

Data Publication in the Meteorological Sciences: the OJIMS project

Sarah Callaghan, Fiona Hewer, Sam Pepler, Paul Hardaker
and Alan Gadian

JISC



**British Atmospheric
Data Centre**

NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE
NATURAL ENVIRONMENT RESEARCH COUNCIL



Fiona's Red Kite

interpreting weather and climate for your needs



**National Centre for
Atmospheric Science**

NATURAL ENVIRONMENT RESEARCH COUNCIL

Why publish data?

- Scientific publication mainly focuses on the analysis, interpretation and conclusions drawn from a given dataset.
- Examining the raw data that forms the dataset is more difficult, as datasets are usually stored in digital media, in a variety of (proprietary or non-standard) formats.
- Peer-review is generally only applied to the methodology and final conclusions of a piece of work, and not the underlying data itself. But if the conclusions are to stand, the data must be of good quality.
- A process of data publication, involving peer-review of datasets would be of benefit to many sectors of the academic community.

Benefits of data publication

- **Data scientists** have to ensure that the data are of high quality and that the associated metadata and documentation are complete and understandable.
- This leaves little time for the analysis of the data required to produce a paper suitable for journal publication.
- Publishing a dataset in a data journal will provide academic credit to data scientists, and without diverting effort from their primary work on ensuring data quality.

Benefits of data publication (2)

- **Funding organisations** want to get the best possible science for their money.
- Running measurement campaigns is expensive, so the more reuse that can be derived from a dataset, the better.
- Publication in a data journal ensures that the dataset is uploaded to a trusted repository where it will be backed up, archived and curated.
 - won't be vulnerable to bit-rot or being lost/stored on obsolete media.
- The peer-review process reassures the funder that the published dataset is of good quality and that the experiment was carried out appropriately.

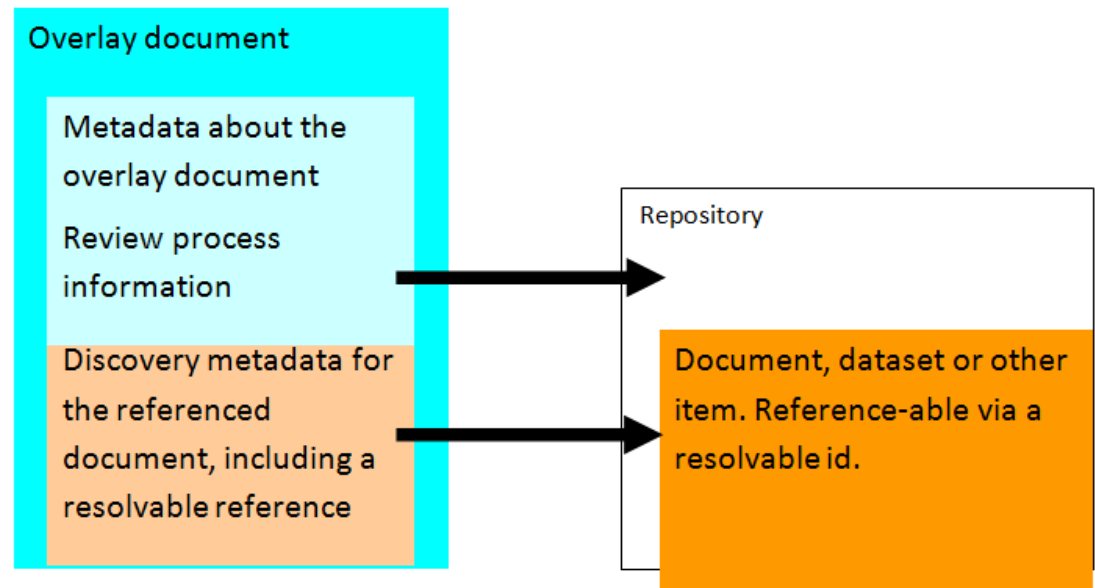
Benefits of data publication (3)

- Demonstrates to the **wider research community** that the datasets are reliable and complete, and therefore the data can be trusted.
- Useful to **researchers outside the immediate field**, as data journals will be a good starting point for information about what sort of data is available and a quick and convenient way of finding out who to contact about accessing them.
 - This will encourage **inter-disciplinary collaboration**, and open up the user base not only for the datasets, but also the data journal and the underlying repositories.
- The availability of published datasets will make it easier to validate conclusions through the reanalysis of those datasets.
- Helps show transparency in the scientific process, improving **public accountability**.

What is an overlay journal?

- Overlay journals sit on top of, and make use of, the content stored in other pre-existing repositories.
- The overlay journal database consists of a number of overlay documents.
- These are structure documents created to annotate another resource with information on the quality of the resource.

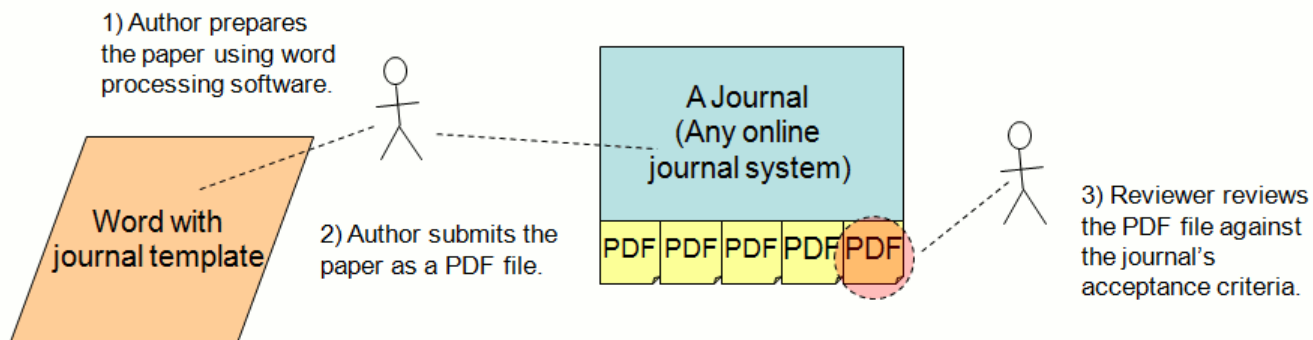
The overlay document has three basic elements:



Overlay documents

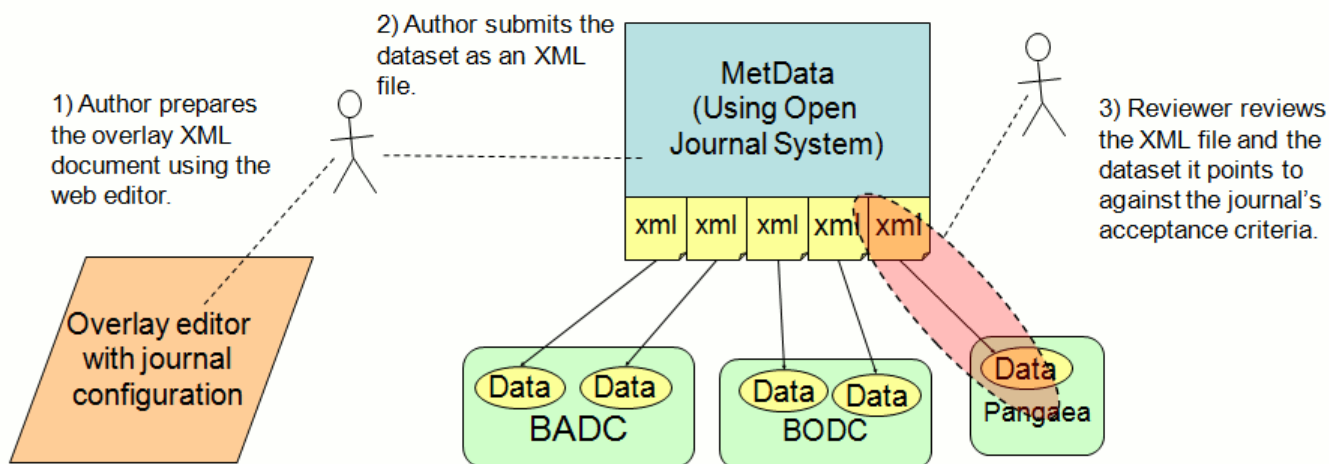
- can be treated as any other documents in an electronic system
- can provide added-value information about the resource they refer to
 - for example, a star-rating given by readers, or a series of review comments.
- Overlay journals do not actually store the datasets they reference
 - they simply store overlay documents about the datasets which contain links to the datasets.
- Overlay journals are not only limited to data publication; they can be applied to other objects which can be stored in a repository, but which might not be so easy to reproduce in print,
 - e.g. video or multimedia files.
- An overlay journal could look at other journal-published and unpublished papers
 - its overlay document could allow users of the overlay journal to award star ratings to the paper to which it refers.
- The underlying technology for such an overlay journal remains the same.

The traditional online journal model



Submission procedure for datasets

Overlay journal model for publishing data



Survey of users

- At the NCAS Conference in Bristol of 8-10 December 2008, conference delegates were invited to complete a survey to investigate the potential implications for the meteorological sciences should a Journal of Meteorological Data be created and operated.
 - **85** delegates from **24** institutions responded to the survey (more than a third of the total number of delegates attending the conference).
 - Respondents were mainly university-based scientists from the fields of atmospheric composition and chemistry, atmospheric physics, dynamical meteorology and climate science.
 - **46%** of respondents were less experienced scientists having less than three years experience of research work.
 - **25%** of respondents had more than 10 years experience.

What they said: Data journal

- **69%** agreed that they would like to access data from an RMetS Journal.
- **67%** agreed that they were more likely to deposit their data in a data centre if they can obtain academic credit through a data journal.
- **Almost all respondents** were users or creators of meteorological data of some kind. Data from experimental campaigns was more commonly used and created by these NCAS delegates than data from General Circulation Models, other numerical models, operational systems, and instrument and observing facilities.
- The evidence of increasing collaboration suggests an **emerging market** for provision of meteorological data to non-meteorologists.
- **91%** of respondents had never heard of the only existing data journal in this area, ESSD <http://earth-system-science-data.net/>, which is aimed at all environmental sciences .

What they said: open access subject repository

The concept of the Open-Access Repository includes both a new subject-based repository and overlay mechanics to search and access it and other repositories.

- **70%** of delegates rated at least one of the repository features as a great idea that they would use.
- **71%** said the most appealing feature of the Repository was a “Single web site to search many repositories” .
- **31%** said they would use a “New repository for pre-print, post print, grey literature, podcasts, web-page archives, datasets and software” . A further **65%** said they might use this feature.
- **12-18%** said they would use features such as “User rating of articles,” “supplementary information e.g. videos, discussion group open forum,” and “user comments and tags for items”

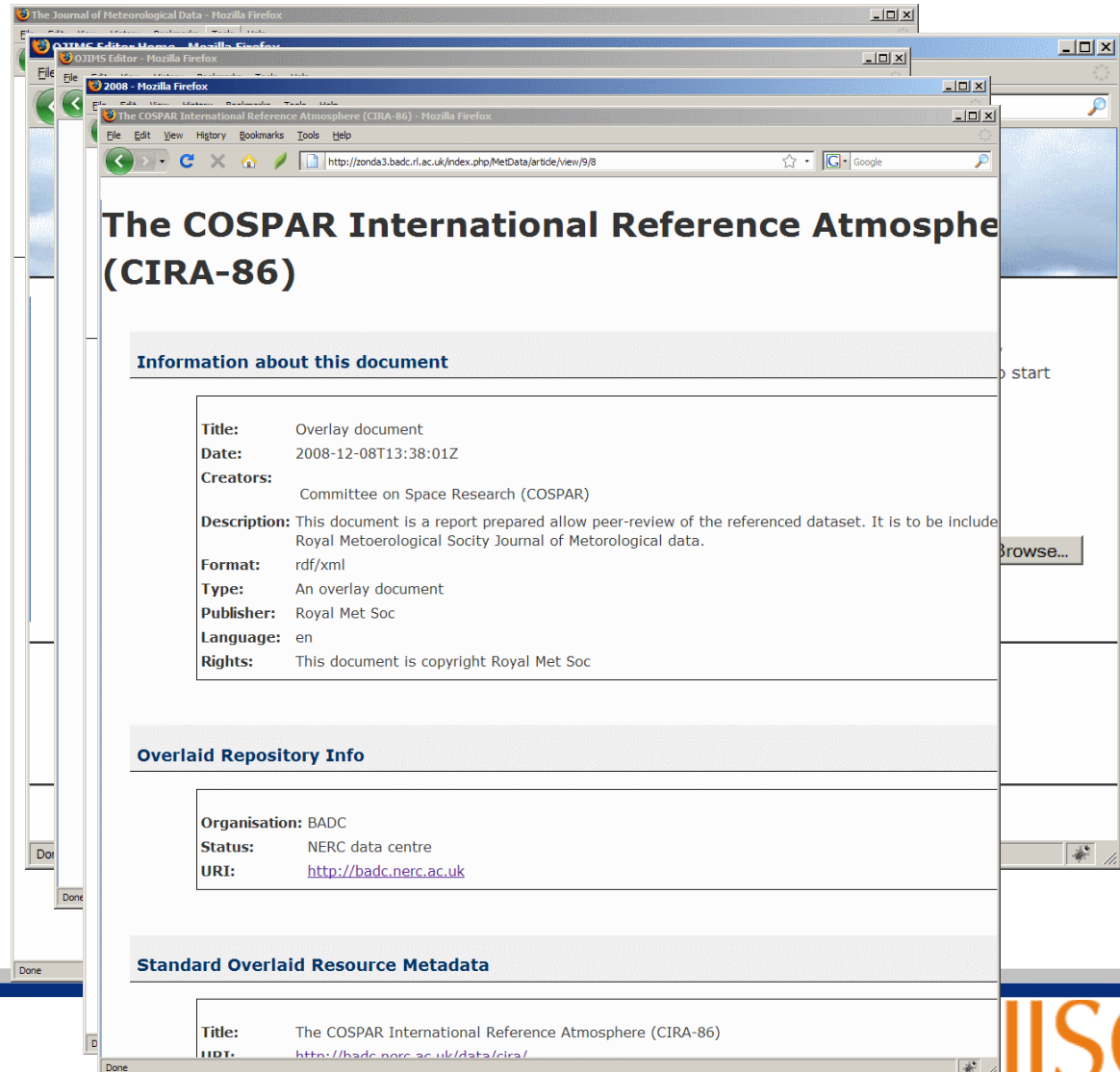
What they said: open access subject repository (2)

- **58%** of respondents created grey literature, which might be stored in the new repository.
- **79%** write papers for peer-reviewed journals.
- **19%** use repositories as their most common method for getting the full text of articles (a further **28%** use them occasionally to do so) and only **38%** use institutional repositories to archive their articles.

The Repository can't become a single, comprehensive source of information for the meteorological sciences unless it attracts unprecedented volumes of deposits in its new repository, or there is a substantial increase in archiving to existing repositories.

Data Journal Demonstrator

- Part of the OJIMS project was developing a demonstration data journal.
- This was done using Open Journal Systems (OJS) because of its open source nature and the ease of adaption



The screenshot displays a Mozilla Firefox browser window showing the OJIMS interface. The main content area displays the title "The COSPAR International Reference Atmosphere (CIRA-86)" and a section titled "Information about this document".

Information about this document

Title:	Overlay document
Date:	2008-12-08T13:38:01Z
Creators:	Committee on Space Research (COSPAR)
Description:	This document is a report prepared allow peer-review of the referenced dataset. It is to be include Royal Metroerological Society Journal of Meteorological data.
Format:	rdf/xml
Type:	An overlay document
Publisher:	Royal Met Soc
Language:	en
Rights:	This document is copyright Royal Met Soc

Overlaid Repository Info

Organisation:	BADC
Status:	NERC data centre
URI:	http://badc.nerc.ac.uk

Standard Overlaid Resource Metadata

Title:	The COSPAR International Reference Atmosphere (CIRA-86)
URI:	http://badc.nerc.ac.uk/data/cira/

Project achievements

- Developed a business case for data journals on behalf of the academic publishing community.
- Developed some of the software technologies required to run a data journal.
- Developed and ran a document repository (<http://cedadocs.badc.rl.ac.uk/>) , which is now a resource freely available to members of the atmospheric science community and a source for documentation about the datasets stored at the BADC.

Conclusions

- It's good for organisations to have a document repository capable of storing grey literature (web pages, project reports, pictures, video etc.) as well as journal papers.
- There is a strong need for a method for publishing data (a data journal) in the meteorological and atmospheric sciences.
 - Publication of data will ensure that datasets are of good quality and will provide data scientists with academic credit for having created the datasets and uploading them to a data repository where they can be archived and curated.
- There is a desire to have an overlay repository which can serve as a single point of search for numerous institutional repositories.
 - At the moment it is felt that the institutional repositories don't have a critical mass of documentation stored in them to make them a key place to search.

Further information

- The OJIMS website with full survey reports:
 - <http://proj.badc.rl.ac.uk/ojims>
- Ariadne articles:
 - “Overlay Journals and Data Publishing in the Meteorological Sciences” <http://www.ariadne.ac.uk/issue60/callaghan-et-al/>
 - “How to Publish Data Using Overlay Journals: The OJIMS Project” <http://www.ariadne.ac.uk/issue61/callaghan-et-al/>
- Sarah Callaghan – sarah.callaghan@stfc.ac.uk