

Metadata for Data Citation and Discovery

Alex Ball

DCC/UKOLN, University of Bath

6 July 2012



Except where otherwise stated, this work is licensed under Creative Commons Attribution 2.5 Scotland:
<http://creativecommons.org/licenses/by/2.5/scotland/>

Funded by **JISC**

Outline

Citation, discovery and reuse

Data citation

Motivation

Metadata

Data discovery

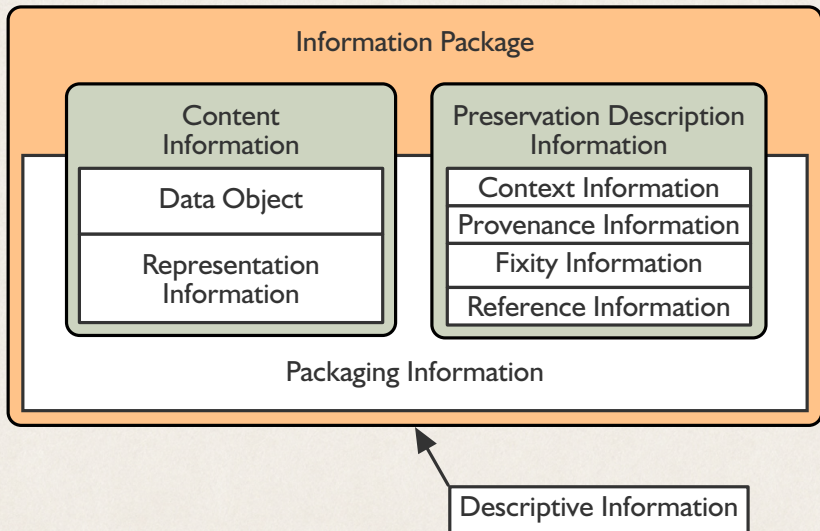
Motivation

Metadata

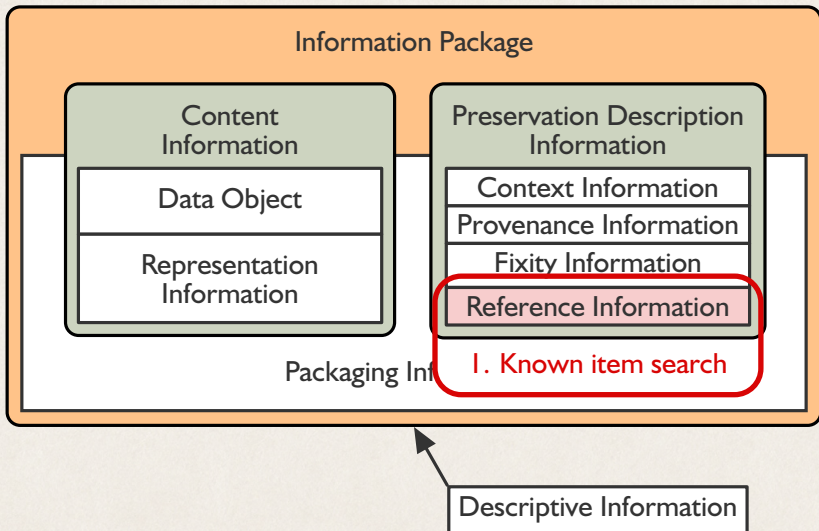
Issues and challenges

Recommendations

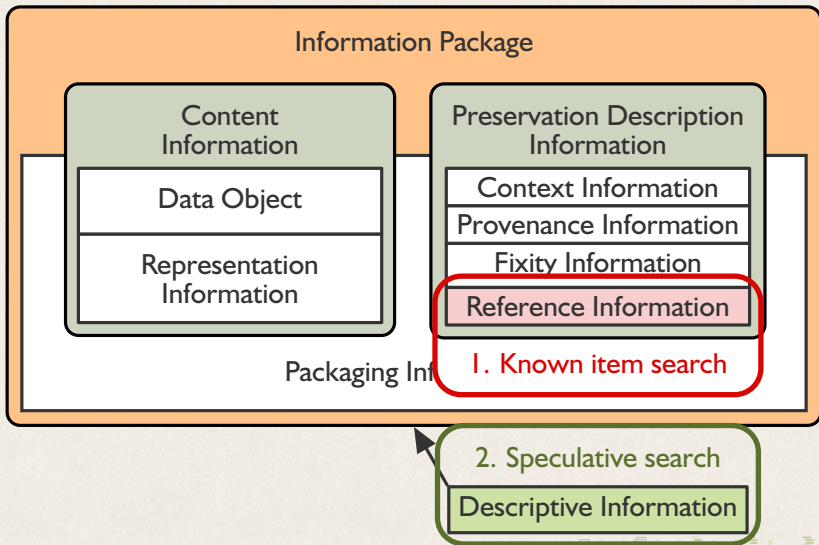
Layers of metadata



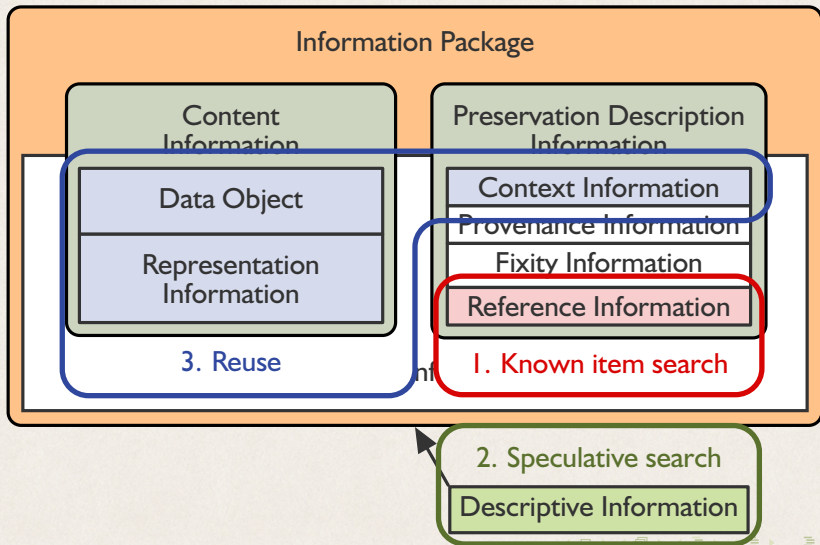
Layers of metadata



Layers of metadata



Layers of metadata



What's great about journal papers?

- ▶ Awareness raising
- ▶ Protection from plagiarism
- ▶ Verification of results
- ▶ Basis for future research
- ▶ Reward models
- ▶ Permanent access

What's great about journal papers?

- ▶ Awareness raising
- ▶ Protection from plagiarism
- ▶ ~~Verification of results~~
- ▶ ~~Basis for future research~~
- ▶ Reward models
- ▶ Permanent access

Data citations provide. . .

- ▶ Visibility for data
- ▶ Protection from plagiarism
- ▶ Possibility for verification of results
- ▶ Data on which to base future research
- ▶ Possibility for reward models
- ▶ Access

Citation styles

- ▶ Altman, M. & King, G. (2007). A proposed standard for the scholarly citation of quantitative data. *D-Lib Magazine*, 13(3/4). doi:10.1045/march2007-altman
- ▶ Lawrence, B. N., Jones, C. M., Matthews, B. M. & Pepler, S. J. (2008, February 1). *Data publication* (Claddier Project Report No. 3). BADC. Retrieved from <http://purl.org/oai/oai:epubs.cclrc.ac.uk:work/43641>
- ▶ Green, T. (2010, February). *We need publishing standards for datasets and data tables*. OECD Publishing. doi:10.1787/787355886123
- ▶ Starr, J. & Gastl, A. (2011). isCitedBy: A metadata scheme for DataCite. *D-Lib Magazine*, 17(1/2). doi:10.1045/january2011-starr

Citation styles

Four data citation styles: which elements do they use?

Altman and King (2007): Dataverse

Lawrence et al. (2008): BADC

Green (2010): OECD

Starr and Gastl (2011): DataCite

Citation styles

Author

Altman and King (2007): Dataverse

- ▶ Sidney Verba.

NORC [Producer];

Lawrence et al. (2008): BADC

- ▶ Iwi, A. and B. N. Lawrence

Green (2010): OECD

- ▶ OECD

Starr and Gastl (2011): DataCite

- ▶ Irino, T; Tada, R

Citation styles

Publication date

Altman and King (2007): Dataverse

- ▶ Sidney Verba. 1998.

NORC [Producer];

Lawrence et al. (2008): BADC

- ▶ Iwi, A. and B. N. Lawrence (2004).

Green (2010): OECD

- ▶ OECD (2009),

(Accessed on 14 September 2009)

Starr and Gastl (2011): DataCite

- ▶ Irino, T; Tada, R (2009):

Citation styles

Title

Altman and King (2007): Dataverse

- ▶ Sidney Verba. 1998. "U.S. and Russian Social and Political Participation Data," NORC [Producer];

Lawrence et al. (2008): BADC

- ▶ Iwi, A. and B. N. Lawrence (2004). A 500 year control run of HadCM3.

Green (2010): OECD

- ▶ OECD (2009), "Key short-term indicators", Main Economic Indicators (Accessed on 14 September 2009)

Starr and Gastl (2011): DataCite

- ▶ Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797.

Citation styles

Version

Altman and King (2007): Dataverse

- ▶ Sidney Verba. 1998. "U.S. and Russian Social and Political Participation Data," NORC [Producer];

Lawrence et al. (2008): BADC

- ▶ Iwi, A. and B. N. Lawrence (2004). A 500 year control run of HadCM3. **Version 1.**

Green (2010): OECD

- ▶ OECD (2009), "Key short-term indicators", Main Economic Indicators (Accessed on 14 September 2009)

Starr and Gastl (2011): DataCite

- ▶ Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. **V.2.**

Citation styles

Feature

Altman and King (2007): Dataverse

- ▶ Sidney Verba. 1998. "U.S. and Russian Social and Political Participation Data," NORC [Producer];

Lawrence et al. (2008): BADC

- ▶ Iwi, A. and B. N. Lawrence (2004). A 500 year control run of HadCM3. [GridSeries, <http://ndg.nerc.ac.uk/csml2/GridSeries>] Version I.

Green (2010): OECD

- ▶ OECD (2009), "Key short-term indicators", Main Economic Indicators (Accessed on 14 September 2009)

Starr and Gastl (2011): DataCite

- ▶ Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. V.2.

Citation styles

Resource type

Altman and King (2007): Dataverse

- ▶ Sidney Verba. 1998. "U.S. and Russian Social and Political Participation Data," NORC [Producer]; data set [Type (DC)]

Lawrence et al. (2008): BADC

- ▶ Iwi, A. and B. N. Lawrence (2004). A 500 year control run of HadCM3. [GridSeries, <http://ndg.nerc.ac.uk/csml2/GridSeries>] Version I.

Green (2010): OECD

- ▶ OECD (2009), "Key short-term indicators", Main Economic Indicators (database). (Accessed on 14 September 2009)

Starr and Gastl (2011): DataCite

- ▶ Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. V.2. Dataset.

Citation styles

Publisher

Altman and King (2007): Dataverse

- ▶ Sidney Verba. 1998. "U.S. and Russian Social and Political Participation Data," NORC [Producer]; data set [Type (DC)] **ICPSR** [Distributor].

Lawrence et al. (2008): BADC

- ▶ Iwi, A. and B. N. Lawrence (2004). A 500 year control run of HadCM3. [GridSeries, <http://ndg.nerc.ac.uk/csml2/GridSeries>] Version I. **BADC**.

Green (2010): OECD

- ▶ OECD (2009), "Key short-term indicators", Main Economic Indicators (database). (Accessed on 14 September 2009)

Starr and Gastl (2011): DataCite

- ▶ Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. V.2. **Geological Institute, University of Tokyo**. Dataset.

Citation styles

Identifier

Altman and King (2007): Dataverse

- ▶ Sidney Verba. 1998. "U.S. and Russian Social and Political Participation Data," [hdl:1902.4/00754](https://doi.org/10.1902.4/00754)
NORC [Producer]; data set [Type (DC)] ICPSR [Distributor].

Lawrence et al. (2008): BADC

- ▶ Iwi, A. and B. N. Lawrence (2004). A 500 year control run of HadCM3. [GridSeries, <http://ndg.nerc.ac.uk/csm12/GridSeries>] Version 1. BADC. [urn:badc.nerc.ac.uk_coapec500yr](https://nbn-resolving.org/urn:nbn:uk:nerc:badc-coapec500yr)

Green (2010): OECD

- ▶ OECD (2009), "Key short-term indicators", Main Economic Indicators (database). [doi: 10.1787/data-00039-en](https://doi.org/10.1787/data-00039-en)
(Accessed on 14 September 2009)

Starr and Gastl (2011): DataCite

- ▶ Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. V.2. Geological Institute, University of Tokyo. Dataset. [doi:10.1594/PANGAEA.726855](https://doi.org/10.1594/PANGAEA.726855).

Citation styles

Location

Altman and King (2007): Dataverse

- ▶ Sidney Verba. 1998. "U.S. and Russian Social and Political Participation Data," [hdl:1902.4/00754](https://hdl.handle.net/1902.4/00754)
NORC [Producer]; data set [Type (DC)] ICPSR [Distributor].

Lawrence et al. (2008): BADC

- ▶ Iwi, A. and B. N. Lawrence (2004). A 500 year control run of HadCM3. [GridSeries, <http://ndg.nerc.ac.uk/csm12/GridSeries>] Version 1. BADC. urn:badc.nerc.ac.uk_coapec500yr [Available from <http://badc.nerc.ac.uk/data/coapec500yr>].

Green (2010): OECD

- ▶ OECD (2009), "Key short-term indicators", Main Economic Indicators (database). doi: 10.1787/data-00039-en <http://dx.doi.org/10.1787/data-00039-en> (Accessed on 14 September 2009)

Starr and Gastl (2011): DataCite

- ▶ Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. V.2. Geological Institute, University of Tokyo. Dataset. doi:10.1594/PANGAEA.726855. <http://dx.doi.org/10.1594/PANGAEA.726855>

Citation styles

Unique Numeric Fingerprint

Altman and King (2007): Dataverse

- ▶ Sidney Verba. 1998. "U.S. and Russian Social and Political Participation Data," hdl:1902.4/00754
UNF:3:ZNQRI14053UZq389x0Bffg?== NORC [Producer]; data set [Type (DC)] ICPSR [Distributor].

Lawrence et al. (2008): BADC

- ▶ Iwi, A. and B. N. Lawrence (2004). A 500 year control run of HadCM3. [GridSeries, <http://ndg.nerc.ac.uk/csm12/GridSeries>] Version 1. BADC. urn:badc.nerc.ac.uk_coapec500yr [Available from <http://badc.nerc.ac.uk/data/coapec500yr>].

Green (2010): OECD

- ▶ OECD (2009), "Key short-term indicators", Main Economic Indicators (database). doi: 10.1787/data-00039-en <http://dx.doi.org/10.1787/data-00039-en> (Accessed on 14 September 2009)

Starr and Gastl (2011): DataCite

- ▶ Irino, T; Tada, R (2009): Chemical and mineral compositions of sediments from ODP Site 127-797. V.2. Geological Institute, University of Tokyo. Dataset. doi:10.1594/PANGAEA.726855. <http://dx.doi.org/10.1594/PANGAEA.726855>

Key citation elements

- ▶ Author
- ▶ Publication date
- ▶ Title
- ▶ Location

Key citation elements

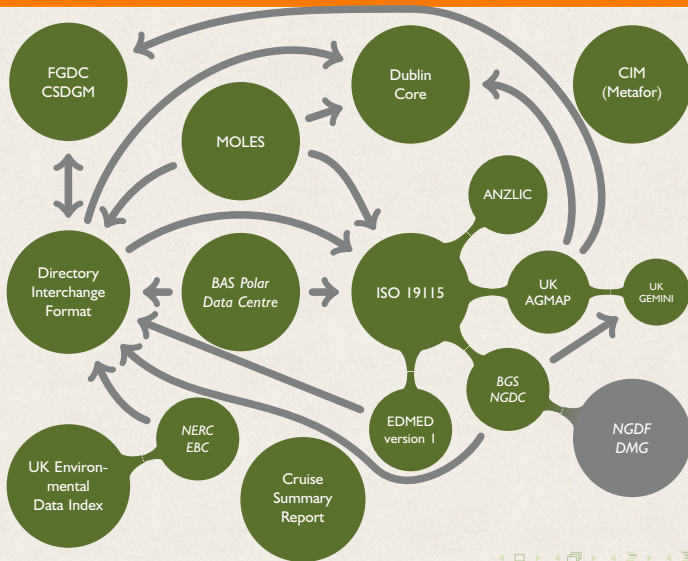
- ▶ Author
- ▶ Publication date
- ▶ Title
- ▶ Location (= identifier)

Motivation for a data discovery service

Needed to realise some benefits of data reuse:

- ▶ Maximum return on funder investment
- ▶ Reduced duplication of effort
- ▶ Reduced costs of data collection
- ▶ Broader scope of possible research

Geospatial/Environmental Data



Search

[Search](#)

[Clear Search](#)

Type

[Dataset](#) (5180)
[Simulation](#)
[Metadata](#) (3470)
[Software](#) (87)

Project

[CCSM](#) (2752)
[CMIP5](#) (1617)
[GeoMIP](#) (8)
[NARCCAP](#) (116)
[NCL](#) (47)
[...Show More](#)

Institute

[NCAR](#) (1461)
[NSF-DOE-](#)
[NCAR](#) (356)

Model

[CCSM3](#) (1807)
[CCSM4](#) (1486)
[CESM1-BGC](#) (180)
[CESM1-CAM5](#) (83)
[CESM1-](#)
[FASTCHEM](#) (36)
[...Show More](#)

Experiment

[1pctCO2](#) (52)
[B04.10](#) (3)
[B04.16](#) (4)
[B04.19](#) (4)
[B04.23](#) (4)
[...Show More](#)

Frequency

[3-Hourly](#) (203)
[6-Hourly](#) (10)
[Daily](#) (82)
[Fixed \(time-](#)
[independent\)](#) (136)
[Monthly](#) (1368)
[...Show More](#)

Product

[forcing](#) (4)
[hist](#) (326)
[other](#) (1)
[output1](#) (1372)
[output2](#) (245)
[...Show More](#)

Realm

[Aerosol](#) (199)
[Atmosphere](#) (484)
[Atmospheric](#)
[chemistry](#) (111)
[Climate](#) (137)
[Earth system](#) (565)
[...Show More](#)

CF Variable

[10m wind speed](#) (5)
[2m air](#)
[temperature](#) (6)
[2m reference](#)
[specific humidity](#) (6)
[2m reference](#)
[temperature](#) (6)
[2m relative](#)
[humidity](#) (6)
[...Show More](#)

Variable Name

[ADVS](#) (15)
[ADVS_ISOP](#) (8)
[ADVS_SUBM](#) (8)
[ADVT](#) (15)
[ADVT_ISOP](#) (8)
[...Show More](#)

Ensemble

[r0i0p0](#) (136)
[r10i1p1](#) (27)
[r10i2p1](#) (54)
[r11i1p1](#) (12)
[r12i1p1](#) (12)
[...Show More](#)

Physical Domain

[Atmosphere](#) (794)
[Earth system](#) (138)
[Ice](#) (481)
[Land](#) (652)
[Ocean](#) (763)

[Home](#) | [Search](#) | [Data](#) | [Account](#) | [About](#) | [Contact Us](#) | [Login](#)

Website powered by Science Gateway Framework (version 2.0.5-SNAPSHOT-20120702-010127)

[Privacy Policy](#) | [Terms of Use](#)

About the DCS

Search for data

Results

You are searching for...

documents containing
geoscientific information (in
topic category)

**443 results returned in 0.63
seconds.**

[Return to your results](#)

[Edit this search](#)

Bookmark this
record with:

Download

The metadata is available for
downloading in the following
formats:

[NERC format \(CSV\)](#)

[NERC format \(XML version 0.7\)](#)

This metadata complies with
INSPIRE and GEMINI 2 standards.

Details

 [Provide feedback](#)

Title: Chemistry of iron in freshwaters of Northwest England

Abstract: Chemical composition of freshwater samples from sites in Northern England. Measurements of pH, dissolved major ions (Na, Mg, K, Ca, Cl, NO₃, SO₄), dissolved organic carbon (DOC), dissolved Al, Fe(II) and total Fe, and measurements of Al, Fe(II) and total Fe on samples following dialysis. A view service identifying sample locations is also available

Temporal coverage: These data span the period from **Thursday 25 September 2003 to Thursday 26 February 2004** inclusive.

Resource locator:

information

[Further Information](#) - Preprint of peer-reviewed paper describing the results. Contains details of all analytical methods used.

miscellaneous

[Map View Service](#) - This resource has a related map view service

[Documents](#) - Documents available to assist with re-use of this dataset.

[Data download](#) - Download this data

Conditions for access and use constraints: CEH must be acknowledged in all resultant publications

Responsible party - Distributor

Centre for Ecology & Hydrology

Individual name: Steve Lofts

Address: Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, OX10 8BB, United Kingdom

Email: [enquiries\(at\)ceh\(dot\)ac\(dot\)uk](mailto:enquiries(at)ceh(dot)ac(dot)uk)

Geographic extent



The metadata covers the area West - East from -4.1158° to -1.4383° and North - South from 54.7783° to 53.6226°. [Search](#) for all metadata within this area.

 [KML](#)

RESEARCH DATA
AUSTRALIA

SEARCH

Advanced Search

Home Australian Ocean Data Network Collection RAN XBT Temperature Profile D...

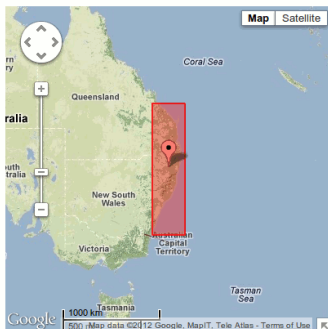


RAN XBT Temperature Profile Data - HMAS PERTH

ProjectID: 98004PE From: 1998-01-27 To: 1998-03-12

This dataset contains temperature profile data recorded using expendable BaThermograph probes (XBT) and a digital recording system. The dataset contains 50 observations obtained during RAN cruise 98004PE (HMAS PERTH). Source file: 98004PE.xml

Coverage:



Time Period:
From 1998 To 1998

Subjects:

ANZSRC

STATISTICS
MATHEMATICAL PHYSICS
NUMERICAL AND
COMPUTATIONAL MATHEMATICS
EARTH SCIENCES
OTHER ENVIRONMENT

Keywords

Oceans | Ocean Temperature |
Water Temperature

Identifiers:

Local: AODN:f311f2bb-3937-499e-8dcb-8f3d502b9709

<http://services.ands.org.au/home/orca/rda/>

Access

Rights

Rights statement

The citation to be used in publications which use this data should follow the format: RAN HYDROGRAPHY AND METOC BRANCH, <http://www.metoc.gov.au> accessed at dd-MMM-YYYY

Rights statement

Attribution 2.5 Australia

Rights statement

While every effort has been made to ensure the accuracy of

Connections

Researchers

- Data Management Officer

ANDS Suggested Links

10628 Collections with matching subjects

Common data discovery metadata elements

Identification

- ▶ Dataset Name (15)
- ▶ Dataset Version (4)
- ▶ Dataset Date (13)
- ▶ Dataset Identifier (12)
- ▶ Metadata Scheme Name (7)
- ▶ Metadata Scheme Version (5)
- ▶ Metadata Record Date (10)
- ▶ Metadata Record Identifier (3)

Responsibility

- ▶ Project/Study/Series Name (9)
- ▶ Project/Study/Series Status (4)
- ▶ Rights/Restrictions (14)
- ▶ Agent (15)
- ▶ Agent Contact Details (11)

Common data discovery metadata elements

Archiving

- ▶ Location (15)
- ▶ File Format(s) (10)
- ▶ Storage Medium (6)
- ▶ Size (7)
- ▶ Data Quality Information (5)
- ▶ Data Preview (4)
- ▶ Dataset Language (7)
- ▶ Dataset Status (8)

Spatiotemporal Coverage

- ▶ Spatial Extent (12)
- ▶ Spatial Resolution (7)
- ▶ Temporal Extent (15)
- ▶ Temporal Resolution (5)

Common data discovery metadata elements

Topical Coverage and Derivation

- ▶ Dataset Type (12)
- ▶ Subject/Keywords (13)
- ▶ Abstract/Summary/Description (14)
- ▶ Parameters Used (6)
- ▶ Methodology/Instrumentation (8)
- ▶ Processing Steps (6)
- ▶ Related Datasets (11)
- ▶ Derived Publications (11)

Source: Ball, A. (2009). *Scientific Data Application Profile Scoping Study Report*. University of Bath, UKOLN.

<http://www.ukoln.ac.uk/projects/sdapss/>

Most common data discovery metadata elements

- ▶ Dataset Name (15)
- ▶ Dataset Date (13)
- ▶ Dataset Identifier (12)
- ▶ Metadata Record Date (10)
- ▶ Rights/Restrictions (14)
- ▶ Agent (15)
- ▶ Agent Contact Details (11)
- ▶ Location (15)
- ▶ File Format(s) (10)
- ▶ Spatial Extent (12)
- ▶ Temporal Extent (15)
- ▶ Dataset Type (12)
- ▶ Subject/Keywords (13)
- ▶ Abstract/Summary/Description (14)
- ▶ Related Datasets (11)
- ▶ Derived Publications (11)

Most common data discovery metadata elements

- ▶ **Dataset Name** (15)
- ▶ **Dataset Date** (13)
- ▶ **Dataset Identifier** (12)
- ▶ **Metadata Record Date** (10)
- ▶ **Rights/Restrictions** (14)
- ▶ **Agent** (15)
- ▶ **Agent Contact Details** (11)
- ▶ **Location** (15)
- ▶ **File Format(s)** (10)
- ▶ **Spatial Extent** (12)
- ▶ **Temporal Extent** (15)
- ▶ **Dataset Type** (12)
- ▶ **Subject/Keywords** (13)
- ▶ **Abstract/Summary/Description** (14)
- ▶ **Related Datasets** (11)
- ▶ **Derived Publications** (11)

Most common data discovery metadata elements

- ▶ **Dataset Name** (15)
- ▶ Dataset Date (13)
- ▶ Dataset Identifier (12)
- ▶ Metadata Record Date (10)
- ▶ Rights/Restrictions (14)
- ▶ **Agent** (15)
- ▶ Agent Contact Details (11)
- ▶ **Location** (15)
- ▶ File Format(s) (10)
- ▶ Spatial Extent (12)
- ▶ **Temporal Extent** (15)
- ▶ Dataset Type (12)
- ▶ Subject/Keywords (13)
- ▶ Abstract/Summary/Description (14)
- ▶ Related Datasets (11)
- ▶ Derived Publications (11)

Most common data discovery metadata elements

- ▶ **Dataset Name** (15)
- ▶ **Dataset Date** (13)
- ▶ **Dataset Identifier** (12)
- ▶ **Metadata Record Date** (10)
- ▶ **Rights/Restrictions** (14)
- ▶ **Agent**¹ (15)
- ▶ **Agent Contact Details** (11)
- ▶ **Location**² (15)
- ▶ **File Format(s)** (10)
- ▶ **Spatial Extent** (12)
- ▶ **Temporal Extent** (15)
- ▶ **Dataset Type** (12)
- ▶ **Subject/Keywords** (13)
- ▶ **Abstract/Summary/Description** (14)
- ▶ **Related Datasets** (11)
- ▶ **Derived Publications** (11)

1. DataCite uses Creator, Publisher, Contributor
2. DataCite holds this separately
3. DataCite also has Dataset Version (4), Dataset Language (7), Size (7)

Attributing datasets to many contributors

giardine.etal2011sda-suppl.xls - LibreOffice Calc

File Edit View Insert Format Tools Data Window Help

Calibri 11 B I U

Researcher ID <http://dx.doi.org/10.1038/ng.785>

#	A	B	C	D	E	F	G	H	I	J	K
	#dbID	HGVS name	dbSNP ss#	dbSNP rs#	OMIM	Swiss-Prot	Researcher ID	PMID	Common name		
5	ALOX5AP_00001	NM_001629.2:c.323+154T>C	-	rs4468448	-	-	Hb Var (A-2391-2010)	17918249	-		
6	ALOX5AP_00002	NM_001629.2:c.323+3269T>C	-	rs4769058	-	-	Hb Var (A-2391-2010)	17918249	-		
8	ALOX5AP_00003	NM_001629.2:c.324-3699A>G	-	rs9508834	-	-	Hb Var (A-2391-2010)	17918249	-		
9	AQP9_00001	NM_020980.3:c.835A>G	-	rs1867380	-	-	Hb Var (A-2391-2010)	17918249	-		
10	ARG2_00002	NM_001172.3:c.860-11T>C	-	rs10483802	-	-	Hb Var (A-2391-2010)	18275000	-		
11	ARG2_00001	NM_001172.3:c.860-426C>A	-	rs10483801	-	-	Hb Var (A-2391-2010)	18275000	-		
12	ASS1_00001	NM_000050.4:c.495+1473C>T	-	rs590086	-	-	Hb Var (A-2391-2010)	17918249	-		
13	ASS1_00002	NM_000050.4:c.597+18A>G	-	rs652313	-	-	Hb Var (A-2391-2010)	17918249	-		
14	ASS1_00003	NM_000050.4:c.838+2190A>G	-	rs12555797	-	-	Hb Var (A-2391-2010)	17918249	-		
15	ASS1_00004	NM_000050.4:c.839-88A>T	-	rs543048	-	-	Hb Var (A-2391-2010)	17918249	-		
16	ATRX_00086	NM_000489.3:c.20+1G>A	-	-	-	-	Hb Var (A-2391-2010)	12858175	-		
17	ATRX_00003	NM_000489.3:c.109C>T	-	-	-	-	Hb Var (A-2391-2010)	10632111	-		
18	ATRX_00004	NM_000489.3:c.187G>T	-	-	-	-	Hb Var (A-2391-2010)	18409179	-		
19	ATRX_00005	NM_000489.3:c.236C>G	-	-	-	-	Hb Var (A-2391-2010)	12858175	-		
20	ATRX_00087	NM_000489.3:c.242+2T>C	-	-	-	-	Hb Var (A-2391-2010)	16266892	-		
21	ATRX_00088	NM_000489.3:c.370G>T	-	-	-	-	Hb Var (A-2391-2010)	16376512	-		
22	ATRX_00089	NM_000489.3:c.390_391insA	-	-	-	-	Hb Var (A-2391-2010)	18409179	-		
23	ATRX_00090	NM_000489.3:c.413_414delIAA	-	-	-	-	Hb Var (A-2391-2010)	11449489	-		
24	ATRX_00006	NM_000489.3:c.521G>A	-	-	-	-	Hb Var (A-2391-2010)	19055664	-		
25	ATRX_00007	NM_000489.3:c.524G>A	-	-	-	-	Hb Var (A-2391-2010)	10204841	-		
26	ATRX_00091	NM_000489.3:c.528_529insCAA	-	-	-	-	Hb Var (A-2391-2010)	9326931	-		
27	ATRX_00093	NM_000489.3:c.536A>G	-	-	-	-	Hb Var (A-2391-2010)	8968741	-		
28	ATRX_00008	NM_000489.3:c.565C>G	-	-	-	-	Hb Var (A-2391-2010)	16813605	-		
29	ATRX_00009	NM_000489.3:c.568C>G	-	-	-	-	Hb Var (A-2391-2010)	9326931	-		
30	ATRX_00010	NM_000489.3:c.568C>T	-	-	-	-	Hb Var (A-2391-2010)	11449489	-		
31	ATRX_00011	NM_000489.3:c.569C>T	-	-	-	-	Hb Var (A-2391-2010)	16763962	-		
32	ATRX_00012	NM_000489.3:c.576G>C	-	-	-	-	Hb Var (A-2391-2010)	9326931	-		
33	ATRX_00013	NM_000489.3:c.576G>C	-	-	-	-	Hb Var (A-2391-2010)	14592816	-		

Variant Submission Information Microattribution Information Phenotype Information Variant Frequency Information

Sheet 2 / 4 PageStyle_Microattribution Information STD Sum=0 103%

Dynamic datasets

Two types:

- ▶ Revised datasets

- ▶ Expanding datasets

Dynamic datasets

Three strategies:

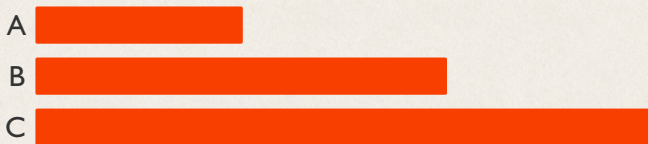
1. Differentiate versions by access date rather than ID



2. Take time slices



3. Take snapshots



Recommendations: metadata for citation

▶ **Author**

- ▶ Record roles, identifiers and contact details as well as names.

▶ **Publication date**

▶ **Title**

- ▶ It helps to avoid confusion if this is different from the article title.

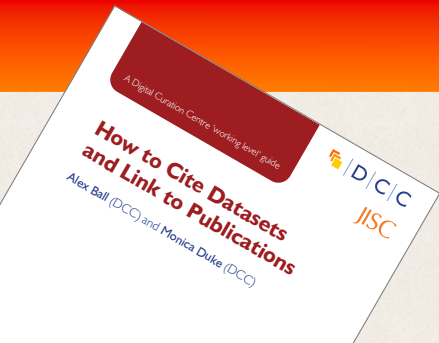
▶ **Location/Identifier**

- ▶ Express identifiers in location (`http`) form if possible.
- ▶ Use DOIs for 'published' data and Handles, ARKs, PURLs, etc. for other data.
- ▶ Use different DOIs for different versions.
- ▶ Location given should be a catalogue record/landing page for the data.
- ▶ Giving the **Publisher** or **Host Archive** as well provides some recourse if the link breaks.

Recommendations: metadata for discovery

In addition to the citation metadata, these should also be recorded where they apply:

- ▶ **Contributors**
- ▶ **Abstract/Summary/Description**
- ▶ **Subject/Keywords**
- ▶ **Rights/Restrictions**
- ▶ **Spatial Coverage**
- ▶ **Temporal Coverage**
- ▶ **Derived Publications**
- ▶ **Related Datasets**
- ▶ **Resource Type**
- ▶ **File Format(s)**
- ▶ **Important Dates**
Creation, Submission, Acceptance, Use By...
- ▶ **Language***
- ▶ **Version***
- ▶ **Size***
- ▶ **Metadata Record Date**



<http://www.dcc.ac.uk/resources/how-guides/>



Thank you for your attention

DCC Website: <http://www.dcc.ac.uk/>

Alex Ball: <http://www.ukoln.ac.uk/ukoln/staff/a.ball/>

8th International Digital Curation Conference

“Infrastructure, Intelligence, Innovation: driving the Data Science agenda”

14–16 January 2013, Amsterdam

<http://www.dcc.ac.uk/events/idcc13>