JISC

PREPARDE

D4.1 Roadmap to tighter linking between journal publications and datasets, including data visualisation checks and interface improvements, for review processes and enhanced publications.

D4.2 Worked and operational examples of cross-linking between publications and datasets.

Project Information			
Project Identifier	*To be completed by JISC*		
Project Title	PREPARDE: Peer REview for Publication & Accreditation of Research Data in the Earth sciences		
Project Hashtag	#preparde		
Start Date	1 July 2012	**End Date**	30 June 2013
Lead Institution	University of Leicester		
Project Director	Dr Jonathan Tedds		
Project Manager	Dr Sarah Callaghan		
Contact email	sarah.callaghan@stfc.ac.uk		
Partner Institutions	University of Leicester British Atmospheric Data Centre (BADC) US National Centre for Atmospheric Research (NCAR) California Digital Library (CDL) Digital Curation Centre (DCC) University of Reading Wiley Faculty of 1000 Ltd		
Project Webpage URL	http://proj.badc.rl.ac.uk/preparde/wiki		
Programme Name	*Managing Research Data*		
Programme Manager	Simon Hodson		

Document Information	
Author(s)	Sarah Callaghan

Project Role(s)	Project Manager		
Date		**Filename**	
URL	*If this report is on your project web site*		
Access	☐ Project and JISC internal		☒ General dissemination

Document History		
Version	**Date**	**Comments**
1	21 June 2013	First draft
2	15 July 2013	Second draft
3	12 August 2013	Final draft
4	30 August 2013	Addition of some more examples for the sharing of metadata and "data behind the graph" types of cross-linking.

Crosslinking between journal publications and data centres

Introduction

This document discusses crosslinking between journal publishers and data repositories for the purposes of data publication. It identifies a number of possible routes for crosslinking and discusses the issues and blockers associated with them.

Each topic is broken down into the same subsections, which are:

- Type of crosslinking
- Reason for crosslinking
- Current procedures
- How to implement this crosslink in Geoscience Data Journal (GDJ)
- How to roll out this crosslink to other journals
- Further work and issues

In all cases, the business case for crosslinking needs to be made; in other words the payback as a result of the linking needs to be proportional to the effort involved in making the link. In most cases crosslinking improves visibility of both the dataset catalogue page and the journal article.

This document combines deliverables D4.1 (Roadmap to tighter linking between journal publications and datasets, including data visualisation checks and interface improvements, for review processes and enhanced publications) and D4.2 (Worked and operational examples of cross-linking between publications and datasets) of the PREPARDE project.

1. Crosslinking using DOIs

Type of crosslinking

Data citation, using DOIs as persistent, actionable links.

Reason for crosslinking

For a data journal, permanent linking to the dataset which is the subject of the data article is essential.

Data citation also allows datasets to become part of the formal scientific record, and allows citation metrics for the datasets to be gathered. These metrics can then provide an indication of the dataset’s impact. The citations themselves can be used to track what other uses the dataset is being put to.

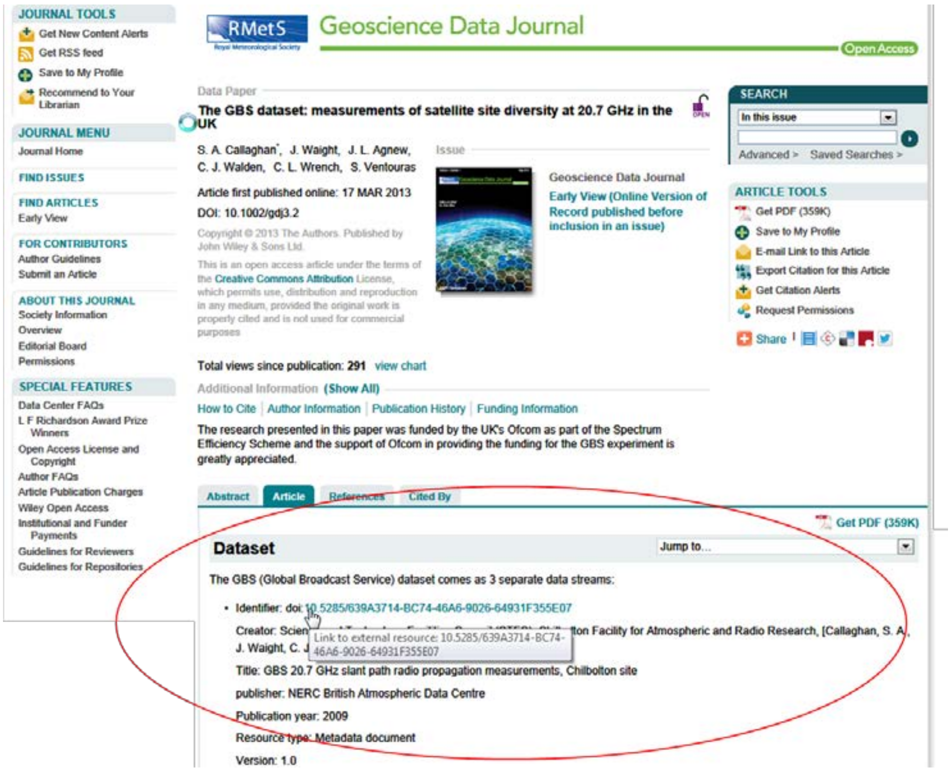
Current procedures

At this time, DOIs in citations are primarily used for journal articles, and the research culture is such that datasets are rarely cited. However, several groups have come together to promote the use of data citation. DataCite is one of these, and acts as a minting authority and registry for the assignment of DOIs to datasets.

The BADC (along with the other NERC environmental data centres, and other national, international, institutional and subject-based repositories) are collaborating with DataCite in order to mint DOIs for datasets held in their archives.

How to implement this crosslink in Geoscience Data Journal (GDJ)

Figure 1 shows the main page of a data paper in GDJ. As can be seen, the dataset details are shown at the start of the Data Paper to make dataset prominent as the focus of the article. Core elements of the DataCite metadata schema are also displayed so that the details of the dataset are both machine- and human-readable.



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

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

Issue

Geoscience Data Journal

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

Total views since publication: 291 view chart

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.

Dataset

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

- Identifier: doi: 10.5285/639A3714-BC74-46A6-9026-64931F355E07

  Creator: Scien ton Facility for Atmospheric and Radio Research, [Callaghan, S. A., J. Waight, C. J

  Link to external resource: 10.5285/639A3714-BC74-46A6-9026-64931F355E07

  Title: GBS 20.7 GHz slant path radio propagation measurements, Chilbolton site

  publisher: NERC British Atmospheric Data Centre

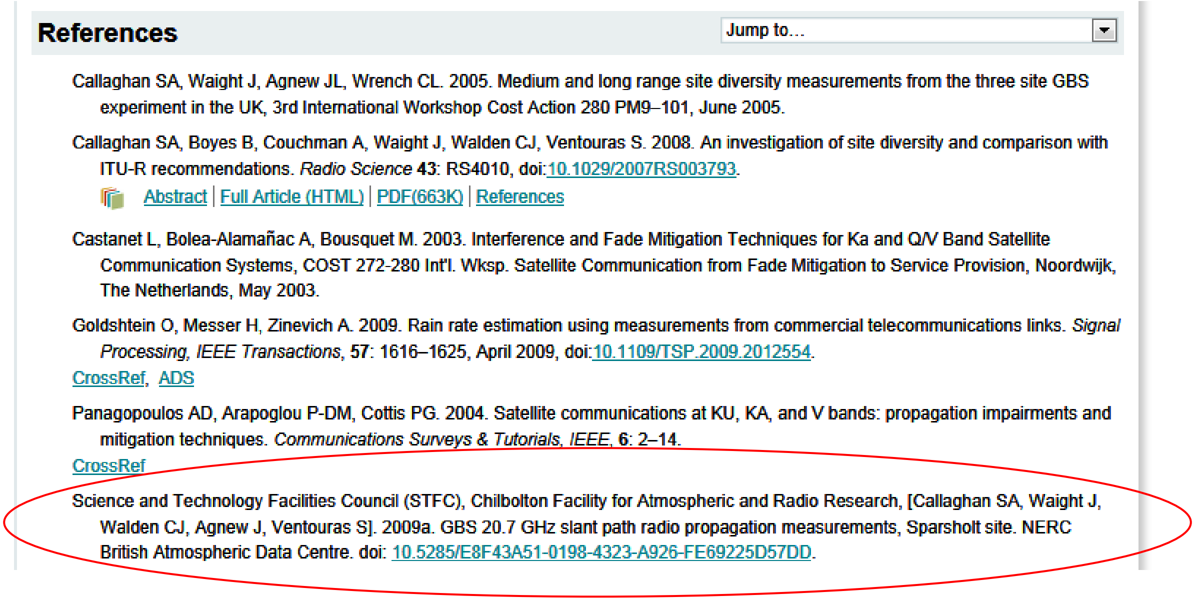
  Publication year: 2009

  Resource type: Metadata document

  Version: 1.0

Figure 1: Screenshot of data article online in GDJ. (http://onlinelibrary.wiley.com/doi/10.1002/gdj3.2/abstract)

Figure 2 shows the dataset details in the reference list. The dataset is included as a full reference in the reference list to give it equal weight to other publications, and to allow it to be picked up by citation tracking mechanisms, which only operate on the references list.



References

Callaghan SA, Waight J, Agnew JL, Wrench CL. 2005. Medium and long range site diversity measurements from the three site GBS experiment in the UK, 3rd International Workshop Cost Action 280 PM9–101, June 2005.

Callaghan SA, Boyes B, Couchman A, Waight J, Walden CJ, Ventouras S. 2008. An investigation of site diversity and comparison with ITU-R recommendations. *Radio Science* **43**: RS4010, doi:10.1029/2007RS003793.
Abstract | Full Article (HTML) | PDF(663K) | References

Castanet L, Bolea-Alamañac A, Bousquet M. 2003. Interference and Fade Mitigation Techniques for Ka and Q/V Band Satellite Communication Systems, COST 272-280 Int'l. Wksp. Satellite Communication from Fade Mitigation to Service Provision, Noordwijk, The Netherlands, May 2003.

Goldshtein O, Messer H, Zinevich A. 2009. Rain rate estimation using measurements from commercial telecommunications links. *Signal Processing, IEEE Transactions*, **57**: 1616–1625, April 2009, doi:10.1109/TSP.2009.2012554.
CrossRef, ADS

Panagopoulos AD, Arapoglou P-DM, Cottis PG. 2004. Satellite communications at KU, KA, and V bands: propagation impairments and mitigation techniques. *Communications Surveys & Tutorials, IEEE*, **6**: 2–14.
CrossRef

Science and Technology Facilities Council (STFC), Chilbolton Facility for Atmospheric and Radio Research, [Callaghan SA, Waight J, Walden CJ, Agnew J, Ventouras S]. 2009a. GBS 20.7 GHz slant path radio propagation measurements, Sparsholt site. NERC British Atmospheric Data Centre. doi: 10.5285/E8F43A51-0198-4323-A926-FE69225D57DD.

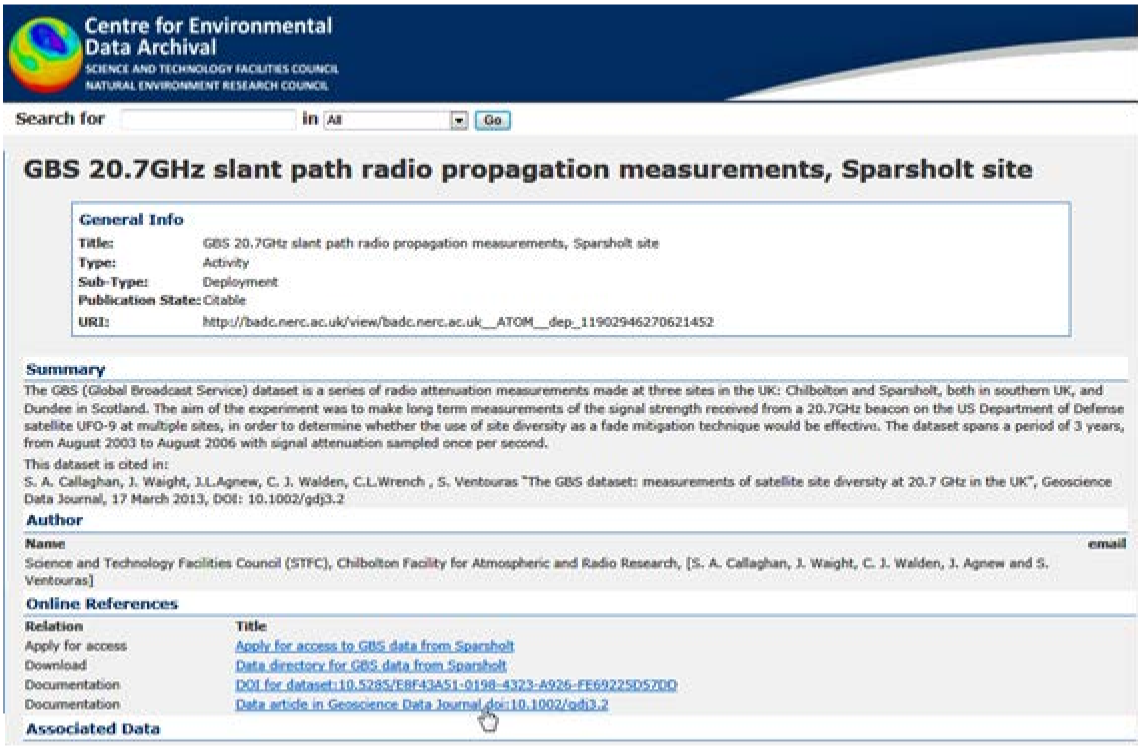
Figure 2: Dataset citation in the reference list. (http://onlinelibrary.wiley.com/doi/10.1002/gdj3.2/references)

The core metadata elements chosen are also appropriate to traditional reference structure, e.g. author, publication year, title, publisher. This follows DataCite recommendations for citation of datasets.

The dataset identifier at the start of the paper is hyperlinked to the dataset landing page using a DOI search. If a DOI is not provided (e.g. other unique identifier such as accession code has been used) then the URL can be hard-coded instead.

In the reference list, the reference is linked in the conventional manner – Wiley Online Library (Wiley's online publishing platform) automatically detects that there is a DOI in the reference and uses the DOI resolver service (http://dx.doi.org) to hyperlink to the cited material (in this case, the dataset). In cases where the dataset does not have a DOI the link would not be inserted by these automatic means, but the reference details should be sufficient for the reader to be able to identify the data source and find this manually. Accession number based URLs would be hyperlinked, however.

To create the return link from the dataset to the data article, Geoscience Data Journal sends an auto-generated email to inform the data repository (in the case of this example, BADC) when the data article is published (providing the DOI of original dataset and DOI of the data article). The data repository then manually (or automatically) updates its dataset landing page with a link to the published Data Paper (figure 3).


Centre for Environmental
Data Archival
SCIENCE AND TECHNOLOGY FACILITIES COUNCIL
NATURAL ENVIRONMENT RESEARCH COUNCIL
Search for
in All
Go

GBS 20.7GHz slant path radio propagation measurements, Sparsholt site

General Info

Title:	GBS 20.7GHz slant path radio propagation measurements, Sparsholt site
Type:	Activity
Sub-Type:	Deployment
Publication State:	Citable
URI:	http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dep_11902946270621452

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.

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 access to GBS data from Sparsholt
Download	Data directory for GBS data from Sparsholt
Documentation	DOI for dataset:10.5285/E8F43A51-0198-4323-A926-FE69225D57DD
Documentation	Data article in Geoscience Data Journal doi:10.1002/gdj3.2

Associated Data

Figure 3: Landing page of the published dataset, showing the citation and link back to the GDJ data article. (doi:10.5285/E8F43A51-0198-4323-A926-FE69225D57DD)

How to roll out this crosslink to other journals

The benefit of this crosslinking process is that it takes advantage of the already existing mechanisms to turn article DOIs into hyperlinks in the online version of the journal article. As DOIs are functionally identical regardless of what they identify, no new tools need to be created or processes put in place to manage or create the crosslinks.

Further work and issues

Note that this email method for informing the data repository that a dataset held in their archive has been cited is ultimately not scalable (figure 4). It is for this reason that we propose a registry to act as an intermediary between data centres and journal publishers (figure 5)

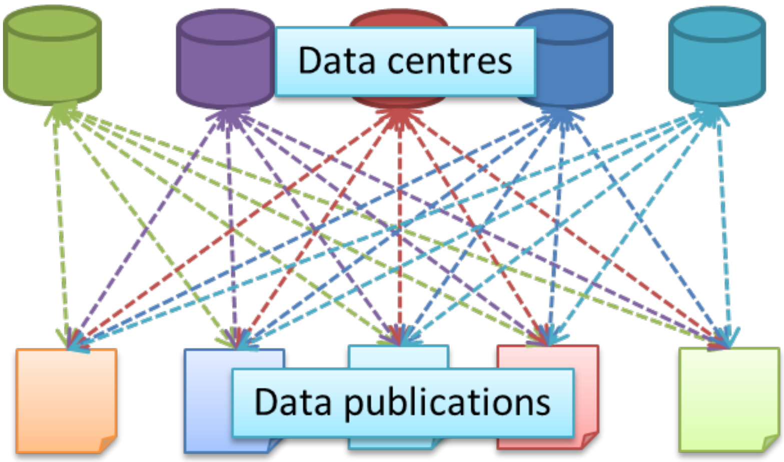


Figure 4: Multiplication of links required for journals and repositories to interact individually.

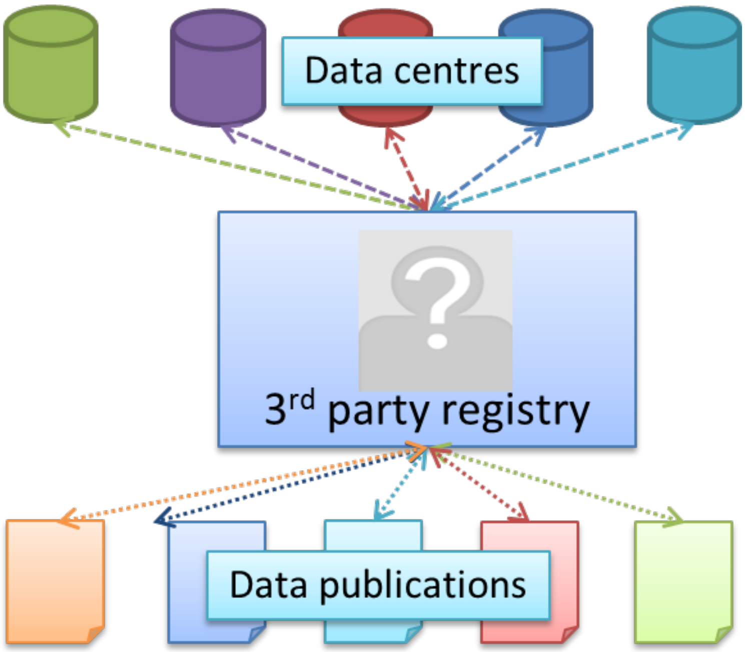


Figure 5: Interactions between data repositories and journals as mediated by a 3rd party registry.

At this time, no such registry exists. However, some of the aspects that would be required are met by the DataCite metadata store (figure 6). DataCite have standardised a set of bibliometric metadata that they require 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 . This search is also available as an API using Solr Search Handler for the API calls, the endpoint is: http://search.datacite.org/api. Given a DOI, a journal can then easily find the DOI standard metadata.

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

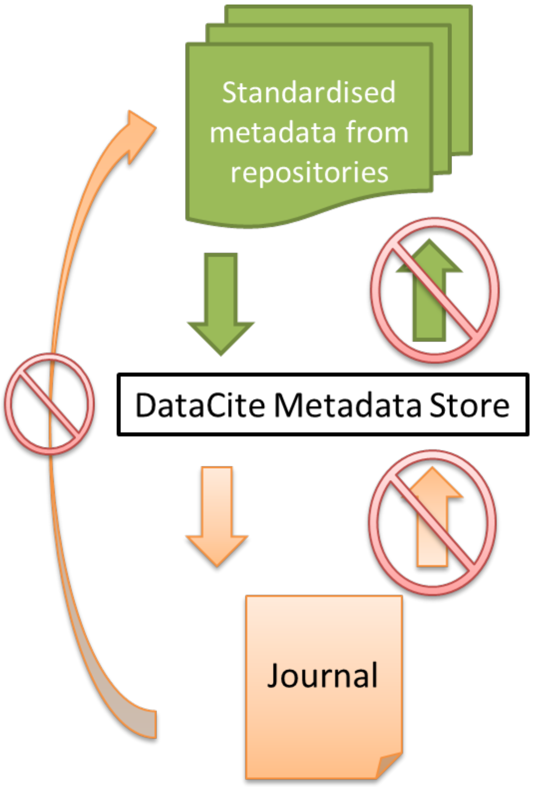


Figure 6: The DataCite metadata store as a potential registry and mediator between journals and data repositories.

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

The OpenAIRE repository (http://www.openaire.eu) has also been suggested as a potential registry to link between datasets and publications, seeing as they are aiming to collect this linking information as part of their core business[1].

There is a need for research on whether or not links from dataset records to data articles are followed by users. This can be done to a certain extent by the journal looking at referrals to identify what proportion of these come from the dataset landing page. It is worth noting that in general, links to pages also help with Google page rankings, improving the discoverability of both the article and the dataset. Note though that Google's current policy means that most regular Google searches

[1] "Creating a robust, participatory service for the cross-linking of peer-reviewed scientific publications and associated datasets is the principal goal of OpenAIREplus." http://www.openaire.eu/en/component/content/article/76-highlights/326-openaireplus-press-release

will demote journal articles in search results. This statement may therefore only be true of Google Scholar searches.

2. Data repository banner ads

Type of crosslinking

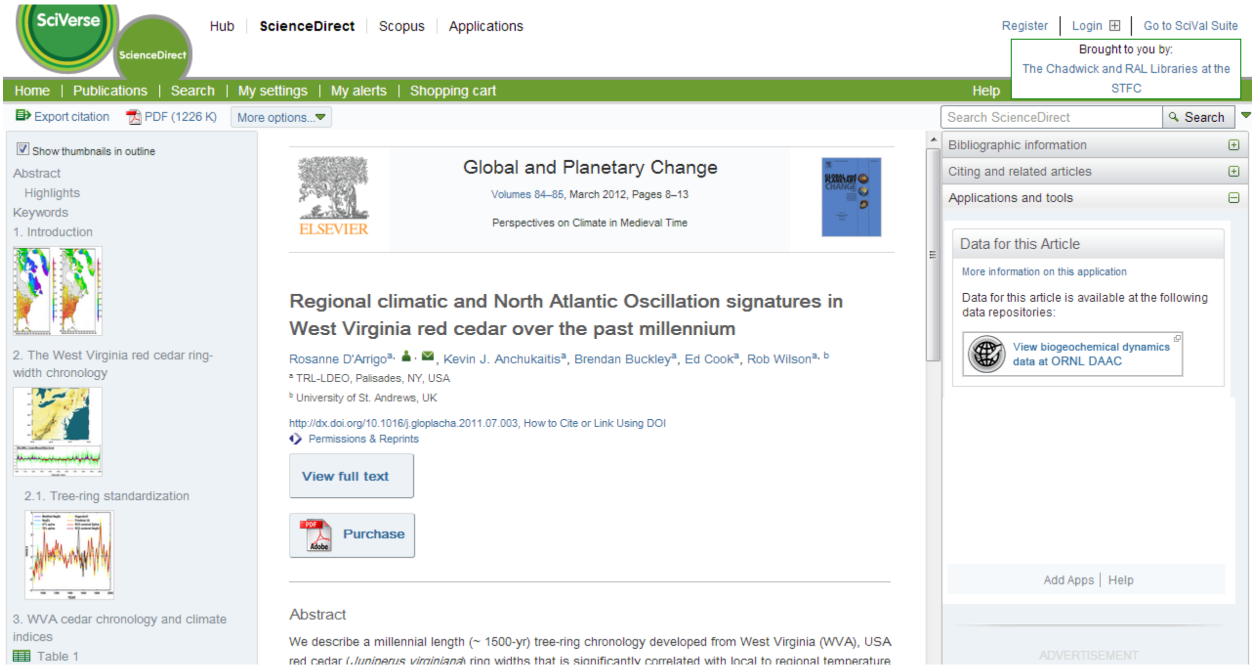
For articles where data repositories are explicitly mentioned (even if a dataset is not formally cited), a banner ad and link to the data repository could be placed on the article page.

Reason for crosslinking

For situations where datasets are referred to in the text, but not explicitly cited, and a data repository is identified as the source of those datasets, a link back to the data repository could be implemented. This would allow readers of the article to get quickly to the repository hosting the data, where they could search for the data or other information.

Current procedures

This has been implemented in some journals though is not common practice. For example, Elsevier collaborates with selected data repositories to show banner links next to relevant articles on ScienceDirect[2] (figure 7).


SciVerse
ScienceDirect
Hub | ScienceDirect | Scopus | Applications
Register | Login ⊞ | Go to SciVal Suite
Home | Publications | Search | My settings | My alerts | Shopping cart
Help
Export citation
PDF (1226 K)
More options...
Search ScienceDirect
Search
☑ Show thumbnails in outline
Abstract
Highlights
Keywords
1. Introduction
2. The West Virginia red cedar ring-width chronology
2.1. Tree-ring standardization
3. WVA cedar chronology and climate indices
Table 1
ELSEVIER
Bibliographic information ⊞
Citing and related articles ⊞
Applications and tools ⊟
Data for this Article
More information on this application
View full text
Purchase
Add Apps | Help
ADVERTISEMENT

Brought to you by:
The Chadwick and RAL Libraries at the STFC

Global and Planetary Change
Volumes 84–85, March 2012, Pages 8–13
Perspectives on Climate in Medieval Time

Regional climatic and North Atlantic Oscillation signatures in West Virginia red cedar over the past millennium

Rosanne D'Arrigo[a], Kevin J. Anchukaitis[a], Brendan Buckley[a], Ed Cook[a], Rob Wilson[a, b]

[a] TRL-LDEO, Palisades, NY, USA

[b] University of St. Andrews, UK

http://dx.doi.org/10.1016/j.gloplacha.2011.07.003, How to Cite or Link Using DOI

Permissions & Reprints

Abstract

We describe a millennial length (~ 1500-yr) tree-ring chronology developed from West Virginia (WVA), USA red cedar (*Juniperus virginiana*) ring widths that is significantly correlated with local to regional temperature

Data for this article is available at the following data repositories:

View biogeochemical dynamics data at ORNL DAAC

Figure 7: example banner link in a ScienceDirect article (http://www.sciencedirect.com/science/article/pii/S0921818111001159)

[2] http://www.elsevier.com/about/content-innovation/database-linking Accessed 15 July 2013

How to implement this crosslink in Geoscience Data Journal *(GDJ)*

The data article would need to be text mined for strings such as flags, accession numbers or the names of data repositories. If a string is found matching the name of a repository, then a pre-generated banner could be added to the paper sidebar.

If the article refers to a dataset using a DOI, then the banner could link directly to the DOI landing page. If not it should link to the main page of the repository.

How to roll out this crosslink to other journals

The mechanism for other journals would be the same as GDJ.

Further work and issues

For efficient text mining, a taxonomy and controlled vocabulary list of repository names and identifiers (such as accession numbers) would need to be created.

Webpage real estate tends to be congested, so research would have to be done to determine whether a fixed ad or a flyover image and link would be more appropriate.

A relationship would be needed between the journal and the repository to ensure that the artwork/logo used for the ad is up to date, and the link used in the banner ad goes to the appropriate place in the repository.

3. Geographical maps

Type of crosslinking

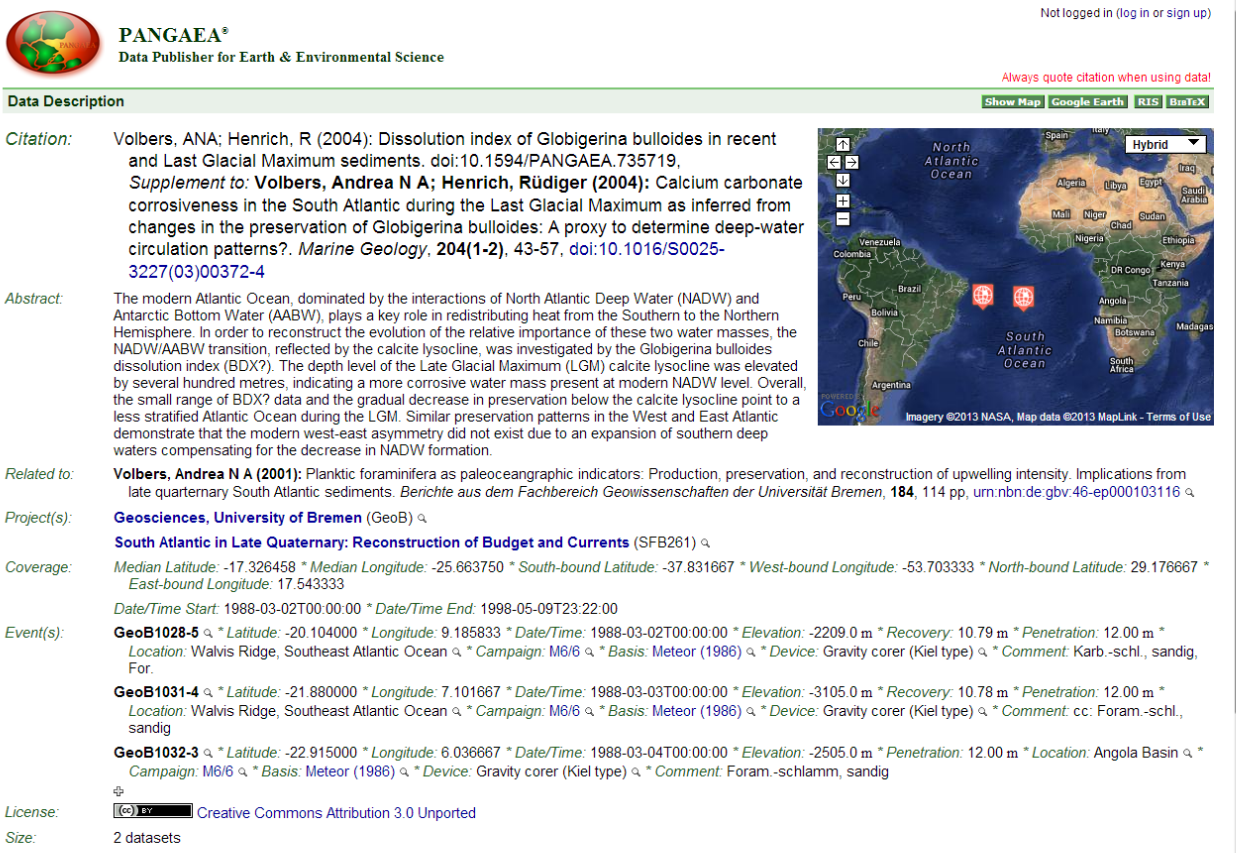
Where geolocation data for a dataset is present in the dataset’s metadata in the repository (or in the DataCite metadata), plotting the locations given on an interactive map.

Reason for crosslinking

This is an added feature for the article’s readers, allowing them to quickly and easily see at a glance where the observational data were measured. For papers referring to multiple datasets with geolocation metadata, this would allow the reader to see the relative positions of the locations.

Current procedures

This form of mapping is currently done by Pangaea (figure 8) as a standard part of their dataset catalogue and DOI landing pages. Note that the map uses Google Maps as its base layer.



Data Description

Citation: Volbers, ANA; Henrich, R (2004): Dissolution index of Globigerina bulloides in recent and Last Glacial Maximum sediments. doi:10.1594/PANGAEA.735719,
Supplement to: **Volbers, Andrea N A; Henrich, Rüdiger (2004):** Calcium carbonate corrosiveness in the South Atlantic during the Last Glacial Maximum as inferred from changes in the preservation of Globigerina bulloides: A proxy to determine deep-water circulation patterns?. *Marine Geology*, **204(1-2)**, 43-57, doi:10.1016/S0025-3227(03)00372-4

Abstract: The modern Atlantic Ocean, dominated by the interactions of North Atlantic Deep Water (NADW) and Antarctic Bottom Water (AABW), plays a key role in redistributing heat from the Southern to the Northern Hemisphere. In order to reconstruct the evolution of the relative importance of these two water masses, the NADW/AABW transition, reflected by the calcite lysocline, was investigated by the Globigerina bulloides dissolution index (BDX?). The depth level of the Late Glacial Maximum (LGM) calcite lysocline was elevated by several hundred metres, indicating a more corrosive water mass present at modern NADW level. Overall, the small range of BDX? data and the gradual decrease in preservation below the calcite lysocline point to a less stratified Atlantic Ocean during the LGM. Similar preservation patterns in the West and East Atlantic demonstrate that the modern west-east asymmetry did not exist due to an expansion of southern deep waters compensating for the decrease in NADW formation.

Related to: **Volbers, Andrea N A (2001):** Planktic foraminifera as paleoceangraphic indicators: Production, preservation, and reconstruction of upwelling intensity. Implications from late quarternary South Atlantic sediments. *Berichte aus dem Fachbereich Geowissenschaften der Universität Bremen*, **184**, 114 pp, urn:nbn:de:gbv:46-ep000103116

Project(s): **Geosciences, University of Bremen** (GeoB)
South Atlantic in Late Quaternary: Reconstruction of Budget and Currents (SFB261)

Coverage: *Median Latitude:* -17.326458 * *Median Longitude:* -25.663750 * *South-bound Latitude:* -37.831667 * *West-bound Longitude:* -53.703333 * *North-bound Latitude:* 29.176667 * *East-bound Longitude:* 17.543333
Date/Time Start: 1988-03-02T00:00:00 * *Date/Time End:* 1998-05-09T23:22:00

Event(s): **GeoB1028-5** * *Latitude:* -20.104000 * *Longitude:* 9.185833 * *Date/Time:* 1988-03-02T00:00:00 * *Elevation:* -2209.0 m * *Recovery:* 10.79 m * *Penetration:* 12.00 m * *Location:* Walvis Ridge, Southeast Atlantic Ocean * *Campaign:* M6/6 * *Basis:* Meteor (1986) * *Device:* Gravity corer (Kiel type) * *Comment:* Karb.-schl., sandig, For.

GeoB1031-4 * *Latitude:* -21.880000 * *Longitude:* 7.101667 * *Date/Time:* 1988-03-03T00:00:00 * *Elevation:* -3105.0 m * *Recovery:* 10.78 m * *Penetration:* 12.00 m * *Location:* Walvis Ridge, Southeast Atlantic Ocean * *Campaign:* M6/6 * *Basis:* Meteor (1986) * *Device:* Gravity corer (Kiel type) * *Comment:* cc: Foram.-schl., sandig

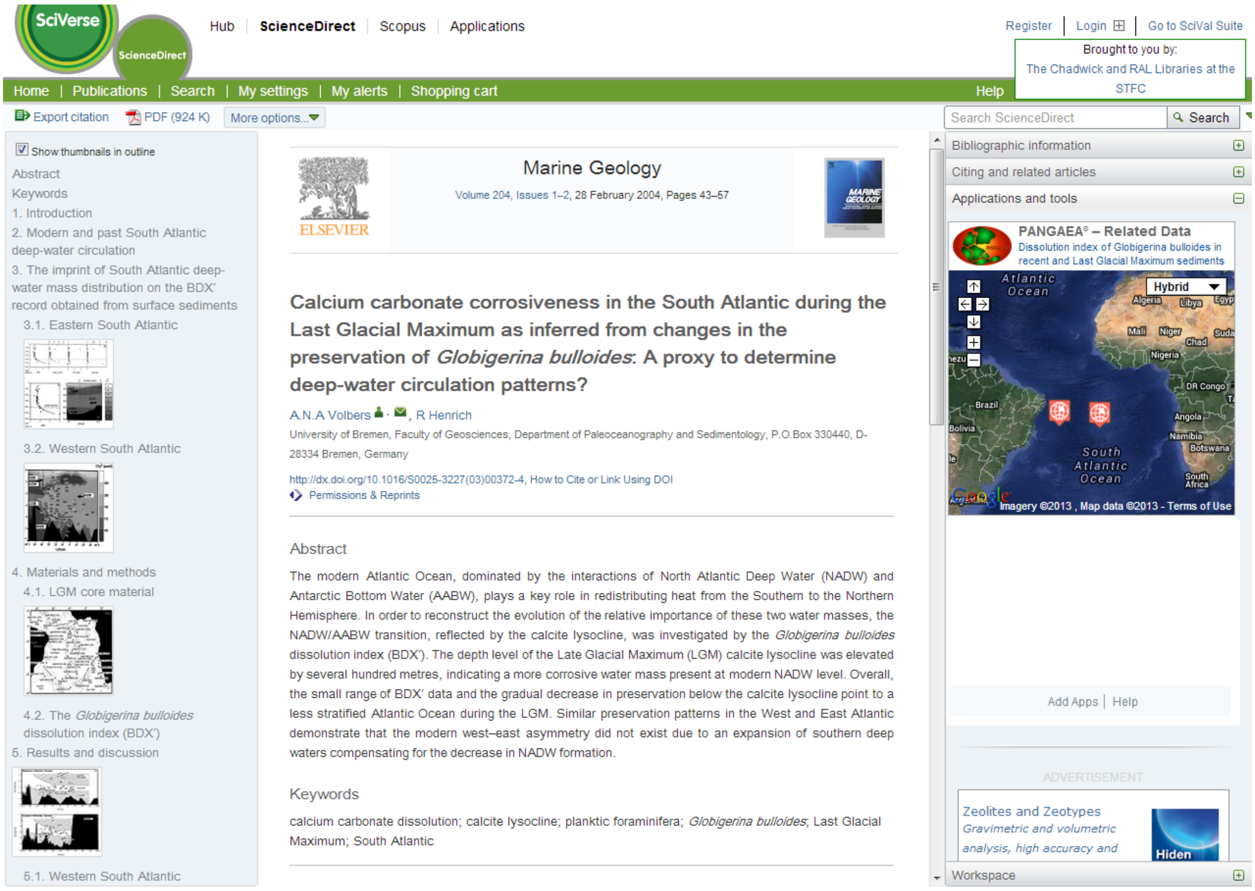
GeoB1032-3 * *Latitude:* -22.915000 * *Longitude:* 6.036667 * *Date/Time:* 1988-03-04T00:00:00 * *Elevation:* -2505.0 m * *Penetration:* 12.00 m * *Location:* Angola Basin * *Campaign:* M6/6 * *Basis:* Meteor (1986) * *Device:* Gravity corer (Kiel type) * *Comment:* Foram.-schlamm, sandig

License: Creative Commons Attribution 3.0 Unported

Size: 2 datasets

Figure 8 – example mapping of geolocation metadata in the Pangaea data repository landing page. (http://doi.pangaea.de/10.1594/PANGAEA.735719)

Figure 9 shows an example of geolocation data from Pangaea being shown in the Elsevier ScienceDirect article webpage. (Note that this dataset is the same as is shown in figure 8.)


SciVerse
ScienceDirect
Hub
ScienceDirect
Scopus
Applications
Register
Login
Go to SciVal Suite
Brought to you by:
The Chadwick and RAL Libraries at the STFC
Home
Publications
Search
My settings
My alerts
Shopping cart
Help
Export citation
PDF (924 K)
More options...
Search ScienceDirect
Search
☑ Show thumbnails in outline
Abstract
Keywords
1. Introduction
2. Modern and past South Atlantic deep-water circulation
3. The imprint of South Atlantic deep-water mass distribution on the BDX' record obtained from surface sediments
3.1. Eastern South Atlantic
3.2. Western South Atlantic
4. Materials and methods
4.1. LGM core material
4.2. The Globigerina bulloides dissolution index (BDX')
5. Results and discussion
5.1. Western South Atlantic
ELSEVIER
Bibliographic information
Citing and related articles
Applications and tools
PANGAEA® – Related Data
Dissolution index of Globigerina bulloides in recent and Last Glacial Maximum sediments
Atlantic Ocean
Hybrid
Algeria
Libya
Egyp
Mali
Niger
Chad
Suda
Nigeria
DR Congo
Brazil
Bolivia
Angola
Namibia
Botswana
South Atlantic Ocean
South Africa
Google
Imagery ©2013 , Map data ©2013 - Terms of Use
Add Apps
Help
ADVERTISEMENT
Zeolites and Zeotypes
Gravimetric and volumetric analysis, high accuracy and
Hiden
Workspace

Marine Geology

Volume 204, Issues 1–2, 28 February 2004, Pages 43–57

Calcium carbonate corrosiveness in the South Atlantic during the Last Glacial Maximum as inferred from changes in the preservation of *Globigerina bulloides*: A proxy to determine deep-water circulation patterns?

A.N.A Volbers, R Henrich

University of Bremen, Faculty of Geosciences, Department of Paleoceanography and Sedimentology, P.O.Box 330440, D-28334 Bremen, Germany

http://dx.doi.org/10.1016/S0025-3227(03)00372-4, How to Cite or Link Using DOI

Permissions & Reprints

Abstract

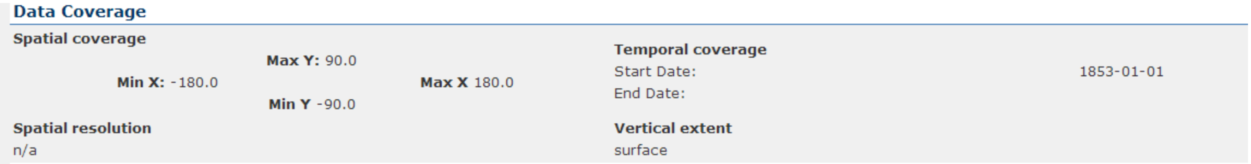
The modern Atlantic Ocean, dominated by the interactions of North Atlantic Deep Water (NADW) and Antarctic Bottom Water (AABW), plays a key role in redistributing heat from the Southern to the Northern Hemisphere. In order to reconstruct the evolution of the relative importance of these two water masses, the NADW/AABW transition, reflected by the calcite lysocline, was investigated by the *Globigerina bulloides* dissolution index (BDX'). The depth level of the Late Glacial Maximum (LGM) calcite lysocline was elevated by several hundred metres, indicating a more corrosive water mass present at modern NADW level. Overall, the small range of BDX' data and the gradual decrease in preservation below the calcite lysocline point to a less stratified Atlantic Ocean during the LGM. Similar preservation patterns in the West and East Atlantic demonstrate that the modern west–east asymmetry did not exist due to an expansion of southern deep waters compensating for the decrease in NADW formation.

Keywords

calcium carbonate dissolution; calcite lysocline; planktic foraminifera; *Globigerina bulloides*; Last Glacial Maximum; South Atlantic

Figure 9: example Elsevier article on ScienceDirect displaying geolocation metadata on a map for the dataset referred to in the article.

The BADC collect and display spatial data in the dataset catalogue pages (figure 9), but do not as yet show it in the form of a map (though plans are underway to do this in the near future).



Data Coverage

Spatial coverage

Max Y: 90.0

Min X: -180.0

Max X 180.0

Min Y -90.0

Temporal coverage

Start Date: 1853-01-01

End Date:

Spatial resolution

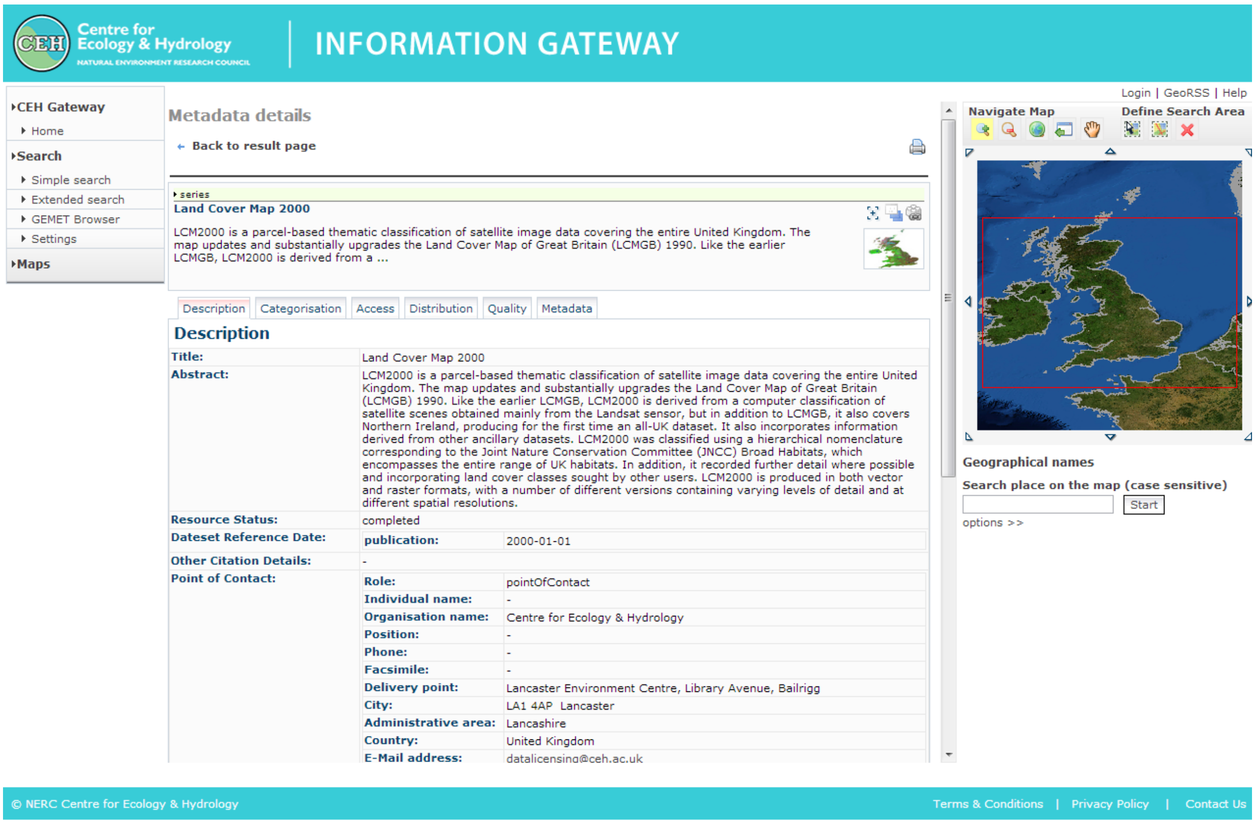
n/a

Vertical extent

surface

Figure 10: Data coverage metadata for the Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current) dataset at http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_ukmo-midas

Another example of this mapping can be seen in the Centre for Ecology and Hydrology’s Information Gateway (figure 11).


Centre for Ecology & Hydrology
NATURAL ENVIRONMENT RESEARCH COUNCIL
INFORMATION GATEWAY
Login | GeoRSS | Help
CEH Gateway
Home
Search
Simple search
Extended search
GEMET Browser
Settings
Maps
Navigate Map
Define Search Area
Geographical names
Search place on the map (case sensitive)
Start
options >>
© NERC Centre for Ecology & Hydrology
Terms & Conditions | Privacy Policy | Contact Us

Metadata details

← Back to result page

series

Land Cover Map 2000

LCM2000 is a parcel-based thematic classification of satellite image data covering the entire United Kingdom. The map updates and substantially upgrades the Land Cover Map of Great Britain (LCMGB) 1990. Like the earlier LCMGB, LCM2000 is derived from a ...

Description | Categorisation | Access | Distribution | Quality | Metadata

Description

Title:	Land Cover Map 2000	
Abstract:	LCM2000 is a parcel-based thematic classification of satellite image data covering the entire United Kingdom. The map updates and substantially upgrades the Land Cover Map of Great Britain (LCMGB) 1990. Like the earlier LCMGB, LCM2000 is derived from a computer classification of satellite scenes obtained mainly from the Landsat sensor, but in addition to LCMGB, it also covers Northern Ireland, producing for the first time an all-UK dataset. It also incorporates information derived from other ancillary datasets. LCM2000 was classified using a hierarchical nomenclature corresponding to the Joint Nature Conservation Committee (JNCC) Broad Habitats, which encompasses the entire range of UK habitats. In addition, it recorded further detail where possible and incorporating land cover classes sought by other users. LCM2000 is produced in both vector and raster formats, with a number of different versions containing varying levels of detail and at different spatial resolutions.	
Resource Status:	completed	
Dateset Reference Date:	**publication:**	2000-01-01
Other Citation Details:	-	
Point of Contact:	**Role:**	pointOfContact
	Individual name:	-
	Organisation name:	Centre for Ecology & Hydrology
	Position:	-
	Phone:	-
	Facsimile:	-
	Delivery point:	Lancaster Environment Centre, Library Avenue, Bailrigg
	City:	LA1 4AP Lancaster
	Administrative area:	Lancashire
	Country:	United Kingdom
	E-Mail address:	datalicensing@ceh.ac.uk

Figure 10: Land Cover Map 2000 metadata and spatial coverage bounding box (on map) from CEH's information gateway.

How to implement this crosslink in Geoscience Data Journal (GDJ)

There are two potential ways of doing this cross-link.

The first option is to query the dataset's geolocation metadata as it is stored in the host repository. Once those metadata have been extracted from the repository for the dataset, they can be ingested as part of the dataset metadata as stored in GDJ, and used to plot the dataset spatial extent on a third party map (e.g. Google maps). The repository would need to provide an API, or some other method for GDJ to query its metadata catalogue. Note too that different repositories have different metadata schemas, so would require GDJ to create multiple methods of collecting the metadata, which would not be scaleable.

The second option involves a recent (and as of the date of this report, unconfirmed) development in the DataCite metadata schema (version 3.0) which recommends the use of the GeoLocation property (with point and box sub-properties). This would enable GDJ to pull geolocation data directly from the DataCite metadata store along with the other metadata properties (such as title, creator, etc.) that it already ingests from there, allowing GDJ to collect geolocation metadata from multiple repositories without having to map from the different repository metadata schemas to a standard GDJ schema. Potentially this could allow the spatial information of datasets from different data repositories to be shown on the same map.

Clicking on the map could send the reader to the dataset landing page, or allow other features such as zooming in on an area. Another use case would be to allow journal readers to search for papers based on the geographic location or spatial extent of the datasets referred to in the papers.

We believe it is best for GDJ to ingest the minimum metadata necessary for citation, and all other metadata to remain in a single source, i.e. the data centre. GJD would then use an API to look up elements of the metadata that it wanted to make use of, for example geolocation metadata to display a map.

How to roll out this crosslink to other journals

This type of crosslink is mainly of use to journals in the geosciences where location information about the dataset is meaningful. However, this interactive viewer idea could be used in a wide range of other scientific fields where an interactive display of some parameters would be of interest to the reader.

Further work and issues

The main concern with this crosslink is the proliferation of different methods to get the required geolocation data from the many different repositories. Standardisation is the key to enabling this, and the new Geolocation property in the DataCite metadata schema is a promising first step. The EU's INSPIRE directive could also be a route for standardisation, though we believe that the DataCite standard is likely to be more accessible and easier to take up.

As with all of these interactive features, care must be taken to ensure that the value to the journal reader outweighs the effort involved in implementing them.

4. Pulling metadata from the data repository into journal workflows

Type of crosslinking

Pre-publication metadata sharing between repositories and journals.

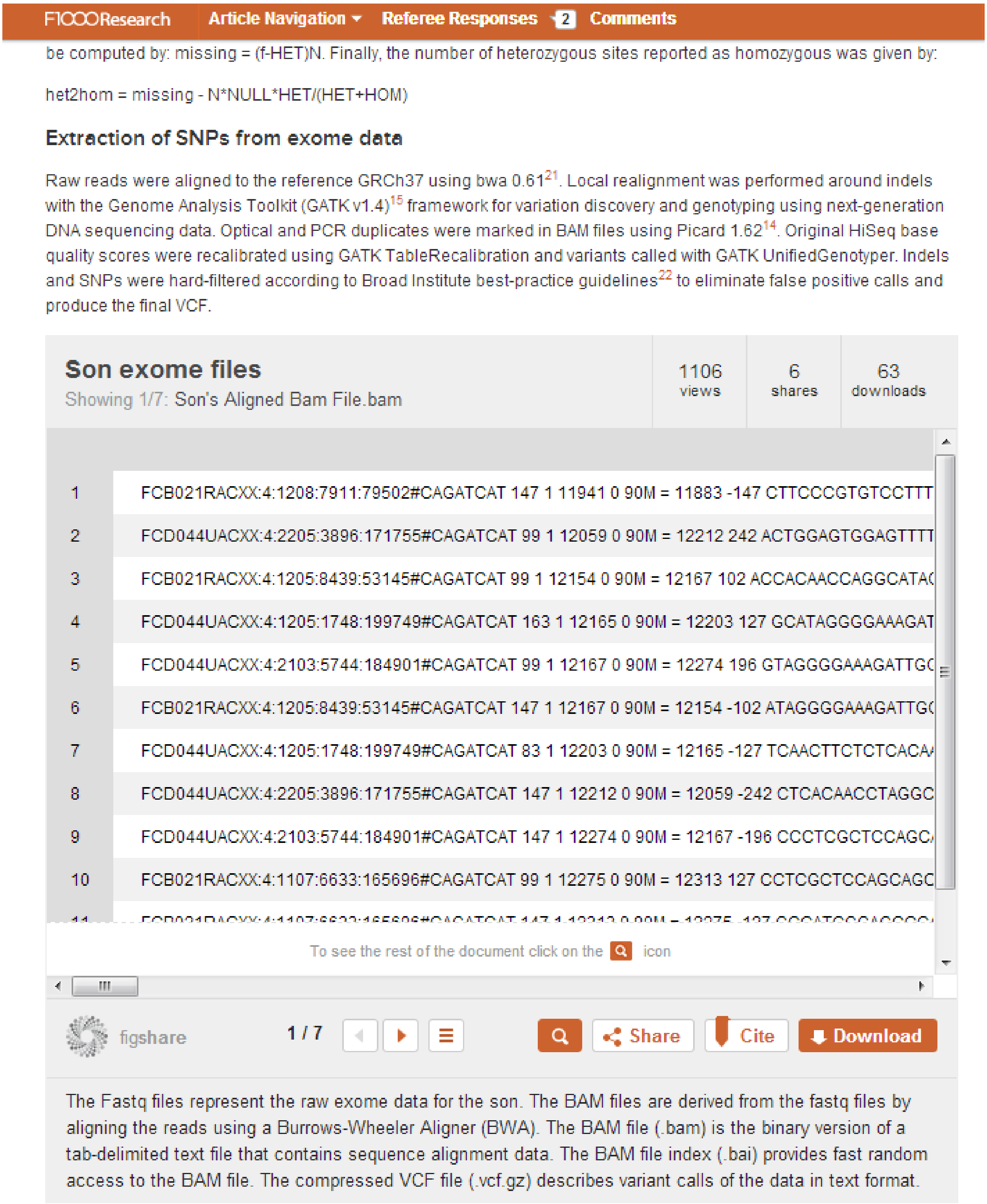
Reason for crosslinking

This sharing of metadata between repositories and journals at the article submission stage is to reduce the effort needed by the article author to input metadata about the dataset into the article. Copying the dataset metadata across to the journal submission system means that the author doesn't have to type the same information into multiple locations (the journal and the repository) and also allows them to check if the metadata stored in the data repository is correct.

It is conceivable that for data repositories that ask for significant amounts of metadata, a tool could be produced which would automatically generate a first draft data article in a highly structured format. At this time however, the interest in data publication is not such that the effort required to generate this tool is warranted. It is also likely that this tool would operate in the repository and would produce a downloadable document suitable for editing in a word processing software and then uploading (in appropriate format) to the journal submission site.

Current procedures

An example of sharing metadata between the repository and the journal can be seen in figure 11, where the figshare widget in the article not only provides access to the dataset used in the article, but also provides repository metadata about the dataset, namely number of views, shares, downloads etc. Figshare provide F1000Research with a line of xml code to install the widget on the paper webpage. This is mainly done by email at the time of writing, but figshare is said to be working on an API for automatic creation of the widget code.



be computed by: missing = (f-HET)N. Finally, the number of heterozygous sites reported as homozygous was given by:

het2hom = missing - N*NULL*HET/(HET+HOM)

Extraction of SNPs from exome data

Raw reads were aligned to the reference GRCh37 using bwa 0.61[21]. Local realignment was performed around indels with the Genome Analysis Toolkit (GATK v1.4)[15] framework for variation discovery and genotyping using next-generation DNA sequencing data. Optical and PCR duplicates were marked in BAM files using Picard 1.62[14]. Original HiSeq base quality scores were recalibrated using GATK TableRecalibration and variants called with GATK UnifiedGenotyper. Indels and SNPs were hard-filtered according to Broad Institute best-practice guidelines[22] to eliminate false positive calls and produce the final VCF.

Son exome files

Showing 1/7: Son's Aligned Bam File.bam

1106	6	63
views	shares	downloads

```
FCB021RACXX:4:1208:7911:79502#CAGATCAT 147 1 11941 0 90M = 11883 -147 CTTCCCGTGTCCTTT
FCD044UACXX:4:2205:3896:171755#CAGATCAT 99 1 12059 0 90M = 12212 242 ACTGGAGTGGAGTTTT
FCB021RACXX:4:1205:8439:53145#CAGATCAT 99 1 12154 0 90M = 12167 102 ACCACAACCAGGCATA(
FCD044UACXX:4:1205:1748:199749#CAGATCAT 163 1 12165 0 90M = 12203 127 GCATAGGGGAAAGAT
FCD044UACXX:4:2103:5744:184901#CAGATCAT 99 1 12167 0 90M = 12274 196 GTAGGGGAAAGATTG(
FCB021RACXX:4:1205:8439:53145#CAGATCAT 147 1 12167 0 90M = 12154 -102 ATAGGGGAAAGATTG(
FCD044UACXX:4:1205:1748:199749#CAGATCAT 83 1 12203 0 90M = 12165 -127 TCAACTTCTCTCACA/
FCD044UACXX:4:2205:3896:171755#CAGATCAT 147 1 12212 0 90M = 12059 -242 CTCACAACCTAGGC
FCD044UACXX:4:2103:5744:184901#CAGATCAT 147 1 12274 0 90M = 12167 -196 CCCTCGCTCCAGC/
FCB021RACXX:4:1107:6633:165696#CAGATCAT 99 1 12275 0 90M = 12313 127 CCTCGCTCCAGCAGC
```

To see the rest of the document click on the 🔍 icon

The Fastq files represent the raw exome data for the son. The BAM files are derived from the fastq files by aligning the reads using a Burrows-Wheeler Aligner (BWA). The BAM file (.bam) is the binary version of a tab-delimited text file that contains sequence alignment data. The BAM file index (.bai) provides fast random access to the BAM file. The compressed VCF file (.vcf.gz) describes variant calls of the data in text format.

Figure 11: Example Figshare widget embedded in an F1000Research paper (http://f1000research.com/articles/1-3/v1) The widget provides access to the data in figshare, as well as providing repository metadata about the dataset, namely number of views, shares, downloads etc.

How to implement this crosslink in Geoscience Data Journal *(GDJ)*

Although not currently done, a simple manual workflow could be implemented. The author could input minimal dataset information such as DOI and the journal's editorial office or production team could use the DOI to locate the metadata and add the necessary information into the journal article.

As a first step, a restricted list of data centres would need to be compiled, and access to this service limited to datasets held in those data centres. The data centre would need to provide an API or other mechanism for the journal to ingest the metadata into the journal submission system. Also, there would need to be a standardised mapping between the repository metadata and the journal metadata, for example core elements of the DataCite metadata schema.

How to roll out this crosslink to other journals

The issues would be the same for other data journals as they are for GDJ.

Further work and issues

This method of crosslinking, though potentially very useful to authors, is the least mature of the list, and would require significant software development time from both the repository and journal sides.

Again, it requires many-to-many relationships to be built up to map the dataset metadata appropriately, which is not scalable in the long term, though a third party registry and common standards for dataset metadata could go a long way to alleviate this. Standards would also allow automatic ingestion and mapping of metadata.

Journal publishers often have multiple editorial systems in place, which are supplied by a third party and in use by other publishers, so making changes to these editorial systems would be difficult and time consuming.

There is also a question of how much dataset metadata reviewers expect to see on the journal site. Potentially, it would be less confusing for the reviewers and editorial staff to see the dataset metadata on the repository site, rather than mixed in with the article metadata.

5. "Data behind the graph"

Type of crosslinking

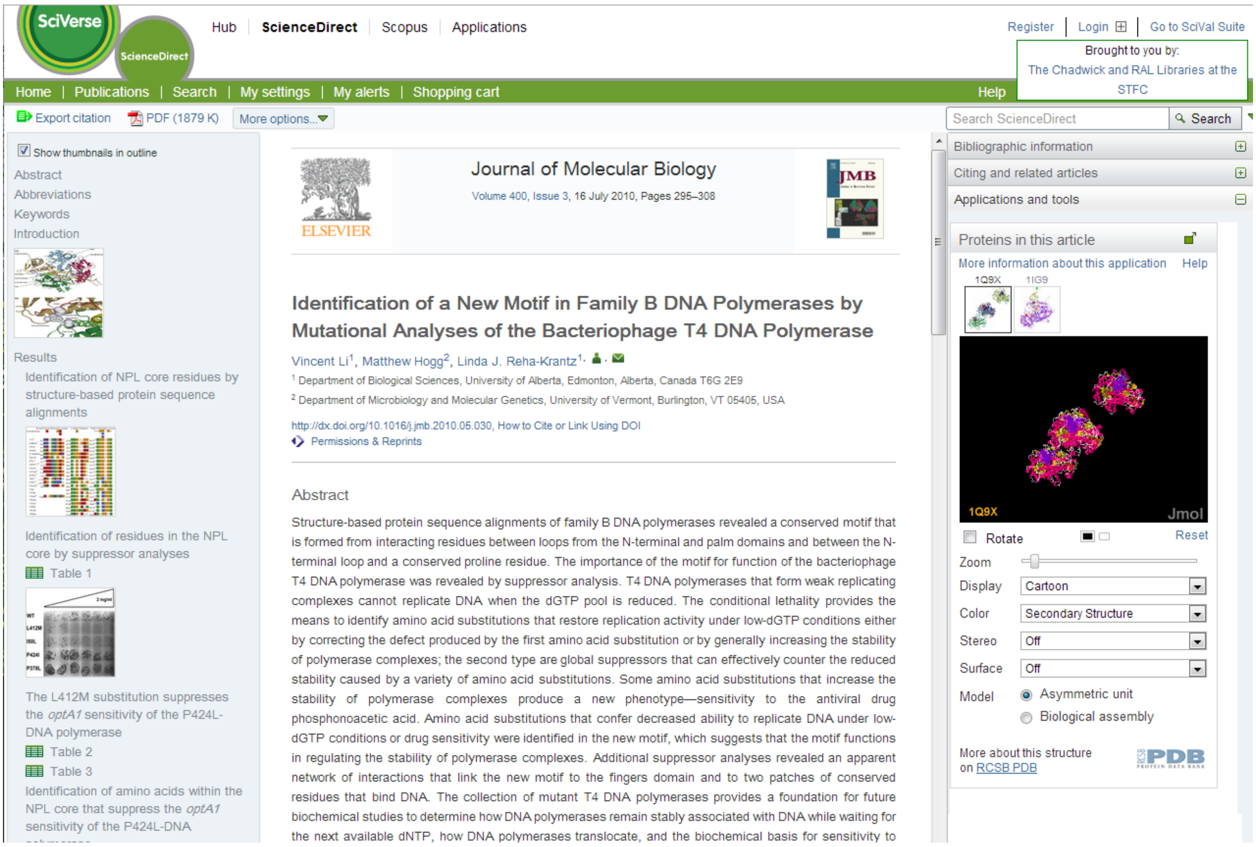
Where there is a plot in the journal article, clicking on it would redirect the reader to the subset of the data used to create that plot.

Reason for crosslinking

One of the main aims of data publication is to make the science in articles more easily verifiable and reproducible by making the data underlying the article visible. Making the data behind the graph more accessible also makes it easier for other researchers to do direct comparisons with previously published results.

Current procedures

Not implemented at this time, though some journals have interactive diagrams or refer readers to supplementary material. See figure 12 for an example of an interactive viewer for proteins in the Journal of Molecular Biology.


SciVerse
ScienceDirect
Hub | ScienceDirect | Scopus | Applications
Register | Login ⊞ | Go to SciVal Suite
Home | Publications | Search | My settings | My alerts | Shopping cart
Help
Export citation
PDF (1879 K)
More options...
Search ScienceDirect
Search
Bibliographic information ⊞
Citing and related articles ⊞
Applications and tools ⊟
ELSEVIER
JMB
1Q9X
1IG9
Jmol
PDB PROTEIN DATA BANK

Brought to you by:
The Chadwick and RAL Libraries at the STFC

☑ Show thumbnails in outline

Abstract
Abbreviations
Keywords
Introduction
Results
Identification of NPL core residues by structure-based protein sequence alignments
Identification of residues in the NPL core by suppressor analyses
Table 1
The L412M substitution suppresses the *optA1* sensitivity of the P424L-DNA polymerase
Table 2
Table 3
Identification of amino acids within the NPL core that suppress the *optA1* sensitivity of the P424L-DNA

Journal of Molecular Biology
Volume 400, Issue 3, 16 July 2010, Pages 295–308

Identification of a New Motif in Family B DNA Polymerases by Mutational Analyses of the Bacteriophage T4 DNA Polymerase

Vincent Li[1], Matthew Hogg[2], Linda J. Reha-Krantz[1],
[1] Department of Biological Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2E9
[2] Department of Microbiology and Molecular Genetics, University of Vermont, Burlington, VT 05405, USA

http://dx.doi.org/10.1016/j.jmb.2010.05.030, How to Cite or Link Using DOI
Permissions & Reprints

Abstract

Structure-based protein sequence alignments of family B DNA polymerases revealed a conserved motif that is formed from interacting residues between loops from the N-terminal and palm domains and between the N-terminal loop and a conserved proline residue. The importance of the motif for function of the bacteriophage T4 DNA polymerase was revealed by suppressor analysis. T4 DNA polymerases that form weak replicating complexes cannot replicate DNA when the dGTP pool is reduced. The conditional lethality provides the means to identify amino acid substitutions that restore replication activity under low-dGTP conditions either by correcting the defect produced by the first amino acid substitution or by generally increasing the stability of polymerase complexes; the second type are global suppressors that can effectively counter the reduced stability caused by a variety of amino acid substitutions. Some amino acid substitutions that increase the stability of polymerase complexes produce a new phenotype—sensitivity to the antiviral drug phosphonoacetic acid. Amino acid substitutions that confer decreased ability to replicate DNA under low-dGTP conditions or drug sensitivity were identified in the new motif, which suggests that the motif functions in regulating the stability of polymerase complexes. Additional suppressor analyses revealed an apparent network of interactions that link the new motif to the fingers domain and to two patches of conserved residues that bind DNA. The collection of mutant T4 DNA polymerases provides a foundation for future biochemical studies to determine how DNA polymerases remain stably associated with DNA while waiting for the next available dNTP, how DNA polymerases translocate, and the biochemical basis for sensitivity to

Proteins in this article
More information about this application Help

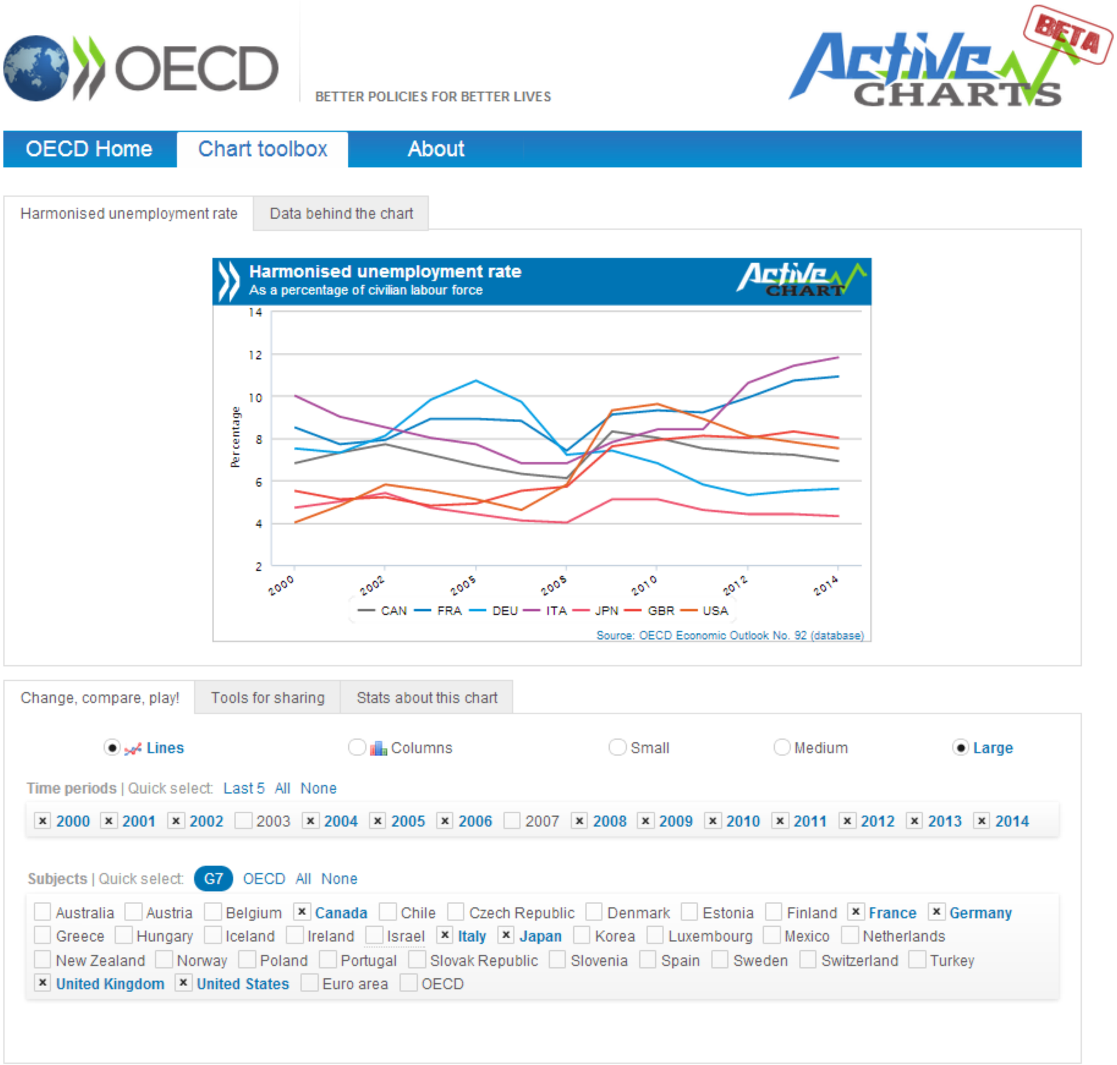
□ Rotate Reset
Zoom
Display: Cartoon
Color: Secondary Structure
Stereo: Off
Surface: Off
Model: ● Asymmetric unit ○ Biological assembly

More about this structure on RCSB PDB

Figure 12: example article with interactive viewer for proteins referred to in the article. (http://www.sciencedirect.com/science/article/pii/S002228361000522X)

The OECD, as part of their ActiveCharts.org project, have provided an interactive site for re-mixing visualising and sharing various statistics from the OECD's databases. Figure 13 shows an example of one of the graphs, while figure 14 shows the raw data used to create it. The remixed graph can be shared and saved as a variety of formats for input into presentations etc. and can also be embedded into webpages.

The functionality in ActiveCharts.org could be integrated into journal paper webpages, though at this time, embedding an active chart into a webpage only provides the published chart – the user has to click on the shared chart's link to get back to ActiveCharts.org, where they can then modify the visualisation of the data.



● Lines ○ Columns ○ Small ○ Medium ● Large

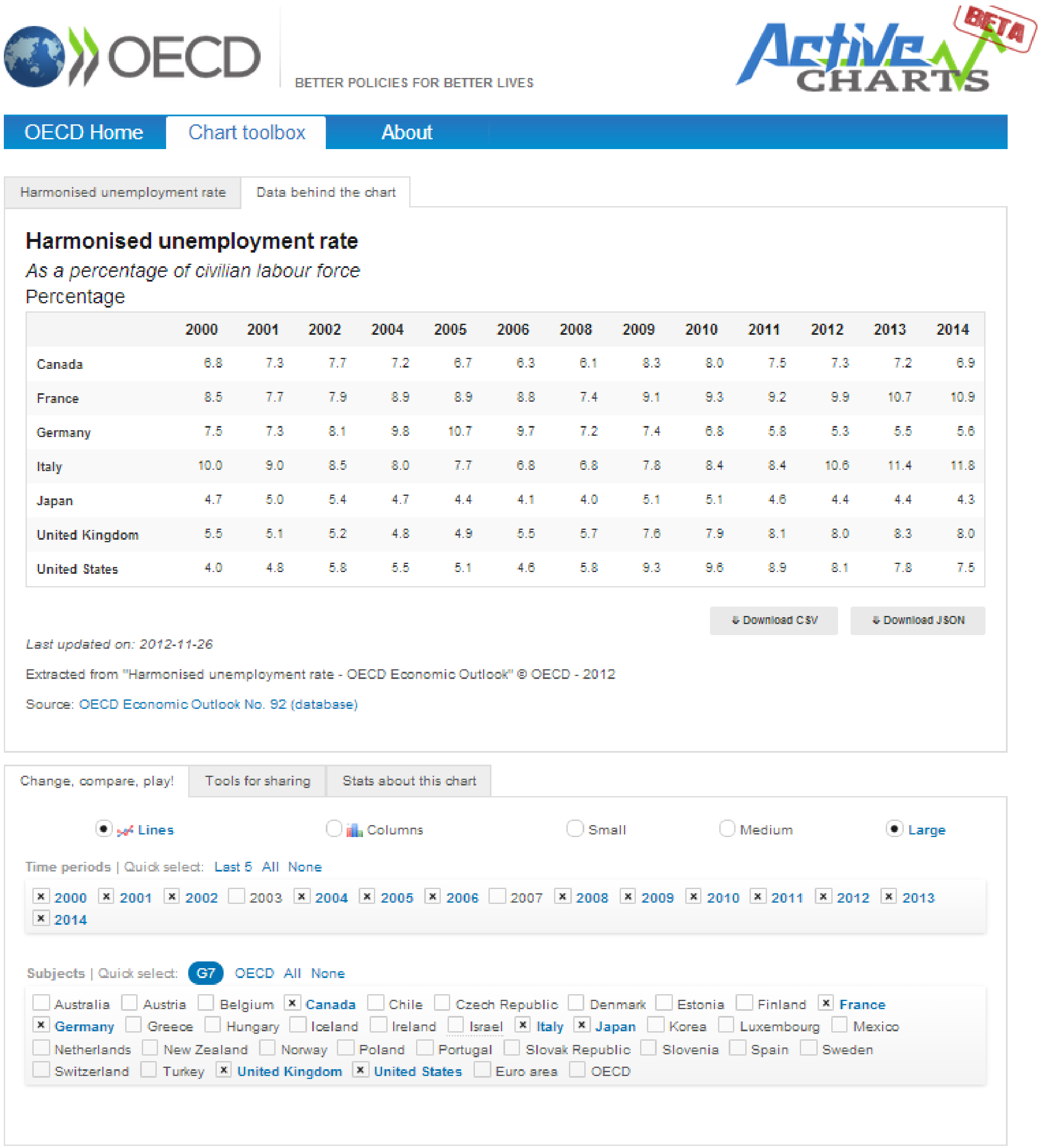
Time periods | Quick select: Last 5 All None

☒ 2000 ☒ 2001 ☒ 2002 ☐ 2003 ☒ 2004 ☒ 2005 ☒ 2006 ☐ 2007 ☒ 2008 ☒ 2009 ☒ 2010 ☒ 2011 ☒ 2012 ☒ 2013 ☒ 2014

Subjects | Quick select: G7 OECD All None

☐ Australia ☐ Austria ☐ Belgium ☒ Canada ☐ Chile ☐ Czech Republic ☐ Denmark ☐ Estonia ☐ Finland ☒ France ☒ Germany
☐ Greece ☐ Hungary ☐ Iceland ☐ Ireland ☐ Israel ☒ Italy ☒ Japan ☐ Korea ☐ Luxembourg ☐ Mexico ☐ Netherlands
☐ New Zealand ☐ Norway ☐ Poland ☐ Portugal ☐ Slovak Republic ☐ Slovenia ☐ Spain ☐ Sweden ☐ Switzerland ☐ Turkey
☒ United Kingdom ☒ United States ☐ Euro area ☐ OECD

Figure 13: Active Chart created and displayed (http://activecharts.org/share/a7dd3bae149b2aba5b8f0d895e00d364) , featuring user selection tick boxes to display/hide data and replot it.



Harmonised unemployment rate

As a percentage of civilian labour force

Percentage

	2000	2001	2002	2004	2005	2006	2008	2009	2010	2011	2012	2013	2014
Canada	6.8	7.3	7.7	7.2	6.7	6.3	6.1	8.3	8.0	7.5	7.3	7.2	6.9
France	8.5	7.7	7.9	8.9	8.9	8.8	7.4	9.1	9.3	9.2	9.9	10.7	10.9
Germany	7.5	7.3	8.1	9.8	10.7	9.7	7.2	7.4	6.8	5.8	5.3	5.5	5.6
Italy	10.0	9.0	8.5	8.0	7.7	6.8	6.8	7.8	8.4	8.4	10.6	11.4	11.8
Japan	4.7	5.0	5.4	4.7	4.4	4.1	4.0	5.1	5.1	4.6	4.4	4.4	4.3
United Kingdom	5.5	5.1	5.2	4.8	4.9	5.5	5.7	7.6	7.9	8.1	8.0	8.3	8.0
United States	4.0	4.8	5.8	5.5	5.1	4.6	5.8	9.3	9.6	8.9	8.1	7.8	7.5

Last updated on: 2012-11-26

Extracted from "Harmonised unemployment rate - OECD Economic Outlook" © OECD - 2012

Source: OECD Economic Outlook No. 92 (database)

● Lines ○ Columns ○ Small ○ Medium ● Large

Time periods | Quick select: Last 5 All None

☒ 2000 ☒ 2001 ☒ 2002 ☐ 2003 ☒ 2004 ☒ 2005 ☒ 2006 ☐ 2007 ☒ 2008 ☒ 2009 ☒ 2010 ☒ 2011 ☒ 2012 ☒ 2013 ☒ 2014

Subjects | Quick select: G7 OECD All None

☐ Australia ☐ Austria ☐ Belgium ☒ Canada ☐ Chile ☐ Czech Republic ☐ Denmark ☐ Estonia ☐ Finland ☒ France ☒ Germany ☐ Greece ☐ Hungary ☐ Iceland ☐ Ireland ☐ Israel ☒ Italy ☒ Japan ☐ Korea ☐ Luxembourg ☐ Mexico ☐ Netherlands ☐ New Zealand ☐ Norway ☐ Poland ☐ Portugal ☐ Slovak Republic ☐ Slovenia ☐ Spain ☐ Sweden ☐ Switzerland ☐ Turkey ☒ United Kingdom ☒ United States ☐ Euro area ☐ OECD

Figure 14: Data behind the graph shown in figure 13.

How to implement this crosslink in Geoscience Data Journal (GDJ)

GDJ is a data journal, and so is primarily concerned with publishing information about datasets without the need for drawing conclusions from the data. For this reason the figures in GDJ are likely to be examples of representative sections of the dataset being published.

At a simple level, for situations where a graph shows, for example, a single day's worth of measurements, it would be possible for the user to click through and download the file containing that day's worth of measurements. This is assuming that the repository can offer download of the data at that resolution, which may not be the case. The link to that particular file would have to be managed carefully, as it may not be appropriate to assign a DOI to a single file.

How to roll out this crosslink to other journals

For other journals, figures are more likely to be generated from processed data stored on the researcher's local workstation, and will probably not be ingested into a data repository in a formal way. Some repositories (e.g. figshare) allow files to be deposited which could contain the data used to generate a single figure, and assign DOIs to those files. In that case, it would be possible to link from the figure to the DOIed data file. It is also possible to imagine a mixed ecosystem in the future, where repository-managed data, cross-linked with research articles, exists alongside small, specific, image-related datasets that are hosted alongside, and much more closely bound to, the articles themselves.

Further work and issues

This method of crosslinking relies on authors being willing and able to submit the exact data subsets they used to create each figure, and therefore involves extra work for them in producing the article, both in producing the subsets, but also archiving them properly.

Figures submitted to a journal for publication will be transformed by the publication process as a matter of course.

Recommendations and Conclusions

This report has outlined a number of potential methods for crosslinking between journals and repositories for the purposes of data publication. The first, linking using DOIs, is the most established and has been demonstrated in GDJ.

There are three main recommendations forthcoming from this work:

1. Standardisation of metadata
2. Use of DOIs and data citation
3. Role of a centralised, 3rd party registry

Standardisation of metadata

For crosslinking to be scalable across multiple journals and data repositories, automatic processes for the linking and sharing of metadata need to be developed. These processes require common standards which are applicable across a wide range of research domains.

The project therefore recommends the use of the DataCite metadata schema as a common metadata kernel for sharing and exchanging dataset metadata.

Use of DOIs and data citation

As DOIs are persistent and actionable links, and are commonly used across the majority of publishers for linking from one paper to another, it is recommended to use them for linking articles to data.

This linking should be done in the context of formal data citation, where other information about the dataset is given, including the creators, title, publishers and date of publication. This project recommends the DataCite citation structure given in the DataCite metadata schema v2.2 (http://schema.datacite.org)

Citations of data should be included in the references list of the article, and the author guidelines for the journal should be updated to request authors cite the datasets used in their article (preferably using DOIs).

Role of a centralised, 3rd party registry

There is a role for a centralised, 3rd party registry and metadata broker in data publication to simplify the process of passing information between data repositories and journals. Instead of the journal having to set up a line of communication with every data repository hosting data mentioned in the journal article (and vice versa), the journal could set up one way of communicating with the central registry. The repositories could then do the same, reducing the effort involved in maintaining multiple links to many different sources.

As yet this registry does not exist, though some existing initiatives (DataCite, OpenAIRE) provide some aspects of the service that this registry would be needed to provide. Although not data related, CrossRef also provide some aspects of this registry service.