Collaborative development of a multi-disciplinary research infrastructure for vocabulary creation, management, publication, discovery, access and re-use

Rowan Brownlee
Data Specialist
Australian National Data Service
rowan.brownlee@ands.org.au
http://orcid.org/0000-0002-1955-1262

Adrian Burton
Director, Services
Australian National Data Service
adrian.burton@ands.org.au
http://orcid.org/0000-0002-8099-7538

Jane Frazier
Data Librarian
Australian National Data Service
jane.frazier@ands.org.au
http://orcid.org/0000-0002-2374-4964

Richard Walker
Web Applications Project Officer
Australian National Data Service
richard.walker@ands.org.au
http://orcid.org/0000-0002-8446-0113

Abstract
The use of controlled vocabularies increases the value of data resources by improving discovery, interoperability, and re-usability. The Australian National Data Service (ANDS), in partnership with members of Australian research organisations, is developing a user-driven vocabulary service to support Australian research groups, universities, and research libraries to find and use published controlled vocabularies, as well as to create, manage, share and publish new controlled vocabularies. This paper follows the evolution of the development of the service. It provides a technical overview of the systems involved as well as a proposed model for support, outreach and responding to community need.
Introduction
Research, education, and government organisations in Australia and internationally are using or developing controlled vocabulary services. ANDS was approached by a group of research organisations requiring access to an Australian controlled vocabulary service. Working collaboratively with these partners to meet their needs, ANDS launched a service in September 2015.

The primary purpose of this paper is to present the ANDS Vocabulary Service, placing it in the context of controlled vocabularies and the needs of ANDS’s partners. The paper also seeks to contribute to discussions within the library community regarding controlled vocabulary services, and the question of the suitability of the ANDS service for use by this community. The ANDS service was developed in the first instance to meet the needs of a range of Australian research organisations. ANDS wants to understand the requirements of other organisations regarding controlled vocabularies, so as to inform future service development.

The paper commences with an overview of controlled vocabularies and their uses, and then presents the background and development of the ANDS service, including a discussion of the technical components of the service. The service support, promotion and outreach facets of the project are also presented.

Definition of a controlled vocabulary
Harpring defines a controlled vocabulary as “an organized arrangement of words and phrases used to index content and/or to retrieve content through browsing or searching. It typically includes preferred and variant terms and has a defined scope or describes a specific domain” (Harpring, 2010). A controlled vocabulary may be a simple list of terms (like an authority file), or a more complex organisation of terms with, for example, definitions, translations, or the expression of broader, narrower, and matching relationships.

To increase legibility, the rest of the paper frequently abbreviates the term “controlled vocabulary” to “vocabulary”.

Control of vocabulary terms
The value of control of vocabulary terms is simply expressed by The Confederation of Open Access Repositories (COAR) (via Wikipedia). The use of controlled vocabularies “ensures that everyone is using the same word to mean the same thing...” and that “…controlled vocabularies if applied consistently offer a uniform standard terminology for describing content, and assist with resource discovery, sharing and reuse of metadata and content, as well as integration between different systems”. To this end, COAR is developing controlled vocabularies to support the standardised description of research outputs (Confederation of Open Access Repositories, 2015).
Controlled vocabulary use: examples and benefits

Cultural aggregation services: multilingual discovery
The Europeana.eu portal provides access to millions of books, paintings, films, museum objects, and archival records that have been digitised throughout Europe (Europeana, 2015). Metadata from 2,000 institutions is supplied in many languages and expressed using a variety of controlled vocabularies. To aid search within the portal, contributed metadata is mapped to the Europeana Data Model, and efforts are made to align different labels for shared concepts. (Charles, 2014).

Libraries: breaking down information silos
The Linked Data 4 Libraries (LD4L) project is a collaboration of the Cornell University Library, the Harvard Library Innovation Lab and the Stanford University Libraries. The aim of the project is to improve the experience of users searching for scholarly information resources. Each institution holds many types of records about scholarly resources, which are locally created and maintained in separate management systems without being accessible to each other. These local silos also do not reference related information resources held by other institutions, and their locally managed descriptive terms and authorities are not necessarily linked to internationally accessible equivalents (Linked Data for Libraries 2015a).

By linking local controlled vocabularies and authorities to independently-maintained subject headings and internationally-managed unique identifier systems for authors, institutions, journals, events, and geographic entities, the project intends to connect information silos across the three institutions to enable sharing of everything that can be known about a resource.

Research: communication and shared understanding
Controlled vocabularies are important for research and scholarly communication, since these rely on precise concepts with shared and structured terminology (Brazma et al., 2001). The ability to replicate and test an experiment or communicate and verify a conclusion requires clear description and communication of concepts about which there is a shared understanding of meaning.

For example, Geoscience Australia contributes to the establishment and maintenance of nationally acceptable common standards for geosciences. In association with the Australasian Controlled Vocabularies working group and the IUGS Commission for the Management and Application of Geoscience Information, there is a suite of controlled vocabularies to support quality science and interoperable geoscience data. Some of the vocabularies include geological timescale, simple lithology, geological unit type, fault type foliation type and metamorphic facies (Geoscience Australia, 2015a) (Geoscience Australia, 2015b).

Research: data reuse, connecting, and collaborating
As a simple example, a tabular dataset in a spreadsheet would typically contain column headings describing the content of each table. A third party wishing to use this data would need to know what those headings mean in order to make sense of the data. To support data reuse, the data creator may supply a data dictionary to accompany the dataset.
Controlled terminology is also vital when seeking to relate datasets to each other, whether this is a relatively simple join across two datasets, or a meta-analysis involving the bringing together of multiple datasets that may have been created at different time periods or in different geographic regions. Wherever datasets are linked or merged, the connections need to be made at points that are known to be common (Berardini et al., 2004).

Data captured by the Integrated Marine Observing System (IMOS) are documented using an agreed community metadata schema, and most data are made available to the public via data infrastructure developed by eMarine Information Infrastructure (eMII) (eMarine Information Infrastructure, 2016). This infrastructure also serves members of the broader Australian marine community who participate in a collaborative venture, the Australian Ocean Data Network (AODN). To support the discovery and exchange of data using this infrastructure, IMOS and AODN collaborators rely on formalised descriptions of their data. These descriptions involve the use of controlled vocabularies, which underpin the indexing of data during dataset registration in data delivery systems, as well as provide the content for search facets available within the IMOS data discovery portal (Integrated Marine Observing System, 2016), (Australian Ocean Data Network, 2015). The use of controlled vocabularies also allows IMOS to integrate internationally with other initiatives such as ODIP (Ocean Data Interoperability Platform) (Ocean Data Interoperability Platform, 2014).

Health and medicine: patient outcomes and policy development

Medical doctors use controlled vocabulary terms and definitions within their diagnostic and records management systems to ensure precision and consistency. Australia is attempting to gain agreement on sets of terms that would enable the implementation of shared electronic patient records to improve patient outcomes and provide greater convenience (NEHTA, 2015).

The Australian Institute of Health and Welfare (AIHW) is a “a major national agency set up by the Australian Government … to provide reliable, regular and relevant information and statistics on Australia's health and welfare.” Its data and reports are used by the Australian government “in discussing, debating, and making policy decisions on health, housing and community services matters.”(Australian Institute of Health and Welfare, 2015)

AIHW maintains national data dictionaries as a reference of standardised terms and protocols used for collecting data that informs policy development (Metadata online registry 2015a). AIHW promotes the use of data description standards, arguing benefits including the following

- consistency of content and definition
- avoidance duplication and diversity of solutions
- reduction in cost of data development
- ability to access and maintain data that is comparable across many different data collections.

(Metadata online registry, 2015b) (Metadata online registry, 2015c) (Metadata online registry, 2015d)
ANDS involvement with controlled vocabularies

As a part of the National Collaborative Research Infrastructure Strategy (NCRIS), ANDS has a cross-facility and cross-sector role to increase the value of research data. Given the value and power that controlled vocabularies demonstrate in the preceding examples, ANDS is promoting and supporting the use of controlled vocabularies by the Australian research sector. ANDS promotes the use of internationally standard terminology as part of the strategic direction of data discovery, connectivity, and reuse (Australian National Data Service, 2011). The ANDS Vocabulary Service aims to:

- facilitate integration of Australian data into global and national frameworks (e.g., Oceans Data Interoperability Portal, National Environmental Information Infrastructure)
- promote reuse of standard terminology in research data within and across NCRIS facilities, research organisations, and research data providers
- facilitate and improve data generation, discovery, delivery, analysis, aggregation, integration, and synthesis
- provide a shared terminology management and publication platform across NCRIS
- make it easy for ANDS partners to add value to their data
- avoid inefficient ‘balkanisation’ of research data
- respond to the sector by documenting standard vocabularies used by Australian research organisations (Research Vocabularies Australia, 2015)

To these ends, ANDS has worked with a group of key stakeholders to establish:

- a discovery portal to make it easier to find controlled vocabularies relevant to research in Australia
- a repository with services to make it easier to access, download, and re-use controlled vocabularies
- an editing suite to facilitate the creation, management, and publication of vocabularies in human and machine-readable standardised formats

These systems are complemented by a set of support and outreach functions to make a complete service package.

In developing the service, ANDS has sought alignment with the needs of key stakeholders. Through early discussion, it became apparent that there was common and fundamental interest in the use of SKOS (W3C, 2012) for capturing and expressing controlled vocabularies. Accordingly, the ANDS service focuses at the level of controlled vocabulary (and not, for example, at the more complex level of ontology and the W3C Web Ontology Language OWL, which support the expression of a far greater range and number of specific relationships (W3C, 2013)).

SKOS is a means of encoding a wide variety of types of controlled vocabulary. The emergence and adoption of this and other related standards (such as the query language SPARQL) by vocabulary creators, and the corresponding availability of a spectrum of software tools, have greatly enabled standards-based approaches to vocabulary services.
Evolution of the ANDS Vocabulary Service

First version of the service
In 2012, following discussions with a number of research agencies, ANDS integrated a basic “vocabulary portal” into the Research Data Australia web site (Research Data Australia, 2015), with some limited support for browsing and publishing vocabulary metadata. Metadata included basic properties such as title, description, publisher, and subject keywords, as well as links to external sites that hosted the vocabulary data, but there was no support for any type of importing or direct upload of vocabulary data.

Vocabulary Create and Manage Project (VOCRAM)
More recent discussions with ANDS stakeholders revealed that interest in working with controlled vocabularies had increased to the point where a joint project would be worthwhile. A project was funded by ANDS as an informal collaboration involving staff from the Atlas of Living Australia (Atlas of Living Australia 2016), Bureau of Meteorology (Bureau of Meteorology 2015), CSIRO (CSIRO 2015c), Integrated Marine Observing System (Integrated Marine Observing System, 2016b), Terrestrial Ecosystem Research Network (Terrestrial Ecosystem Research Network, 2015), and ANDS. All parties collaborated to enable access to a vocabulary management tool and service for their own needs, and to enhance the potential for wider community uptake.

The project partners agreed on a conceptual workflow for vocabulary creation, management, and publishing to be enabled by project outcomes. This included a discussion of tasks and responsibilities to be undertaken by types of users including vocabulary editors and system administrators.

Exploration of workflow informed the development of requirements. Each partner contributed descriptions of desired features, which were assigned levels of priority. From an extensive initial list, the top 15 requirements were selected. These included:

- The ability to create, manage and edit controlled vocabularies and concepts
- Support for 'roles' that can create terms, review/moderate suggested terms, track terms, and view vocabulary status and change logs
- The ability to connect to other existing vocabulary services for publication purposes
- A user-friendly interface for non-experts, but with advanced functionality to support expert users

More information about requirements is available online (Australian National Data Service, 2016).

The top requirements were matched against a set of commercial and open-source software products. Software evaluation was undertaken through desktop research, contact with vendors and development communities, and end-user and technical testing. Among the contenders, the open-source product VocBench (VocBench, 2015) and the commercial product PoolParty (PoolParty, 2015) stood out as best supporting the needs of the ANDS partners.
Based on a series of technical and user evaluations, PoolParty was selected and a contract executed between Semantic Web Company and ANDS for a remotely-hosted instance of the PoolParty Enterprise Server. This licence enables users of the ANDS service to make full use of PoolParty for vocabulary creation and management, including support for a simple editorial workflow for suggesting and approving concepts.

Project partners contributed sample controlled vocabularies to test workflow and provided feedback on the PoolParty software. Work continued on developing service and user support materials and plans for engagement, outreach and capability development and the service went live during September 2015. As of 18 August 2015, there were 22 vocabularies within the portal (either described, or described and hosted).

**Vocabulary publication, browsing, and linked data**

The Research Data Australia web site incorporates, and itself relies on, a vocabulary publication service that makes available a small number of core controlled vocabularies; in particular, two of the Australian Bureau of Statistics's Standard Research Classification (ANZSRC) vocabularies (Australian Bureau of Statistics 2008) and ANDS's own Registry Interchange Format – Collections and Services (Australian National Data Service 2015). These vocabularies are made available through an instance of the SISSVoc Vocabulary Service (Solid Earth and Environment Grid, 2015). SISSVoc is a publication service specifically for vocabularies in RDF, and works best for vocabularies that specify concepts using SKOS (W3C, 2012).

SISSVoc makes a vocabulary available in a number of ways. Firstly, the vocabulary can be browsed in a web browser. If the vocabulary's concepts have been arranged in a hierarchy (i.e., by specifying narrower and broader terms), this hierarchy can be traversed by following links in the browser. There are also some basic search functions provided. Secondly, the vocabulary is accessible using the Linked Data API (GitHub, 2015a), which supports machine-to-machine access via a flexible query mechanism that generates output in a variety of formats (in particular, XML and JSON) that are convenient for consumption by other services.

In theory, ANDS could offer to host and publish controlled vocabularies produced by third parties using the existing SISSVoc service. In practice, there were no processes in place to support this, configuration of SISSVoc is complex, and in-house expertise was not yet well developed.

Soon after the initiation of the VOCRAM project, ANDS was approached by IMOS to provide a hosting and publication service for some of its controlled vocabularies. This proved to be the impetus, and provided a focus needed for further development of the service.

The first step involved a review of the vocabulary publication “landscape”. Tools such as Skosmos (Skosmos, 2015) and the Linked Data Registry (GitHub, 2015b) were considered as part of this review. Skosmos has a particularly attractive user interface, and provides an API for linked data. The Linked Data Registry was considered to be not yet mature enough to use as a basis for further development.
However, in the meantime, CSIRO deployed its own instance of this software (CSIRO, 2015b).

On the basis of an existing investment in SISSVoc, and the maturity of the underlying technology, ANDS decided to continue to use it as the underlying technology of the publication service, and to extend it where necessary to support the particular needs of IMOS and other partners. In future, this may mean adding support for other publication systems.

Vocabulary discovery
As part of the redevelopment of the vocabulary portal, ANDS considered alternate discovery service offerings. Of particular interest were the Datahub catalogue (Datahub, 2015) and the curated service Linked Open Vocabularies (LOV) (LOV, 2015). The ANDS Research Data Australia website was also considered as an existing discovery service.

Research Data Australia (RDA) is ANDS’s catalogue service for research data, with over one hundred thousand entries provided by over one hundred Australian organisations, mostly in the university, government, and cultural sectors. RDA is an in-house product, and it made sense for ANDS to leverage the existing investment in RDA as the new vocabulary portal was developed.

Technical development of the ANDS Vocabulary Service
The following diagram presents a high-level view of the components of the ANDS Vocabulary Service.
The diagram shows the user-accessible and machine-accessible components at the left, and proceeds to the right through to the “under-the-hood” components. Certain details of the architecture have been omitted. In particular, Apache HTTP Server and Tomcat are used to implement and host most of the components.

The services are provided through four distinct user interfaces:
- the editor (PoolParty),
- the editor administration tool,
- the discovery portal, and
- advanced browsing (SISSVoc).

In addition, there are three types of machine-to-machine interfaces provided:
- the Linked Data API (SISSVoc),
- the vocabulary widget, and
- SPARQL endpoints.

Each of these is discussed below.

**Editor**
Vocabulary editing functions are provided by PoolParty. The ANDS instance of PoolParty is “branded” with an ANDS-specific hostname, but otherwise is “off-the-shelf”. The ANDS Services Team is responsible for creating and administering user accounts, and addressing other support requests.

**Editor Administration Tool**
Users of the PoolParty editor may wish to make fine adjustments to the RDF that the editor produces, particularly where vocabulary data has been created with another tool and then imported into PoolParty.

For these cases, ANDS developed an ‘Editor Administration Tool’ as a support tool for users of the PoolParty editor. It is implemented in Java. The Tool is intended for use by ANDS support staff, but also by advanced end users who have a sound knowledge of SKOS.

A user logs in to the Editor Administration Tool via personalised PoolParty username and password, and is then presented with a list of PoolParty vocabularies, and a list of queries and updates that can be performed. The user can select any combination of projects and queries/updates, and click the “Submit” button to perform the set of queries/updates on the set of selected vocabularies.

**Discovery portal**
As mentioned previously, the portal that provides discovery services is based on the code underlying the Research Data Australia web site. It is implemented using PHP, with the CodeIgniter framework providing the core infrastructure (CodeIgniter, 2015). Users can log in using one of a number of authentication servers (typically either AAF or a social login), and use a simple “content management system” to define vocabulary metadata and, if available, to upload the vocabulary data itself, either by transfer from PoolParty or direct file upload. The vocabulary metadata is indexed.
using Apache Solr (Lucene, 2015). If a vocabulary is provided as SKOS, either by being imported from PoolParty, or by being directly uploaded from a file, concept labels are extracted and stored in Solr, so that they can be used as search terms. In addition, the portal provides a basic tree-like browse facility for SKOS vocabularies.

**Browsing and linked data**
For more advanced browsing and linked data access to SKOS vocabularies, ANDS has deployed an instance of SISSVoc, which is customised to support a number of specific needs of ANDS partners. SISSVoc provides both HTML interfaces for access by human users with a web browser, and XML and JSON interfaces for consumption by other services, including the ANDS Vocabulary widget.

**Vocabulary Widget**
Some users may wish to make use of a published vocabulary by incorporating it into their own web sites; for example, as a pick list. To enable this, ANDS provides a “Vocabulary Widget” built on top of the SISSVoc linked data endpoints. When included in a web page, the Widget displays a vocabulary as a browsable tree, and offers a basic search facility. Because of its reliance on SISSVoc, only SKOS vocabularies work well when displayed through the Widget.

**SPARQL endpoints**
For advanced users of RDF, ANDS provides direct access to the underlying SPARQL endpoints of published controlled vocabularies. A link to a vocabulary's endpoint is displayed as part of the vocabulary's entry in the portal.

**Database**
The portal stores vocabulary metadata in a MySQL database. The database is also an important part of the portal's interface with the Vocabulary Toolkit, and it supports the implementation of special SPARQL endpoints for the versions of published controlled vocabularies that are marked as the “current” versions.

**Vocabulary Toolkit and the Repository**
The basis of the repository is the Sesame Server and its Application Programming Interface (Sesame, 2015). Portal users can upload vocabulary data in formats such as CSV and PDF, but RDF formats (such as RDF/XML and Turtle) are treated specially. As mentioned above, RDF data has concepts extracted for indexing. In addition, users can select an option for their RDF data to be imported into Sesame, and, if desired, that SISSVoc endpoints be created. If the data is imported into Sesame, the SPARQL endpoint is enabled.

This functionality is provided to the portal by a Vocabulary Toolkit, implemented in Java and using a number of associated Java technologies. The Toolkit uses the Sesame API to create and manage controlled vocabularies stored in the repository's Sesame Server, as well as for transforming vocabulary data as needed.
Using the services, and obtaining the source code

ANDS also provides the source code for the various components as open source under the Apache License from the ANDS GitHub repositories (GitHub, 2015c). The software is contained in three repositories:

- ANDS-Registry-Core is the portal
- ANDS-Vocabs-Toolkit is the portal-repository Toolkit interface
- ANDS-Vocabs-Editor-Admin is the Editor Administration Tool for working with controlled vocabularies stored in PoolParty

Capability development of controlled vocabulary services

Understanding the landscape
Amongst other aims, the ANDS Vocabulary Service exists to support members of the Australian research community in their use of controlled vocabularies. ANDS will track individuals and groups that are active in the creation, management, publication, and use of vocabularies. In addition, the ANDS service will be informed by agencies that are responsible for vocabularies, including those agencies that manage well-established vocabularies that they wish to continue editing and disseminating, and those agencies that publish vocabularies and/or may use vocabularies originating elsewhere. At a broad level, ANDS will monitor vocabulary-related development at research-intensive institutions and tertiary institutions, acknowledging variations in the needs and capabilities of partners involved in the creation, management, publication, and use of vocabularies.

ANDS engages with Australian research institutions on a broad range of research data management issues to promote the value of structured data collections that are managed, connected, findable, and reusable. This includes engagements with NCRIS data intensive and e-research facilities, Commonwealth science agencies and data aggregating departments and Australian research universities. ANDS vocabulary-related engagements will be part of the broader data-related engagements and services that constitute its relationship with the sector.

Because of its history of wide-ranging engagement with the Australian research community, ANDS has opportunities to facilitate the identification and access to disciplinary and multi-disciplinary controlled vocabularies across its current and future partners, and to play a role in bringing together institutions with common vocabulary-related needs and goals in order to allow those institutions to support and learn from one another. For example, agencies such as Geoscience Australia (GA) and the Australian Institute of Health and Welfare (AIHW) are major providers of vocabularies in their designated fields, and ANDS may be suited to promoting the use of GA and AIHW vocabularies amongst tertiary institutions and research-intensive facilities. (Australian Institute of Health and Welfare 2015) (Geoscience Australia, 2015a)
ANDS maintains regular communication with the libraries, e-research offices, information technology departments and similar groups within Australian research organisations and ANDS will leverage those existing networks to understand or discover vocabulary-related capabilities and needs in the institutions with which it engages.

**Support and promotion**

ANDS will facilitate new and continued uptake of the ANDS Vocabulary Service. This includes support at three levels: awareness raising, working-level, and expert-level support. Awareness-raising support includes providing broad information on how controlled vocabularies can facilitate connections between and reuse of research data, and how this can be enabled by the ANDS Vocabulary Service. Working-level support includes information about how a specific partner may make general use of one or many of the tools in the Vocabulary Service to support their specific needs. Expert-level support includes technical information about how a specific partner may make use of advanced tools; for example, how to query a vocabulary via its Vocabulary Service endpoint or how to incorporate the Vocabulary Widget within a specific data capture or description tool.

The ANDS Vocabulary Service will promote past and present ‘success stories’ about uptake of the service and its benefit to other ANDS partners, the facilitation and promotion of appropriate communities of practice to lend additional support, and ‘pollination’ between partners in the form of bringing together partners with similar interests and needs. In addition, ANDS will identify opportunities for partner Vocabulary Service tool and content use, and will broker or provide guidance on subject-specific controlled vocabularies. Finally, liaisons will provide partners with opportunities to provide feedback on their engagement and on their Vocabulary Service tool and content use. This feedback will be collected and examined and will support user-driven future development of the Vocabulary Service.

Recent engagement activities include the following:

- Health librarians at Monash University wish to access a particular thesaurus, which is not available in machine-readable format. There may be a role for ANDS to work with content users and providers to improve access.
- The med.data project (Med.data.edu.au, 2015) aims to describe and make available Australian medical research institute-generated research data, which is not currently available. The team building med.data plans to make use of ANZSRC, MESH, AIHW subjects and other controlled vocabularies for describing and searching for data in the system. The team has expressed a desire to make use of the ANDS Vocabulary Service in support of this.
- TERN has proposed that ANDS make the NASA GCMD science keywords accessible within the ANDS service to facilitate and promote these standard terms among the TERN facilities and use them as filtering in the TERN portal (Global Change Master Directory, 2015).
- Geoscience Australia has a vocabulary publishing service, and discussions are underway with ANDS for a pilot activity to explore the relationship between these services at a policy and technical level.
• ANDS/National Computational Infrastructure (NCI) discussions are underway to explore use of the ANDS Vocabulary Widget within NCI data capture & description tools (NCI, 2015).

Consultancy
Once an engagement with a partner begins, ANDS will provide customised support based on the needs, capabilities, and desires of the partner. In order to provide customised support, ANDS has developed workflows to enable it to understand fully the vocabulary-related needs, capabilities, and desires of the partner, and to capture information about account, user-role workflow management, and other administrative needs. ANDS will provide access to both general service documentation (including information about project and account naming conventions) and content-related support materials. Content-related support materials include information on:

• How to provide controlled vocabulary content and description in the Vocabulary Service
• How to convert that content into a machine-readable format
• How to use the Service to discover and make use of vocabularies already in the system
• Technical resources for how to create, manage, version, query vocabularies
• Technical resources for how to perform machine-to-machine tasks using those vocabularies

Conclusion
Examples discussed in this paper illustrate how controlled vocabularies enable data re-use, facilitate international collaboration, and drive complex information services for researchers. Controlled vocabularies can contribute to better research as well as better discovery and re-use of research materials.

In collaboration with a number of research institutions, the Australian National Data Service has developed a service to promote and facilitate the use of globally and nationally standard terminology. This service, developed as part of the National Collaborative Research Infrastructure, is available for use by Australian research organisations, including libraries, universities, research institutes, collecting organisations and government agencies. The service has the potential to be an important part of the systemic information infrastructure for research, with potential applications more broadly for business, public policy, industry and education. Common standardised concepts, easily shared within and between sectors, represent an enabling infrastructure layer for an innovative society in the information age. In presenting the ANDS Vocabulary Service, ANDS seeks to increase its engagement with and support for the library community, and to foster discussion about the community’s use of controlled vocabularies.
**Glossary**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAF</td>
<td>Australian Access Federation Providers of the Rapid Connect authentication service.</td>
<td><a href="https://rapid.aaf.edu.au/">https://rapid.aaf.edu.au/</a></td>
</tr>
<tr>
<td>DBpedia</td>
<td>Crowd-sourced community effort to extract structured information from Wikipedia.</td>
<td><a href="http://wiki.dbpedia.org/">http://wiki.dbpedia.org/</a></td>
</tr>
<tr>
<td>GEMET</td>
<td>General Multilingual Environmental Thesaurus</td>
<td><a href="https://marinemetadata.org/references/gemet">https://marinemetadata.org/references/gemet</a></td>
</tr>
<tr>
<td>LCSH</td>
<td>Library of Congress Subject Headings</td>
<td><a href="http://id.loc.gov/authorities/subjects.html">http://id.loc.gov/authorities/subjects.html</a></td>
</tr>
<tr>
<td>MESH</td>
<td>Medical Subject Headings</td>
<td><a href="https://www.nlm.nih.gov/mesh/">https://www.nlm.nih.gov/mesh/</a></td>
</tr>
<tr>
<td>PHP</td>
<td>PHP: Hypertext Preprocessor Programming language for implementing the “back end” of web applications.</td>
<td><a href="https://php.net/">https://php.net/</a></td>
</tr>
<tr>
<td>RDA</td>
<td>Research Data Australia</td>
<td><a href="https://researchdata.ands.org.au/">https://researchdata.ands.org.au/</a></td>
</tr>
<tr>
<td>RD-A</td>
<td>Research Data Alliance</td>
<td><a href="https://rd-alliance.org/">https://rd-alliance.org/</a></td>
</tr>
<tr>
<td>RDF</td>
<td>Resource Description Framework. A generic information modelling framework.</td>
<td><a href="http://www.w3.org/RDF/">http://www.w3.org/RDF/</a></td>
</tr>
<tr>
<td>RVA</td>
<td>Research Vocabularies Australia</td>
<td><a href="https://vocabs.ands.org.au/">https://vocabs.ands.org.au/</a></td>
</tr>
<tr>
<td>SKOS</td>
<td>Simple Knowledge Organization System</td>
<td><a href="http://www.w3.org/2004/02/skos/">http://www.w3.org/2004/02/skos/</a></td>
</tr>
<tr>
<td>Solr</td>
<td>Software library providing the search capabilities to RVA.</td>
<td><a href="http://lucene.apache.org/solr/">http://lucene.apache.org/solr/</a></td>
</tr>
<tr>
<td>SPARQL</td>
<td>SPARQL Protocol and RDF Query Language</td>
<td><a href="http://www.w3.org/TR/sparql11-overview/">http://www.w3.org/TR/sparql11-overview/</a></td>
</tr>
<tr>
<td>URI</td>
<td>Uniform Resource Identifier</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td>XML</td>
<td>Extensible Markup Language. As RDF/XML, a popular encoding of RDF data.</td>
<td></td>
</tr>
<tr>
<td></td>
<td><a href="http://www.w3.org/TR/rdf-syntax-grammar/">http://www.w3.org/TR/rdf-syntax-grammar/</a></td>
<td></td>
</tr>
</tbody>
</table>
References


Australian National Data Service 2016, Vocabulary Create and Manage project outputs. For public access, viewed 10 January 2016 [https://drive.google.com/drive/u/0/folders/0B57_4U1eo8d3aVNYUY1SG1pS00](https://drive.google.com/drive/u/0/folders/0B57_4U1eo8d3aVNYUY1SG1pS00)


GitHub, 2015b, UKGovLD/ukl-registry-poc, viewed 17 August 2015
https://github.com/UKGovLD/ukl-registry-poc/

GitHub, 2015c, Australian National Data Service, viewed 17 Aug. 2015
https://github.com/au-research/

Global Change Master Directory, 2015, Learn About GCMD, viewed 18 August 2015
http://gcmd.nasa.gov/learn/index.html

Harpring, P. 2010, Introduction to Controlled Vocabularies: Terminologies for Art, Architecture, and Other Cultural Works (Getty Publications), Getty.edu, viewed 21 August 2015
http://www.getty.edu/publications/virtuallibrary/160606018X.html


https://docs.google.com/document/d/1JvjrWMTpMiH7WnuieNqcT0zpJAXUPo6x4uMBj1pEx0Y/edit#heading=h.bz78ahr21sjd


Linked Data for Libraries 2015a, Linked Data for Libraries, viewed 17 August 2015
https://www.ld4l.org/

https://wiki.duraspace.org/pages/viewpage.action?pageId=41354028

Linked Open Vocabularies, 2015, Linked Open Vocabularies, viewed 17 August 2015
http://lov.okfn.org/dataset/lov/


Med.data.edu.au, 2015, med.data.edu.au, viewed 10 January 2016
http://med.data.edu.au/

Metadata Online Registry, 2015a, National Data Dictionaries, viewed 17 August 2015, http://meteor.aihw.gov.au/content/index.phtml/itemId/274816


Research Data Alliance, 2015, Vocabulary Services Interest Group, viewed 17 August 2015 https://rd-alliance.org/node/47826


Skosmos, 2015, Skosmos, viewed 17 August 2015 http://skosmos.org/


Endnote

¹ A metamorphic facies is a set of metamorphic mineral assemblages that were formed under similar pressures and temperatures. (Wikipedia, 2015)