Data publication – policies and procedures from the PREPARDE project

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

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/

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
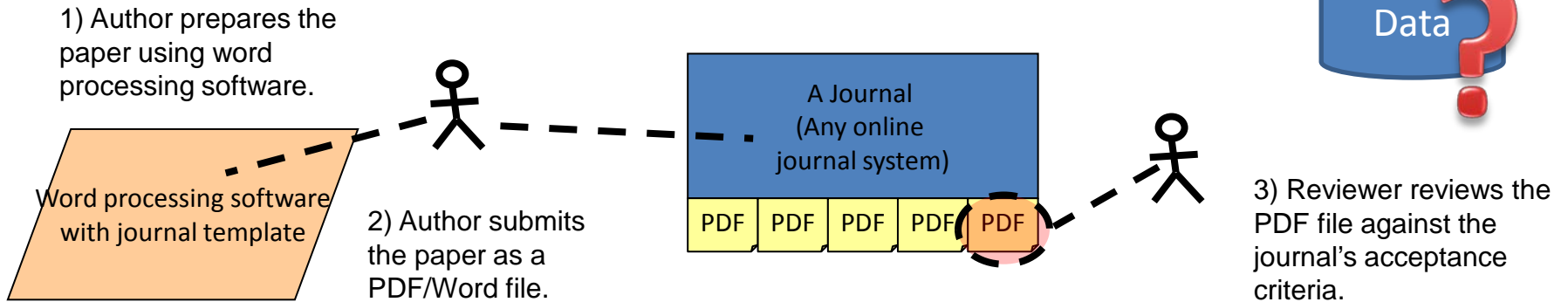- Project Start Date: 1st July 2012
- Project End Date: 31st June 2013

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.
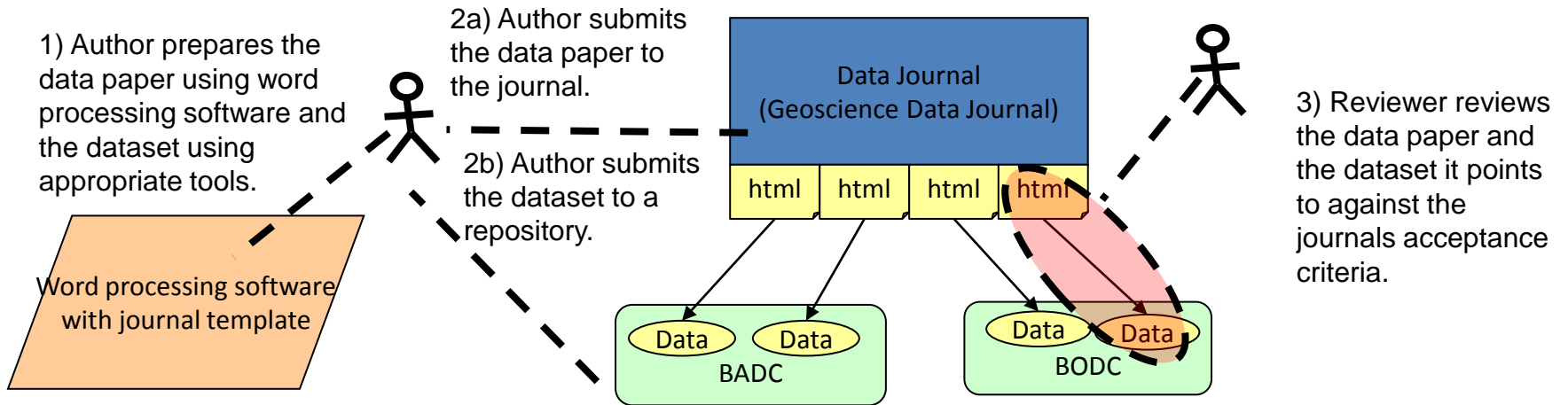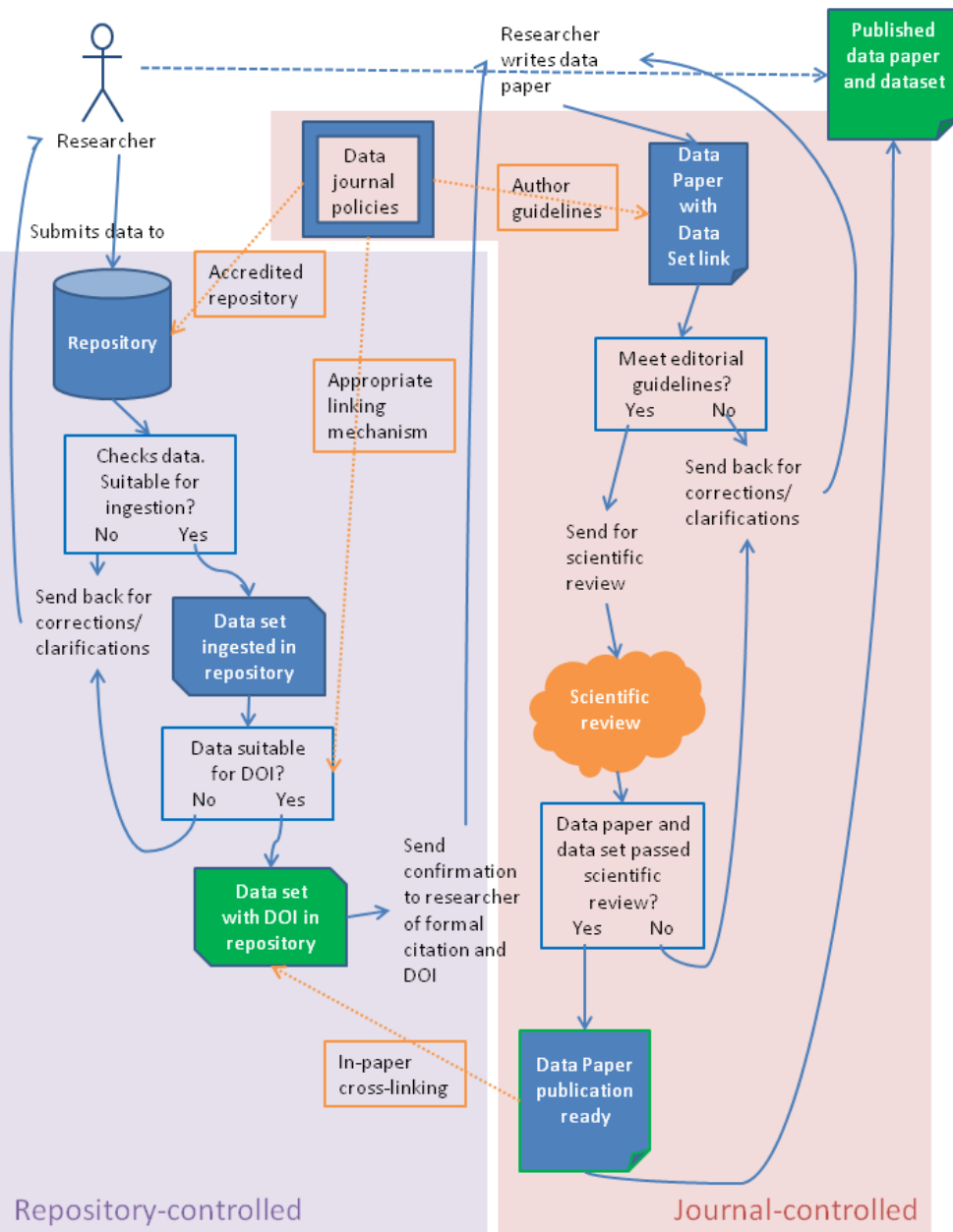 - the when, how and why data was collected and what the data-product is.

How we publish data

The traditional online journal model

1) Author prepares the paper using word processing software.

Word processing software with journal template

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

A Journal (Any online journal system)

PDF | PDF | PDF | PDF | PDF

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

Data **?**

Overlay journal model for publishing data

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

Word processing software with journal template

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

2b) Author submits the dataset to a repository.

Data Journal (Geoscience Data Journal)

html | html | html | html

Data | Data
BADC

Data | Data
BODC

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

JISC · University of Leicester · British Atmospheric Data Centre (NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE NATURAL ENVIRONMENT RESEARCH COUNCIL) · WILEY-BLACKWELL · University of Reading · DCC · CDL (University of California California Digital Library) · F1000Research · NCAR (NATIONAL CENTER FOR ATMOSPHERIC RESEARCH)

PREPARDE topics

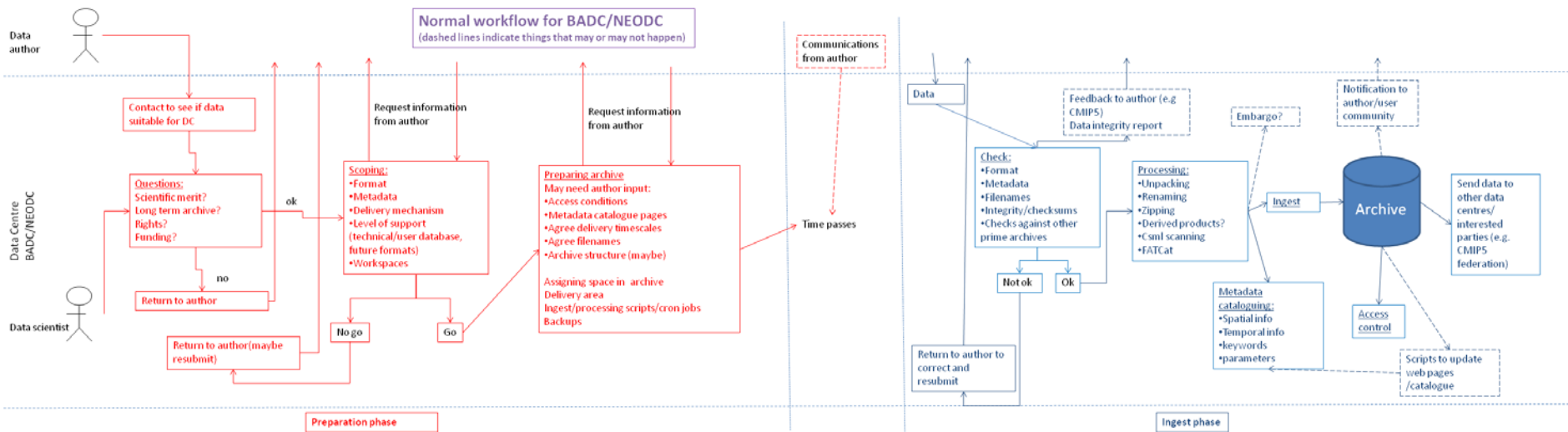Example steps/workflow required for a researcher to publish a data paper

3 main areas of interest (in orange)
1. Workflows and cross-linking between journal and repository
2. Repository accreditation
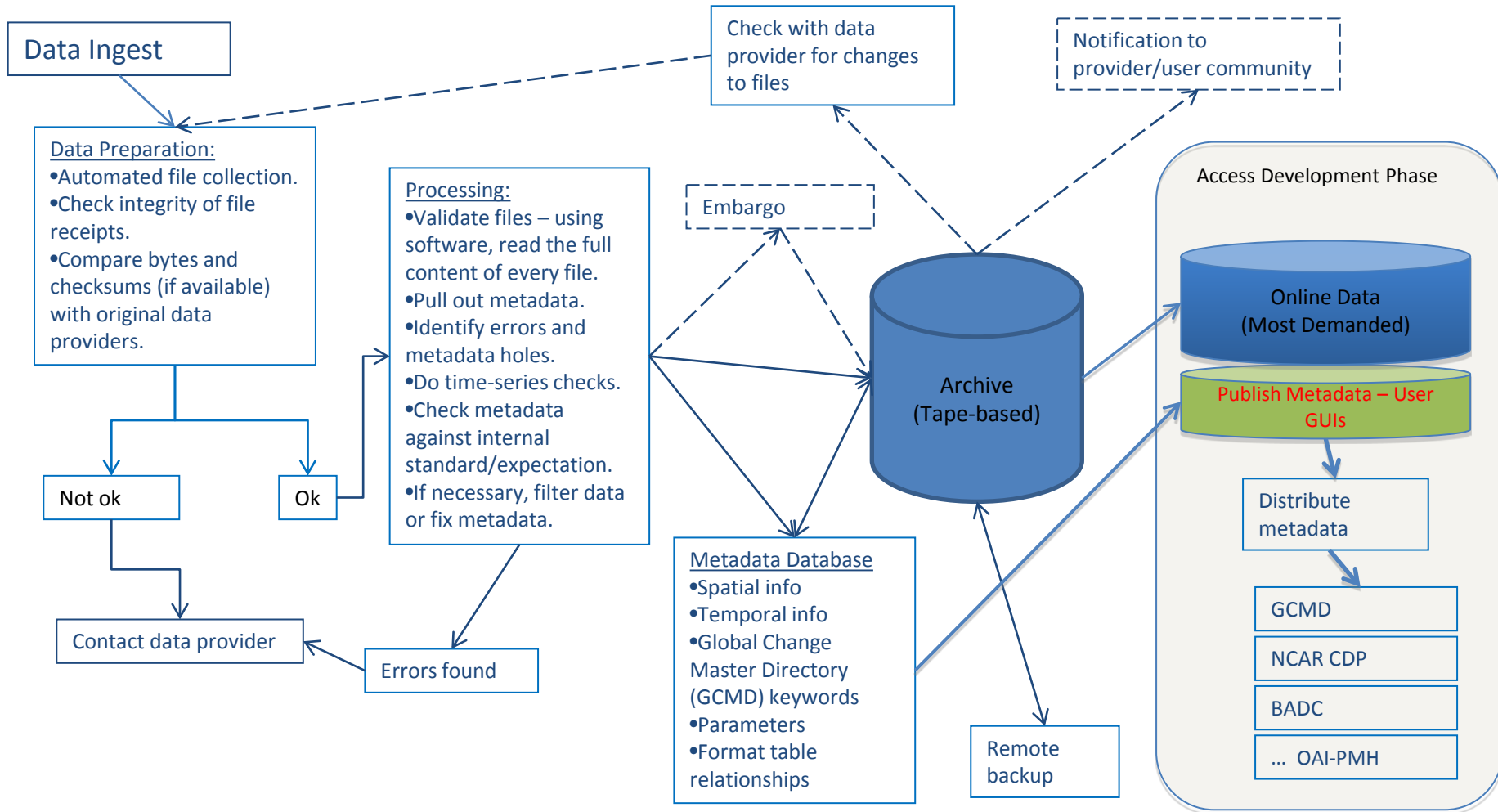3. Scientific peer-review of data

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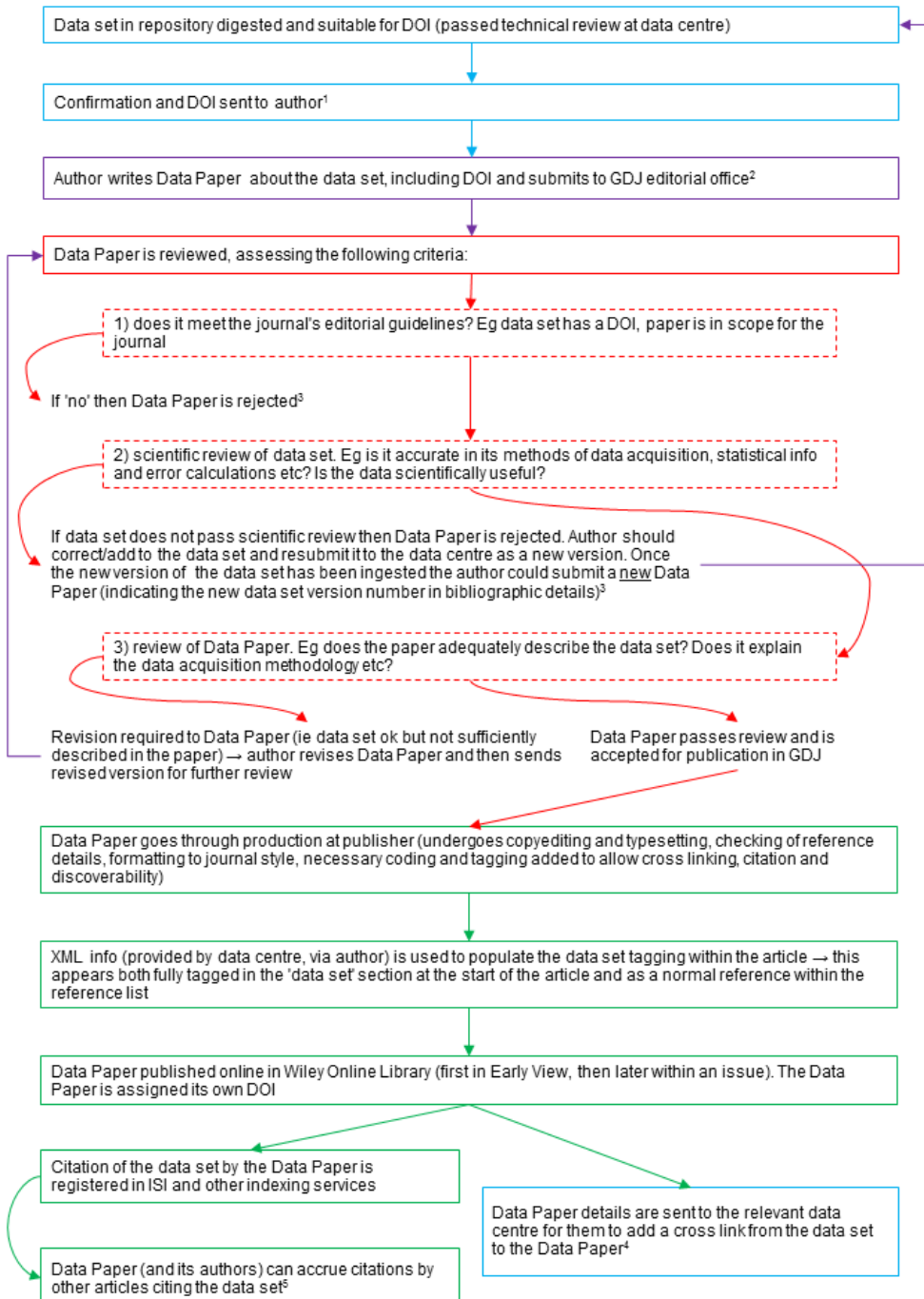
Repository Workflow – NCAR Comp. & Info. Systems Lab Research Data Archive (RDA)
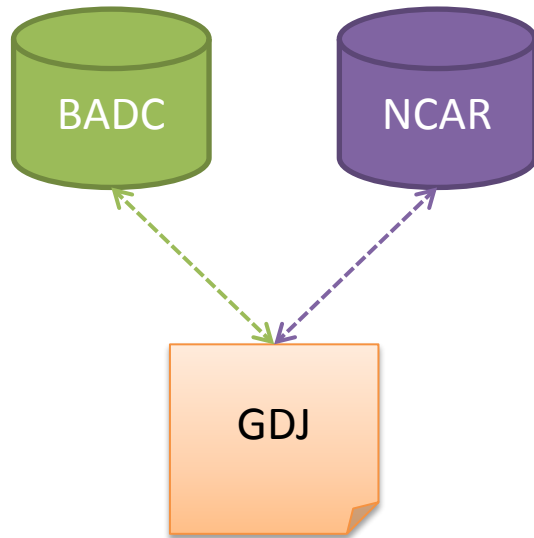
Journal workflow

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

Data Paper (and its authors) can accrue citations by other articles citing the data set[5]
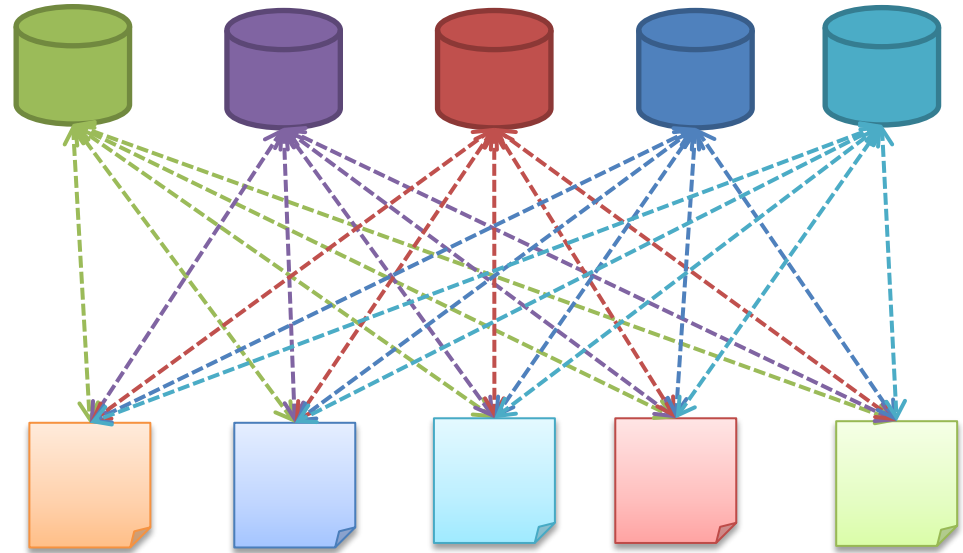
Aim is to minimise effort needed to submit a data paper by taking advantage of already submitted metadata.

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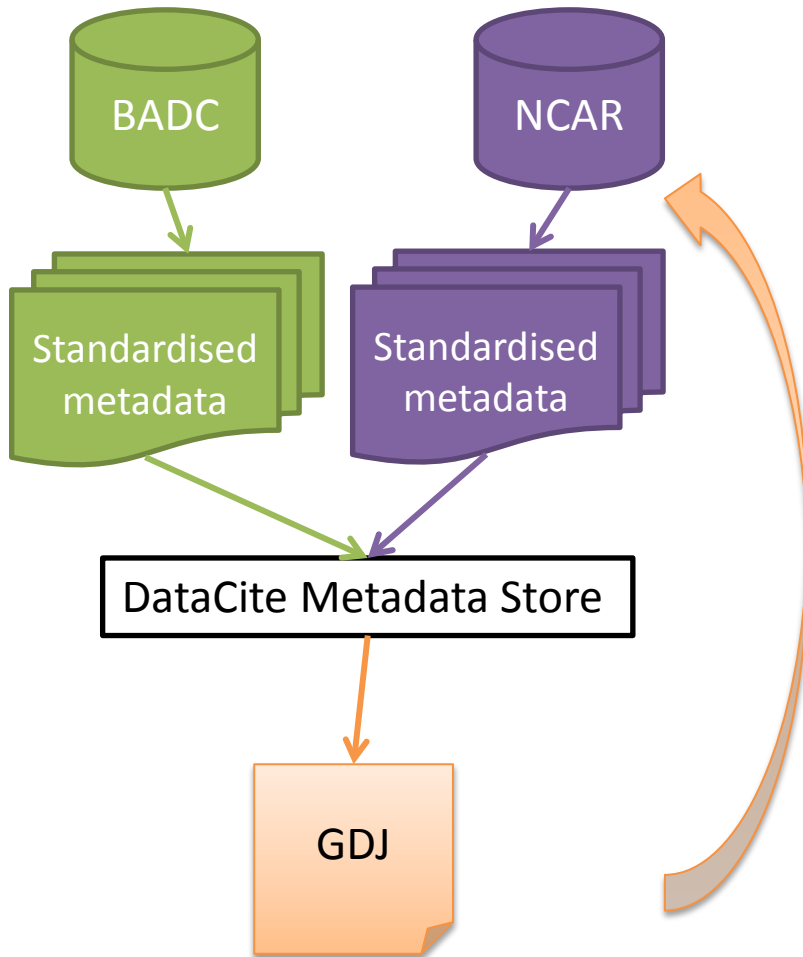
Cross-linking

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

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)

- ***Workshop on cross-linking between data centres and publishers 30th April 2013 at RAL, UK***

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

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

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 Royal Meteorological Society

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

Geoscience Data Journal

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

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

SEARCH
In this issue
Advanced > Saved Searches >

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.

Abstract | **Article** | References | Cited By

Get PDF (359K)

Keywords:
site diversity; radio propagation; fade mitigation techniques

Abstract   Jump to...

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   Jump to...

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

- Identifier: doi: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

  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].

University of Reading

DCC

University of California
CDL
California Digital Library

NCAR
NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Dataset catalogue page (and DOI landing page)

Reference to Data Article

Clickable link to Data Article

Repository accreditation

Link between data paper and dataset is crucial!

- How do data journal editors know a repository is trustworthy?

- How can repositories prove they're trustworthy?

What makes a repository trustworthy?
- Many things: mission, processes, expertise, workflows, history, systems, documentation, …
- Assessing trustworthiness requires assessing the entire repository workflow

- ***PREPARDE / IDCC13 Workshop – report in draft***
- Peer review of data is implicitly peer review of repository
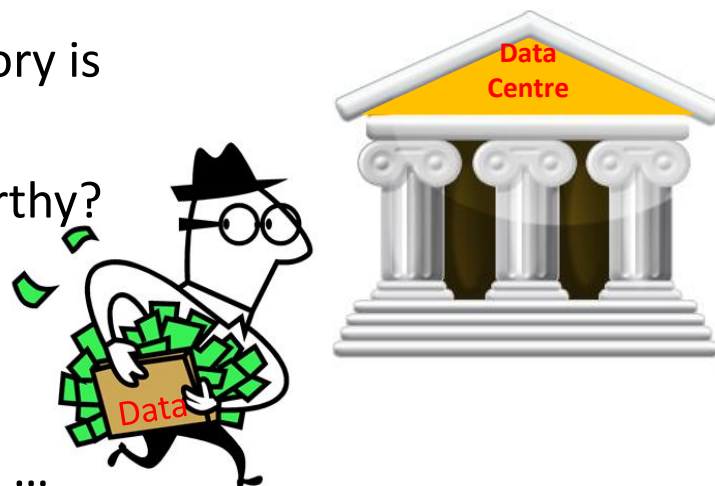
And what does "trustworthy" mean, when you get right down to it?

Repository accreditation schemes:

European Framework for Audit and Certification of Digital Repositories

TrustedDigitalRepository.eu

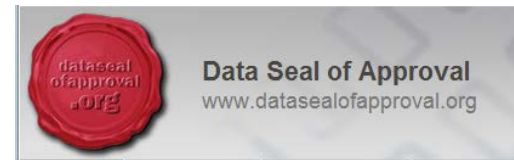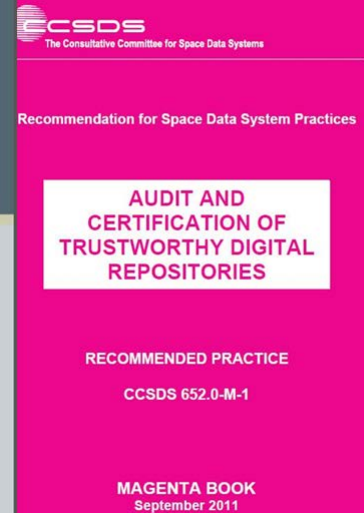**Making the digital world more reliable**

These schemes look at all of the business of running a repository, but don't directly address the issues required for data publication.

Data for publication needs to:
- Be persistent
- Be permanently identified
- Be provided with a landing page
- Have standard publication metadata
- Have accessibility/licensing information

Document at: http://bit.ly/ZhYHZl
Feedback to: https://www.jiscmail.ac.uk/DATA-PUBLICATION

Trustworthy Repositories Audit & Certification: Criteria and Checklist

OCLC

Contents:

Introduction
Establishing Audit and Certification Criteria
Towards an International Audit & Certification Process
Using this Checklist for Audit & Certification
Applicability of Criteria
Relevant Standards, Best Practices & Controls
Terminology
Audit and Certification Criteria
Organizational Infrastructure
Digital Object Management
Technologies, Technical Infrastructure & Security
Audit Checklist
Glossary
Appendices

Version 1.0
February 2007

CCSDS
The Consultative Committee for Space Data Systems

Recommendation for Space Data System Practices

AUDIT AND CERTIFICATION OF TRUSTWORTHY DIGITAL REPOSITORIES

RECOMMENDED PRACTICE

CCSDS 652.0-M-1

MAGENTA BOOK
September 2011

Data Seal of Approval
www.datasealofapproval.org

DataCite

Helping you to find, access, and reuse data

ICSU WORLD DATA SYSTEM

JISC · University of Leicester · British Atmospheric Data Centre · NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE · NATURAL ENVIRONMENT RESEARCH COUNCIL · WILEY-BLACKWELL · University of Reading · DCC · CDL California Digital Library · F1000Research · NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

Peer-review of data

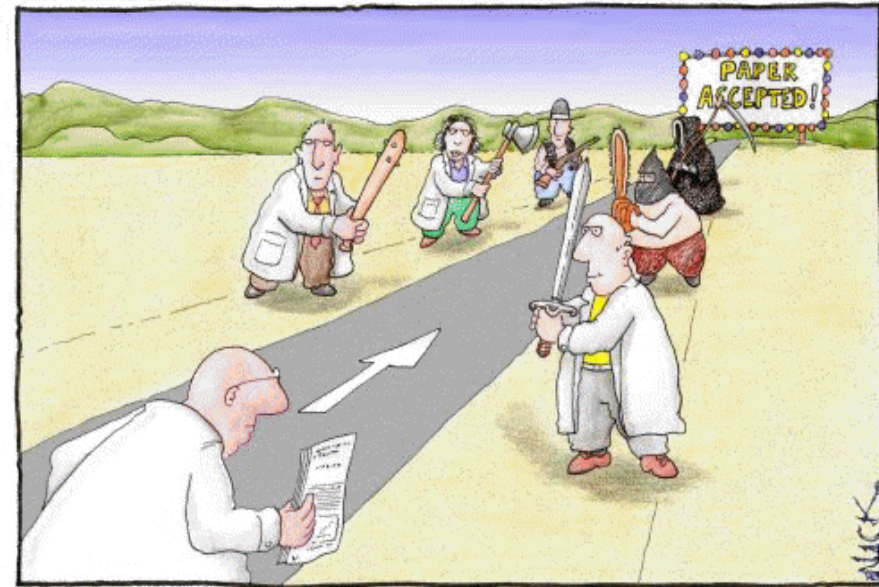Summary Recommendations from Workshop at the British Library, 11 March 2013.

Workshop attendees included funders, publishers, repository managers and other interested parties.

Draft recommendations put up for discussion and feedback from audience captured.

3 main parts to recommendations:

- Connecting data review with data management planning

- Connecting scientific, technical review and curation

- Connecting data review with article review

Feedback from the community still welcome!

Most scientists regarded the new streamlined peer-review process as 'quite an improvement.'

http://libguides.luc.edu/content.php?pid=5464&sid=164619

Document at: http://bit.ly/DataPRforComment

Feedback to: https://www.jiscmail.ac.uk/DATA-PUBLICATION

Please! Tell us what you think

Always happy to get input from others!

#preparde

sarah.callaghan@stfc.ac.uk

@sorcha_ni

data-publication@jiscmail.ac.uk

Workshop on cross-linking between data centres and publishers 30th April 2013 at Rutherford Appleton Laboratory, UK

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