



METADATA Plenary Session

RDA P5 San Diego

research data sharing without barriers
rd-alliance.org

Agenda

- Introduction to Metadata & Principles of Metadata Groups
- Use Case Template
- Standards Directory
- Provenance
- Plan
- Conclusion
- Discussion
- TAB Liaison:
- Secretariat Liaison:



- **Introduction to Metadata & Principles of Metadata Groups**
- Use Case Template
- Standards Directory
- Provenance
- Plan
- Conclusion
- Discussion



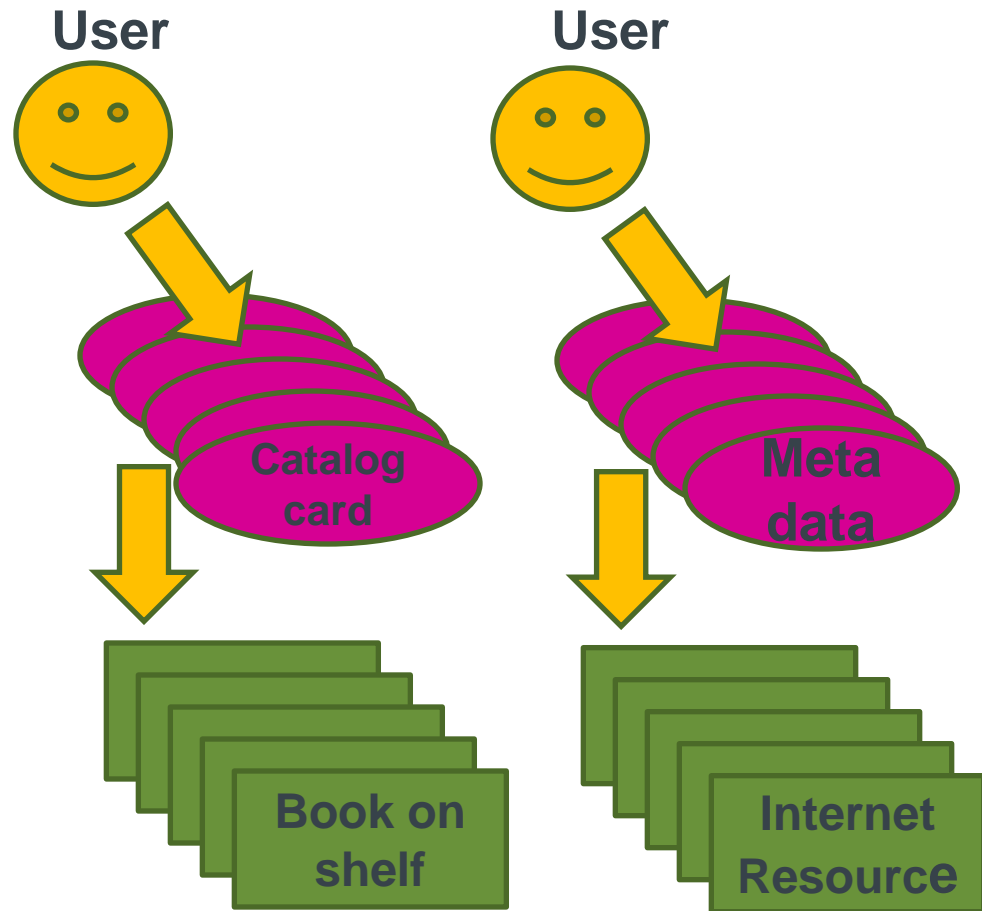
Introduction to Metadata and Metadata Principles

Keith G Jeffery

research data sharing without barriers
rd-alliance.org

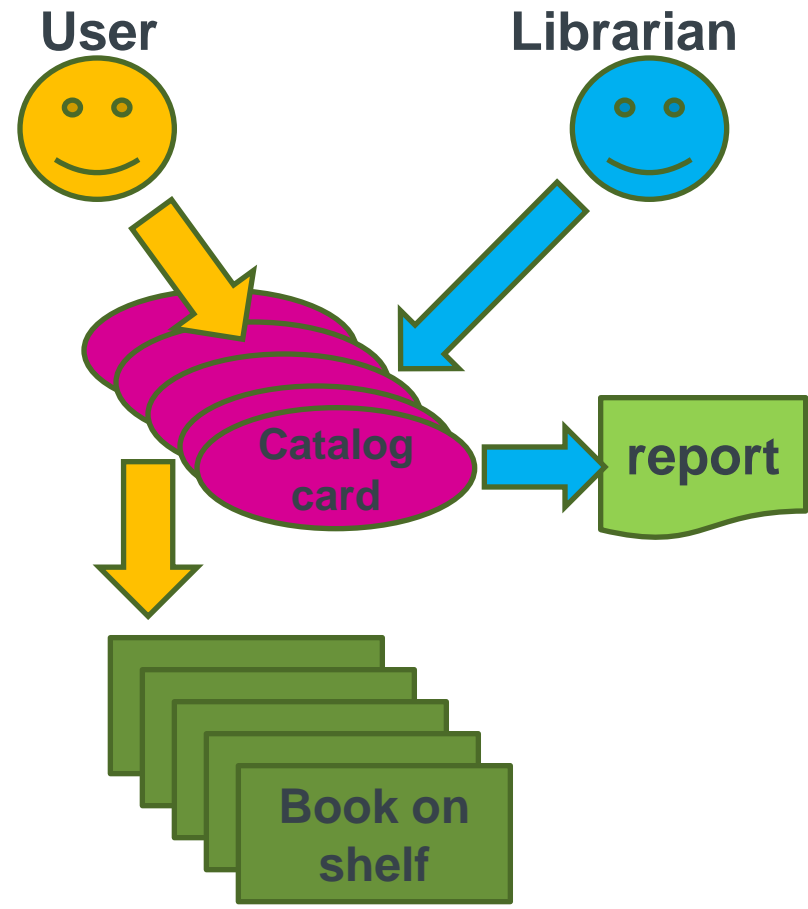
Metadata

- Data about data (DCMI definition)
 - Unhelpful!
- Analogy of user of library
- Somehow describes internet resources for the end-user



Metadata

- Consider a library
 - Catalogue cards
 - Books on shelves
- To researcher or reader the catalogue cards are **metadata**
 - Describe the book and point to where it is on the shelf
 - Descriptive and navigational metadata
- To librarian catalogue cards are **data**
 - use catalogue cards to count number of books on 'information technology
- **Take Home Message: So do not distinguish data and metadata except by how used**



Open Data: First-cut list of GENERAL elements

- Unique Identifier (for later use including citation)
- Location (URL)
- Description
- Keywords (terms)
- Temporal coordinates
- Geospatial coordinates
- Originator (organisation(s) / person(s))
- Project
- Facility / equipment
- Quality
- Availability (licence, persistence)
- Provenance
- Citations
- Related publications (white or grey)
- Related software
- Schema
- Medium / format

**This list from experience.
What have we missed?
Any other general required metadata elements?**

Open Data: Relationships not Elements ⁸

- Unique Identifier (for later use including citation)
- Location (URL)
- Description
- Keywords (terms)
- Temporal coordinates
- Geospatial coordinates
- Originator (organisation(s) / person(s))
- Project
- Facility / equipment
- Quality
- Availability (licence, persistence)
- Provenance
- Citations
- Related publications (white or grey)
- Related software
- Schema
- Medium / format

**These are NOT
properties or elements,
they are relationships**

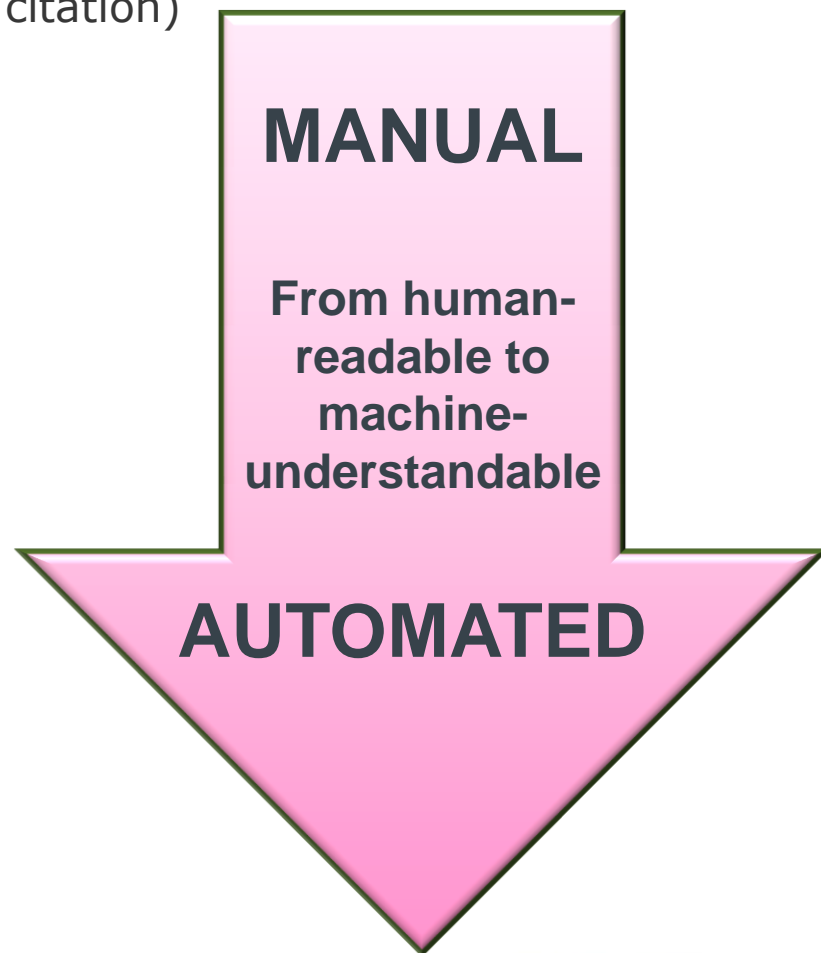
- Discovery
 - Find digital objects of interest
 - Find navigation to them
- Context
 - Is the digital object relevant?
 - Is the digital object of appropriate quality?
 - Is the digital object available (rights)?
- Detailed
 - Connecting the digital object(s) to application software
 - E.g. schema

Open Data: Purposing the elements

- Unique Identifier (for later use including citation)
- Location (URL)
- Description
- Keywords (terms)
- Temporal coordinates
- Geospatial coordinates
- Originator (organisation(s) / person(s))
- Project
- Facility / equipment
- Quality
- Availability (licence, persistence)
- Provenance
- Citations
- Related publications (white or grey)
- Related software
- Schema
- Medium / format



- Unique Identifier (for later use including citation)
- Location (URL)
- Description
- Keywords (terms)
- Temporal coordinates
- Geospatial coordinates
- Originator (organisation(s) / person(s))
- Project
- Facility / equipment
- Quality
- Availability (licence, persistence)
- Provenance
- Citations
- Related publications (white or grey)
- Related software
- Schema
- Medium / format



Metadata: one classification

- Description
- Location (navigation)
- Contextualisation
- Preservation
- Provenance
- Schema

REQUIREMENTS

- Discovery
- Context
- Detail

PROCESS

- Re-use
- Interoperation

PURPOSE

- The only difference between metadata and data is mode of use
- Metadata is not just for data, it is also for users, software services, computing resources
- Metadata is not just for description and discovery; it is also for contextualisation (relevance, quality, restrictions (rights, costs)) and for coupling users, software and computing resources to data (to provide a VRE)
- Metadata must be machine-understandable as well as human understandable for autonomicity (formalism)
- Management (meta)data is also relevant (research proposal, funding, project information, research outputs, outcomes, impact...)

- Introduction to Metadata & Principles of Metadata Groups
- **Use Case Template**
- Standards Directory
- Provenance
- Plan
- Conclusion
- Discussion



Use Case Template

Rebecca Koskela

DataONE, University of New Mexico

research data sharing without barriers
rd-alliance.org

- Template created by Data in Context IG and Metadata IG
 - Standard format rather than text
 - Captures context-related information
 - Support beyond discovery include indication of data quality
 - Collect specific information needed to align with other activities and projects
- Inaugurated at Dublin (P3) in Metadata Standards Directory WG, Data in Context IG, and Metadata IG sessions

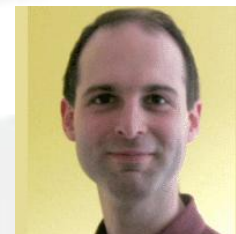
- DICIG:
 - Existing system use cases
 - Future system use cases
- MSDWG: Use cases and comments from
 - Tool developers
 - Data custodians
 - Journal editors, funders
 - PURLs for standards
 - Librarians

- Revised template based on feedback
 - Added comments to fields
 - Example use case; Description of use case and filled in template for that use case
- Groups using template
 - Agricultural directory of data and services
 - Geospatial engineering
 - Materials science synchrotron data
 - Humanities and Social Sciences
 - Cultural heritage

- Email solicitation
 - Localization Microscopy, Nanoscopy
- Additional use cases collected at workshops
 - Collection of steps and processes in the Protein Data Bank data pipeline
 - Discovering a relevant dataset from the Interuniversity Consortium for Political and Social Research (ICPSR)
 - Register a dataset rescued from legacy literature repository and downloading it
- Still collecting use cases
 - File Repository at <https://rd-alliance.org/groups/metadata-ig.html>

- Harmonize template with other WG/IG collecting use cases
- Repository of use cases
 - Searchable
 - Interface for uploading use cases
 - Look for common metadata elements
 - Analyse the use cases to generate 'packages' of metadata elements for each of identified purposes

- Introduction to Metadata & Principles of Metadata Groups
- Use Case Template
- **Standards Directory**
- Provenance
- Plan
- Conclusion
- Discussion



Standards Directory

Alex Ball

research data sharing without barriers
rd-alliance.org

- 1. Develop an **RDA Metadata Standards Directory** listing standards relevant for research data
 - Comprehensive
 - Easy for anyone to contribute or update
- 2. Define and develop **use cases** for research metadata
- 3. Develop a plan for long-term growth and maintenance of the directory

Step One: Expand the DCC Directory

- DCC directory was nearly what MSDWG wanted, but
 - tightly focused on metadata-for-reuse;
 - aimed at UK HE;
 - mostly the work of two people, so incomplete in some disciplines.
- Activity:
 - Survey to enhance the directory: <http://bit.ly/1fToaqd>
 - By late 2013, added:
 - 14 new standards
 - new profiles/extensions
 - 13 new tools for working with metadata

Step Two: Develop Use Cases

- Plenary 3 (Dublin):
 - Tool developers
 - Data custodians
 - Journal editors, funders
 - PURLs for standards
 - Librarians
- Plenary 4 (Amsterdam):
 - Geospatial engineering
 - Materials science synchrotron data
 - Humanities and Social Sciences
 - Cultural heritage

MSDWG: Delivered Products

<http://rd-alliance.github.io/metadata-directory/>

Metadata RDA Metadata Directory	MIDAS-Heritage A British cultural heritage standard for recording information on buildings, archaeological sites, shipwrecks, parks and gardens, battlefields, areas of interest and artefacts. Sponsored by the Forum on Information Standards in Heritage, MIDAS Version 1.1 was released in October 2012.
Edit this page	
Getting Started View the standards View the extensions View the tools View the use cases Browse by subject areas	Summary Edit Standard Website http://www.english-heritage.org.uk/publications/midas-heritage/ Specification http://www.english-heritage.org.uk/content/publications/publicationsNew/guidelines-standards/midas-heritage/midas-heritage-2012-v1_1.pdf Related Vocabularies INSCRIPTION Subjects Arts and Humanities Social and Behavioral Sciences Disciplines Archaeology Architecture Building Conservation Heritage Studies Historical and Philosophical Studies History by Area
Adding standards Adding extensions Adding tools Adding use cases	Extensions Add CARARE metadata schema Edit An application profile of the MIDAS Heritage standard intended for delivering metadata to the CARARE service environment about an organisation's online collections, monument inventory database and digital objects.
github @twitter linkedin facebook	

- Proposed **Metadata Standards Catalogue Working Group** (MSCWG)
 - Develop the Metadata Standards Directory into a Metadata Standards Catalogue...
 - Allow records to be added, searched and retrieved by API.
 - Provide representations of records in machine-understandable form.
 - Develop, with community, recommendations of which standard(s) should be used for which purpose.
 - Provide information on elements defined by each standard, and the purpose they serve.

- Introduction to Metadata & Principles of Metadata Groups
- Use Case Template
- Standards Directory
- **Provenance**
- Plan
- Conclusion
- Discussion



RESEARCH DATA ALLIANCE



Provenance

Dave Durbin

research data sharing without barriers

rd-alliance.org

- Comparison and evaluation of models for data provenance
- Questions of data origins, maintenance of identity through the data life cycle and how to account for modifications
- Interoperable solutions across data models and ontologies
- Serve in an advisory capacity to other RDA WG and IG activities

- Bibliography and Literature Review
- Call for provenance use cases
- Example use case combining PROV, OA and SAM
- Collection of core terminology as a linked data set, using SKOS concepts. <http://purl.org/RDA-Provenance/Concepts/dataConcepts>
- Review of metadata standards in and missing from Metadata Standards Directory
- Applied example of Provenance tracking concepts and considerations to an existing workflow
- Survey of repositories, taxonomy and recommendations

- Collaboration with other RDA Metadata groups (Metadata, Metadata Standards Directory, Data Foundations and Terminology)
- Yesterday's session with Reproducibility

- Introduction to Metadata & Principles of Metadata Groups
- Use Case Template
- Standards Directory
- Provenance
- **Plan**
- Conclusion
- Discussion



Metadata Groups Plan

Keith G Jeffery

research data sharing without barriers
rd-alliance.org

- Use Cases towards Repository
- Metadata Standards Directory towards Catalog
- Towards Recommended Metadata Packages for Purposes

- Production of a proforma agreed upon by MIG, MSDWG, DICIG, RDPIG and sent out to all other RDA groups (but especially the domain-specific groups) to collect:
 - use cases including the purposes for which the metadata is used;
 - the metadata elements
 - required for each use case purpose;
 - commonly provided as metadata along with datasets within the domain;
- Repository of use cases: setting up a repository to store the use cases and metadata requirements and provisioning;

Metadata Standards Directory Towards Catalog

- Metadata Standards Directory
 - Encourage deposition of the metadata standards in the directory of metadata standards with MSDWG;
 - Human useable
- Metadata Standards Catalog
 - Convert to a catalog to be used both by humans and by computer systems
 - Proposed new WG : MSDWG → MSCWG

Towards Recommended Metadata Packages for Purposes (1)

- Analysis of metadata and purposes:
 - Analysis of the use case repository and MSDWG directory of standards to identify
 - the metadata required to satisfy the purpose(s);
 - commonly provided;
 - The analysis will highlight gaps and commonalities within and across domains.
- Proposal of Metadata Packages:
 - Proposal of packages for the commonly required purposes;
 - A package consists of a group of metadata elements/relationships with formal syntax and declared semantics to allow for machine processing as well as human readability;

Towards Recommended Metadata Packages for Purposes (2)

- Validation: Testing of metadata packages and their fitness for the purposes identified by selected groups in RDA;
- Converter provision: Resource provision (through RDA?) to provide converters between common metadata standards and the canonical packages to allow interoperation with legacy metadata environments.

- You know our plan
- Please communicate with us and work together on it
 - Use cases into repository
 - Standards into Directory (Catalog)
- Bringing benefit to all groups in RDA

- Introduction to Metadata & Principles of Metadata Groups
- Use Case Template
- Standards Directory
- Provenance
- Plan
- **Conclusion**
- Discussion



RESEARCH DATA ALLIANCE

Metadata Groups Conclusion

research data sharing without barriers

rd-alliance.org

- A major goal of RDA is sharing of research datasets.
 - interoperability is required using computer systems to discover, contextualise, select, access, transmit or process datasets.
- Interoperability
 - a user accessing the world through a local / institutional / national portal sees not only local datasets and software but also all relevant datasets and software as if they were local.
- Achieved through metadata
 - Characterising the objects (datasets, software, users, computing resources)
 - techniques to match and map those descriptions
 - generation of converters for the underlying data instances.
- Interoperation among many metadata models
 - preserves the richness of the original schemes
 - uses techniques to establish relationships between attributes in the different schemes (matching and mapping).

- Introduction to Metadata & Principles of Metadata Groups
- Use Case Template
- Standards Directory
- Provenance
- Plan
- Conclusion
- **Discussion**



RESEARCH DATA ALLIANCE



Discussion

moderated by Rebecca Koskela

research data sharing without barriers

rd-alliance.org