Connecting data repositories and publishers for data publication

Sarah Callaghan, Fiona Murphy, Jonathan Tedds, Rob Allan, John Kunze, Rebecca Lawrence, Matthew S. Mayernik, Angus Whyte, and the PREPARDE project team

#preparde

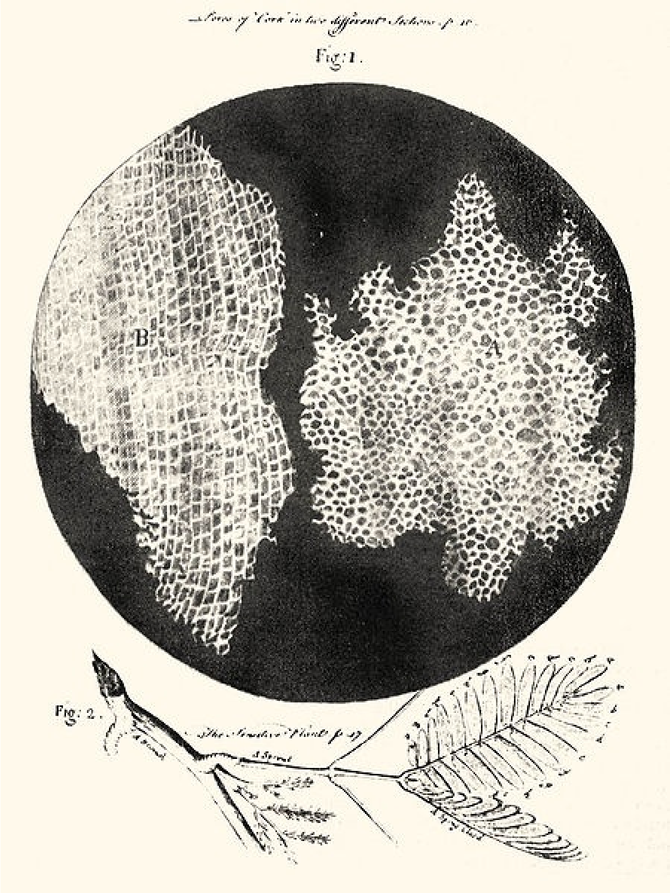
sarah.callaghan@stfc.ac.uk @sorcha_ni

Why link data and publications?

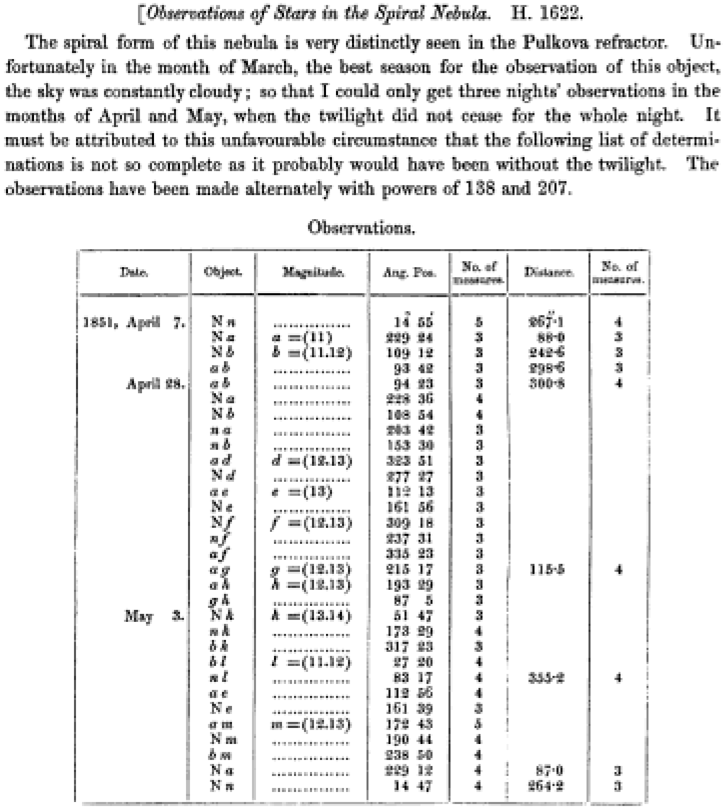
- Data is the foundation of science – without it we can't test our assertions or reproduce our results
- The Internet allows us to link things to other things quickly and easily
- But there are still serious problems to address when it comes to linking data to the scientific record:
  - Data persistence
  - Data and metadata quality
  - Attribution and credit for data producers
  - … and many more

Engraving of printer using the early Gutenberg letter press during the 15th century.
Date unknown - estimate 16th - 19th century
http://commons.wikimedia.org/wiki/File:Gutenberg_press.jpg

Historically, journals have always published data


Pores of Cork in two different Sections p. 10
Fig: 1.
B
A
Fig: 2.
The Sensitive Plant p. 47
A Sprout

Suber cells and mimosa leaves. Robert Hooke, Micrographia, 1665



[*Observations of Stars in the Spiral Nebula.* H. 1622.

The spiral form of this nebula is very distinctly seen in the Pulkova refractor. Unfortunately in the month of March, the best season for the observation of this object, the sky was constantly cloudy; so that I could only get three nights' observations in the months of April and May, when the twilight did not cease for the whole night. It must be attributed to this unfavourable circumstance that the following list of determinations is not so complete as it probably would have been without the twilight. The observations have been made alternately with powers of 138 and 207.

Observations.

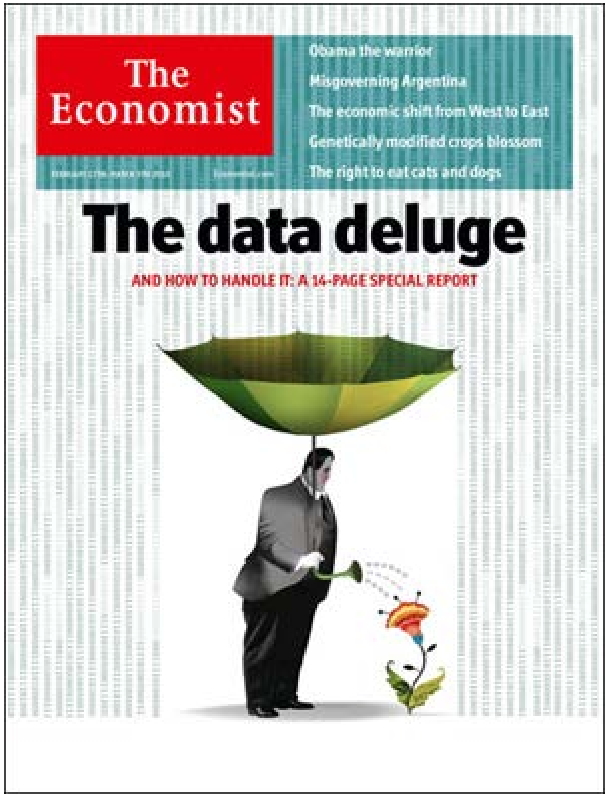
Date.	Object.	Magnitude.	Ang. Pos.	No. of measures.	Distance.	No. of measures.
1851, April 7.	N *n*	....	14° 55′	5	267″·1	4
	N *a*	*a* =(11)	229 24	3	88·0	3
	N *b*	*b* =(11.12)	109 12	3	242·6	3
	a b	....	93 42	3	298·6	3
April 28.	*a b*	....	94 23	3	300·8	4
	N *a*	....	228 36	4		
	N *b*	....	108 54	4		
	n a	....	203 42	3		
	n b	....	153 30	3		
	a d	*d* =(12.13)	323 51	3		
	N *d*	....	277 27	3		
	a e	*e* =(13)	112 13	3		
	N *e*	....	161 56	3		
	N *f*	*f* =(12.13)	309 18	3		
	n f	....	237 31	3		
	a f	....	335 23	3		
	a g	*g* =(12.13)	215 17	3	115·5	4
	a h	*h* =(12.13)	193 29	3		
	g h	....	87 5	3		
May 3.	N *h*	*h* =(13.14)	51 47	3		
	n h	....	173 29	4		
	b h	....	317 23	3		
	b l	*l* =(11.12)	27 20	4		
	n l	....	83 17	4	335·2	4
	a e	....	112 56	4		
	N *e*	....	161 39	3		
	a m	*m* =(12.13)	172 43	5		
	N *m*	....	190 44	4		
	b m	....	238 50	4		
	N *a*	....	229 12	4	87·0	3
	N *n*	....	14 47	4	264·2	3

The Scientific Papers of William Parsons, Third Earl of Rosse 1800-1867

JISC University of Leicester British Atmospheric Data Centre NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE NATURAL ENVIRONMENT RESEARCH COUNCIL WILEY-BLACKWELL University of Reading DCC University of California CDL California Digital Library F1000 FACULTY of 1000 POST-PUBLICATION PEER REVIEW NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

But now… the Data Deluge

“the amount of data generated worldwide...is growing by 58% per year; in 2010 the world generated 1250 billion gigabytes of data”



The Economist

Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
The right to eat cats and dogs

The data deluge

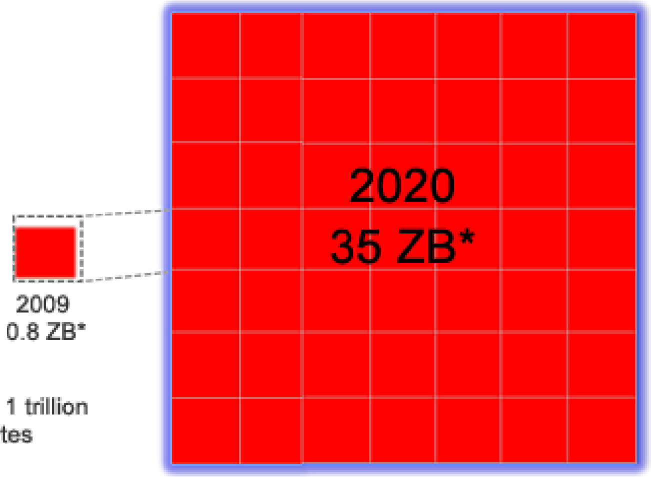
AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

A lot of people are creating a lot of data, and we’re only going to get more of it.

If this is a data deluge – time to start building boats!

The Digital Universe Decade – Are You Ready?
IDCC White Paper, May 2010

Figure 1: The Digital Universe 2009 – 2020
Growing by a Factor of 44



*Zettabyte = 1 trillion gigabytes

Source: IDC Digital Universe Study, sponsored by EMC, May 2010

JISC · University of Leicester · British Atmospheric Data Centre, National Centre for Atmospheric Science, Natural Environment Research Council · Wiley-Blackwell · University of Reading · DCC · University of California CDL California Digital Library · F1000 Faculty of 1000 Post-Publication Peer Review · NCAR National Center for Atmospheric Research

Using citations to link research outputs

- We already have a working method for linking between publications which is
  - commonly used
  - understood by the research community
  - used to create metrics to show how much of an impact something has (citation counts)
  - applied to digital objects (digital versions of journal articles)

- We can extend citation to other things like
  - data
  - code
  - multimedia

http://www.flickr.com/photos/anton41/6588935181/

And the best bit is, we don't need to teach researchers a new method of linking – they cite like they normally would!

JISC University of Leicester British Atmospheric Data Centre NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE NATURAL ENVIRONMENT RESEARCH COUNCIL WILEY-BLACKWELL University of Reading DCC University of California CDL California Digital Library F1000 FACULTY of 1000 POST-PUBLICATION PEER REVIEW NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Reasons for citing and publishing data

http://www.evidencebased-management.com/blog/2011/11/04/new-evidence-on-big-bonuses/

- Pressure from (UK) government to make data from publicly funded research available for free.
  - Scientists want attribution and credit for their work
  - Public want to know what the scientists are doing

- Research funders want reassurance that they're getting value for money
  - Relies on peer-review of science publications (well established) and data (not done yet!)

- Allows the wider research community to find and use datasets, and understand the quality of the data

- Extra incentive for scientists to submit their data to data centres in appropriate formats and with full metadata

PREPARDE: Peer REview for Publication & Accreditation of Research Data in the Earth sciences

- Lead Institution: University of Leicester
- Partners
  - British Atmospheric Data Centre (BADC)
  - US National Centre for Atmospheric Research (NCAR)
  - California Digital Library (CDL)
  - Digital Curation Centre (DCC)
  - University of Reading
  - Wiley-Blackwell
  - Faculty of 1000 Ltd
- Project Lead: Dr Jonathan Tedds (University of Leicester, jat26@le.ac.uk)
- Project Manager: Dr Sarah Callaghan (BADC, sarah.callaghan@stfc.ac.uk)
- Length of Project: 12 months
- Project Start Date: 1st July 2012
- Project End Date: 31st June 2013

Geoscience Data Journal, Wiley-Blackwell and the Royal Meteorological Society

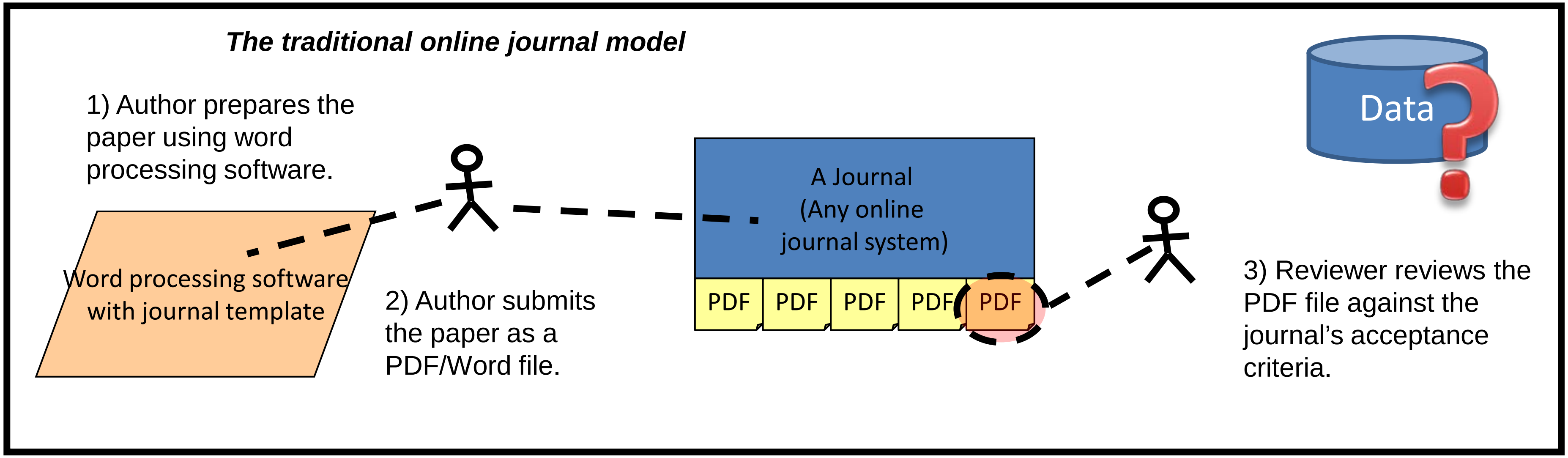
- Partnership formed between Royal Meteorological Society and academic publishers Wiley Blackwell to develop a mechanism for the formal publication of data in the Open Access *Geoscience Data Journal*

- GDJ publishes short data articles cross-linked to, and citing, datasets that have been deposited in approved data centres and awarded DOIs (or other permanent identifier).

- A data article describes a dataset, giving details of its collection, processing, software, file formats, etc., without the requirement of novel analyses or ground breaking conclusions.
    - the when, how and why data was collected and what the data-product is.

RMetS
Royal Meteorological Society

WILEY-BLACKWELL



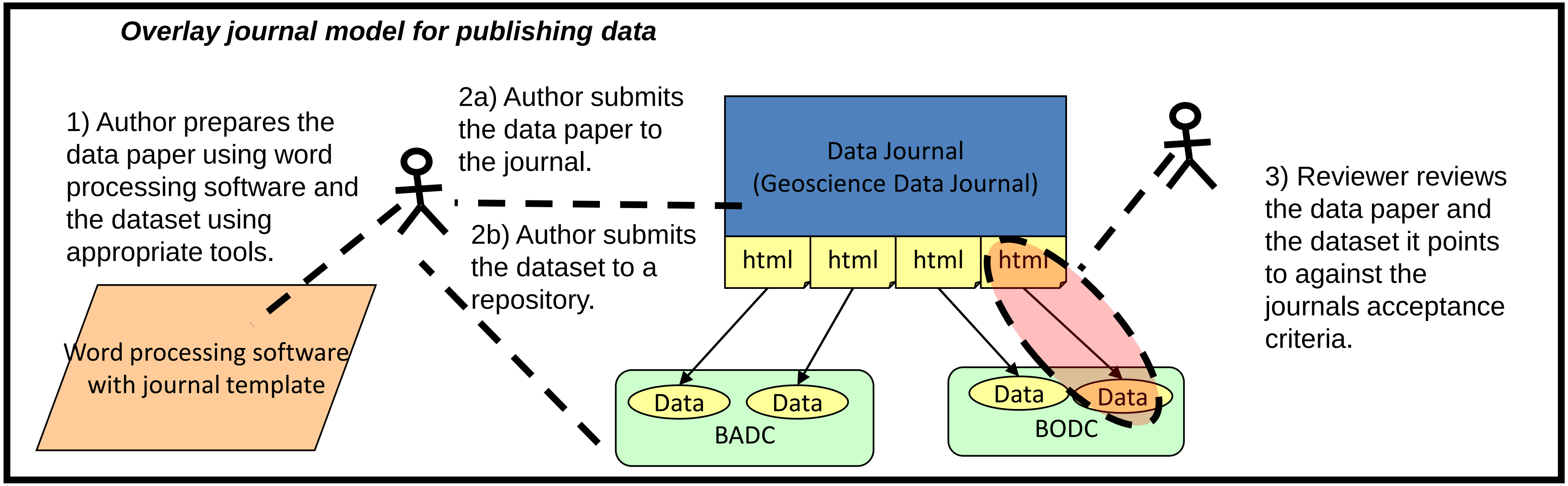
The traditional online journal model

1) Author prepares the paper using word processing software.

2) Author submits the paper as a PDF/Word file.

3) Reviewer reviews the PDF file against the journal's acceptance criteria.

Word processing software with journal template
A Journal (Any online journal system)
PDF
PDF
PDF
PDF
PDF
Data



Overlay journal model for publishing data

1) Author prepares the data paper using word processing software and the dataset using appropriate tools.

2a) Author submits the data paper to the journal.

2b) Author submits the dataset to a repository.

3) Reviewer reviews the data paper and the dataset it points to against the journals acceptance criteria.

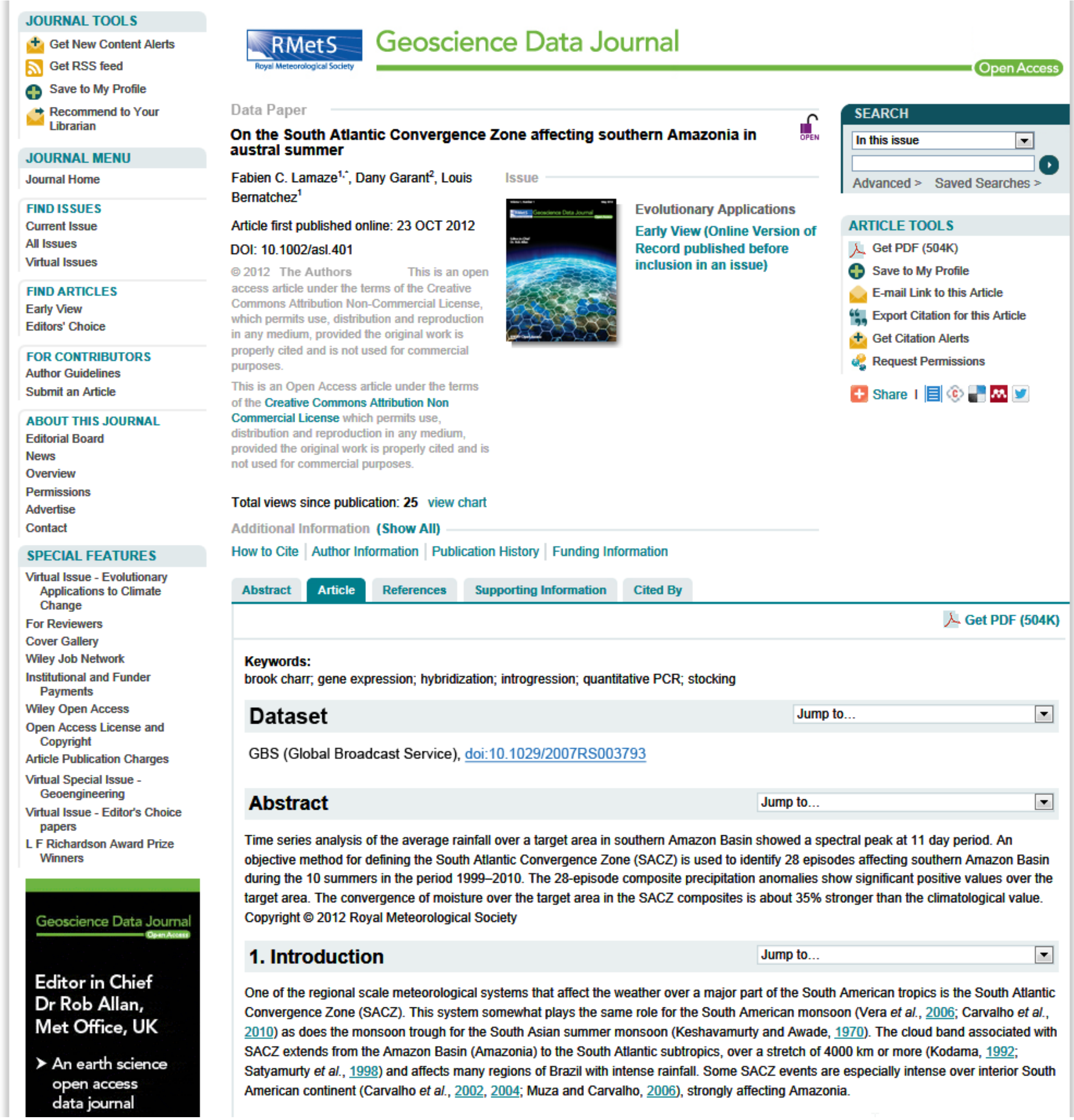
Word processing software with journal template
Data Journal (Geoscience Data Journal)
html
html
html
html
Data
Data
BADC
Data
Data
BODC

JISC University of Leicester British Atmospheric Data Centre NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE NATURAL ENVIRONMENT RESEARCH COUNCIL WILEY-BLACKWELL University of Reading DCC University of California CDL California Digital Library

F1000 FACULTY of 1000 POST-PUBLICATION PEER REVIEW NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Data paper mock-up

Dataset citation is first thing in the paper and is also included in reference list (to take advantage of citation count systems)



JOURNAL TOOLS
- Get New Content Alerts
- Get RSS feed
- Save to My Profile
- Recommend to Your Librarian

JOURNAL MENU
- Journal Home

FIND ISSUES
- Current Issue
- All Issues
- Virtual Issues

FIND ARTICLES
- Early View
- Editors' Choice

FOR CONTRIBUTORS
- Author Guidelines
- Submit an Article

ABOUT THIS JOURNAL
- Editorial Board
- News
- Overview
- Permissions
- Advertise
- Contact

SPECIAL FEATURES
- Virtual Issue - Evolutionary Applications to Climate Change
- For Reviewers
- Cover Gallery
- Wiley Job Network
- Institutional and Funder Payments
- Wiley Open Access
- Open Access License and Copyright
- Article Publication Charges
- Virtual Special Issue - Geoengineering
- Virtual Issue - Editor's Choice papers
- L F Richardson Award Prize Winners

Geoscience Data Journal
Open Access
Editor in Chief
Dr Rob Allan,
Met Office, UK
> An earth science open access data journal

RMetS Royal Meteorological Society

Geoscience Data Journal

Open Access

Data Paper

OPEN

On the South Atlantic Convergence Zone affecting southern Amazonia in austral summer

Fabien C. Lamaze[1,*], Dany Garant[2], Louis Bernatchez[1]

Article first published online: 23 OCT 2012

DOI: 10.1002/asl.401

© 2012 The Authors This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

This is an Open Access article under the terms of the Creative Commons Attribution Non Commercial License which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

Issue

Evolutionary Applications

Early View (Online Version of Record published before inclusion in an issue)

SEARCH

In this issue

Advanced > Saved Searches >

ARTICLE TOOLS
- Get PDF (504K)
- Save to My Profile
- E-mail Link to this Article
- Export Citation for this Article
- Get Citation Alerts
- Request Permissions

Share |

Total views since publication: **25** view chart

Additional Information (Show All)

How to Cite | Author Information | Publication History | Funding Information

Abstract | Article | References | Supporting Information | Cited By

Get PDF (504K)

Keywords:
brook charr; gene expression; hybridization; introgression; quantitative PCR; stocking

Dataset

Jump to...

GBS (Global Broadcast Service), doi:10.1029/2007RS003793

Abstract

Jump to...

Time series analysis of the average rainfall over a target area in southern Amazon Basin showed a spectral peak at 11 day period. An objective method for defining the South Atlantic Convergence Zone (SACZ) is used to identify 28 episodes affecting southern Amazon Basin during the 10 summers in the period 1999–2010. The 28-episode composite precipitation anomalies show significant positive values over the target area. The convergence of moisture over the target area in the SACZ composites is about 35% stronger than the climatological value. Copyright © 2012 Royal Meteorological Society

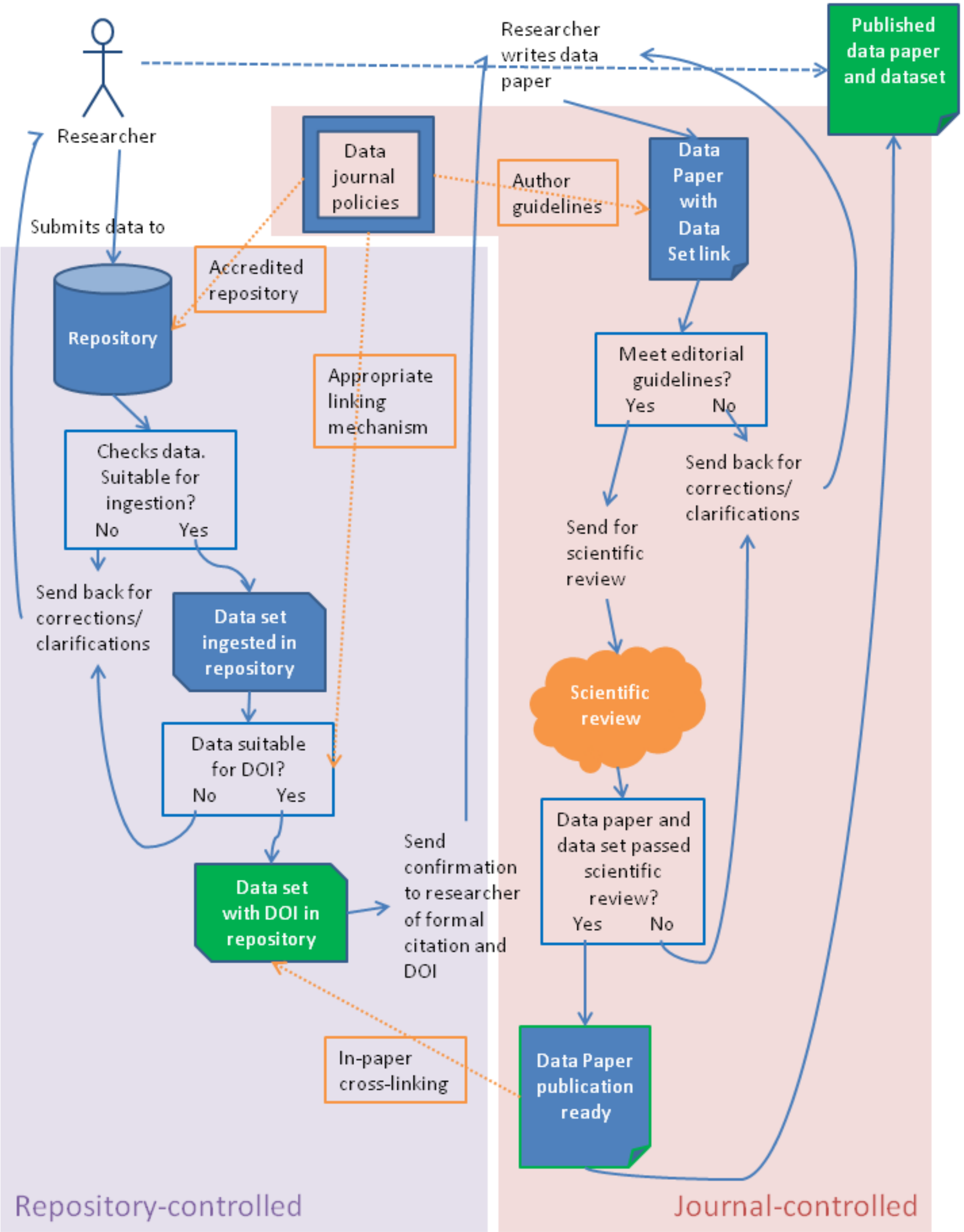
1. Introduction

Jump to...

One of the regional scale meteorological systems that affect the weather over a major part of the South American tropics is the South Atlantic Convergence Zone (SACZ). This system somewhat plays the same role for the South American monsoon (Vera *et al.*, 2006; Carvalho *et al.*, 2010) as does the monsoon trough for the South Asian summer monsoon (Keshavamurty and Awade, 1970). The cloud band associated with SACZ extends from the Amazon Basin (Amazonia) to the South Atlantic subtropics, over a stretch of 4000 km or more (Kodama, 1992; Satyamurty *et al.*, 1998) and affects many regions of Brazil with intense rainfall. Some SACZ events are especially intense over interior South American continent (Carvalho *et al.*, 2002, 2004; Muza and Carvalho, 2006), strongly affecting Amazonia.

JISC University *of* Leicester British Atmospheric Data Centre NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE NATURAL ENVIRONMENT RESEARCH COUNCIL WILEY-BLACKWELL University of Reading DCC University of California CDL California Digital Library

F1000 FACULTY of 1000 POST-PUBLICATION PEER REVIEW NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH


Researcher
Researcher writes data paper
Published data paper and dataset
Submits data to
Data journal policies
Author guidelines
Data Paper with Data Set link
Accredited repository
Repository
Appropriate linking mechanism
Meet editorial guidelines?
Yes
No
Send back for corrections/ clarifications
Checks data. Suitable for ingestion?
No
Yes
Send for scientific review
Send back for corrections/ clarifications
Data set ingested in repository
Scientific review
Data suitable for DOI?
No
Yes
Send confirmation to researcher of formal citation and DOI
Data paper and data set passed scientific review?
Yes
No
Data set with DOI in repository
In-paper cross-linking
Data Paper publication ready
Repository-controlled
Journal-controlled

PREPARDE topics

Example steps/workflow required for a researcher to publish a data paper

3 main areas of interest (in orange)

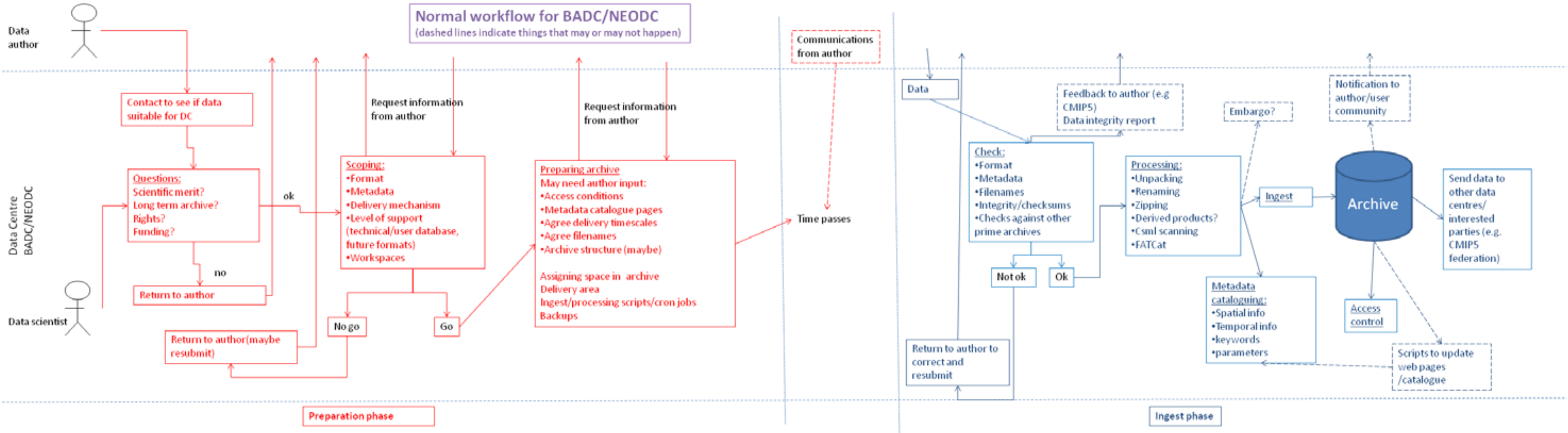
1. Workflows and cross-linking between journal and repository
2. Repository accreditation
3. Scientific peer-review of data

- Division of area of responsibilities between
  - *repository controlled* processes
  - *journal controlled* processes

JISC University of Leicester British Atmospheric Data Centre NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE NATURAL ENVIRONMENT RESEARCH COUNCIL WILEY-BLACKWELL University of Reading DCC University of California CDL California Digital Library F1000 FACULTY of 1000 POST-PUBLICATION PEER REVIEW NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Data repository workflows

- Data centre and journal workflows captured
  - Workflows are very varied! No one-size fits all method
  - Can have multiple workflows in the same data centre, depending on interactions with external sources (“Engaged submitter”/ “Data dumper” / “Third party requester”)



Normal workflow for BADC/NEODC
(dashed lines indicate things that may or may not happen)

Data author

Data Centre BADC/NEODC

Data scientist

Contact to see if data suitable for DC

Questions:
Scientific merit?
Long term archive?
Rights?
Funding?

ok

no

Return to author

Return to author(maybe resubmit)

Request information from author

Scoping:
•Format
•Metadata
•Delivery mechanism
•Level of support (technical/user database, future formats)
•Workspaces

No go

Go

Request information from author

Preparing archive
May need author input:
•Access conditions
•Metadata catalogue pages
•Agree delivery timescales
•Agree filenames
•Archive structure (maybe)

Assigning space in archive
Delivery area
Ingest/processing scripts/cron jobs
Backups

Preparation phase

Communications from author

Time passes

Data

Feedback to author (e.g CMIP5)
Data integrity report

Check:
•Format
•Metadata
•Filenames
•Integrity/checksums
•Checks against other prime archives

Not ok

Ok

Return to author to correct and resubmit

Processing:
•Unpacking
•Renaming
•Zipping
•Derived products?
•Csml scanning
•FATCat

Embargo?

Ingest

Metadata cataloguing:
•Spatial info
•Temporal info
•keywords
•parameters

Notification to author/user community

Archive

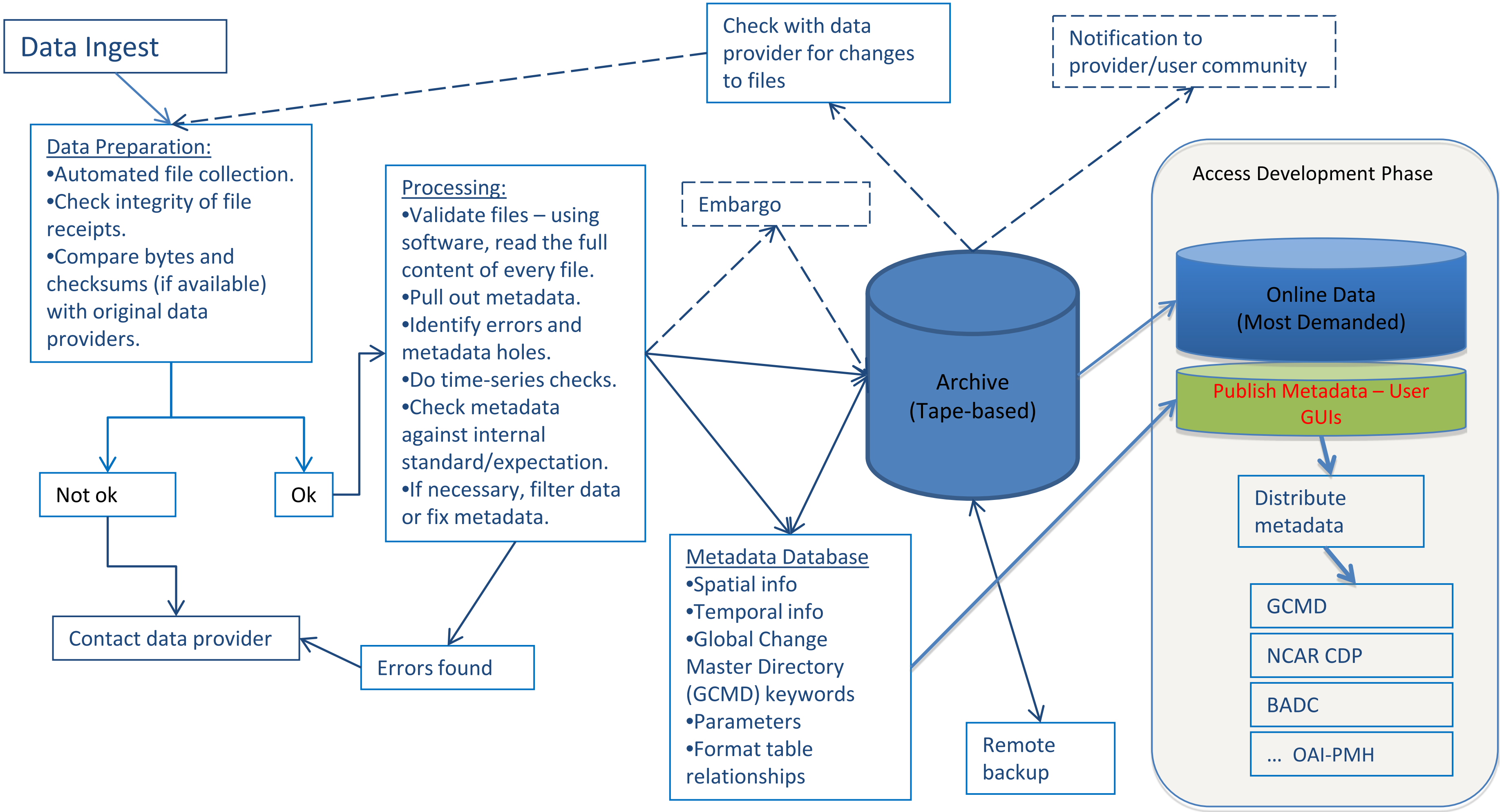
Access control

Scripts to update web pages /catalogue

Send data to other data centres/ interested parties (e.g. CMIP5 federation)

Ingest phase

Repository Workflow – NCAR Comp. & Info. Systems Lab Research Data Archive (RDA)



Data Ingest

Data Preparation:
- Automated file collection.
- Check integrity of file receipts.
- Compare bytes and checksums (if available) with original data providers.

Not ok

Ok

Contact data provider

Processing:
- Validate files – using software, read the full content of every file.
- Pull out metadata.
- Identify errors and metadata holes.
- Do time-series checks.
- Check metadata against internal standard/expectation.
- If necessary, filter data or fix metadata.

Errors found

Check with data provider for changes to files

Notification to provider/user community

Embargo

Archive (Tape-based)

Metadata Database
- Spatial info
- Temporal info
- Global Change Master Directory (GCMD) keywords
- Parameters
- Format table relationships

Remote backup

Access Development Phase

Online Data (Most Demanded)

Publish Metadata – User GUIs

Distribute metadata

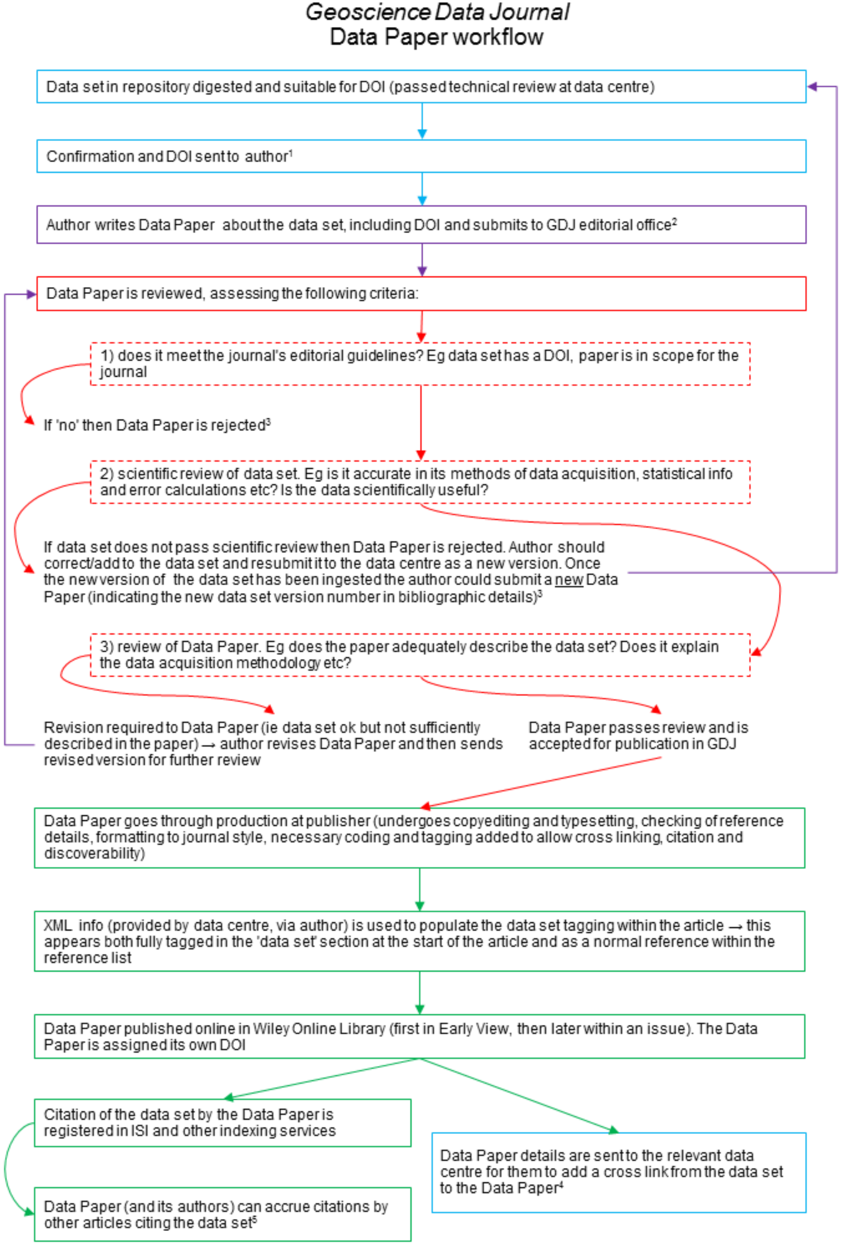
GCMD

NCAR CDP

BADC

… OAI-PMH

JISC University of Leicester British Atmospheric Data Centre NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE NATURAL ENVIRONMENT RESEARCH COUNCIL WILEY-BLACKWELL University of Reading DCC University of California CDL California Digital Library F1000 FACULTY of 1000 POST-PUBLICATION PEER REVIEW NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH



Geoscience Data Journal
Data Paper workflow

Data set in repository digested and suitable for DOI (passed technical review at data centre)

Confirmation and DOI sent to author[1]

Author writes Data Paper about the data set, including DOI and submits to GDJ editorial office[2]

Data Paper is reviewed, assessing the following criteria:

1) does it meet the journal's editorial guidelines? Eg data set has a DOI, paper is in scope for the journal

If 'no' then Data Paper is rejected[3]

2) scientific review of data set. Eg is it accurate in its methods of data acquisition, statistical info and error calculations etc? Is the data scientifically useful?

If data set does not pass scientific review then Data Paper is rejected. Author should correct/add to the data set and resubmit it to the data centre as a new version. Once the new version of the data set has been ingested the author could submit a new Data Paper (indicating the new data set version number in bibliographic details)[3]

3) review of Data Paper. Eg does the paper adequately describe the data set? Does it explain the data acquisition methodology etc?

Revision required to Data Paper (ie data set ok but not sufficiently described in the paper) → author revises Data Paper and then sends revised version for further review

Data Paper passes review and is accepted for publication in GDJ

Data Paper goes through production at publisher (undergoes copyediting and typesetting, checking of reference details, formatting to journal style, necessary coding and tagging added to allow cross linking, citation and discoverability)

XML info (provided by data centre, via author) is used to populate the data set tagging within the article → this appears both fully tagged in the 'data set' section at the start of the article and as a normal reference within the reference list

Data Paper published online in Wiley Online Library (first in Early View, then later within an issue). The Data Paper is assigned its own DOI

Citation of the data set by the Data Paper is registered in ISI and other indexing services

Data Paper (and its authors) can accrue citations by other articles citing the data set[5]

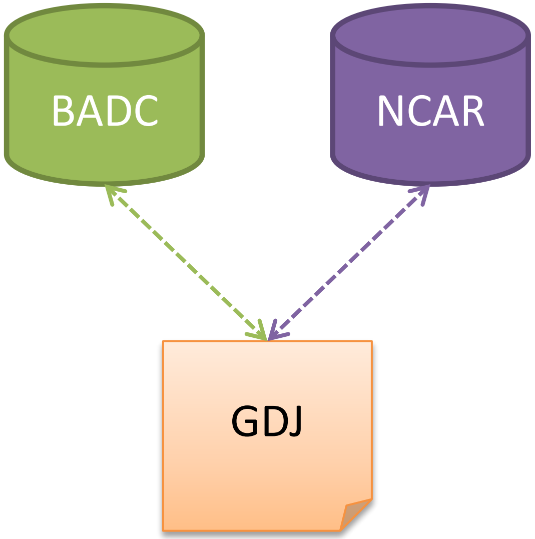
Data Paper details are sent to the relevant data centre for them to add a cross link from the data set to the Data Paper[4]

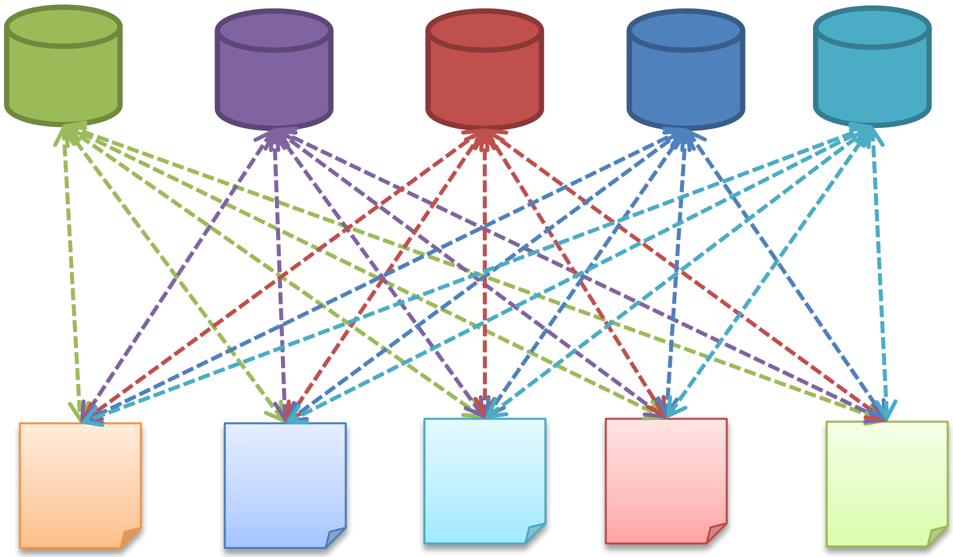
Journal workflow

- Work on comparisons and identification of cross-linking points is continuing.
  - Aim is to minimise effort needed to submit data paper by taking advantage of already submitted metadata.

JISC
University of Leicester
British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL
WILEY-BLACKWELL
University of Reading
DCC
University of California CDL California Digital Library
F1000 FACULTY of 1000 POST-PUBLICATION PEER REVIEW
NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Cross-linking



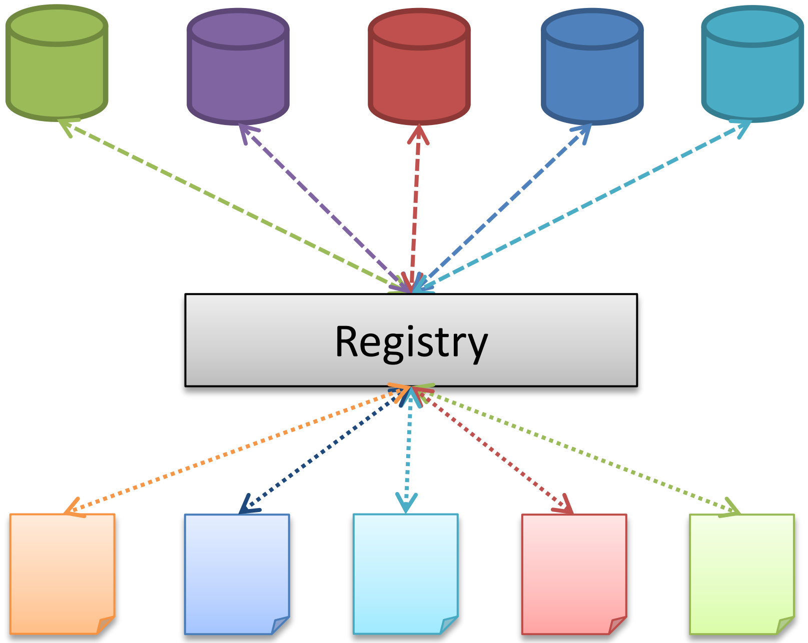


This is what we have to focus on for PREPARDE – demonstrate cross linking between GDJ and BADC (and maybe NCAR)

Unfortunately this direct cross-linking isn't scaleable!

Need for off-the shelf solutions that can work across multiple research domains

Cross-linking – the ideal situation



Registry could provide other functions as well as being an intermediary between journals and data repositories like:

- Certify data centres are “trustworthy”
- Administer linking mechanism
- Provide search and metrics functions

Disadvantages:

- Single point of failure
- Difficulty of standardisation across different research domains

Could OpenAIRE be this registry?

JISC · University of Leicester · British Atmospheric Data Centre, National Centre for Atmospheric Science, Natural Environment Research Council · Wiley-Blackwell · University of Reading · DCC · University of California CDL California Digital Library · F1000 Faculty of 1000 Post-publication peer review · NCAR National Center for Atmospheric Research

Do we have a start?

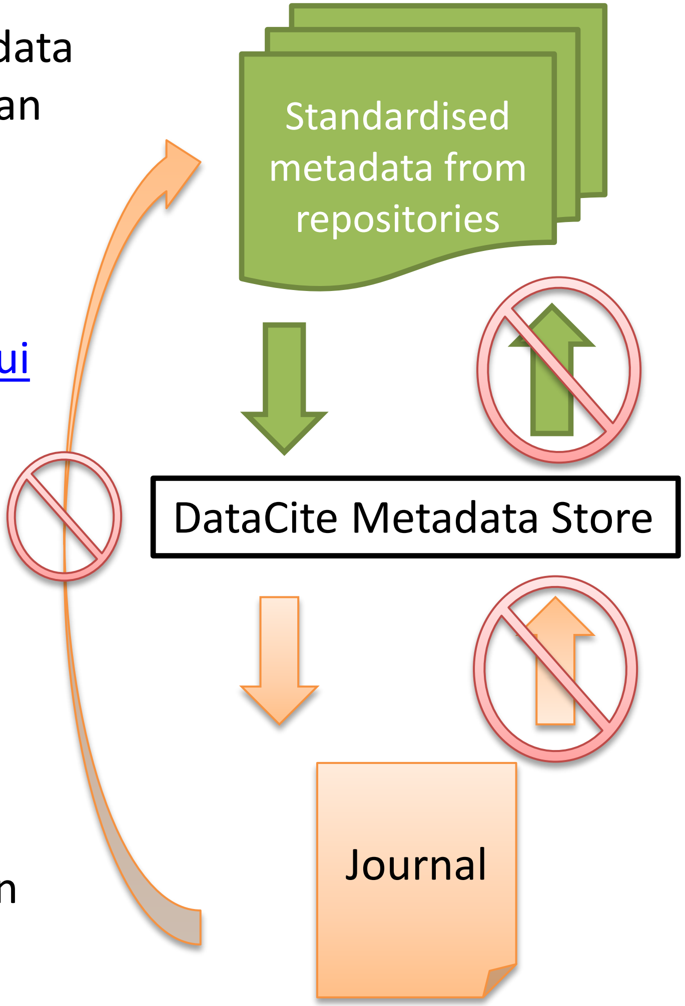
DataCite have standardised a set of bibliometric metadata that have to be submitted before a DOI for a dataset can be minted by a repository.

This metadata is then made openly available via the DataCite metadata search: http://search.datacite.org/ui

Given a DOI, a journal can then easily find the DOI standard metadata.

DataCite also have a content resolver http://data.datacite.org/static/index.html

What's missing is the return link, where the journal can let the repository know that a dataset has been cited (directly or via DataCite)



DataCite Metadata Schema

http://schema.datacite.org/

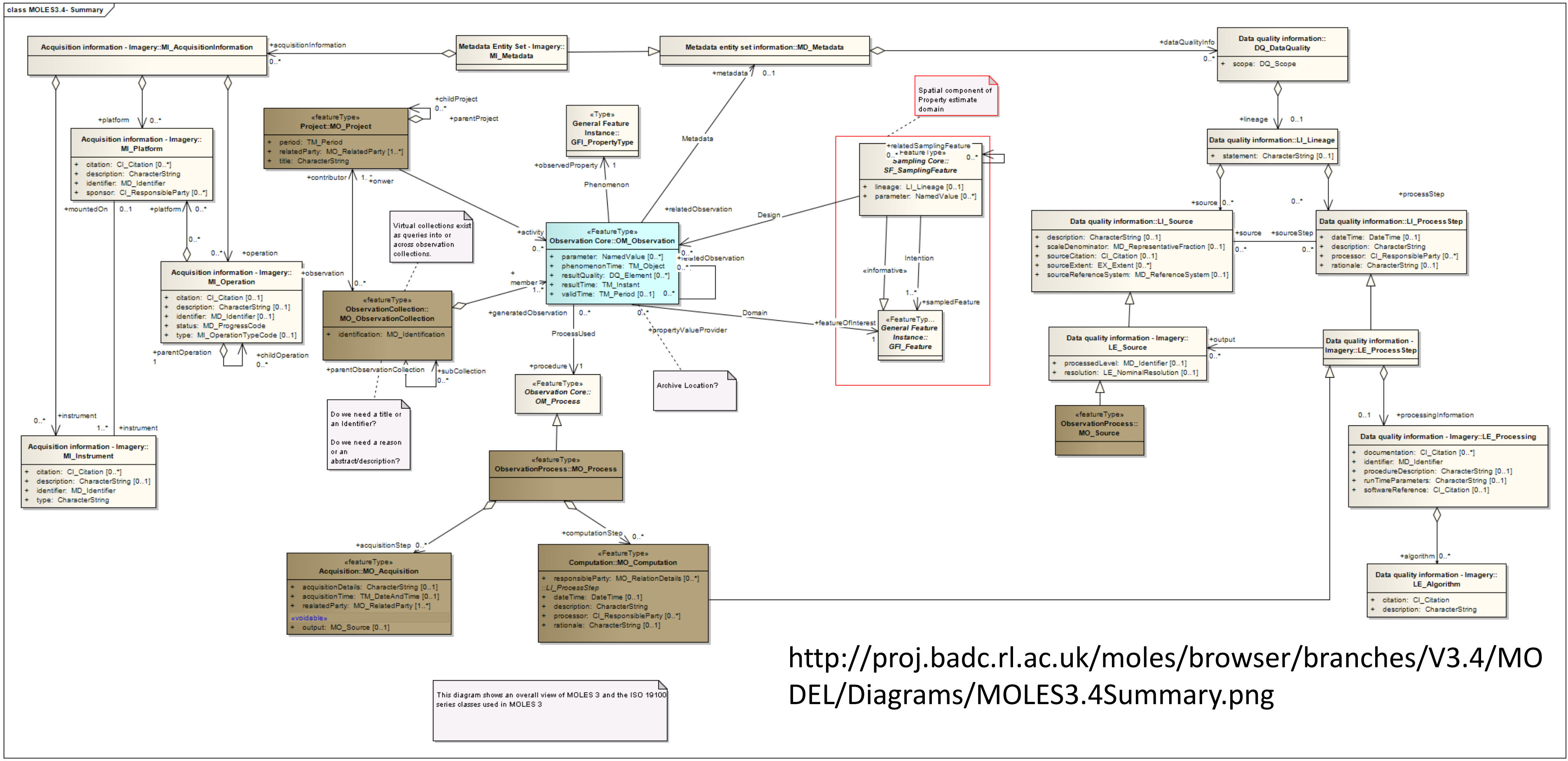
DataCite Mandatory Properties

ID	Property
1	Identifier (with type attribute)
2	Creator (with name identifier attributes)
3	Title (with optional type attribute)
4	Publisher
5	PublicationYear

DataCite Optional Properties

ID	Property
6	Subject (with schema attribute)
7	Contributor (with type and name identifier attributes)
8	Date (with type attribute)
9	Language
10	ResourceType (with description attribute)
11	AlternateIdentifier (with type attribute)
12	RelatedIdentifier (with type and relation type attributes)
13	Size
14	Format
15	Version
16	Rights
17	Description (with type attribute)

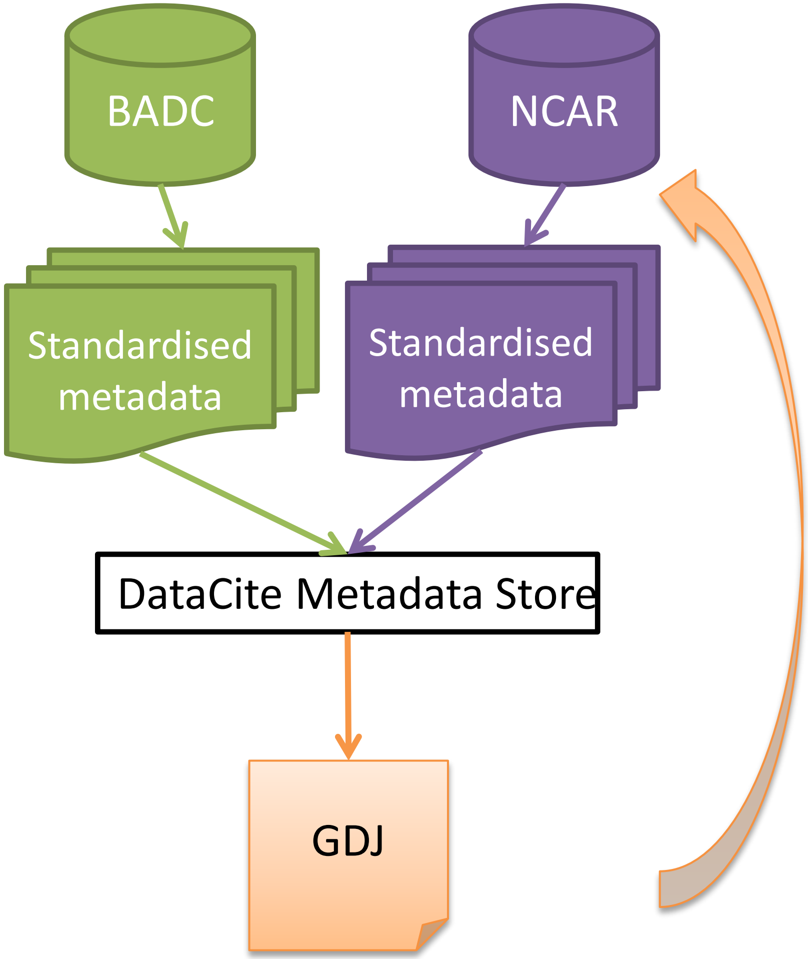
MOLES: Metadata Objects for Linking Environmental Sciences v3.4



class MOLES3.4- Summary
Acquisition information - Imagery::MI_AcquisitionInformation
+acquisitionInformation
0..*
Metadata Entity Set - Imagery::
MI_Metadata
Metadata entity set information::MD_Metadata
+dataQualityInfo
0..*
Data quality information::
DQ_DataQuality
+ scope: DQ_Scope
+metadata
0..1
+platform
0..*
Acquisition information - Imagery::
MI_Platform
+ citation: CI_Citation [0..*]
+ description: CharacterString
+ identifier: MD_Identifier
+ sponsor: CI_ResponsibleParty [0..*]
+mountedOn
0..1
+platform
0..*
«featureType»
Project::MO_Project
+ period: TM_Period
+ relatedParty: MO_RelatedParty [1..*]
+ title: CharacterString
+childProject
0..*
+parentProject
+contributor
1..*
+onwer
«Type»
General Feature
Instance::
GFI_PropertyType
+observedProperty
1
Phenomenon
Metadata
Spatial component of
Property estimate
domain
+relatedSamplingFeature
0..*
«Feature Type»
Sampling Core::
SF_SamplingFeature
0..*
+ lineage: LI_Lineage [0..1]
+ parameter: NamedValue [0..*]
+lineage
0..1
Data quality information::LI_Lineage
+ statement: CharacterString [0..1]
+source
0..*
0..*
+processStep
Virtual collections exist
as queries into or
across observation
collections.
+activity
0..*
«FeatureType»
Observation Core::OM_Observation
+ parameter: NamedValue [0..*]
+ phenomenonTime: TM_Object
+ resultQuality: DQ_Element [0..*]
+ resultTime: TM_Instant
+ validTime: TM_Period [0..1]
0..*
+relatedObservation
0..*
+relatedObservation
Design
Intention
«informative»
1..*
+sampledFeature
«FeatureTyp...
General Feature
Instance::
GFI_Feature
1
+featureOfInterest
Domain
0..*
0..*
Acquisition information - Imagery::
MI_Operation
+ citation: CI_Citation [0..1]
+ description: CharacterString [0..1]
+ identifier: MD_Identifier [0..1]
+ status: MD_ProgressCode
+ type: MI_OperationTypeCode [0..1]
+operation
0..*
+observation
0..*
+parentOperation
1
+childOperation
0..*
«featureType»
ObservationCollection::
MO_ObservationCollection
+ identification: MO_Identification
+
member
1..*
+generatedObservation
0..*
+parentObservationCollection
+subCollection
0..*
ProcessUsed
+propertyValueProvider
+procedure
1
«FeatureType»
Observation Core::
OM_Process
Archive Location?
Data quality information::LI_Source
+ description: CharacterString [0..1]
+ scaleDenominator: MD_RepresentativeFraction [0..1]
+ sourceCitation: CI_Citation [0..1]
+ sourceExtent: EX_Extent [0..*]
+ sourceReferenceSystem: MD_ReferenceSystem [0..1]
+source
+sourceStep
0..*
0..*
Data quality information::LI_ProcessStep
+ dateTime: DateTime [0..1]
+ description: CharacterString
+ processor: CI_ResponsibleParty [0..*]
+ rationale: CharacterString [0..1]
Data quality information - Imagery::
LE_Source
+ processedLevel: MD_Identifier [0..1]
+ resolution: LE_NominalResolution [0..1]
+output
0..*
Data quality information -
Imagery::LE_ProcessStep
«featureType»
ObservationProcess::
MO_Source
0..1
+processingInformation
Data quality information - Imagery::LE_Processing
+ documentation: CI_Citation [0..*]
+ identifier: MD_Identifier
+ procedureDescription: CharacterString [0..1]
+ runTimeParameters: CharacterString [0..1]
+ softwareReference: CI_Citation [0..1]
+algorithm
0..*
Data quality information - Imagery::
LE_Algorithm
+ citation: CI_Citation
+ description: CharacterString
0..*
+instrument
1..*
+instrument
Acquisition information - Imagery::
MI_Instrument
+ citation: CI_Citation [0..*]
+ description: CharacterString [0..1]
+ identifier: MD_Identifier
+ type: CharacterString
Do we need a title or
an Identifier?
Do we need a reason
or an
abstract/description?
«featureType»
ObservationProcess::MO_Process
+acquisitionStep
0..*
+computationStep
0..*
«featureType»
Acquisition::MO_Acquisition
+ acquisitionDetails: CharacterString [0..1]
+ acquisitionTime: TM_DateAndTime [0..1]
+ realatedParty: MO_RelatedParty [1..*]
«voidable»
+ output: MO_Source [0..1]
«FeatureType»
Computation::MO_Computation
+ responsibleParty: MO_RelationDetails [0..*]
::LI_ProcessStep
+ dateTime: DateTime [0..1]
+ description: CharacterString
+ processor: CI_ResponsibleParty [0..*]
+ rationale: CharacterString [0..1]
This diagram shows an overall view of MOLES 3 and the ISO 19100
series classes used in MOLES 3

http://proj.badc.rl.ac.uk/moles/browser/branches/V3.4/MODEL/Diagrams/MOLES3.4Summary.png

What PREPARDE is going to do



We already have a link from the GDJ data article to the data repository – thanks to the DOI.

GDJ can also pull the standard DOI metadata attached to that DOI from the DataCite metadata store

We need to figure out a way so GDJ can inform the repository that their dataset has been cited/published – bearing in mind scaling issues!

Might have to start with a manual work-around.

Tell us what you think

Workshop on cross-linking between data centres and publishers planned for May 2013 at Rutherford Appleton Laboratory, UK

Workshop on peer-review of data planned for March 2013 at the British Library

Always happy to get input from others!

Image Credit: http://bit.ly/9H4qBX

Project website: http://proj.badc.rl.ac.uk/preparde/wiki
Project blog: http://proj.badc.rl.ac.uk/preparde/blog

Thanks! Any questions?

#preparde
sarah.callaghan@stfc.ac.uk
@sorcha_ni

The project is led by the University of Leicester and the support of JISC and NERC in funding the PREPARDE project is gratefully acknowledged.

JISC
University of Leicester
British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL
WILEY-BLACKWELL
University of Reading
DCC
University of California
CDL
California Digital Library
F1000
FACULTY of 1000
POST-PUBLICATION PEER REVIEW
NCAR
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

JISC

University of Leicester

British Atmospheric Data Centre
NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL

WILEY-BLACKWELL

University of Reading

DCC

University of California
CDL
California Digital Library

F1000 FACULTY of 1000
POST-PUBLICATION PEER REVIEW

NCAR
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Composite image from Flickr user bnilsen and Matt Stempeck (NOI), shared under Creative Commons license

Pweez to stand by...we seemz to
be ekzpeerinsing teknikul difikulteez.
ICANHASCHEEZBURGER.COM