Overview of scientific metadata for data publishing, citation, and curation

Alex Ball

UKOLN, University of Bath, UK

22 September 2011

DC-2011
KB, The Hague, Netherlands

Outline

A study of scientific metadata

```
http://www.ukoln.ac.uk/
projects/sdapss/
```

Scientific Data Application Profile Scoping Study Report

Document details

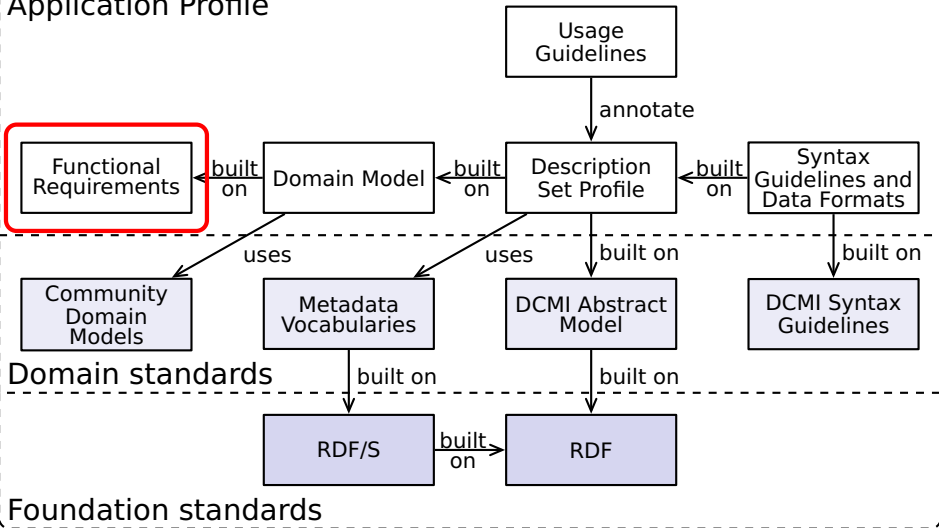|	
Author:	Alexander Ball, UKOLN, University of Bath
Date:	3rd June 2009
Version:	1.1
Document Name:	sdapss.pdf
Notes:	Changes from version 1.0: Typographical corrections made. References added. Conclusions expanded.

Singapore Framework
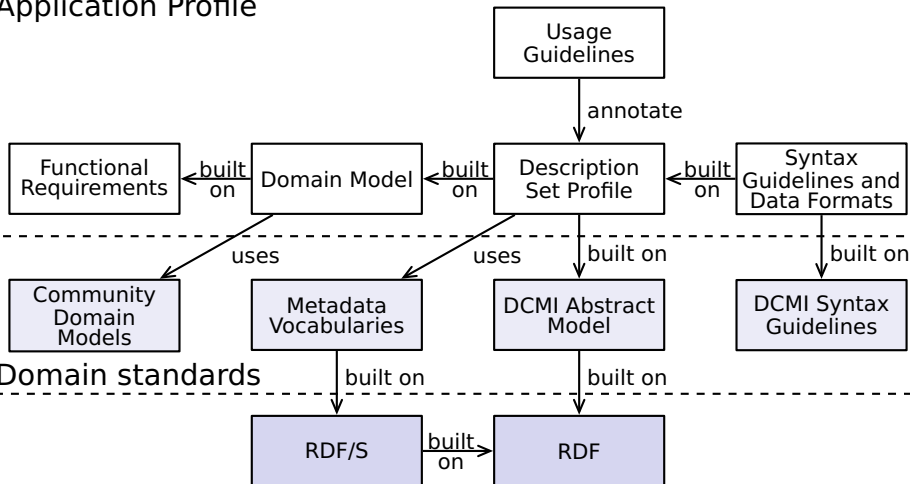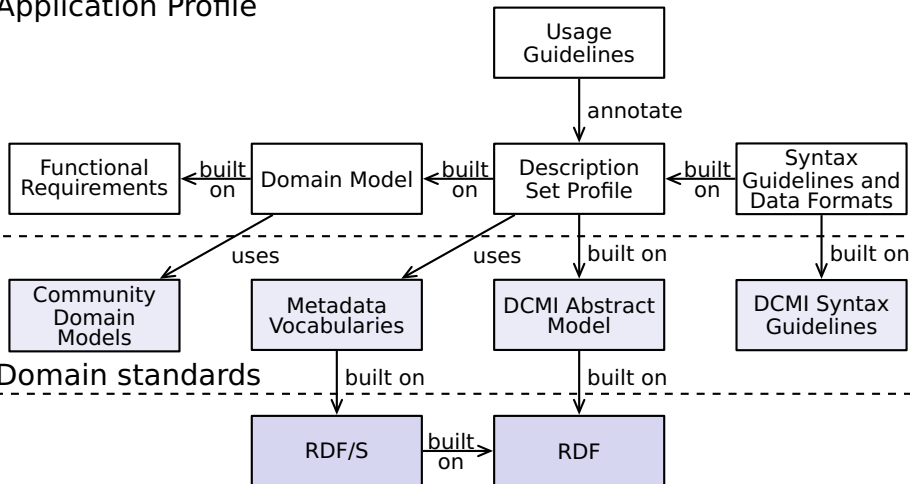
Singapore Framework

Repository landscape

Singapore Framework
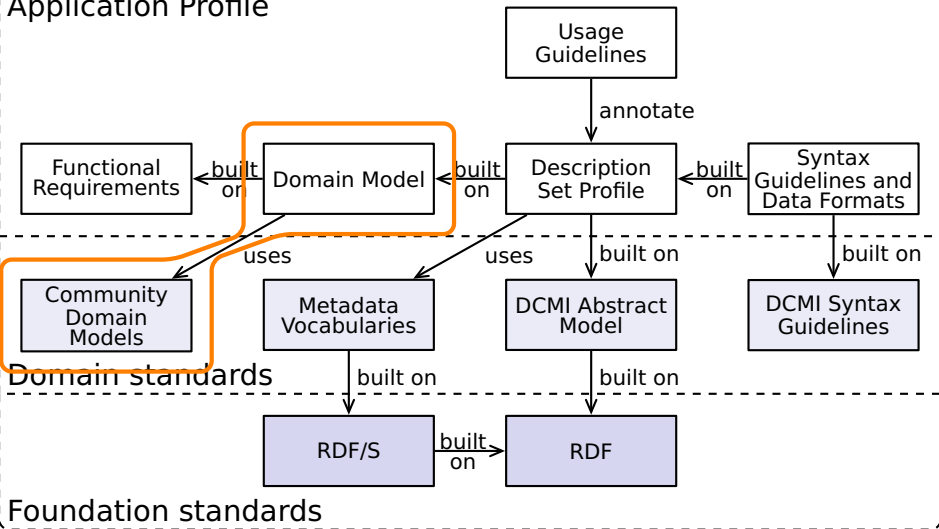
Singapore Framework

Singapore Framework
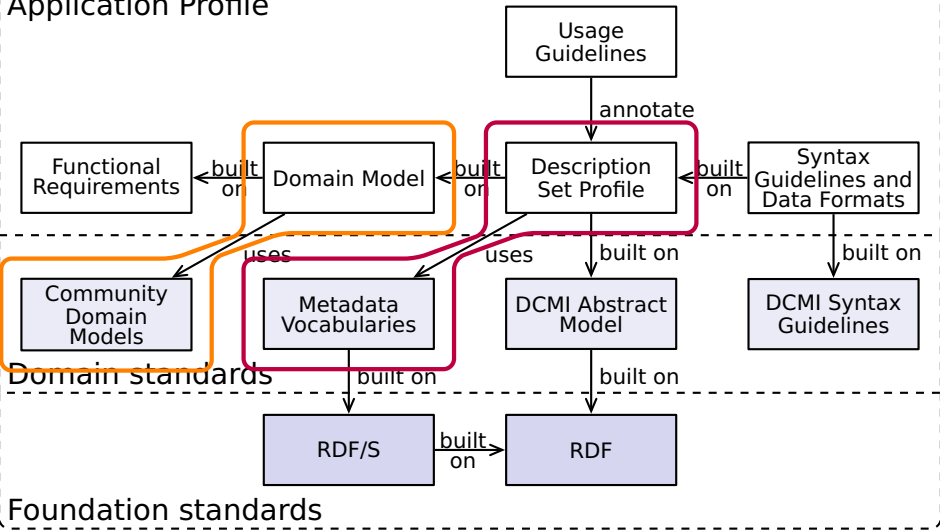
Schemes and profiles in the wild
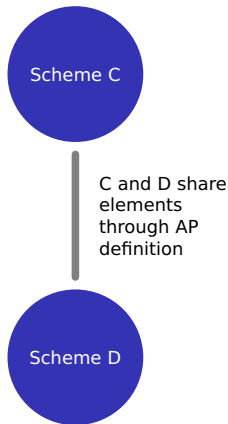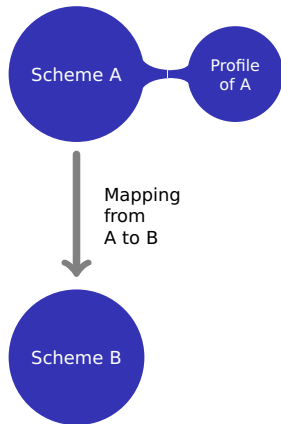
Legend

Geospatial/Environmental Data
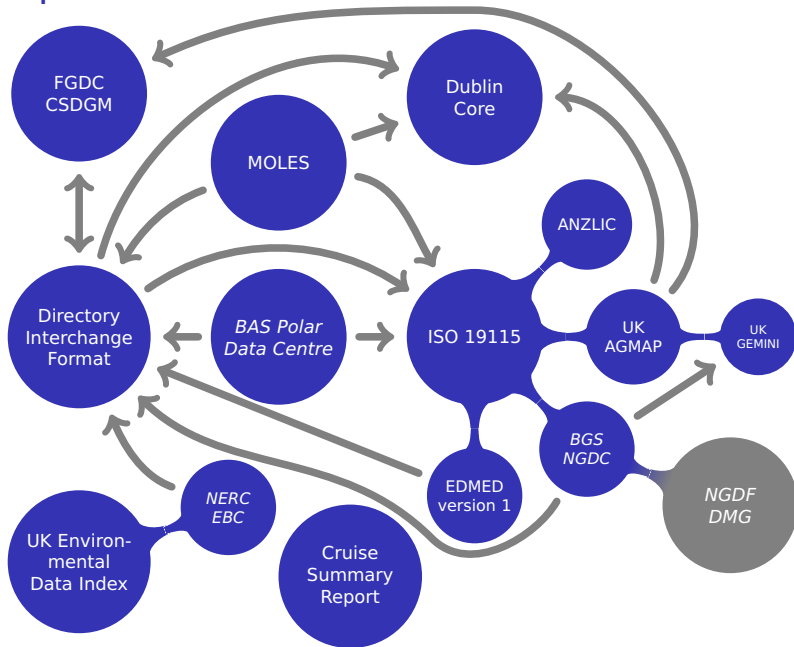
Biology Data
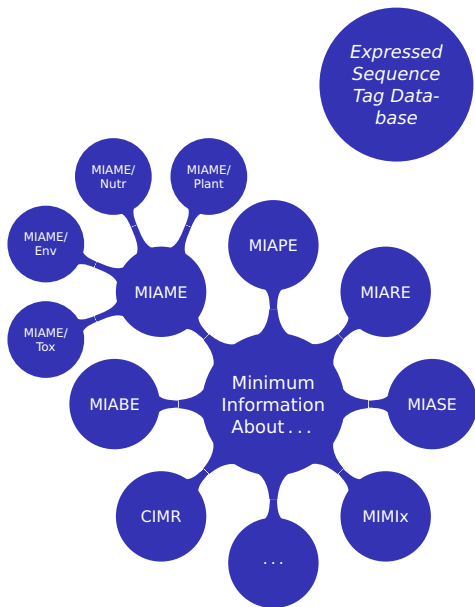
Social Science and Humanities Data
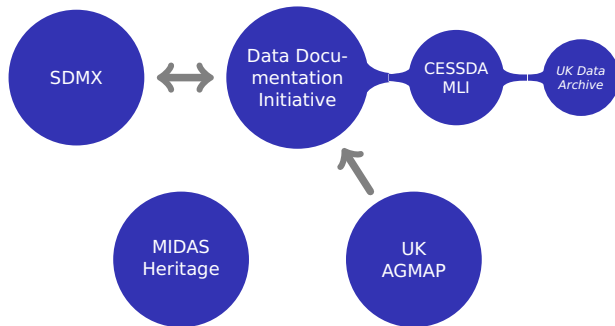
Structural Sciences

General

Comparison of metadata standards and profiles
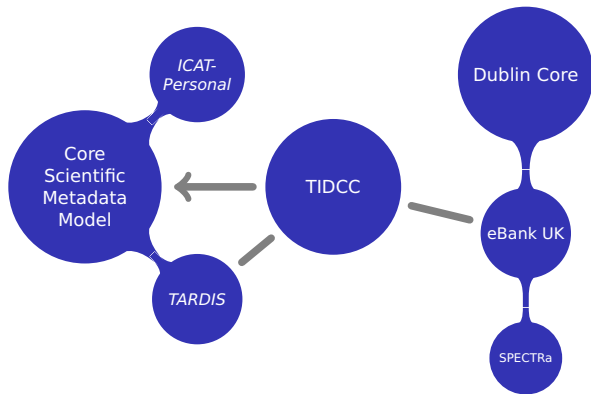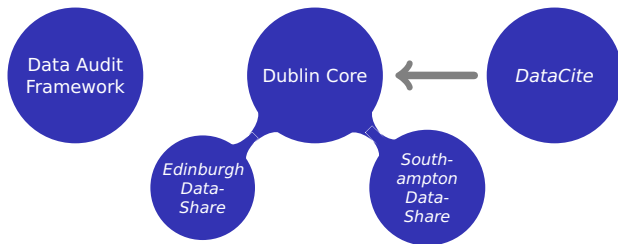
Legend

- The following lists show elements that were common to at least 3 of the 15 standards/profiles studied.
- Elements in red occurred in the most standards/profiles.
- Elements in black occurred in the fewest standards/profiles.

Identification

- Dataset Name
- Dataset Version
- Dataset Date
- Dataset Identifier

- Metadata Scheme Name
- Metadata Scheme Version
- Metadata Record Date
- Metadata Record Identifier

Responsibility

- Project/Study/Series Name
- Project/Study/Series Status
- Rights/Restrictions

- Agent
- Agent Contact Details

Archiving

- Location
- File Format(s)
- Storage Medium
- Size
- Data Quality Information
- Data Preview
- Dataset Language
- Dataset Status

Spatiotemporal Coverage

- Spatial Extent
- Spatial Resolution
- Temporal Extent
- Temporal Resolution

Topical Coverage and Derivation

- Dataset Type
- Subject/Keywords
- Abstract/Summary/ Description
- Parameters Used

- Methodology/ Instrumentation
- Processing Steps
- Related Datasets
- Derived Publications

Is a scientific metadata scheme worth the trouble?

DataCite Metadata Elements

(based on v2.2, in SDAPSS terms)

Identification

- Dataset Identifier
- Dataset Name
- Dataset Date
- Dataset Version
- Metadata Record Date
- *Metadata Record Version* (not in SDAPSS list)

Responsibility

- Agent (Creator/Publisher/Contributor)
- Rights/Restrictions

Archiving

- Dataset Language
- Size
- File Format(s)

Topical Coverage and Derivation

- Subject/Keywords
- Dataset Type
- Related Datasets
- Derived Publications
- Abstract/Summary/Description

Questions?

Further information

Contact Alex Ball at
a.ball@ukoln.ac.uk

Scientific Data Application Profile Scoping Study
available from:
`http://www.ukoln.ac.uk/projects/sdapss/`

Notes

ISO 19115

Full name
: ISO 19115:2003. *Geographic information – Metadata.*

Maintained by
: ISO Technical Committee 211

Usage
: *Lingua franca* for geospatial data; widely profiled. ISO 19139:2007 provides its XML format, Geographic MetaData XML (GMD).

Further information
: `http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020`

ANZLIC

Full name
: Australia and New Zealand Land Information Council Metadata Profile

Maintained by
: ANZLIC – the Spatial Information Council

Usage
: Profile of ISO 19115 for use in Australia and New Zealand, replacing the *ANZLIC Metadata Guidelines* version 2 (2001)

Further information
: `http://www.anzlic.org.au/Publications/Metadata+Project/default.aspx`

BGS NGDC

Maintained by British Geological Survey National Geoscience Data Centre

Usage Application profile for use by the NGDC, based on ISO 19115 with additional elements to support legacy and specialist metadata, including the National Geospatial Data Framework *Discovery Metadata Guidelines*

Further information (NGDC) `http://www.bgs.ac.uk/services/ngdc/home.html`

Further information (NGDF DMG)
`http://www.ngdf.org.uk/Metadata/metguide/metaguide12.pdf`

Back

UK AGMAP

Full name	UK Academic Geospatial Metadata Application Profile
Maintained by	Go-Geo!, EDINA
Usage	Profile of ISO 19115 (and superset of UK GEMINI) for use by the UK higher and further education sectors
Further information	`http://www.gogeo.ac.uk/cgi-bin/mdres.cgi`

UK GEMINI

Full name UK Geo-spatial Metadata Interoperability Initiative Standard

Maintained by Association for Geographic Information

Usage Profile of ISO 19115 for use by the UK Government and national services such as Gigateway

Further information

`http://www.agi.org.uk/uk-gemini/`

EDMED

Full name   European Directory of Marine Environmental Datasets

Maintained by   British Oceanographic Data Centre, EU SeaDataNet Project

Usage   Metadata scheme used by the EDMED database and national databases; the XML form of version 1 (replacing version 0) is a profile of ISO 19115

Further information   `https://www.bodc.ac.uk/data/information_and_inventories/edmed/`

CSR

Full name Cruise Summary Report

Maintained by Intergovernmental Oceanographic Commission, Bundesamt für Seeschiffahrt und Hydrographie/Deutsches Ozeanographisches Datenzentrum

Usage Standard form for reporting observations and the samples collected by oceanographic cruises

Further information `http://www.seadatanet.org/Metadata/CSR`

Back

BAS PDC

Maintained by British Antarctic Survey Polar Data Centre

Usage Application profile for use by the PDC, harmonious with both ISO 19115 and DIF

Further information `http://www.antarctica.ac.uk/bas_research/data/`

Back

DIF

Full name Directory Interchange Format

Maintained by Committee on Earth Observation Satellites International Directory Network (CEOS IDN)

Usage Native metadata scheme for the Global Change Master Directory, hosted by NASA

Further information `http://gcmd.nasa.gov/User/difguide/difman.html`

FGDC CSDGM

Full name	Content Standard for Digital Geospatial Metadata
Maintained by	Federal Geographic Data Committee
Usage	Native metadata scheme for the Geospatial One-Stop data portal and the National Spatial Data Infrastructure Clearinghouse.

Further information

`http://www.fgdc.gov/metadata/csdgm/`

Back

MOLES

Full name
: Metadata Objects for Linking Environmental Sciences

Maintained by
: Centre for Environmental Data Archival (CEDA), Science and Technology Facilities Council (STFC)

Usage
: Native metadata scheme for CEDA and the Natural Environment Research Council's DataGrid

Further information
: http://proj.badc.rl.ac.uk/moles/wiki

Back

UKEDI

Full name UK Environmental Data Index

Maintained by Environmental Information Centre, Centre for Ecology and Hydrology (CEH)

Usage Native metadata scheme for the UKEDI catalogue

Further information `http://ukedi.ceh.ac.uk/`

Back

NEBC

Maintained by Natural Environment Research Council
(NERC) Environmental Bioinformatics Centre

Usage Native metadata scheme used across all
data holdings, loosely based on UKEDI
(further, type-specific metadata is also held)

Further information `http://nebc.nerc.ac.uk/data/`
`standards-and-ontologies/overview`

Back

MIBBI

Full name
: Minimum Information for Biological and Biomedical Investigations

Funded by
: Natural Environment Research Council, Biotechnology and Biological Sciences Research Council

Purpose
: To promote and co-ordinate projects working on reporting guidelines (especially 'Minimum Information' schemes), and drive adoption of these guidelines

Further information
: `http://mibbi.org/index.php/MIBBI_portal`

Back

dbEST

Full name Expressed Sequence Tag Database

Maintained by National Center for Biotechnology Information

Usage Native metadata scheme for EST data in GenBank

Further information `http://www.ncbi.nlm.nih.gov/dbEST/how_to_submit.html`

Back

Dryad

Full name Dryad Metadata Application Profile

Maintained by Dryad Development Team

Usage Native metadata profile for the Dryad repository, mostly drawn from Dublin Core, Darwin Core and BIBO

Further information `https://www.datadryad.org/wiki/Metadata_Profile`

Back

DwC

Full name   Darwin Core standard

Maintained by   Taxonomic Databases Working Group

Usage   Biodiversity informatics standard used by the Global Biodiversity Information Facility, the Atlas of Living Australia, the Ornithological Information System, the Ocean Biogeographic Information System, and others

Further information

http://rs.tdwg.org/dwc/index.htm

DDI

Full name Data Documentation Initiative

Maintained by Data Documentation Initiative Alliance

Usage Metadata standard used or profiled by social science data archives across the globe

Further information http://www.ddialliance.org/

CESSDA MLI

Full name Council of European Social Science Data Archives Minimum Level of Information

Maintained by Council of European Social Science Data Archives

Usage Common base profile of DDI for use by the member archives of CESSDA

Further information `http://www.cessda.org/sharing/managing/3/`

UKDA

Maintained by UK Data Archive

Usage Profile of DDI for use by the UKDA, based on the CESSDA MLI.

Further information `http://www.data-archive.ac.uk/create-manage/document/overview`

Back

SDMX

Full name Statistical Data and Metadata Exchange

Maintained by Statistical Data and Metadata Exchange Initiative

Usage Metadata standard used by the UN, the World Bank, the Organization for Economic Co-operation and Development and others

Further information `http://sdmx.org/`

Back

MIDAS Heritage

Full name
Manual and Data Standard for Monument Inventories (MIDAS) Heritage

Maintained by
English Heritage, Forum on Information Standards in Heritage

Usage
Metadata standard used by UK agencies, record offices, archaeological contractors and academic researchers working with the historic environment

Further information
`http://www.english-heritage.org.uk/professional/archives-and-collections/nmr/heritage-data/midas-heritage/`

Back

Core Scientific Metadata Model

AKA CCLRC Scientific Metadata Model

Maintained by Science and Technology Facilities Council (STFC)

Usage Metadata standard used by STFC's ICAT data catalogue and the Australian Research Enabling Environment (ARCHER) XDMS tool, among others

Further information `https://sourceforge.net/userapps/mediawiki/ericayang/index.php?title=I2S2#Core_Scientific_MetaData_model_.28CSMD.29`

Back

ICAT-Personal

Maintained by Science and Technology Facilities Council (STFC)

Usage Extension of the CSMD model to accommodate data processing stages beyond initial collection

Further information `http://sourceforge.net/apps/mediawiki/icatlite/`

Back

TARDIS

Full name	The Australian Repositories for Diffraction Images
Maintained by	TARDIS Project partners (Monash University, VeRSI, Australian Synchrotron, and others)
Usage	Profile of the CSMD model for Australian crystallographic data
Further information	`http://dx.doi.org/10.1107/S0907444908015540`

Back

eBank UK

Full name eBank UK Metadata Application Profile

Maintained by UKOLN, University of Bath (DCC); University of Southampton; University of Manchester

Usage Dublin Core Application Profile suited to crystallography, later used by the eCrystals Federation Project

Further information `http://www.ukoln.ac.uk/projects/ebank-uk/schemas/profile/`

SPECTRa

Full name	Submission, Preservation and Exposure of Chemistry Teaching and Research Data
Maintained by	University of Cambridge, Imperial College London
Usage	Extension of the eBank profile to organic and computational chemistry
Further information	`http://www.lib.cam.ac.uk/spectra/FinalReport.html`

TIDCC

Full name   Towards an International Data Commons for Crystallography

Maintained by   The Australian Repositories for Diffraction Images (TARDIS) Project, eCrystals Federation Project, DataMINX Project, Australian Research Council's Molecular and Materials Structure Network (MMSN)

Usage   Profile of the CSMD model for Australian crystallographic data

Further information   `http://wiki.ecrystals.chem.soton.ac.uk/images/9/9d/ECrystals-WP4-PM-Final.pdf`

Back

Edinburgh DataShare

Full name DSpace Metadata Schema for Edinburgh
DataShare

Maintained by University of Edinburgh

Usage Profile of Dublin Core and administrative
DSpace metadata for use in Edinburgh's
data repository

Further information `http://www.disc-uk.org/docs/`
`Edinburgh_DataShare_DC-schema1.pdf`

Southampton DataShare

Full name
: DataShare Metadata Schema for ePrints Soton

Maintained by
: University of Southampton

Usage
: Profile of Dublin Core and administrative ePrints metadata for use in Southampton's data repository

Further information
: `http://www.disc-uk.org/docs/sPrints_Soton_Metadata.pdf`

DataCite

Full name	DataCite Metadata Schema for the Publication and Citation of Research Data
Maintained by	DataCite Metadata Working Group
Usage	Scheme for the metadata collected by DataCite when registering a DOI for a database
Further information	`http://schema.datacite.org/`

Back

DAF

Full name Data Asset Framework Audit Form 3B: Data asset management (Optional extended element set)

AKA Data Audit Framework

Maintained by HATII, University of Glasgow (DCC)

Usage Suggested set of data to collect about research data assets when assessing their management

Further information

http://data-audit.eu/documents.html