

# Data citation in the Earth Sciences: the UK perspective

Sarah Callaghan\*  
sarah.callaghan@stfc.ac.uk  
@sorcha\_ni

\*and many others, including members of the PREPARDE and NERC data citation and publication project teams and the CODATA working group on data citation

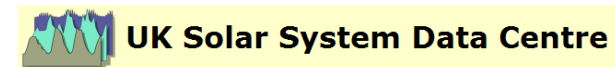
IDCC, San Francisco, 27 Feb 2014

# Who are we and why do we care about data?

The UK's Natural Environment Research Council (NERC) funds six data centres which between them have responsibility for the long-term management of NERC's environmental data holdings.

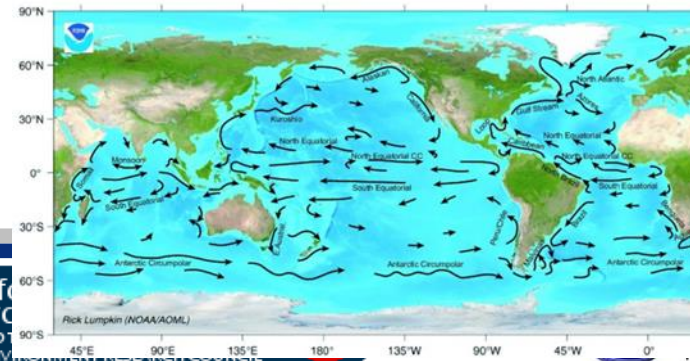
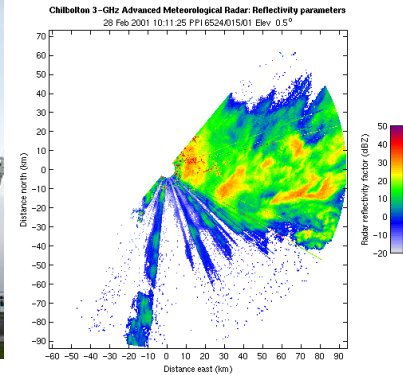
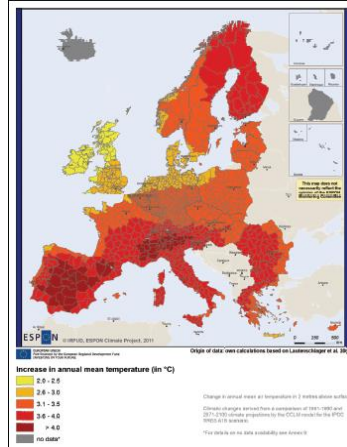
We deal with a variety of environmental measurements, along with the results of model simulations in:

- Atmospheric science
- Earth sciences
- Earth observation
- Marine Science
- Polar Science
- Terrestrial & freshwater science, Hydrology and Bioinformatics



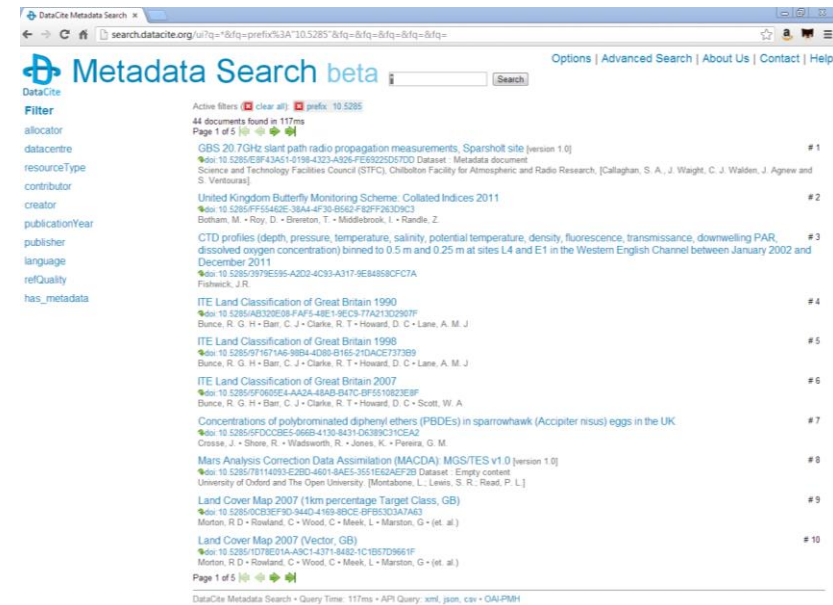
# What types of data do we have?

1. Time series, some still being updated e.g. meteorological measurements
2. Large 4D synthesised datasets, e.g. Climate, Oceanographic, Hydrological and Numerical Weather Prediction model data generated on a supercomputer
3. 2D scans e.g. satellite data, weather radar data
4. 2D snapshots, e.g. cloud camera
5. Traces through a changing medium, e.g. radiosonde launches, aircraft flights, ocean salinity and temperature
6. Datasets consisting of data from multiple instruments as part of the same measurement campaign
7. Physical samples, e.g. fossils



# How we (NERC) cite data

- The NERC data centres have the ability to mint DOIs and assign them to datasets in their archives. We have also produced:
  - guidelines for the data centre on what is an appropriate dataset to cite
  - guidelines for data providers about data citation and the sort of datasets we will cite
  - text in the NERC grants handbook telling grant applicants about data citation
- NERC held datasets have been published in data journals and cited in papers.
- Still plenty of work to do! Not just mechanical processes (e.g. workflows, guidelines) but also changing the culture so that citing and publishing data is the norm.



NERC's guidance on citing data and assigning DOIs can be found at:  
<http://www.nerc.ac.uk/research/sites/data/doi.asp>

# What sort of data can we/will we assign a DOI to?

Dataset has to be:

- Stable (i.e. not going to be modified)
- Complete (i.e. not going to be updated)
- Permanent – by assigning a DOI we're committing to make the dataset available for posterity
- Good quality – by assigning a DOI we're giving it our data centre stamp of approval, saying that it's complete and all the metadata is available

When a dataset is cited that means:

- There will be bitwise fixity
- With no additions or deletions of files
- No changes to the directory structure in the dataset "bundle"

A DOI should point to a *html* representation of some record which describes a *data object* – i.e. a landing page.

Upgrades to versions of data formats will result in new editions of datasets.



Viewing GBS 20.7GHz slant path radio propagation measurements, Chilbolton site

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](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 2011, doi:10.5285/639a3714-bc74-46a6-9026-64931f355e07

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	<a href="#">Apply for access to GBS data from Chilbolton</a>
Download	<a href="#">Data directory for GBS data from Chilbolton</a>
Documentation	<a href="#">DOI for dataset 10.5285/639a3714-bc74-46a6-9026-64931f355e07</a>
Documentation	<a href="#">Data article in Geoscience Data Journal doi:10.1002/gdj3.2</a>

**Associated Data**

Type	Title
Data Production Tool	<a href="#">Chilbolton: GBS receiver</a>
Activity	<a href="#">Chilbolton Facility for Atmospheric and Radio Research (CFARR)</a>
Observation Station	<a href="#">Chilbolton Facility for Atmospheric and Radio Research (CFARR), UK</a>

Dataset catalogue page (and DOI landing page)

Dataset citation

Clickable link to Dataset in the archive



Download PDF

Export

More options...

Search ScienceDirect



Advanced search

## Article outline

 Show full outline

Highlights

Abstract

Keywords

1. Introduction

2. Methods

3. Results and discussion

4. Conclusions

Acknowledgements

Annex 1. Full source information for r...

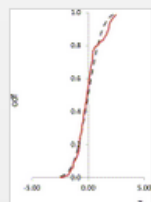
References

## Figures and tables

Table 1

Table 2

Table 3

Learn more about our redesign on our blog. [Click here for details.](#)

ELSEVIER

## Journal of Environmental Radioactivity

Volume 126, December 2013, Pages 314–325



## Evaluating summarised radionuclide concentration ratio datasets for wildlife ☆

M.D. Wood<sup>a</sup>, N.A. Beresford<sup>b</sup>, B.J. Howard<sup>b</sup>, D. Copplestone<sup>c</sup>[Show more](#)<http://dx.doi.org/10.1016/j.jenvrad.2013.07.022>[Open Access](#)

## Highlights

- The approach
- In contrast, t
- We propose
- Available da
- Generic CR

## References

Albrecht et al., 2007 J. Albrecht, M. Abalos, T.M. Rice

**Heavy metal levels in ribbon snakes (*Thamnophis sauritus*) and anuran larvae from the Mobile-Tensaw River Delta, Alabama, USA**

Arch. Environ. Contam. Toxicol., 53 (4) (2007), pp. 647–654

[View Record in Scopus](#) | [Full Text](#) via CrossRef | [Cited By in Scopus \(4\)](#)

Barnett et al., 2013a C.L. Barnett, N.A. Beresford, L.A. Walker, M. Baxter, C. Wells, D. Copplestone

**Element and radionuclide concentrations in representative species of the ICRP's reference animals and plants and associated soils from a forest in north-west England**NERC-Environmental Inf. Data Centre (2013) <http://dx.doi.org/10.5285/e40b53d4-6699-4557-bd55-10d196ece9ea>

Barnett et al., 2013b C.L. Barnett, N.A. Beresford, L.A. Walker, M. Baxter, C. Wells, D. Copplestone

**Transfer parameters for ICRP reference animals and plants collected from a forest ecosystem**

Radiat. Environ. Biophys. (2013) (in press)

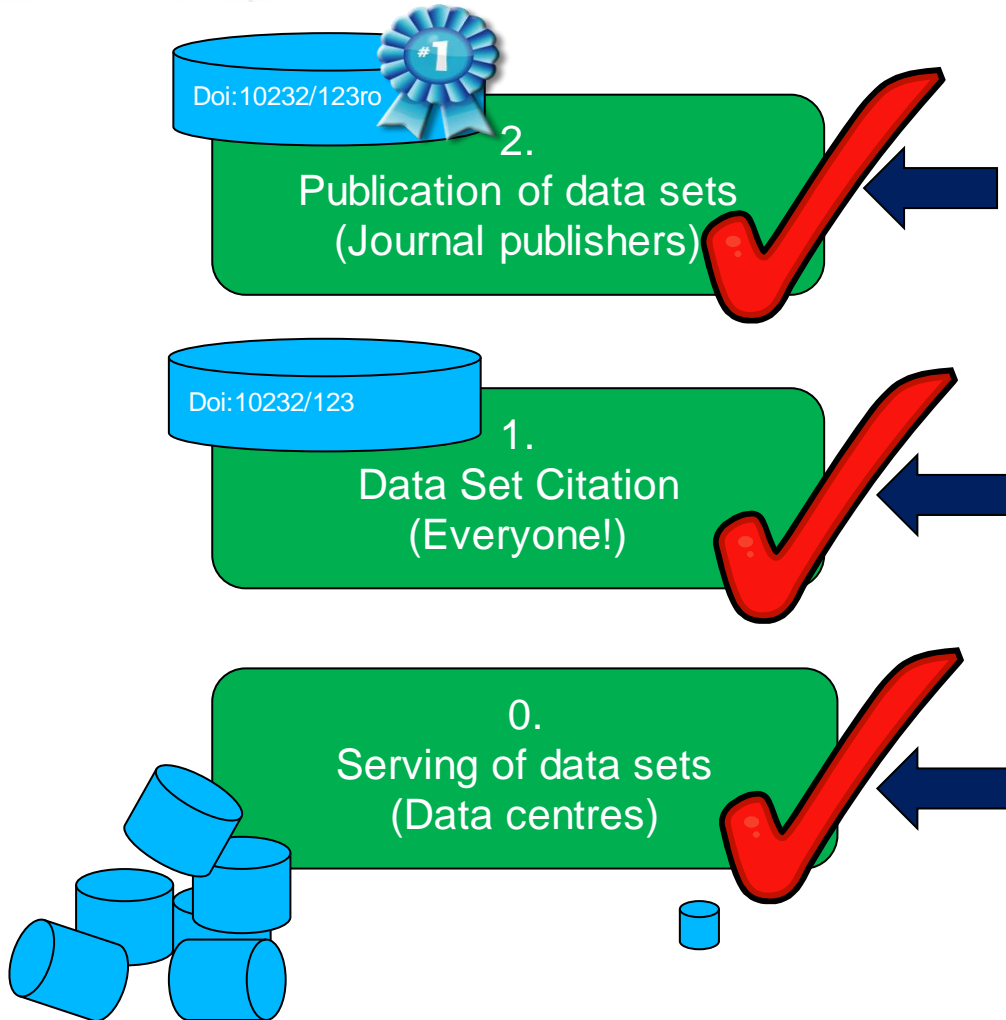
# Another example of a cited dataset



**British Atmospheric  
Data Centre**

NATIONAL CENTRE FOR ATMOSPHERIC SCIENCE  
NATURAL ENVIRONMENT RESEARCH COUNCIL

# What we've done and how we've done it



Data paper has been published in a data journal, linked via DOI to underlying dataset. Formal citations of datasets (also using DOIs) done in standard academic articles.

Can cite using URLs, but we've realised that people don't trust URLs. We're loading DOIs with more meaning than them simply being a persistent identifier – using them to signify completeness and technical quality of the dataset. We're also looking at citation counts as metric for dataset impact.

The day job – take in data and metadata supplied by scientists (often on a on-going basis). Make sure that there is adequate metadata and that the data files are appropriate format. Make it available to other interested parties.