork is an open source catalogue application to manage spatially referenced resources. It provides powerful metadata editing and search functions, as well as an embedded interactive web map viewer. It is currently used in numerous Spatial Data Infrastructure initiatives across the world.

“GeoNetwork has been developed to connect spatial information communities and their data using a modern architecture, which is at the same time powerful and low cost, based on the principles of Free and Open Source Software (FOSS) and International and Open Standards for services and protocols (a.o. from ISO/TC211 and OGC)”.

Geonetwork is targeted as a researcher-orientated metadata solution. Deployments of Geonetwork should couple it with a repository for storing the data described by the metadata; the standard implementation includes data storage, but is not very robust. Oracle is commonly used, but PostgreSQL and MySQL have both been used successfully.

Geonetwork has been implemented by a number of Australian public sector agencies, with enhancements. In particular, the BlueNet MEST, <http://anzlicmet.bluenet.utas.edu.au>, is an enhanced version of GeoNetwork 2.2. Amongst other things, these enhancements provide support for the profiles of the AS/NZS-19115 geographic metadata standard—Marine Community Profile (MCP).

The AODN (Australian Ocean Data Network) ANDS project (EIF023) uses the AODN MEST, which is based in turn on the BlueNet MEST. See: <http://mest.aodn.org.au/geonetwork/srv/en/main.home>. Geonetwork can be used as a harvester to pull together records from other organisations using the same tool: the AODN MEST (Metadata entry and search tool) currently harvests 9 other Geonetwork MESTs.

The application of the BlueNet MEST as the Australian Government Metadata Entry Tool, and the ANZLIC Metadata Entry Tool, is the result of collaboration between the Australian Office for Spatial Data Management (OSDM) <<http://www.osdm.gov.au>>, ANZLIC: the Spatial Information Council <<http://anzlic.org.au>>, GeoScience Australia <<http://www.ga.gov.au>>, and the BlueNet project <<http://www.bluenet.org.au>>.

A crosswalk from the ISO19115 MCP (see <http://www.aodc.gov.au/index.php?id=37>) to RIF-CS has been implemented. The AODN have implemented an OAI-PMH harvest point from the AODN MEST into Research Data Australia.

6. Institutional Solutions under development

The following solutions are currently being developed using ANDS funding.

6.1 University of Queensland

Contact: Nigel Ward, n.ward4@uq.edu.au

The University of Queensland is currently developing an institutional metadata store solution to respond to the needs of Research Data Australia, and compliance requirements from university administration. Metadata for collections, parties, activities and services is stored in a relational database (Postgres). The system provides a RESTful interface for creating, managing and accessing entities via the Apache Abdera implementation of the Atom Publishing Protocol. Individual entities are versioned and have multiple representations (currently Atom, RDF, RIF-CS v1.2 and XHTML representations). The system supports simple publication workflows where entities are either internal, under review, or published.



The Seeding the Commons project, under which the solution is being developed, is documented at <http://itee.uq.edu.au/~eresearch/projects/ands/stc/>. The Atom representation of Research Data used by the store is documented at <http://dataspace.uq.edu.au/doc/atom>

This solution is not funded by the Metadata Stores Program, so it is not necessarily being developed to be redeployable. ANDS will provide updates if this changes.

6.2 University of Sydney

The University of Sydney is currently producing software requirements for an institutional metadata store, to be used federally across the collections; we will provide a more complete description of their solution in August. This solution is not funded by the Metadata Stores program, so it is not necessarily being developed to be redeployable; ANDS will provide updates if this changes.

6.3 Generic Object Store

In our discussion of metadata store types above, we noted that the Metadata Stores program has funded collection-level stores and integrated stores, while object-level stores have been created under Data Capture projects, and aligned to particular instruments. Capturing the metadata early is the right thing to do, but the result has been a proliferation of specialist metadata stores, geared to particular projects and instruments, which are difficult to redeploy.

Many instruments and projects already have infrastructure for object-level metadata stores, but not all do. ANDS believes it is necessary to provide a fallback solution for researchers who don't have access to object-level infrastructure; the need for such a fallback would likely be most pronounced in the humanities.

This fallback would be a generic object-level metadata store:

- it would store descriptions of data objects in a generic schema, applicable across disciplines (for instance, Dublin Core);
- it would be extensible, to allow additional specialist metadata to be recorded about objects;
- it would link object descriptions to the storage for the objects themselves; it should not provide storage for the objects, so that it can be deployed across a range of storage solutions;
- it should provide support for aggregating objects into collections, by imposing some sort of membership graph over object identifiers (for instance, OAI-ORE);
- it should provide a single URI for any collections that it gives an aggregation for;
- it needs not provide separate collection-level descriptions for any collections it aggregates: the object-level store complements collection-level stores (which would use the URI for the collection that the object-level store provides).

The Metadata Stores program is considering funding activity to develop and deploy a generic object-level store. Such a store has some functional overlap with institutional repositories, which also provide generic metadata of data objects; but the store would be a much more lightweight solution.

7. Appendix: Local Metadata Stores

The Metadata Stores program is not directly funding local metadata stores, specific to a research group. It does fund integrated metadata stores (Squirrel, MeCAT) which can be configured for discipline-specific use; but these are still institutional stores: they can be tightly coupled with instruments and collection production, but they can also be coupled more loosely.

However, local metadata stores are crucial to good data management and to populating broad-scope



metadata stores. The Data Capture projects funded by ANDS often involve setting up a local metadata store, specific to the instrument, for that reason. ANDS cannot recommend metadata stores for specific disciplines or projects; however, researchers should consider the following requirements for their local stores.

The local metadata store should:

1. Store metadata in a format which is in common use in the discipline.
2. Store metadata that supports discovery and evaluation of data (e.g. keywords).
3. Store metadata that supports reuse of data (e.g. experimental configuration, interpretation of dependent variables, access rights—these may simply be a link to a separate file or a paper).
4. Export metadata to other formats commonly used in describing metadata, especially in metadata aggregators (note that OAI-PMH requires a feed to be available in Dublin Core.)
5. Support aggregation of metadata (harvesting and/or syndication)—especially for international discipline repositories
6. Support automated gathering of metadata from instruments (e.g. file header), and of related metadata from other databases (e.g. HR systems, grants programs).
7. Integrate in researcher workflows with minimal disruption (e.g. through web services).
8. Allow error checking, validation, and use of constrained vocabularies.
9. Allow metadata describing both collections and objects within collections, if that is appropriate to the discipline.
10. Allow hierarchical organisation of metadata, where appropriate to the discipline (e.g. ordering metadata by project and/or experiment).

Not all metadata store solutions will satisfy all requirements; automated metadata gathering and integration, in particular, are not widespread, and should not automatically disqualify a candidate store. All these features are worth considering in evaluating candidates, and researchers need to work out which features are priorities for them. The highest priorities are likely to be commonly used formats, hierarchical organisation, and aggregation support.

