



## Building a Disciplinary Metadata Standards Directory

Alex Ball<sup>1</sup> Sean Chen<sup>2</sup> Jane Greenberg<sup>3</sup> Cristina Perez<sup>3</sup>  
Keith Jeffery<sup>4</sup> Rebecca Koskela<sup>5</sup>

<sup>1</sup>DCC/UKOLN Informatics, University of Bath

<sup>2</sup>School of Law, Duke University

<sup>3</sup>University of North Carolina, Chapel Hill

<sup>4</sup>EuroCRIS

<sup>5</sup>University of New Mexico

26 February 2014



Except where otherwise stated, this work is licensed under the Creative Commons Attribution 4.0 International licence: <http://creativecommons.org/licenses/by/4.0/>

The Research Data Alliance is supported by the European Commission, the US Government and the Australian Government.

# Outline

Motivation

Purpose of the Working Group

DCC Disciplinary Metadata Catalogue

Phase 1 development

Phase 2 development

Longer term development



# Standards are a Good Thing

- Standard protocols → different systems can communicate
- Standard file formats → different software can process the same files
- Standard metadata → different systems can use/combine data

# Standards are a Good Thing

- Standard protocols → different systems can communicate
- Standard file formats → different software can process the same files
- Standard metadata → different systems can use/combine data

BUT

this only works if everyone uses the same standards:

- Not enough standards → everyone does their own thing
- Too many standards → things get stuck in silos

# Metadata standards: too many or too few?

Different purposes require different metadata → standards geared towards

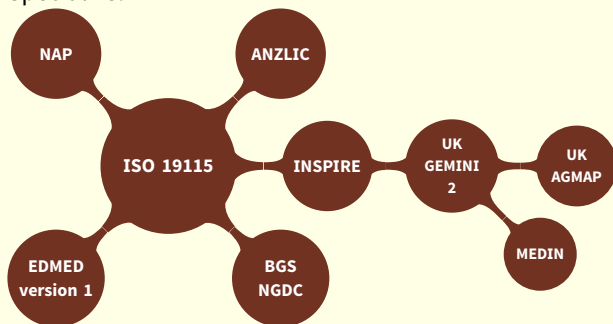
- ▶ discovery
- ▶ preservation
- ▶ packaging and exchange
- ▶ administration and rights management
- ▶ reuse of research data by third parties
  - ▶ Crystallography
  - ▶ Astronomy
  - ▶ Sociology
  - ▶ Psychology
  - ▶ ...

# Keeping it just right

Is there a standard that suits my purpose?

Yes Use it!

Sort of... Specialize!

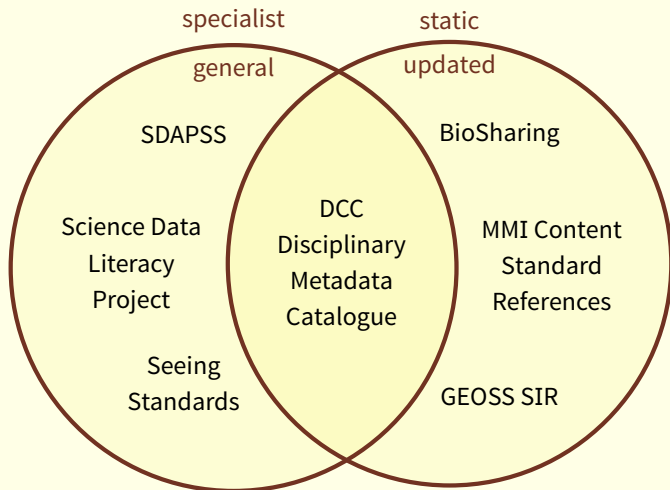


No New standard needed...

# MSDWG Goals

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 plan for long-term growth and maintenance of the directory

# Existing work

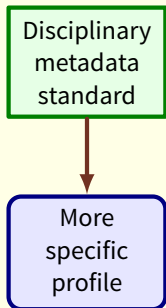




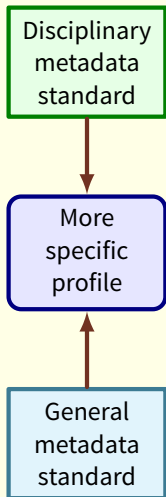
# DCC Catalogue elements

Disciplinary  
metadata  
standard

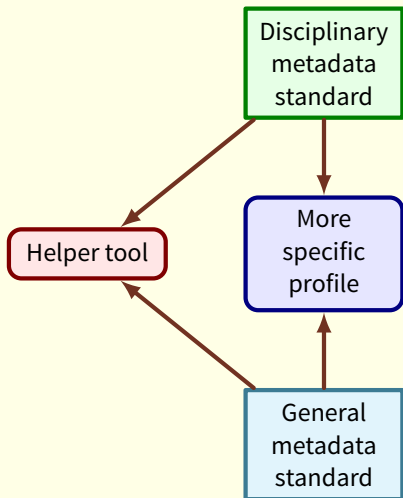
# DCC Catalogue elements



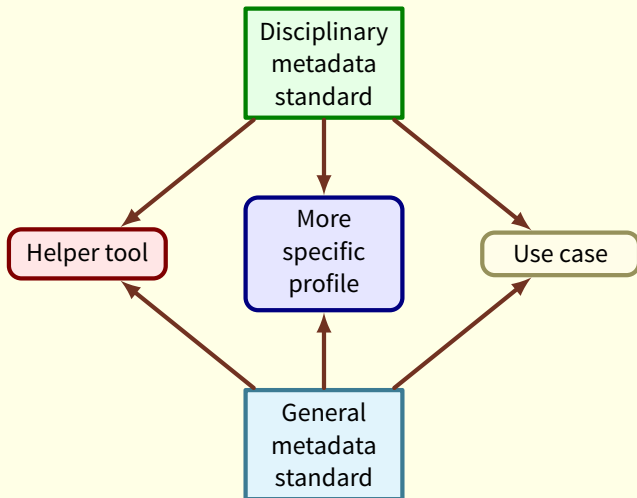
# DCC Catalogue elements



# DCC Catalogue elements



# DCC Catalogue elements



# Record for a metadata standard

## SPASE Data Model

An information model for describing the elements of the heliophysics data environment, and a set of resource types which can be used to describe data along with its scientific context, source, provenance, content and location. It is designed to support a federated data system where data may reside at different locations and may be separated from the metadata which describes it. The preferred expression form is XML.

The Space Physics Archive Search and Extract (SPASE) effort is implemented by the SPASE Consortium which is composed of representatives from the international Heliophysics data community. The Current Release of the data model (2.2.2) was updated in October 2012.

← Description

<b>Mappings</b>	<a href="#">OAI</a>
<b>Related Vocabularies</b>	<a href="#">SPASE Dictionary</a>
<b>Specification</b>	<a href="http://www.spase-group.org/docs/schema/">http://www.spase-group.org/docs/schema/</a>
<b>Standard's website</b>	<a href="http://www.spase-group.org/data/">http://www.spase-group.org/data/</a>

← Key links/  
facts

## Extensions

### [IMPEX Data Model](#)

A simulation extension to the [SPASE](#) data model.

## Tools

### [SPASE Metadata Editor](#)

A web-based editor for generating [SPASE](#) descriptions.

### [SPASE Tools](#)

The [SPASE](#) website's list of tools for working with [SPASE](#) metadata and the [SPASE](#) framework.

## Use Cases

### [NSSDC SPASE Registry](#)

The National Space Science Data Center's registry of [SPASE](#)-described space science mission data.

### [SPASE Inside](#)

The [SPASE](#) website's list of systems that use [SPASE](#) compliant metadata to enable search services.



# Record for a metadata standard

## SPASE Data Model

An information model for describing the elements of the heliophysics data environment, and a set of resource types which can be used to describe data along with its scientific context, source, provenance, content and location. It is designed to support a federated data system where data may reside at different locations and may be separated from the metadata which describes it. The preferred expression form is XML.

The Space Physics Archive Search and Extract (SPASE) effort is implemented by the SPASE Consortium which is composed of representatives from the international Heliophysics data community. The Current Release of the data model (2.2.2) was updated in October 2012.

← Description

<b>Mappings</b>	<a href="#">OAI</a>
<b>Related Vocabularies</b>	<a href="#">SPASE Dictionary</a>
<b>Specification</b>	<a href="http://www.spase-group.org/docs/schema/">http://www.spase-group.org/docs/schema/</a>
<b>Standard's website</b>	<a href="http://www.spase-group.org/data/">http://www.spase-group.org/data/</a>

← Key links/  
facts

## Extensions

### [IMPEX Data Model](#)

A simulation extension to the [SPASE](#) data model.

← Links to  
extensions

## Tools

### [SPASE Metadata Editor](#)

A web-based editor for generating [SPASE](#) descriptions.

### [SPASE Tools](#)

The [SPASE](#) website's list of tools for working with [SPASE](#) metadata and the [SPASE](#) framework.

## Use Cases

### [NSSDC SPASE Registry](#)

The National Space Science Data Center's registry of [SPASE](#)-described space science mission data.

### [SPASE Inside](#)

The [SPASE](#) website's list of systems that use [SPASE](#) compliant metadata to enable search services.

# Record for a metadata standard

## SPASE Data Model

An information model for describing the elements of the heliophysics data environment, and a set of resource types which can be used to describe data along with its scientific context, source, provenance, content and location. It is designed to support a federated data system where data may reside at different locations and may be separated from the metadata which describes it. The preferred expression form is XML.

The Space Physics Archive Search and Extract (SPASE) effort is implemented by the SPASE Consortium which is composed of representatives from the international Heliophysics data community. The Current Release of the data model (2.2.2) was updated in October 2012.

<b>Mappings</b>	<a href="#">OAI</a>
<b>Related Vocabularies</b>	<a href="#">SPASE Dictionary</a>
<b>Specification</b>	<a href="http://www.spase-group.org/docs/schema/">http://www.spase-group.org/docs/schema/</a>
<b>Standard's website</b>	<a href="http://www.spase-group.org/data/">http://www.spase-group.org/data/</a>

## Extensions

### [IMPEX Data Model](#)

A simulation extension to the [SPASE](#) data model.

## Tools

### [SPASE Metadata Editor](#)

A web-based editor for generating [SPASE](#) descriptions.

### [SPASE Tools](#)

The [SPASE](#) website's list of tools for working with [SPASE](#) metadata and the [SPASE](#) framework.

## Use Cases

### [NSSDC SPASE Registry](#)

The National Space Science Data Center's registry of [SPASE](#)-described space science mission data.

### [SPASE Inside](#)

The [SPASE](#) website's list of systems that use [SPASE](#) compliant metadata to enable search services.

← Description

← Key links/  
facts

← Links to  
extensions

← Links to  
tools



# Record for a metadata standard

## SPASE Data Model

An information model for describing the elements of the heliophysics data environment, and a set of resource types which can be used to describe data along with its scientific context, source, provenance, content and location. It is designed to support a federated data system where data may reside at different locations and may be separated from the metadata which describes it. The preferred expression form is XML.

The Space Physics Archive Search and Extract (SPASE) effort is implemented by the SPASE Consortium which is composed of representatives from the international Heliophysics data community. The Current Release of the data model (2.2.2) was updated in October 2012.

<b>Mappings</b>	<a href="#">OAI</a>
<b>Related Vocabularies</b>	<a href="#">SPASE Dictionary</a>
<b>Specification</b>	<a href="http://www.spase-group.org/docs/schema/">http://www.spase-group.org/docs/schema/</a>
<b>Standard's website</b>	<a href="http://www.spase-group.org/data/">http://www.spase-group.org/data/</a>

## Extensions

### [IMPEX Data Model](#)

A simulation extension to the [SPASE](#) data model.

## Tools

### [SPASE Metadata Editor](#)

A web-based editor for generating [SPASE](#) descriptions.

### [SPASE Tools](#)

The [SPASE](#) website's list of tools for working with [SPASE](#) metadata and the [SPASE](#) framework.

## Use Cases

### [NSSDC SPASE Registry](#)

The National Space Science Data Center's registry of [SPASE](#)-described space science mission data.

### [SPASE Inside](#)

The [SPASE](#) website's list of systems that use [SPASE](#) compliant metadata to enable search services.

← Description

← Key links/  
facts

← Links to  
extensions

← Links to  
tools

← Links to  
use cases

# Why not just use the DCC Catalogue?

- ▶ Not obvious how to contribute updates or recommend new entries
- ▶ Part of a larger site, so limited functionality and write access
- ▶ Hosted by single organization
- ▶ Scope not quite broad enough
  - ▶ Designed to support UK Higher Education
  - ▶ Excludes contextual/administrative metadata standards
  - ▶ Excludes preservation metadata standards
  - ▶ Excludes structural metadata standards
  - ▶ Excludes standards relating to non-tabular data: audio, video, interview transcripts, etc.

# Why not just use the DCC Catalogue?

- ▶ Not obvious how to contribute updates or recommend new entries
- ▶ Part of a larger site, so limited functionality and write access
- ▶ Hosted by single organization
- ▶ Scope not quite broad enough
  - ▶ Designed to support UK Higher Education
  - ▶ Excludes contextual/administrative metadata standards
  - ▶ Excludes preservation metadata standards
  - ▶ Excludes structural metadata standards
  - ▶ Excludes standards relating to non-tabular data: audio, video, interview transcripts, etc.

Good starting point, but more work needed.

# Phase 1: survey

## Sent to lists:

- ▶ RDA-All
- ▶ RDA MSDWG
- ▶ RDA Metadata Interest Group
- ▶ EuroCRIS
- ▶ EPOS
- ▶ DC-SAM
- ▶ ASIS&T RDAP summit series
- ▶ DataONE
- ▶ ESIP
- ▶ STFC
- ▶ ...

### Metadata Directory Collection

In cooperation with the UNC SILS Metadata Research Center and the Research Data Alliance students at UNC SILS are gathering information about metadata standards that apply to scientific data. The aim of this is achieve a short term goal of the Research Data Alliance's Metadata Standards Directory Working Group to:

Develop a prototype wiki-based directory (RDA Metadata Directory) listing metadata standards applicable to scientific data. The initial emphasis will be on widely-used and domain community- endorsed metadata standards and schemas with significant interoperability / re-use capability.

Information submitted to this project will be integrated with similar information from the Digital Curation Center to prototype and build a sustainable platform for sharing and exposing information about metadata standards.

If you have previously submitted information to our collection: check the Digital Curation Center's directory of standards for science data at: <http://www.dcc.ac.uk/resources/metadata-standards>

Please submit one standard per form. At the end of the questionnaire a link to another form submission is available.

\* Required

**Standard Name \***

**Description of the Metadata Standard**  
May be a URL to about page or free text summary.

**Metadata Standard URL \***  
URL for the standard's home page

**Domain of the Metadata Standard \***  
Domains where the standard is in use

Biology

Earth Science

Physical Science

**18 contributions (14 new standards, 4 extensions) 13 new tools for generating metadata**

**DCC @:** <http://www.dcc.ac.uk/resources/metadata-standards>

### STANDARDS ADDED TO DCC directory

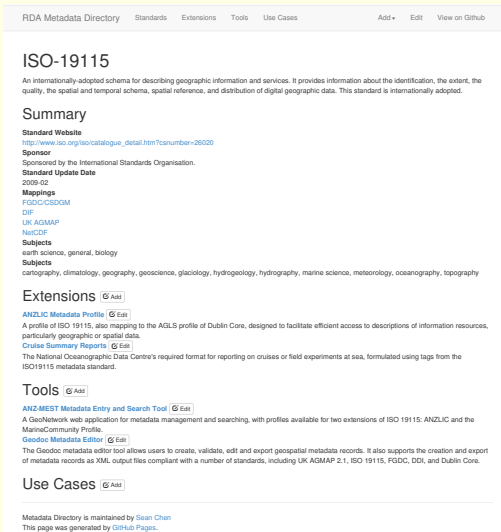
1. ClinicalTrials.gov Protocol Data Element Definitions (DRAFT)
2. FITS (Flexible Image Transport System)
3. GenDMS
4. Genome Metadata
5. GEOSS Standards Registry
6. Journeau
7. Observ-OM
8. Observations and Measurements
9. Open Archives Initiative
10. PROV
11. QuDEX
12. Resource Metadata for the Virtual Observatory
13. WCS - World Coordinate System
14. The Gulf of Mexico Research Initiative Information and Data Cooperative (GRIIDC)

### CONTRIBUTIONS FROM:

Syracuse University  
Purdue University Libraries  
National Center for Ecological Analysis and Synthesis  
DICE Center, UNC-CH  
National Snow and Ice Data Center, Univ. of Colorado  
ResearXis-Discinnet  
University Medical Center Groningen  
Commonwealth Scientific Industrial Research Organisation  
Open Microscopy Environment  
Rensselaer Polytechnic Institute  
Newcastle University  
Uk Data Archive  
US Virtual Astronomical Observatory / Space Telescope Science Institute  
EDINA, The University of Edinburgh  
U. S. Geoscience Information Network  
Met Office, UK  
IMOS/AODN Integrated Marine Observing System/Australian Ocean Data Network  
NRCAN/GeoConnections

# Phase 2: a new platform

- ▶ Pages are generated from simple easy-to-edit text files.
- ▶ Maintainers can edit the pages directly.
- ▶ Others can fork the whole directory, make changes on their copy, and submit a pull request.



The screenshot shows the RDA Metadata Directory website. The navigation bar includes 'Standards', 'Extensions', 'Tools', 'Use Cases', 'Add', 'Edit', and 'View on GitHub'. The main content is for 'ISO-19115', described as an internationally-adopted schema for geographic information and services. It includes sections for 'Summary', 'Standard Website' (with a URL), 'Sponsor' (International Standards Organisation), 'Standard Update Date' (2009-02), 'Mappings' (FGDC/CSDGM, DIF, UK AGMAP, NatCDF), 'Subjects' (earth science, general, biology, cartography, climatology, geography, geoscience, glaciology, hydrogeology, hydrography, marine science, meteorology, oceanography, topography), 'Extensions' (ANZLIC Metadata Profile, ANZ-MEST Metadata Entry and Search Tool, Cruise Summary Reports), 'Tools' (ANZ-MEST Metadata Entry and Search Tool, Geodoc Metadata Editor), and 'Use Cases'. At the bottom, it notes the page is maintained by Sean Chen and generated by GitHub Pages.

# Future developments

- ▶ Extend the directory to serve more use cases:
  - ▶ searching
  - ▶ querying and processing by automated tools
- ▶ ...representation of records in RDF?
- ▶ ...representation of the standards themselves in RDF?
- ▶ Categorize metadata standards by primary application, e.g. discovery, preservation, reuse.



Thank you for your attention

RDA Metadata Standards Directory Working Group web page:

<https://rd-alliance.org/working-groups/metadata-standards-directory-working-group.html>

DCC Disciplinary Metadata:

<http://www.dcc.ac.uk/resources/metadata-standards>

Standards survey: <http://bit.ly/1fToaqd>