

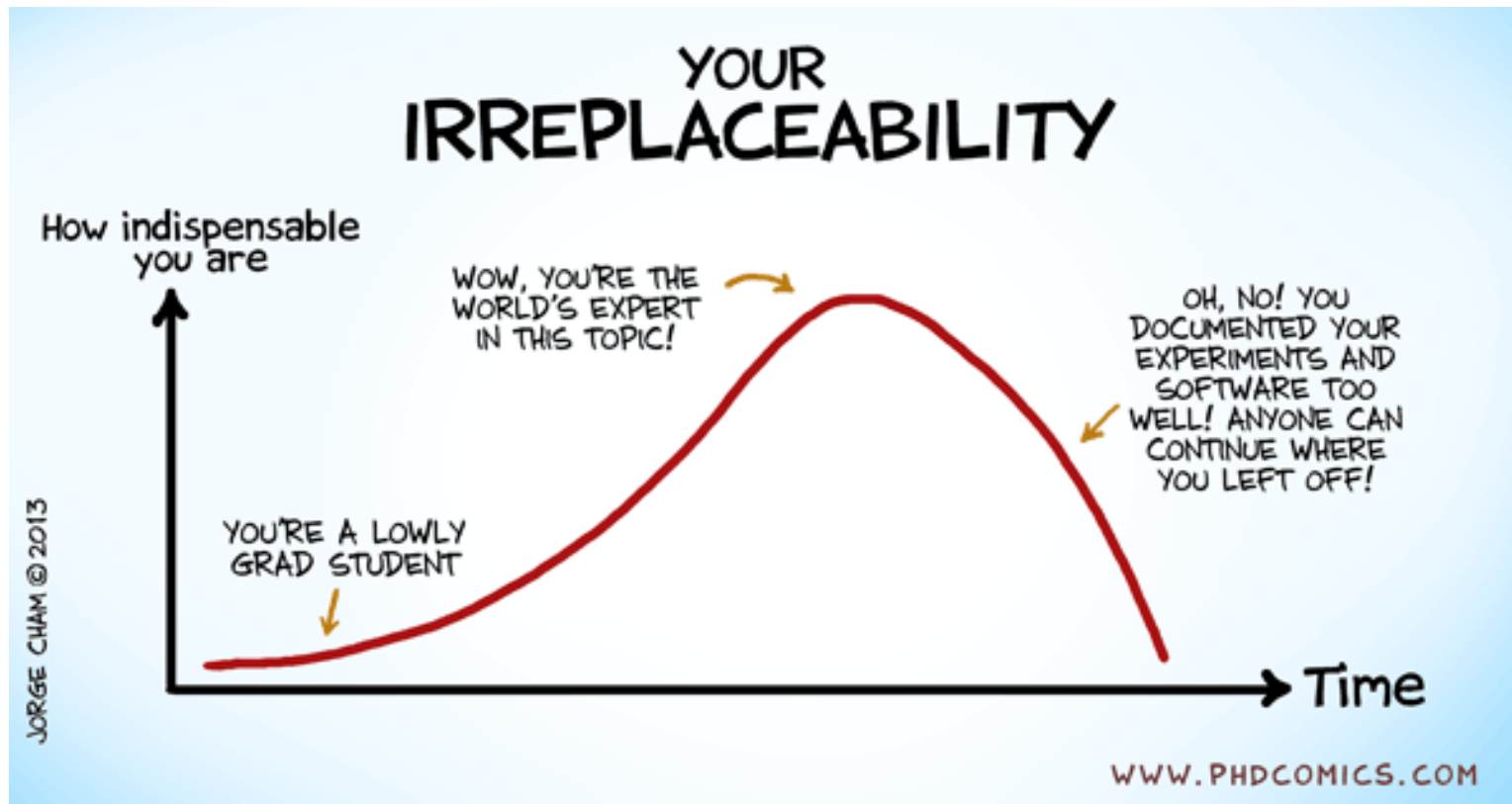
Data centre workflows for publication

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

#preparde

sarah.callaghan@stfc.ac.uk [@sorcha_ni](https://twitter.com/sorcha_ni)

One of the many reasons why convincing people to do data management is hard



Why cite and publish data?

- **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



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

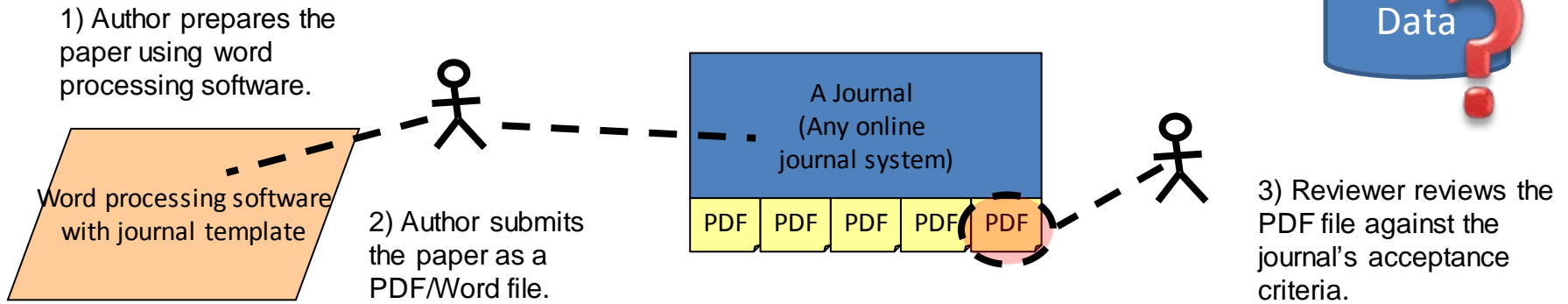
Geoscience Data Journal, Wiley-Blackwell and the Royal Meteorological Society

- 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.

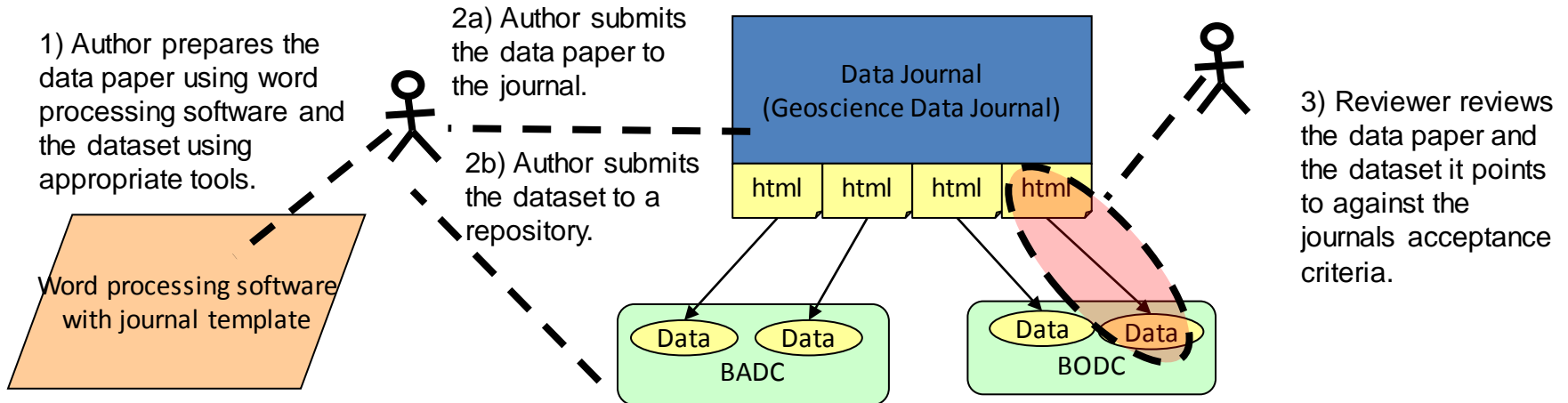


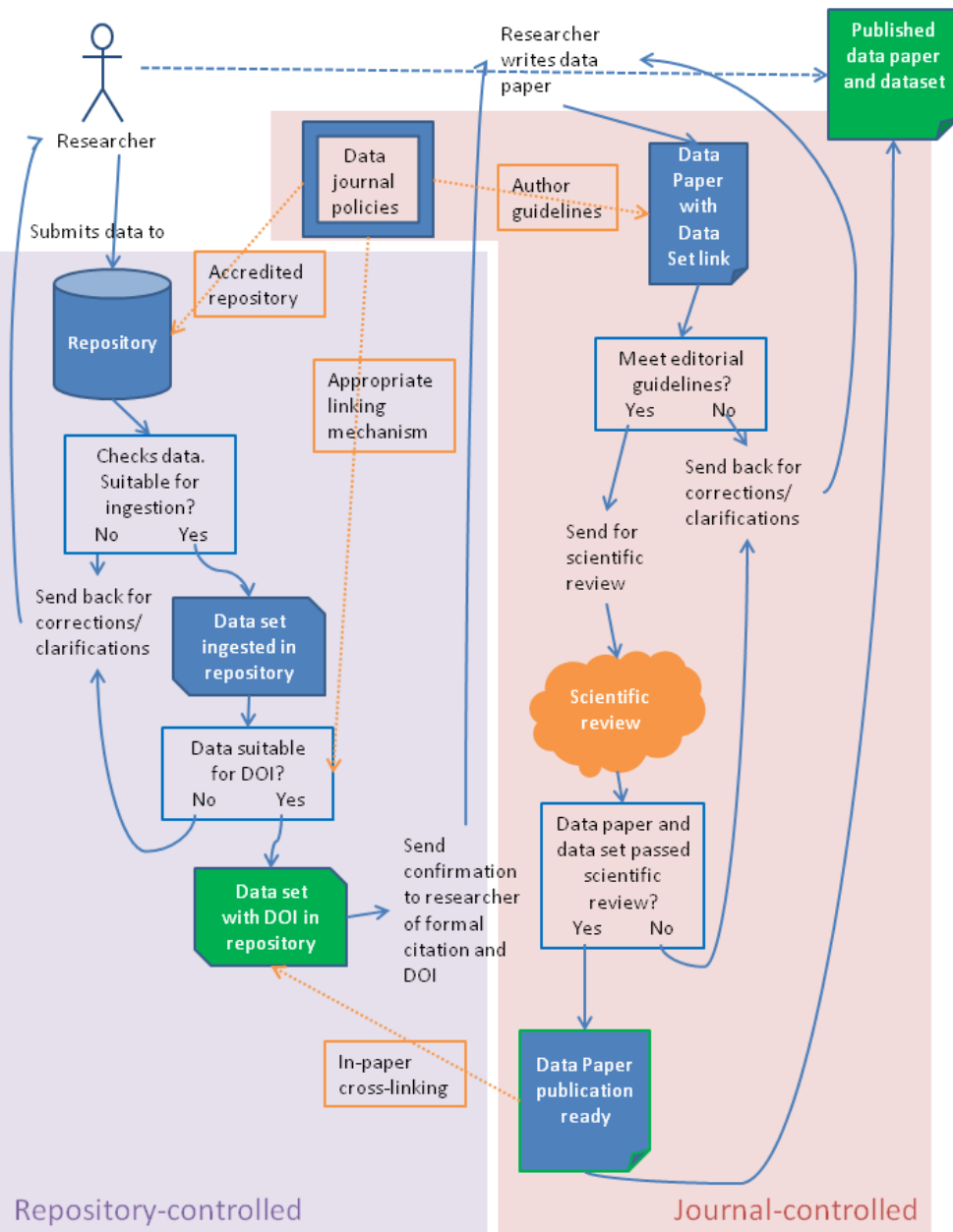
How we publish data

The traditional online journal model



Overlay journal model for publishing data





PREPARDE topics

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

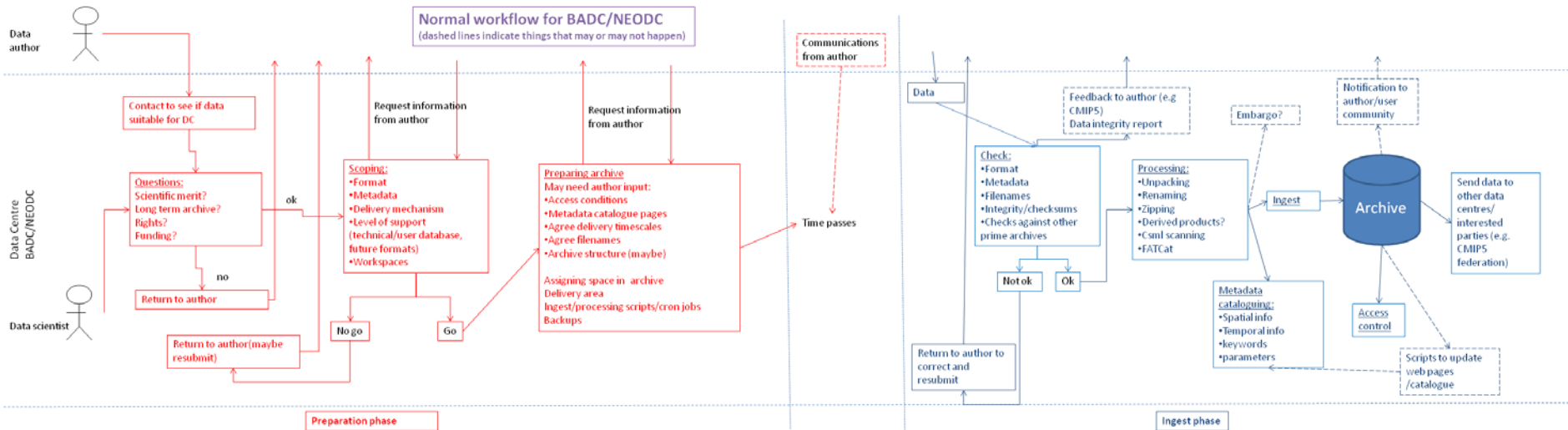
3 main areas of interest (in orange)

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

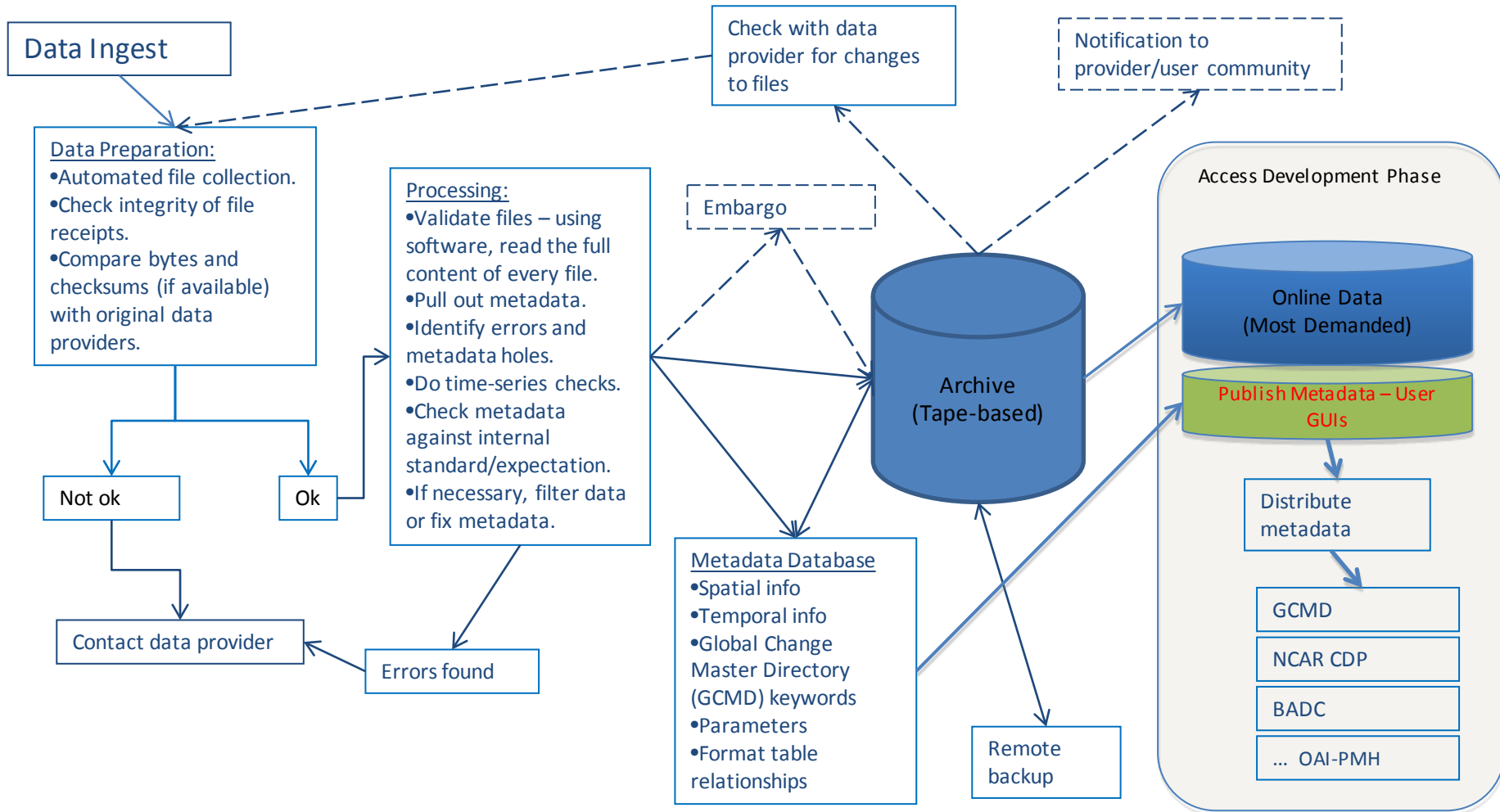
- Division of area of responsibilities between
 - *repository controlled* processes
 - *journal controlled* processes

Data repository workflows

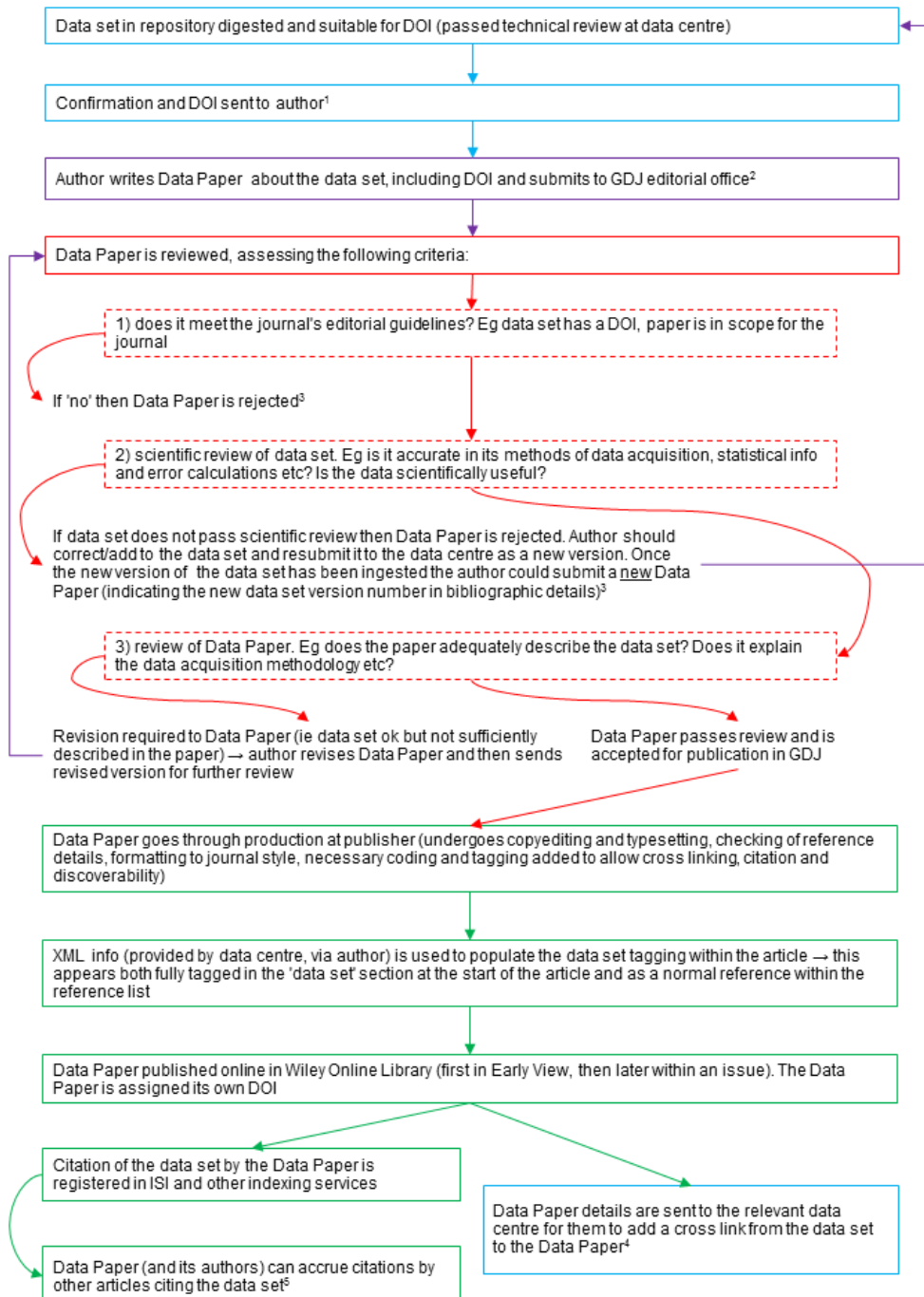
- Data centre and journal workflows captured
 - Workflows are very varied! No one-size fits all method
 - Can have multiple workflows in the same data centre, depending on interactions with external sources (“Engaged submitter”/ “Data dumper”/ “Third party requester”)



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



Geoscience Data Journal Data Paper workflow

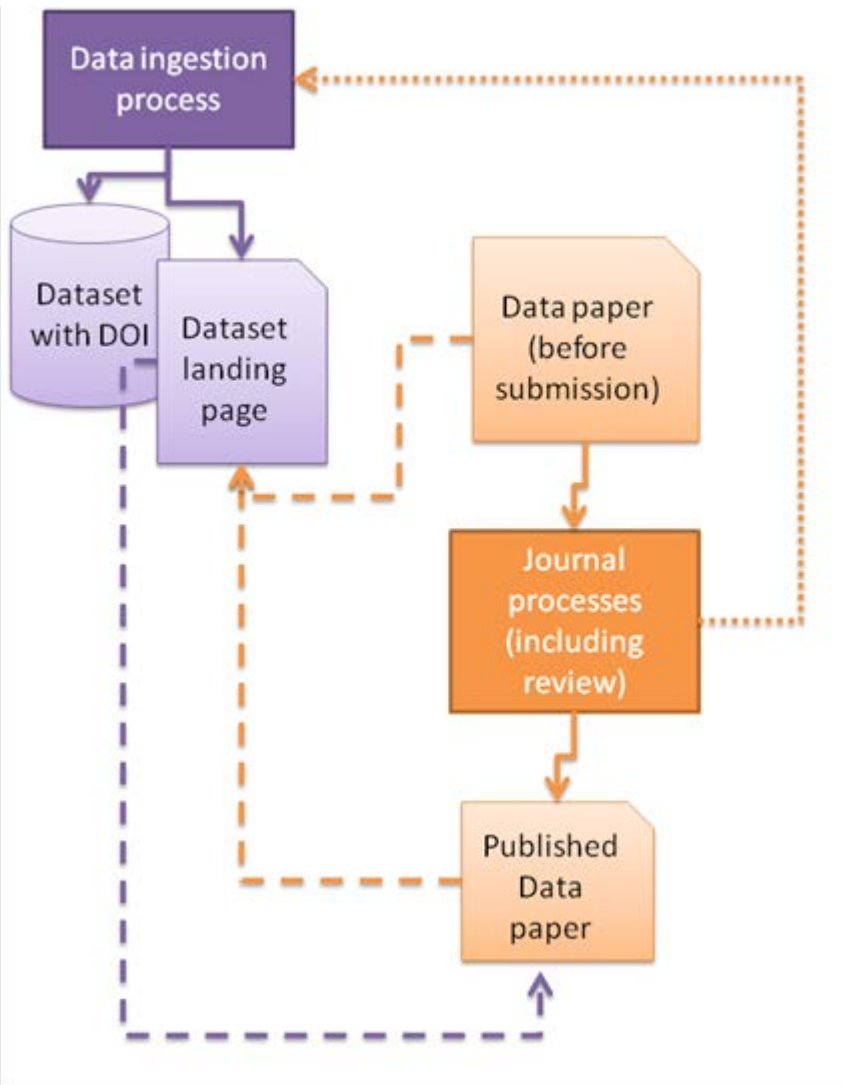


Journal workflow

Aim is to minimise effort needed to submit a data paper by taking advantage of already submitted metadata.

Sharing metadata also ensures that additions/corrections made in one location get propagated through to the others

Generic data publication workflow.



Dashed lines indicate linking (via URL) or citation (via DOI).

Solid lines indicate the results or inputs into processes.

Dotted line indicated where the results of a process need to be fed back into another process.

Journal responsibilities are orange, data centre's are purple

Using citations to link research outputs

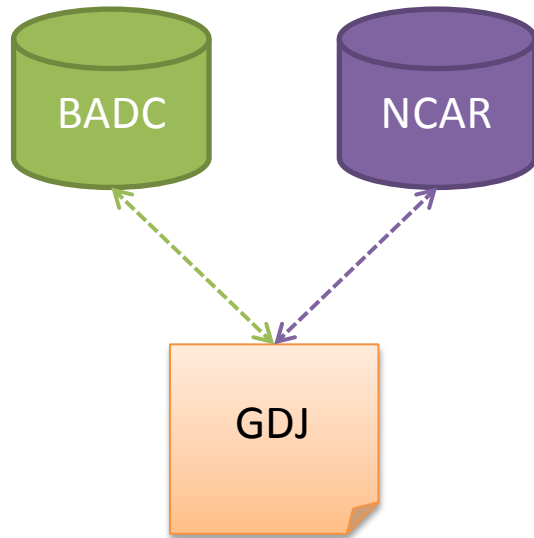
- We already have a working method for linking between publications which is:
 - commonly used
 - understood by the research community
 - used to create metrics to show how much of an impact something has (citation counts)
 - applied to digital objects (digital versions of journal articles)
- We can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

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

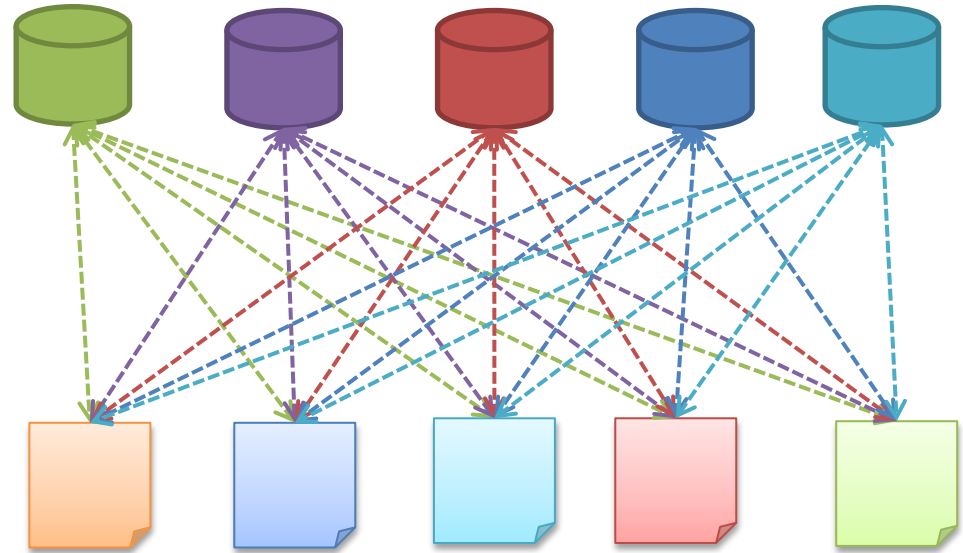


<http://www.naa.gov.au/records-management/capability-development/keep-the-knowledge/index.aspx>

Cross-linking

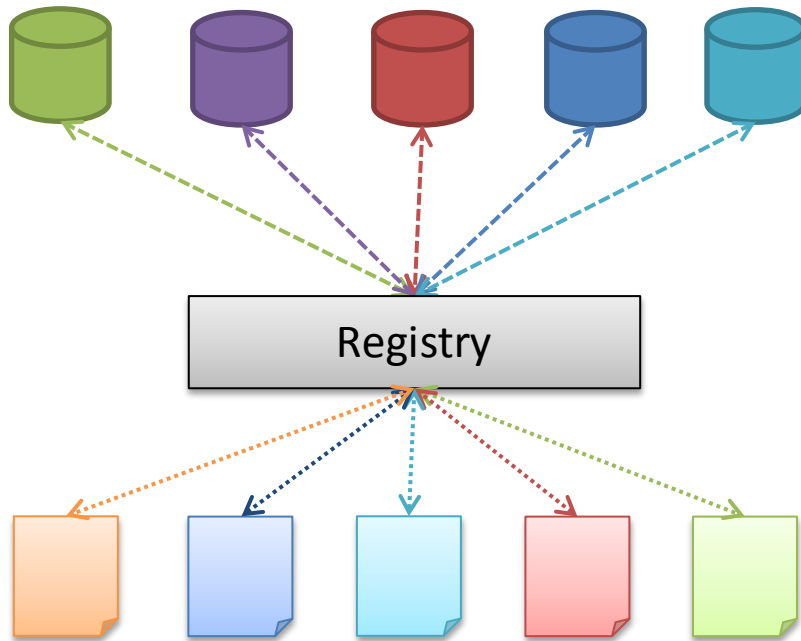


This is what we have to focus on for PREPARDE – demonstrate cross linking between GDJ and a data repository (BADC/NCAR)



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? Could DataCite? Could re3data.org?

Registry would need to be discipline agnostic!

Do we have a start?

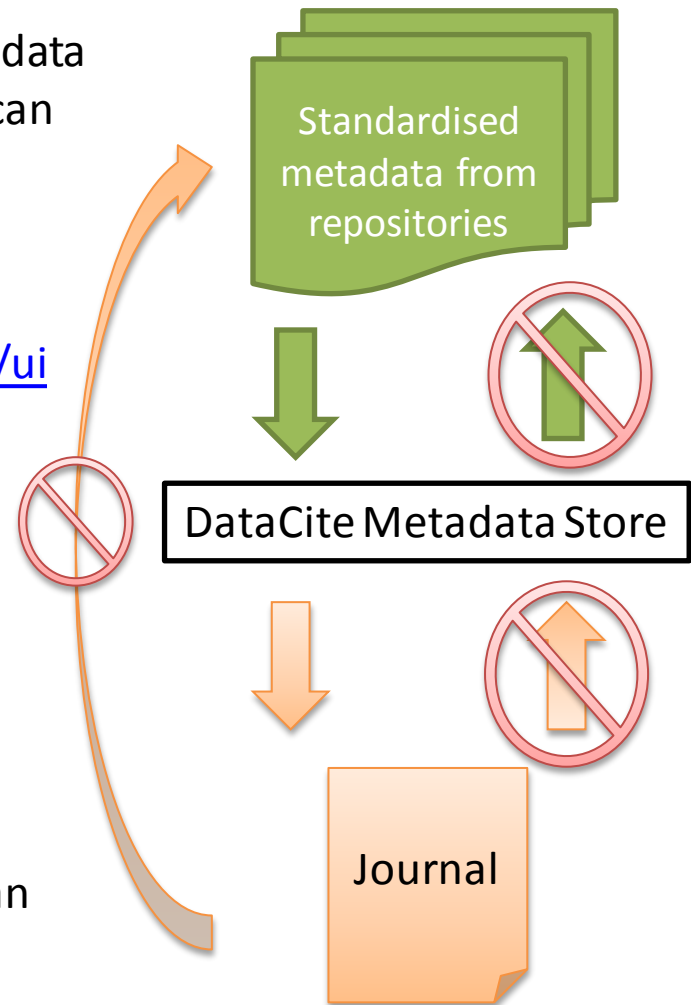
DataCite have standardised a set of bibliometric metadata that have to be submitted before a DOI for a dataset can be minted by a repository.

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/>

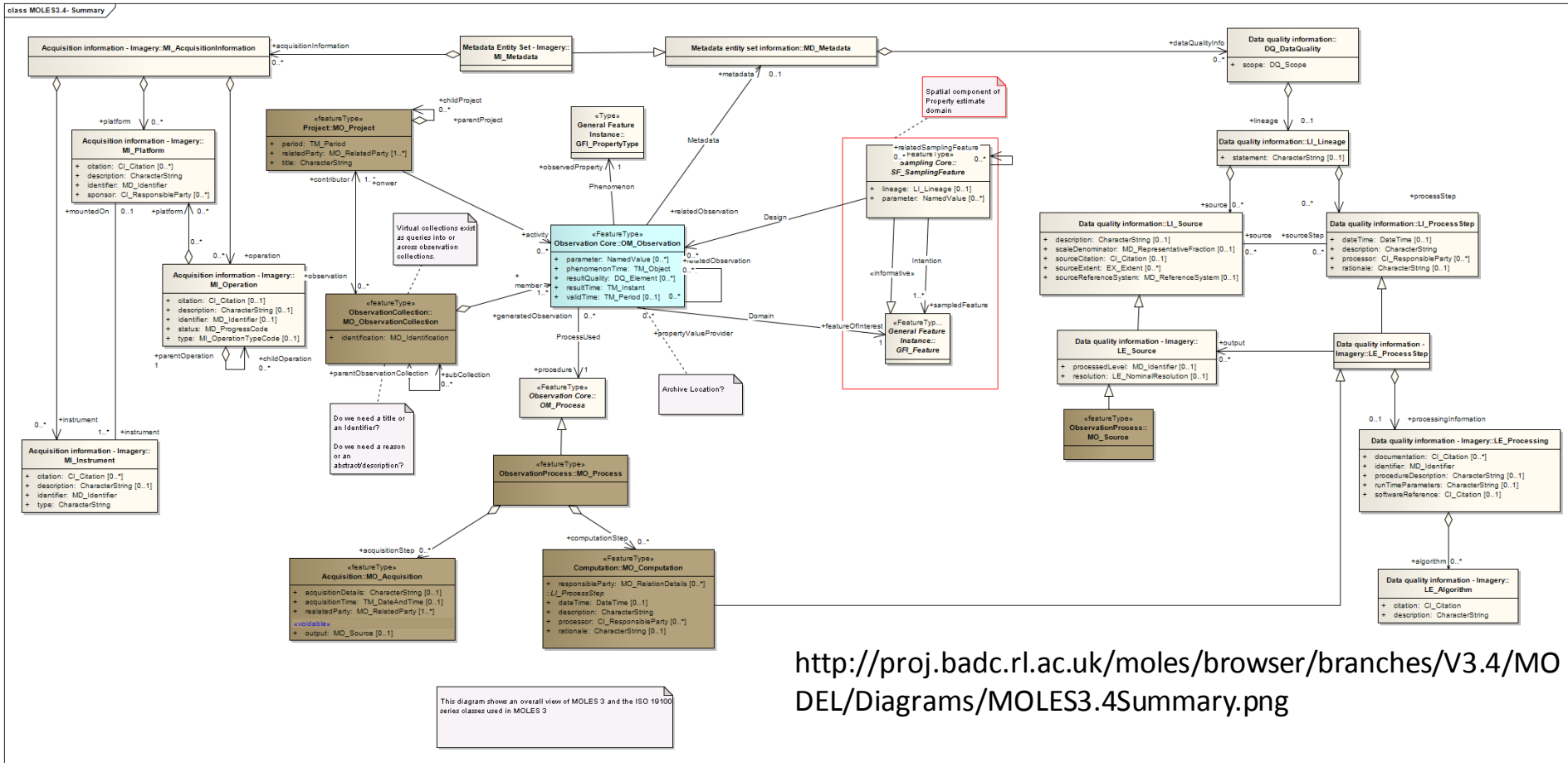
DataCite Mandatory Properties

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

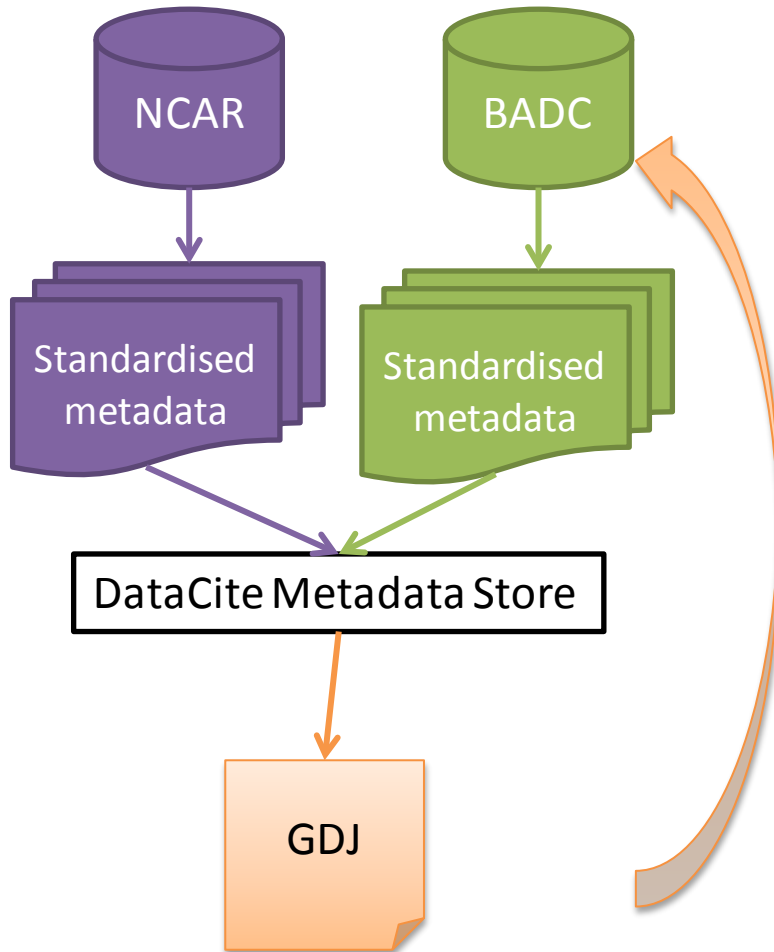
DataCite Optional Properties

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



What PREPARDE has done



- 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
- GDJ needs to inform the repository that their dataset has been cited/published – bearing in mind scaling issues!
- At this time, we have a manual work-around (i.e. email)

The GBS dataset: measure x

onlinelibrary.wiley.com/doi/10.1002/gdj3.2/full

BADC - Trac | METAFOR | Home | Google Mail | BBC NEWS | News Fr... | Sorcha ni gCeallagh... | Add to Wish List | Other bookmarks

WILEY ONLINE LIBRARY

PUBLICATIONS | BROWSE BY SUBJECT | RESOURCES | ABOUT US | The Chadwick & RAL Libraries

Home > Earth Sciences > General & Introductory Earth Sciences > Geoscience Data Journal > Early View > Abstract

JOURNAL TOOLS

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

JOURNAL MENU

Journal Home

FIND ISSUES

Early View

FIND ARTICLES

Early View

FOR CONTRIBUTORS

Author Guidelines

Submit an Article

ABOUT THIS JOURNAL

Society Information

Overview

Editorial Board

Permissions

SPECIAL FEATURES

Data Center FAQs

L.F. Richardson Award Prize Winners

Open Access License and Copyright

Author FAQs

Article Publication Charges

Wiley Open Access

Institutional and Funder Payments

Guidelines for Reviewers

Guidelines for Repositories

RMetS Geoscience Data Journal

Open Access

Data Paper

The GBS dataset: measurements of satellite site diversity at 20.7 GHz in the UK

S. A. Callaghan, J. Waight, J. L. Agnew, C. J. Walden, C. L. Wrench, S. Ventouras

Issue

Article first published online: 17 MAR 2013
DOI: 10.1002/gdj3.2

Copyright © 2013 The Authors. Published by John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes

Additional Information (Show All)

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

The research presented in this paper was funded by the UK's Ofcom as part of the Spectrum Efficiency Scheme and the support of Ofcom in providing the funding for the GBS experiment is greatly appreciated.

SEARCH

In this issue

Advanced > Saved Searches >

ARTICLE TOOLS

- Get PDF (359K)
- Save to My Profile
- E-mail Link to this Article
- Export Citation for this Article
- Get Citation Alerts
- Request Permissions

Share |

Abstract Article References Cited By

Get PDF (359K)

Keywords:
site diversity; radio propagation; fade mitigation techniques

Abstract

The GBS (Global Broadcast Service) dataset is a series of radio attenuation measurements made at three sites in the UK: Chilbolton and Sparsholt, both in southern UK, and Dundee in Scotland. The aim of the experiment was to make long term measurements of the signal strength received from a 20.7 GHz beacon on the US Department of Defense satellite UFO-9 at multiple sites, in order to determine whether the use of site diversity as a fade mitigation technique would be effective. The dataset spans a period of 3 years, from August 2003 to August 2006 with signal attenuation sampled once per second.

Dataset

The GBS (Global Broadcast Service) dataset comes as 3 separate data streams:

- Identifier: [doi:10.5285/639A3714-BC74-46A6-9026-64931F355E07](https://doi.org/10.5285/639A3714-BC74-46A6-9026-64931F355E07)
Creator: Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [Callaghan, S. A., J. Waight, C. J. Walden, J. Agnew and S. Ventouras].
Title: GBS 20.7 GHz slant path radio propagation measurements, Chilbolton site
publisher: NERC British Atmospheric Data Centre
Publication year: 2009
Resource type: Metadata document
Version: 1.0
- Identifier: [doi:10.5285/db8d8981-1a51-4d6e-81c0-cced9b921390](https://doi.org/10.5285/db8d8981-1a51-4d6e-81c0-cced9b921390)
Creator: Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [Callaghan, S. A., J. Waight, C. J. Walden, J. Agnew and S. Ventouras].

Live Data paper!

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

DOI: 10.1002/gdj3.2



Viewing GBS 20.7GHz slant x

badc.nerc.ac.uk/view/badc.nerc.ac.uk_ATOM_dep_11902119479621181

BADC - Trac METAFOR | Home Google Mail BBC NEWS | News Fr... Sorcha ní gCeallagh... Other bookmarks

Centre for Environmental Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL

Search for in All

GBS 20.7GHz slant path radio propagation measurements, Chilbolton site

General Info

Title: GBS 20.7GHz slant path radio propagation measurements, Chilbolton site
Type: Activity
Sub-Type: Deployment
Publication State: Citable
URI: http://badc.nerc.ac.uk/view/badc.nerc.ac.uk_ATOM_dep_11902119479621181

Summary

The GBS (Global Broadcast Service) dataset is a series of radio attenuation measurements made at three sites in the UK: Chilbolton and Sparsholt, both in southern UK, and Dundee in Scotland. The aim of the experiment was to make long term measurements of the signal strength received from a 20.7GHz beacon on the US Department of Defense satellite UFO-9 at multiple sites, in order to determine whether the use of site diversity as a fade mitigation technique would be effective. The dataset spans a period of 3 years, from August 2003 to August 2006 with signal attenuation sampled once per second.

Please cite this dataset as:
 Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [S. A. Callaghan, J. Waight, C. J. Walden, J. Agnew and S. Ventouras], GBS 20.7GHz slant path radio propagation measurements, Sparsholt site, [Internet]. British Atmospheric Data Centre, 2003-2005. 1st April 2014. doi:10.1002/gdj3.2

This dataset is cited in:
 S. A. Callaghan, J. Waight, J.L.Agnew, C. J. Walden, C.L.Wrench , S. Ventouras "The GBS dataset: measurements of satellite site diversity at 20.7 GHz in the UK", Geoscience Data Journal, 17 March 2013, DOI: 10.1002/gdj3.2

Author

Name email
 Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [S. A. Callaghan, J. Waight, C. J. Walden, J. Agnew and S. Ventouras]

Online References

Relation	Title
Apply for access	Apply for to GBS data from Chilbolton
Download	Data directory for GBS data from Chilbolton
Documentation	DOI for dataset:10.5285/620-2714-b71-46-c-0026-64921f355e07
Documentation	Data article in Geoscience Data Journal doi:10.1002/gdj3.2

Associated Data

Type	Title
Data Production Tool	Chilbolton: GBS receiver
Activity	Chilbolton Facility for Atmospheric and Radio Research (CFARR)
Observation Station	Chilbolton Facility for Atmospheric and Radio Research (CFARR), UK

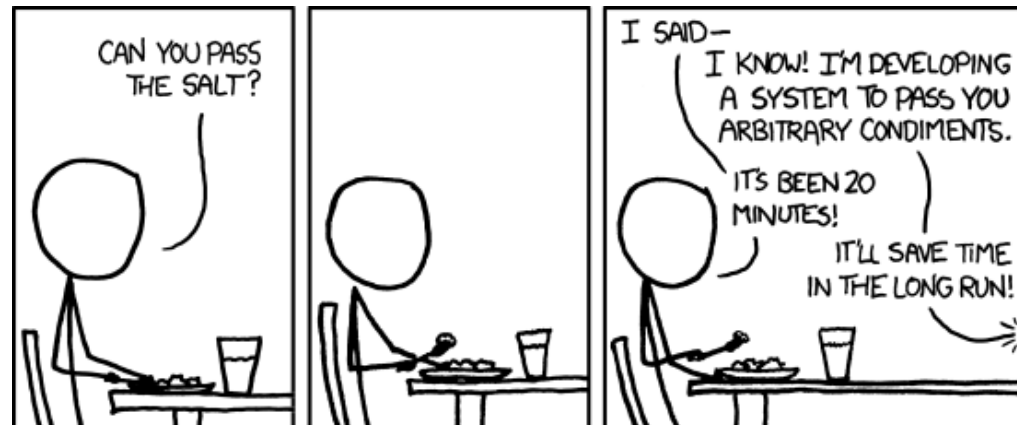
Dataset catalogue page (and DOI landing page)

Reference to Data Article

Clickable link to Data Article

Problems still to solve

- Automatic methods for:
 - (Data) journal informing repository dataset has been cited
 - Repository linking back to paper citing dataset
- Sharing of dataset metadata between repository and journal
 - So paper author doesn't have to repeatedly enter metadata in multiple locations
 - So corrections made in one place can be propagated across
- Centralised registry for cross-linking
 - Deal with scalability issues in direct linking between journals and repositories
- Methods for issuing corrections to data after data paper has been published



THE GENERAL PROBLEM [HTTP://XKCD.COM/974/](http://xkcd.com/974/)

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.

