Metadata: PHRN Guidelines

Responsible Officer:	PHRN Chief Executive
Contact Officer:	Manager Policy & Client Services Dr Felicity Flack (08) 6389 7302
Superseded Documents:	Nil
Review:	Initial Draft
Associated PHRN Documents	PHRN Metadata Framework PHRN Project Participants Metadata Consultation: Final Report PHRN Researcher Consultations

1. Executive Summary

1.1 Focus/Policy Statement

The purpose of the Metadata: Population Health Research Network (PHRN) Guidelines is to provide guidance to PHRN project participants ("project participants") regarding standards and best practices that may be applicable to the management of, access to, and interoperability of, metadata throughout the network.

1.2 Audience

The content of the Metadata: PHRN Guidelines is primarily intended for project participants, in particular, those staff with metadata responsibilities.

1.3 Implementation

To provide data users with an interest or experience in research using linked data the ability to access metadata about population-related datasets within Australia, the PHRN Management Council has recognised:

- The need for a document which provides guidance to project participants regarding the management of, access to, and interoperability of, metadata available within the network;

- The need for flexibility given the organisational environments, resourcing issues and different business requirements of individual project participants and partners collaborating with the PHRN;

- The need to take into account the existing information management and policies, practices, processes and infrastructure of individual project participants that will be involved in the management of, access to, and interoperability of, metadata.

1.4 Review

The Metadata: PHRN Guidelines will be reviewed annually by the PHRN Management Council to take into account elements such as developments in metadata principles and standards, legislative and regulatory requirements.

2. Metadata: An introduction

2.1 Definition

Metadata can be described as data about data. In the context of research using data linkage metadata is structured information that describes, explains, and locates core datasets which are routinely linked by data linkage units in Australia.

2.2 Uses for metadata

The primary uses of metadata include:

- Resource discovery in terms of identifying and understanding what data is available and its relevance to the researcher;

- Administrative control;

- Security;

- Information management;

- Rights management;
- Preservation.

Project participants may wish to consider all of the uses for metadata as part of their organisation's metadata life cycle management.

2.3 Types of metadata

2.3.1 Descriptive metadata

Descriptive metadata describes core datasets which are routinely linked by data linkage units in Australia for purposes such as discovery and identification. Descriptive metadata may answer the following questions:

- Where did the data come from?
- Why was it collected?
- Who collected it, when and where?
- What are the content, quality and context of the dataset?
- What instruments/technologies were used to collect the data, and how were they set up?
- What has been done to the data since it was collected?
- What reference documents, such as data dictionaries and quality assurance statements, are relevant to this dataset?

2.3.2 Structural metadata

Structural metadata describes how a person or a computer should read the data of core datasets which are routinely linked by data linkage units in Australia. Structural metadata may answer the following questions:

- How is the data set up?
- What formats and versions of formats are used?
- How is the database configured?
- How does it relate to other data?

2.3.3 Administrative metadata

Administrative metadata provides information to help manage core datasets which are routinely linked by data linkage units in Australia, such as when and how it was created, file type and other technical information, and who can access it. There are two subsets of administrative data: rights management metadata and preservation metadata.

- Rights and access metadata provides information about access and usage rules. Rights and access metadata may answer the following questions:
  - Who is allowed to view, edit or otherwise modify the data or the metadata and under what conditions?
  - Who has some kind of authority over the data?
  - Who has the authority to change the rules?
  - Are there costs associated with access?

- Who has accessed the data, and what have they in turn done with it?
- Under what licence is the data being made available?

- Preservation metadata helps build a sense of trust in the data and allows for the data to be used long into the future. Preservation metadata may answer the following questions:
  - Is the data authentic, authoritative and original?
  - Has there been any restructuring e.g. due to software and file-formats changing?
  - What software has been used to access the data in the past?

2.3.4 Project-specific scientific metadata

Project-specific scientific metadata is information that is very specific to research using linked data and is needed to use and interpret the data extracts. This information is provided to the research team for each approved project, generally by the data linkage unit or by the data custodians directly. Project-specific scientific metadata can contain descriptive, structural and administrative metadata information and may answer the following questions:

- What metadata is important in interpreting the data extracts?
- How should a data user interpret the results of the linkage?
- What is the quality of the linkage?

2.4 Benefits of using metadata

The use of metadata to describe core datasets which are routinely linked by data linkage units in Australia provides many benefits:

- Facilitate resource discovery and access:
  - Provides data users with the knowledge about how to access datasets;
  - Informs data users about datasets including name, location, approvals required, custodian contact details, quality of the data;
  - Provide links back to source metadata for information about variables, mode and method of collection, changes over time;
  - Flexibility in searching and browsing to support usage;
  - Aids scholarship, research and collaboration;
  - Assists with dissemination of information.

- Metadata protects investment in data:
  - Mitigates effect of staff turnover and loss of organisational memory when individuals depart;
  - Makes implicit knowledge explicit;
  - Allows reuse to increase return on investment;
  - Provides documentation of data sources and quality;
  - Supports management of datasets including asset and information management.

- Metadata helps data users understand data:

- Provides standardisation and consistency in terminology and attribution;
- Focuses on key elements of data;
- Helps data users determine the data's fitness for use;
- Facilitates data transfer and interpretation by new data users.

- Metadata improves robustness of research:
  - Informs data users' research questions and study design/protocol;
  - Potential to improve Australian data users' competitiveness when applying for funding.

- Metadata can limit liability:
  - Helps prevent data from being inappropriately used or provides protection if data is inappropriately used;
  - Assists in the minimisation of disclosure risks in the context of use of sensitive personal information.

- Metadata is evidence of data custodianship;

- Metadata reduces the time data linkage units spend verbally communicating information to data users about the availability of linked or linkable data:
  - Data users do not have to keep asking data linkage units questions;
  - Data users will contact the data linkage unit's client services team with more focused and sophisticated queries;
  - Supports client services of data linkage units.

- Metadata cuts overall costs:
  - Allows automation of tools which ease overall burden and cost of data population and maintenance.

- Metadata publishing builds a foundation for the future:
  - Implementing good metadata practices today will prepare the PHRN for tomorrow.

It is important to note that the realisation of the above benefits can only occur if the metadata involved is of high quality, which requires that the metadata 1) adequately describes the data; 2) distinguishes between different data collections; 3) has full population of each metadata instance; and 4) has been created consistently (Queensland Government, Metadata Management Guideline, 2010).

2.5 Metadata and data linkage organisations

In Australia, metadata is created, maintained and published by data custodians, third parties and data linkage units. The data linkage unit acts as a conduit between the data user and the data custodian. It is the data linkage units' role to provide client services to data users wishing to access and use core datasets which are routinely linked by data linkage units. Data linkage units provide metadata to data users through verbal or written communication, including via the data linkage unit's website. This method of providing metadata is also commonly practiced internationally.

3. PHRN Guidelines approach

3.1 Why does the PHRN need metadata guidelines?

There are a number of reasons why the PHRN needs metadata guidelines:

- There is no national approach to metadata across the network and current approaches across jurisdictions and sectors vary widely;

- The network will provide data users with the opportunity to link data from across jurisdictions and sectors. Therefore consistent and interoperable metadata would facilitate the searching of, exchange and sharing of data from these different sources during the project development phase;Data users prefer unmediated access to online metadata, expect seamless information delivery, expect to interact more and more with information and rely on keyword searching (State Library of New South Wales 2012: 2);

- Data users rely heavily on the internet to obtain metadata and may be less tolerant of poor design, content and navigation of metadata reference documents than other users (Smith 2010);

- PHRN guidelines will synthesise information regarding standards, best practices and tools to manaqe metadata in the context of research using linked data. They will aid data linkage unit staff to assist custodians in the development, management of and access to their metadata wherever possible and appropriate.

3.2 Focus/Policy statement

The purpose of the Metadata: PHRN Guidelines is to provide guidance to project participants regarding standards and best practices that may be applicable to the management of, access to, and interoperability of, metadata throughout the network.

3.3 Consultation

Data users (100) and project participant representatives (19) were consulted to inform the development of the Metadata: PHRN Guidelines. The Australian Bureau of Statistics was also provided with an opportunity to provide input to this document.

3.4 Stakeholder conflicts

A defining feature of the network is the multiple stakeholders. These include data users and project participants and often their requirements conflict. Data users are looking for the most information possible, available quickly, easily and at minimal cost, and with limited obstacles. Project participants are trying to increase and develop further documentation and reduce workload for client services staff.

3.5 Audience

The content of the Metadata: PHRN Guidelines is primarily intended for project participants, in particular, those staff with metadata responsibilities.

The scope does not include control over the management, storage and publication of metadata by data custodians. It is expected that they will be governed by their own organisational policies and procedures.

3.6   Implementation

To provide data users with an interest or experience in research using linked data the ability to access metadata about population-related datasets within Australia, the PHRN Management Council has recognised:

- The need for a document which provides guidance to project participants regarding the management of, access to, and interoperability of, metadata available within the network;

- The need for flexibility given the organisational environments, resourcing issues and different business requirements of individual project participants and partners collaborating with the PHRN;

- The need to take into account the existing information management and policies, practices, processes and infrastructure of individual project participants that will be involved in the management of, access to, and interoperability of, metadata.

3.7   Review

The Metadata: PHRN Guidelines will be reviewed annually by the PHRN Management Council to take into account elements such as developments in metadata principles and standards, legislative and regulatory requirements.

The review will consider feedback from each project participant and interested parties. Project participants may wish to consider reviewing the Metadata: PHRN Guidelines annually to determine their metadata life cycle management plan's fitness for purpose.

4.   Metadata management

4.1   Metadata planning

Project participants may wish to consider establishing effective planning processes to ensure the implementation and review of metadata remains consistent with the project participants and data users' requirements. Metadata planning may involve the following activities:

- Identify business requirements for metadata;

- Conduct stakeholder analysis

- Identify metadata management roles and responsibilities;

- Ensure metadata custodianship is included in relevant policies;

- Identify all sources and suppliers of metadata and ensure that metadata is collectively sufficient to meet all identified business requirements;

- Identify the appropriate standards based metadata schemas that will be required to meet business requirements including support, legislative, regulatory and operational requirements;

- Identify opportunities for automatic metadata authorship to ensure that metadata is created consistently;

- Identify the appropriate mechanisms (i.e. controlled vocabularies, syntax encoding systems or taxonomies) that can be used to ensure that metadata is created consistently;

- Identify strategies and tools which may be used to support more efficient processing and recording of metadata such as automating repetitive tasks i.e. update holdings and use of templates and guidelines;

- Identify training needs for metadata management and use;

- Identify risks and barriers to metadata use and develop strategies to mitigate risks;

- Identify appropriate metadata quality assurance processes;

- Develop or leverage rules/governance/guidelines/data entry standards to support the application of metadata;

- Identify technical and security requirements for metadata handling;

- Map metadata to a common/standard metadata profile to ensure alignment and compliance with the project participant's metadata policies;

- Where possible, set in place mechanisms that will record metadata usage and purpose of use by data users and others.

4.2 Metadata creation and acquisition

To facilitate management of, access to and interoperability of PHRN information and services, project participants may wish to consider establishing consistent description and classification of information through the implementation of metadata. Metadata creation or acquisition may involve the following activities:

- Create metadata for core datasets which are routinely linked where it does not currently exist;

- Create metadata using metadata schemas that are interoperable with agreed standards;

- Ensure that project participants' metadata meets the minimum requirements of the agreed standards and that required for business;

- Use standards based metadata schemas when creating metadata;

- Use metadata schemas that meet business requirements and reflect the context of the information being described;

- Use mechanisms such as controlled vocabulary, taxonomy, thesaurus, syntax encoding standards and machine-generation to ensure that metadata is created consistently;

- Automate the process of metadata authorship where possible;

- Ensure that where an extension of elements for metadata schemas is required to meet business requirements, that this extension is implemented according to the metadata extension methodology defined in the agreed standard;

- Create a Metadata Application Profile which outlines the format and content for describing core datasets which are routinely linked. A Profile generally complies with an agreed metadata standard and consists of a template which assists in capturing metadata.

4.3 Metadata organisation and storage

To facilitate management of, access to and interoperability of PHRN information and services, project participants may wish to consider establishing effective and reliable organisation and storage of metadata.

4.4 Metadata access

To facilitate management of, access to and interoperability of PHRN information and services, project participants may wish to consider the establishment of accessible metadata where appropriate. Metadata access may involve the following activities:

- Ensure that metadata describing core datasets which are routinely linked is accessible to all data users;

- Ensure that metadata is accessible to the maximum extent possible without compromising policies that place restrictions on access.

4.5 Metadata harvesting

Metadata harvesting involves the aggregation of metadata records. Project participants may wish to consider metadata harvesting given the following benefits may be derived from metadata harvesting:

- Increase the visibility of your core datasets which are routinely linked and make relationships between datasets discoverable;

- Provide social benefit to the Australian research community by making core datasets which are routinely linked discoverable from multiple locations;

- Provide multiple access points to core datasets which are routinely linked;

- Maximise return on investment in creating the resources and associated metadata;

- Contribute to the national information infrastructure;

- Expose core datasets which are routinely linked and available for linked data research to the international data linkage community;

- Support the open access idea of knowledge sharing;

- Encourage and support collaboration and interoperability.

4.6 Metadata maintenance

To facilitate management of, access to and interoperability of PHRN information and services, project participants may wish to consider:

- Optimising and rationalising metadata which may include monitoring new sources of metadata for existing or new collections that may be published;

- Continuing to create and update agreed metadata elements which will include establishing processes to identify and incorporate updates undertaken by data custodians;

- Utilisation of automated processes for maintenance of metadata where possible.

4.7 Metadata personnel

4.7.1 Roles and responsibilities

Operational responsibility for managing metadata of core datasets which are routinely linked within each jurisdiction ultimately resides with the project participant in conjunction with the relevant data custodians.

Project participants may wish to consider that personnel responsible for managing metadata may benefit from an understanding of:

- Metadata principles and practices;

- Information and records management;

- Security management principles and practices;

- Risk management;

- Legislation governing the use and disclosure of information.

4.7.2 Metadata training

Project participants may wish to consider providing personnel with metadata roles and responsibilities with education and awareness training commensurate to the level of responsibility assigned.

The following organisations may provide metadata training and education courses:

- Council of Australasian Archives and Records Authority;

- Records and Information Management Professionals Australasia;

- Australian and New Zealand Society of Indexers Inc.

4.8 Metadata assessment and review

To facilitate management of, access to, and interoperability of PHRN held information and services, project participants may wish to consider conducting periodic reviews of their metadata. Metadata assessment and review may involve the following activities:

- Annual review of metadata operations with the goal of streamlining workflows and procedures in order to make the most efficient use of available resources;

- Monitoring the usage and use of metadata (refers to the amount of use and the purpose of use respectively);

- Reviewing the scope and nature of project participants' metadata roles and responsibilities to ensure that they are appropriate to their organisation's business requirements;

- Reviewing metadata appointments and delegations to ensure that they are appropriate, current, assigned and agreed by responsible parties;

- Assessing opportunities to extend the use of metadata;

- Reviewing the metadata quality and completeness;

- Reviewing the alignment and continuing fitness for purpose of existing metadata and metadata schemas to ensure that project participants' metadata continues to meet business requirements;

- Reviewing new versions of existing standards that have been released, or new "commonly used" standards that have emerged, and assessing whether these would offer sufficient business benefit to warrant moving to them;
- Monitoring and recording the cost of maintaining metadata.

5. Metadata standards

Metadata requires agreed-upon standards. For example, creators and collectors of metadata need to agree on the concepts, metadata elements, language, spelling, data formats, place names etc to use in documenting the metadata. Agreed metadata standards make it easier for data users to search and compare metadata results.

It is becoming more common as a leading practice to separate standards related to concepts and common terminology from standards related to technical implementation (e.g. information technology formats), although the two sets of standards should (and can) work together.

As a simple illustration, the concept of "Person" may be common across many domains, including Health and Education, even though

- The two domains are interested in different "subsets" of persons:
  - eg "patients", "health professionals" compared with "students", "teachers", "parents"
- There are other concepts relevant to the different domains which are (more or less) "domain specific" (e.g. "medical procedure" compare with "school bus").

Certain characteristics of persons (e.g. date of birth, sex, given name, family name) can also be defined on a common basis across domains – which will aid in comparing and integrating data – even if the information is represented using different, domain specific (or agency specific), technical standards for data and metadata.

This approach would, however, distinguish the term "Person" as denoting a different concept in the case of legislation where corporations, in addition to human beings, were in scope of the term. Where appropriate, simple semantic standards could further show that the concept of "Person" related to Education was a narrower concept (subset) compared with the concept of "Person:" in Law.

The Metadata Online Registry (METeOR), based on the model from ISO 11179, already documents many of these "cross domain" – and "domain specific" – concepts together with standards for categorizing and representing data related to these concepts.

When it comes to common concepts relevant to the definition, management and use of research datasets generically (e.g. "units", "populations", "variables", "classifications", "records", record relationships/record linkage) the internationally agreed Generic Statistical Information Model (GSIM) defines concepts, common terms, relationships between concepts, standard characteristics and so on. The Structures Group and the Concepts Group within GSIM are the groups which are most likely to be relevant for project participants in PHRN.

A further potential advantage from using GSIM as a conceptual standard for describing data is that there are then technical standards, and tools, that implement these concepts (eg in terms of record linkages).

Metadata standards are often discipline specific. The most common metadata standard for cataloguing online resources used in health and medicine is the Dublin

Core which is a metadata element set intended to facilitate the discovery of electronic resources. For example in Australia, health web sites now using the Dublin Core include the Australian Department of Health and Ageing (http://www.health.gov.au/), the Better Health Channel (http://www.betterhealth.vic.gov.au/), the National Health and Medical Research Council (http://www.nhmrc.health.gov.au/) and Health Insite (http://www.healthinsite.gov.au/) (Darmoni 1999).

AGLS is an Australian Standard (AS 5044-2010) Application Profile on Dublin Core. AGLS was updated in 2010, including greater support for Linked Data/Semantic Web, recognising that the internet is no longer just a medium for publishing human-readable documents.

Table 1 refers to some commonly used health and medicine metadata standards.

Table 1: Health and medicine metadata standards

Metadata model	Description	Metadata standard	Other tools
Medical core metadata (MCM)	The MCM has developed a set of standard metadata schema for health and medicine internet resources.	Dublin Core (IETF RFC 5013, ANSI/NISO Standard Z39.85-2007, and ISO Standard 15836:2009)	The (US) National Library of Medicine's MeSH thesaurus to describe subject content.
Catalog and Index for French Speaking Health Sites (CISMeF)	CISMeF is devoted to indexing and describing French language Internet resources in the areas of medicine and health.	Dublin Core (IETF RFC 5013, ANSI/NISO Standard Z39.85-2007, and ISO Standard 15836:2009)	The (US) National Library of Medicine's MeSH thesaurus to describe keywords scheme.
National Environmental Public Health Tracking Network (EPHTN)	The EPHTN (US) is a system of integrated health, exposure, and hazard information and data from a variety of national, state, and city sources.	Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (FGDC-STD-001-1998)	
Metadata Online Registry (METeOR)	Australia's repository for national metadata standards for health, housing and community services statistics and information.	ISO 11179	A range of nationally endorsed data definitions and tools for creating new definitions based on existing already-endorsed components.
National	A comprehensive	Over 20 standards	

Metadata model	Description	Metadata standard	Other tools
Health Information Infrastructure (NHII)	knowledge-based network of interoperable systems of clinical, public health, and personal health information developed by the US Government.	incorporated including HL7, NDF-RT, SNOMED-CT, LOINC, DICOM, IEEE 1073, NCPDP SCRIPT.	

It is generally preferable to use or adapt existing standards because an existing standard can offer:

- Interoperability: Helps ensure metadata records associated with one resource can be accessed, accurately interpreted and subsequently used by a system or integrated with metadata records associated with other resources;

- Cost saving: The schema, usage guidelines and vocabularies are already defined;

- Access to help and advice: An existing standard will have an established community of users so it will be easy for a project participant to access help and advice about how to use the standard;

- Usability: If users are familiar with an existing standard this may increase data users' ability to search metadata;

- Resource discovery: Metadata could be more easily searched by data users;

- Sustainability: Use of existing standards may enable metadata to be maintained by project participants in the long-term.

5.1 Metadata schema

There are hundreds of metadata schemas. A metadata schema describes how the metadata is set up and provides an overall structure to the metadata. Each metadata schema serves a purpose. The use of standard metadata schema ensures the use of a consistent set of metadata elements.

5.2 Metadata consistency

The names of metadata elements may be variable across different metadata schemas despite having the same meaning or purpose. Consistent standards, including schemas, taxonomies and vocabularies for metadata elements is central to achieving systematic and standardised metadata, as well as enabling data users to search results across metadata.

5.3 Metadata interoperability

Whilst it is acknowledged that different metadata standards may be used in practice within different host organizations of the data custodians and project participants, metadata interoperability across the PHRN is supported and encouraged to facilitate the searching of, exchange and sharing of data from different jurisdictions and sectors for research.

5.4   Metadata and obligation

Some metadata schemas stipulate which metadata elements are mandatory or optional. If metadata elements are designated as mandatory and a metadata record contains no metadata for one of the mandatory elements, then that record is considered not to be a true and correct metadata record according to that metadata schema. This can be used to assess the quality of metadata records within a metadata catalogue. The impact of having metadata records that are not true and complete is that not all relevant information will be discovered by the search process.

6.   References

ANDS. 2011a. *Metadata: ANDS Guide (Working Level)*, Australian National Data Service. Retrieved from http://www.ands.org.au/guides/metadata-working.html on 8 March 2013.

Darmoni, S.J. 1999. *CISMeF: Catalog and Index of French-language Health Internet resources. A quality-controlled subject gateway*. Retrieved from http://www.chu-rouen.fr/cismef/cismefeng.html on 1 August 2013.

NISO. 2004. *Understanding metadata*. NISO Press, Bethesda, Maryland.

PHRN Program Office. 2013. *Consultations with PHRN Project Participants about Metadata: Final Report*. Program Office, Perth, Western Australia.

PHRN Program Office 2012. *Researcher Consultations: Final Report*. Program Office, Perth, Western Australia.

State Library of New South Wales. 2012. *Metadata framework: Principles and standards for cataloguing*. State Library of New South Wales, Sydney, New South Wales.

Smith, S. 2010. *Website research*. Retrieved from http://www.websitecriteria.com/website_research/User_expectations_online.html on 20 May 2013.

The State of Queensland (Department of Public Works). 2010. Metadata management guideline. Queensland Government Enterprise Architecture, Brisbane, Queensland.

7.   Appendix A – Common Terms

In addition to the common terms described below, also refer to the PHRN Glossary.

Term	Description
Data Custodian	The organisation or agency which is responsible for the collection, use and disclosure of information in that dataset. The data custodian is responsible for contributing to the guidelines and approval processes on the use of the data, including involvement with ethics committees and input to the protocols surrounding data use.
Data Linkage Activities	Activities to be carried out by the project participants as described in the Project Plans in the PHRN funding agreements.

Term	Description
Data User	A person who uses data in the course of a linkage project. This includes investigators, analysts and others who work for a range of organisations including academic institutions and government organisations.
Element	Item/term for use in metadata description.
Information Asset	As per ISO 27005, an information asset is a physical or logical item or service that has value to project participants and therefore requires protection. This includes: • People (e.g. employees, contractors, third parties, data users); • Information (e.g. data linkage keys, demographic data, metadata, research applications, policies, procedures, guidelines, knowledge); • Hardware (e.g. servers, desktops, laptops, CDs, USB keys); • Software (e.g. encryption software, data linkage applications); • Facilities (e.g. offices, computer room, cabinets, data centre, utilities); • Intangibles (e.g. memberships, licenses, agreements).
Linked Data	Information thought to belong to the same entity held in different data sources that is integrated using a unique ID for the purposes of monitoring, management, evaluation or research
Metadata	Commonly described as data about data.
Ontology	A formal model that allows knowledge to be represented for a specific domain. An ontology describes the types of things that exist (classes), the relationships between them (properties) and the logical ways those classes and properties can be used together (axioms).
PHRN Participant	A group that includes all Project Participants and participant organisations as described in the PHRN Funding Agreement.
PHRN Project Participant ("project participant")	A party to a PHRN Participant's Agreement who is approved by the Commonwealth and is directly involved in data linkage activities.
PHRN System	Information system used by a project participant for the purpose of performing data linkage.
Schema	A metadata schema describes how the metadata is set up and provides an overall structure to the metadata.

Term	Description
Standard	Commonly used and consistently applied formats or processes, which are measurable, well documented and endorsed.
Taxonomy	A formal representation of relationships between items in a hierarchical structure.

8. Appendix B: Useful Resources

8.1 General resources

- AGLS is an Australian Standard (AS 5044-2010) Application Profile on the Dublin Core Metadata Initiative. AGLS was updated in 2010, including greater support for Linked Data/Semantic Web, recognising the internet is no longer just a medium for publishing human-readable documents.

  http://www.agls.gov.au/

- Australian National Data Service (ANDS): ANDS leads the creation of a cohesive national collection of research resources.

  http://www.ands.org.au/index.html

  - Metadata Awareness Level

    http://www.ands.org.au/guides/metadata-awareness.html

  - Metadata Guide Working Level

    http://www.ands.org.au/guides/metadata-working.html

- Data Documentation Initiative (DDI): DDI provides detailed schemas for documenting research datasets including record relationships and linkages.

  http://www.ddialliance.org/

- Dublin Core Metadata Initiative (DCMI): The DCMI supports shared innovation in metadata design and best practices across a broad range of purposes and business models.

  http://dublincore.org/

- Federal Geographic Data Committee (FGDC): The FGDC develops procedures and assists in the implementation of a distributed discovery mechanism for national digital geospatial data.

  http://www.fgdc.gov/

  ANZLIC Metadata Profile: ANZLIC — the Spatial Information Council is the peak intergovernmental organisation providing leadership in the collection, management and use of spatial information in Australia and New Zealand. The ANZLIC Metadata Profile was created to facilitate efficient access to descriptions of information resources, and in particular geographic (or spatial) data from these two countries. http://spatial.gov.au/resources/home#metadata

- Generic Statistical Information Model (GSIM): GSIM provides a common conceptual model for describing research datasets and their components (eg "units", "populations", "variables", "classifications", "records")

  http://www1.unece.org/stat/platform/display/metis/GSIM+v1.0

- International Organization for Standardization (ISO): ISO develop and publish international standards.

  http://www.iso.org/iso/home.htm

- JISC: JISC is a registered charity and work on behalf of UK higher education, further education and skills to champion the use of digital technologies.

  http://www.jisc.ac.uk/

- National Information Standards Organization (NISO): NISO is a non-profit association accredited by the American National Standards Institute that identifies, develops, maintains, and publishes technical standards to manage information.

  http://www.niso.org/home/

- UKOLN: UKOLN is a centre of expertise which advises on digital infrastructure, information policy and data management. UKOLN provide resources and services to the higher and further education sectors including Web journals and other publications, Web services and tools, innovation support, research and development, and events management.

  http://www.ukoln.ac.uk/

8.2 Metadata repositories

- Australian Institute of Health and Welfare (AIHW)'s METeOR: METeOR is Australia's repository for national metadata standards for health, housing and community services statistics and information.

  http://meteor.aihw.gov.au/content/index.phtml/itemId/181162

- Manitoba Centre for Health Policy's (MCHP). MCHP publishes a Concept Dictionary and Glossary as well as Data Repository Documentation.

  http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/resources/concept_dictionary.html

  http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departmental_units/mchp/resources/repository/documentation.html

- Population Data BC's MetaData Central: MetaData Central is a web-based documentation system that includes data dictionaries, linkage rates and metadata.

  http://www.popdata.bc.ca/researchers

- WHO Indicator and Measurement Registry (IMR): IMR provides a central source of metadata of health-related indicators used by WHO and other organizations. It promotes interoperability by drawing on standards such as SDMX (Statistical Data and Metadata eXchange), ISO 11179, DDI and Dublin core

  http://www.who.int/gho/indicator_registry/en/

- UK cancergrid: cancergrid is an initiative involving scientists at the Universities of Oxford and Cambridge, working together to reduce the cost of clinical research, and to increase its value through effective data sharing. The project has produced an ISO11179-compliant metadata registry known as the cancerGrid metadata registry (cgMDR). The registry has proved effective as a lightweight, desktop solution, targeted at the day-to-day needs of cancer researchers, data managers, and software developers. It allows management

and reuse of semantic metadata data elements during software development within an institution or collaborative group before or alongside the processes of wider review and adoption.

www.cancergrid.org

8.3 Glossaries, Ontologies and Controlled Vocabularies

- Documentation for health/medical types

  http://schema.org/docs/meddocs.html

- SNOMED CT, developed by the International Health Terminology Standards Development Organisation (IHTSDO), provides the most comprehensive, multilingual clinical healthcare terminology in the world.

  http://www.ihtsdo.org