

BRITISH  
LIBRARY

## The British Library Perspective on Datasets

Adam Farquhar  
Head of Digital Library Technology  
The British Library

Research Data Manager Forum, Manchester  
30 April 2009

# The British Library:

*‘This is the life blood of research and innovation’*

Science and Innovation Investment Framework 2004-2014, H.M. Treasury (2004)

**Information infrastructure**

**2.23** The growing UK research base must have ready and efficient access to information of all kinds – such as experimental data sets, journals, theses, conference proceedings and patents. **This is the life blood of research and innovation.**

National library of the UK.  
Serves researchers, business, libraries, education & the general public

Collection includes over 2m sound recordings, 5m reports, theses and conference papers, the world’s largest patents collection (c.50m)

Collection fills over 600km of shelving and grows at 11km per year  
30 Tb of digital material growing rapidly

Business and IP Centre:  
Providing inspiration, and enabling protection of creative capital and business development



**Helping people advance knowledge to enrich lives**

The largest document supply service in the world. Secure e-delivery and ‘just in time’ digitisation enables desktop delivery within 2 hours

Generates value to the UK economy each year of 4.4 times public funding

GIA Funding 08/09:  
£94.8m operational,  
£12m capital

Other funding secured 07/08:  
c.£33m

3 main sites in London and Yorkshire. Circa 2,000 staff

## Supporting research

### Science, Technology & Medicine

- Document Supply service provides 1.4m articles/year primarily to scientists
- Renewed engagement with researchers using digital content and online services
- In-depth focus on biomedicine and energy/environment
- Collection includes journals, patents, theses and more, and is updated by some 9,000 articles every day

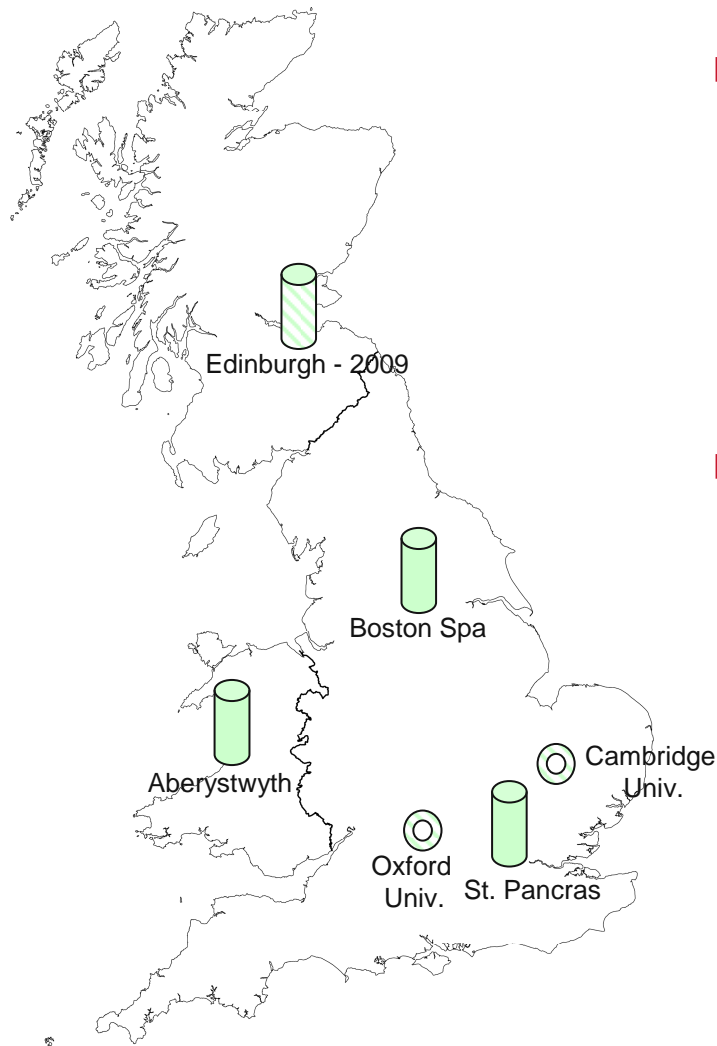
### Social Sciences

- A significant international collection of books, journals, reports, theses, official publications and other materials
- A unique collection of grey literature, of special interest to practitioners and theoreticians
- Research collaboration with ESRC

### Arts & Humanities

- Greatest research collection of its kind in the world
- World-class curatorial expertise by subject, medium and geographical area
- BL has been developing world-leading e-innovations for past decade (e.g. International Dunhuang Project) and building a significant corpus of digitised texts
- Research collaboration with AHRC, British Academy and HEIs

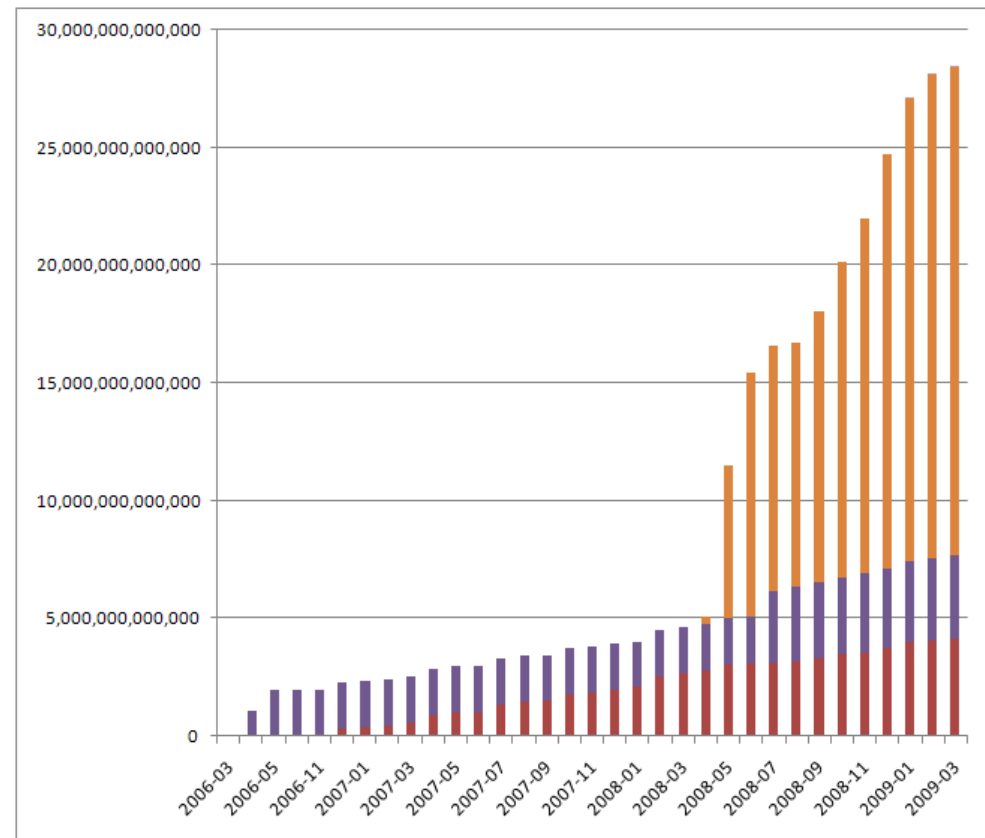
## Building the 21<sup>st</sup> Century Digital Library Infrastructure



- BL Digital library system
  - Large scale, highly resilient digital store (complete online copies at each location)
  - Continuous validation & correction
  - Long term digital storage for BL content & eLegal deposit/distribution
- Long term access (digital preservation)
  - Leading EU-funded digital preservation project 'Planets' (16 UK & international partners)
  - Developing cost models and case studies with UCL ('Life' project)
  - Microsoft Office Open XML in next version of MS Office

## Digital Library System: Progress So Far

- Live Content Streams
  - Sound Archives
  - Voluntary Digital Donations
  - Nineteenth Century Digitised Books
  - Born Digital Newspapers
- Storage
  - >440,000 Digital Items
  - >30 Terabytes of Content
- Coming soon
  - eJournals
  - Digitised Newspapers



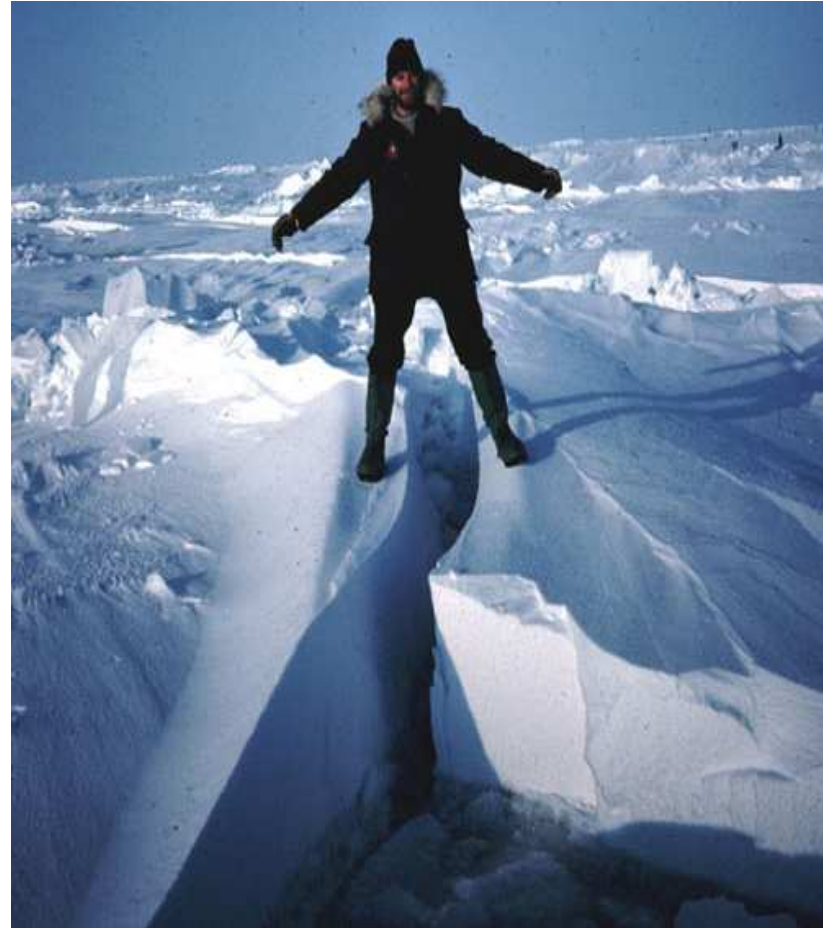
## Datasets: some persistent problems

A widening gap in the scientific record between published research and the data that underlies it

- Published work held by libraries
- Datasets held by data centres
- No widely used method to identify datasets
- No widely used method to cite datasets
- No effective way to link between datasets and articles

As a result, datasets are

- Difficult to discover
- Difficult to access
- Second-class citizens in the scientific record



## Datasets in the scholarly record (OECD White Paper)

- 45% of journal publishers provide access to datasets associated with journal articles they publish
- But there are no rules about how to publish, present, cite, or otherwise catalogue datasets

### Citation

Main mortality estimate  
Settler mortality is calculated as the ratio of mortality rates of European-born soldiers, sailors, and bishops when stationed in colonies. It measures the effects of local diseases on people without inherited or acquired immunities. **Source: Acemoglu et al. (2001), based on Curtin (1989) and other sources.**

### Citation

Tertiary school enrollment: School enrollment, tertiary (% of gross). **Source: Barro and Lee (2000) and their databases**

## Persistent identifiers: A key component of the solution

In order to provide global access to data sets and their metadata through existing catalogues we must have a method to persistently identify them!

This enables:

- Citation
- Increased visibility
- Easier re-use and verification
- Enhanced impact tracking for data producers (Citation Index)
- Support for the Brussels declaration on STM publishing
- Reduced duplication of research data
- Stimulation of new research



## Dataset citation using Digital Object Identifiers (DOIs)

The DOI system offers an easy way to connect the article with the underlying data

Several organisations have started to assign DOIs to datasets

- IUCR, ICPSR, OECD through CrossRef
- Pangea, Mare, and others through TIB

### Dataset

G.Yancheva, N. R. Nowaczyk et al (2007)  
Rock magnetism and X-ray fluorescence spectrometry analyses on sediment cores of the Lake Huguang Maar, Southeast China, PANGAEA

[doi:10.1594/PANGAEA.587840](https://doi.org/10.1594/PANGAEA.587840)

### Article

G. Ycheva, N. R. Nowaczyk et al (2007)  
Influence of the intertropical convergence zone on the East Asian monsoon  
Nature 445, 74-77

[doi:10.1038/nature05431](https://doi.org/10.1038/nature05431)

Cites



## Key facts about DOI

### Usage

- >35m DOIs have been assigned
- >2m resolutions each month

### Organizational

- Not-for-profit International DOI Foundation (IDF)
- Provides social infrastructure
- Includes autonomous registration authorities
- Registration is carried out in co-operation with a publication agent
- Publication agents are responsible for the content (e.g., quality assurance and storage)

### Technical

- A DOI Name is a persistent identifier used to cite and link resources
  - Linked to an object, as an entity – not to its location
  - The location may change, but the DOI remains the same
- The DOI System holds metadata about objects including their URL
- Resolution redirects the user from a DOI name to the URL

## Strengths and weaknesses of DOI

DOIs have some strong advantages

- Accepted by researchers and scientists
- Mature infrastructure
- Put datasets on the same playing field as articles

But perceived as

- Expensive
  - The current IDF business model favours larger registration agencies
- Publisher oriented
  - The largest registration agency is the publisher-oriented CrossRef

## Joint Registration Agency

Organisations with the national science library role are working together to establish a European and global infrastructure to support researchers by providing methods for them to locate, identify, and cite research datasets with confidence

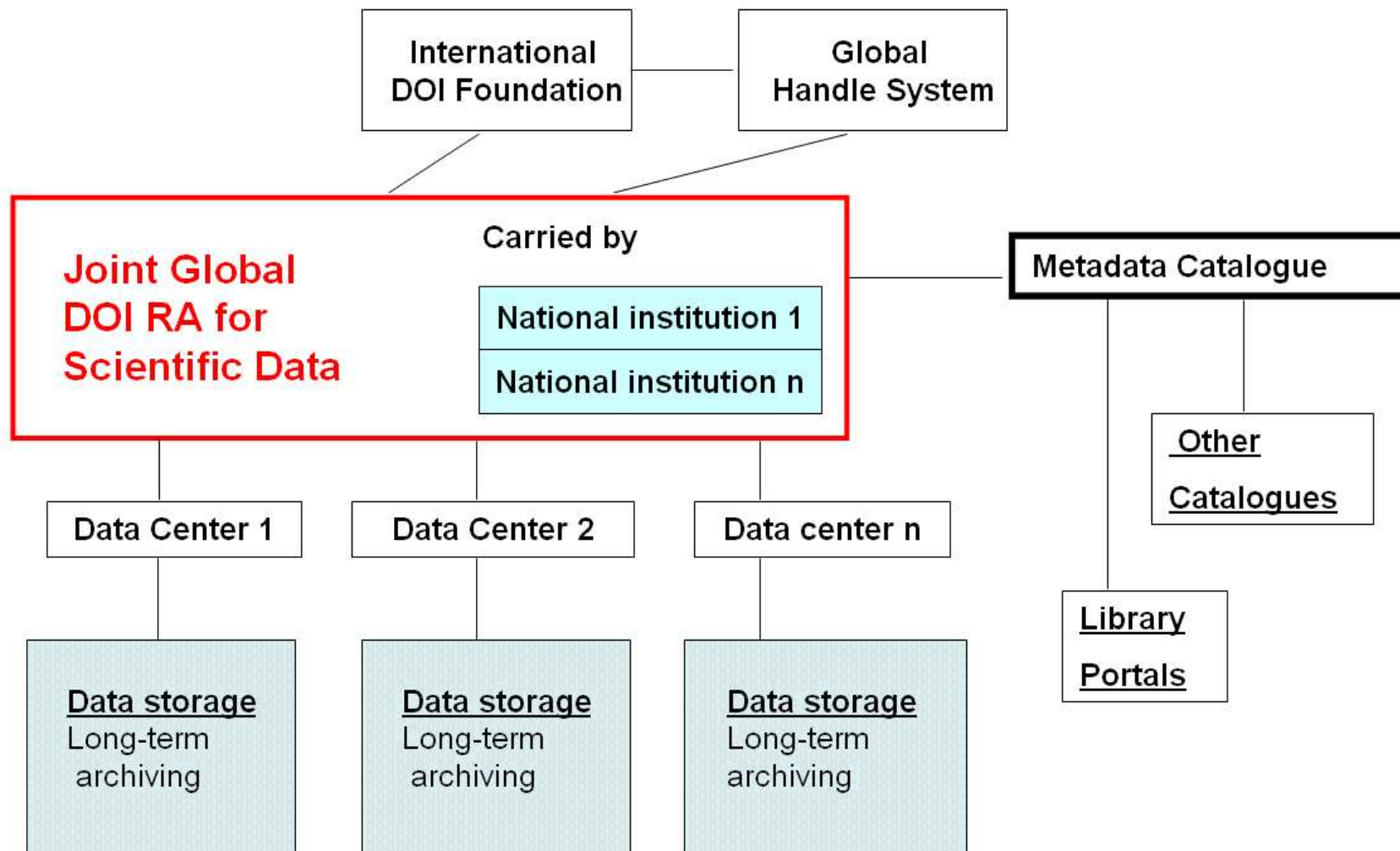
### The JRA

- Maintains the resolution infrastructure
- Maintains a searchable database of metadata
- Manages the identifiers over the long term
- Establish and share best practice

Publishing agents (data centres, research institutes) are responsible for:

- Quality assurance
- Content storage and access
- Creating the identifier
- Creating and updating metadata

# JRA Structure



## Typical workflow (Data Centre)

- Data Centre registers with the JRA
- Data Centre ingests a dataset and assigns an identifier
- Data Centre registers the dataset by submitting an XML file containing relevant bibliographic metadata and the URL for the dataset's access page
  - Metadata drawn from ISO 690-2 for referencing electronic information
    - author
    - title
    - size
    - edition
    - language
    - publisher
    - publishing date
    - publishing place

## Typical workflow (2)

### Author

- Includes citation using the DOI, just like an article

### Reader

- Follows the resolvable link that includes the DOI (or searches for it), just like an article
- Reaches a unique landing page for the dataset
  - Open to every reader
  - Includes the DOI and metadata to help the reader decide if the dataset will help
  - May need to take additional steps to access the dataset

Home + Recent Actions | Browse | Search | My Settings | Alerts | Help

**Quick Search** All fields  Author

search tips Journal/book title  Volume  Issue  Page  Clear  Go  [Advanced Search](#)

**Marine Chemistry**

Volume 95, Issues 1-2, 2 May 2005, Pages 51-63

Font Size:  -  +

Article | [Figures/Tables](#) | [References](#) | [PDF \(931 K\)](#)

[Thumbnails](#) | [Full-Size Images](#)

doi:10.1016/j.marchem.2004.06.040  
[Cite or Link Using DOI](#)

Copyright © 2004 Elsevier B.V. All rights reserved.

## Changes in the concentration of iron in different size fractions during an iron enrichment experiment in the open Southern Ocean

Jun Nishioka<sup>a</sup>, Shigenobu Takeda<sup>b</sup>, Hein J.W. de Baar<sup>c</sup>, Peter L. Croot<sup>c, d</sup>, Marie Boye<sup>c</sup>, Patrick Laan<sup>c</sup> and Klaas R. Timmermans<sup>c</sup>

<sup>a</sup>Central Research Institute of Electric Power Industry, 1646 Abiko, Abiko-shi, Chiba 270-1194, Japan

<sup>b</sup>Department of Aquatic Bioscience, University of Tokyo, Bunkyo, Tokyo 113-8657, Japan

<sup>c</sup>Royal Netherlands Institute for Sea Research, 1790 AB Den Burg, Texel, The Netherlands

<sup>d</sup>Marine Biogeochemie, Institute für Meereskunde und Universität Kiel, D-24105 Kiel, Germany

Received 2 September 2003; revised 17 March 2004; accepted 16 June 2004. Available online 30 December 2004.

### Abstract

An in situ iron enrichment experiment was carried out in the Southern Ocean Polar Frontal Zone and fertilized a patch of water within an eddy of the Antarctic Circumpolar Current (EisenEx, Nov. 2000). During the experiment, a physical speciation technique was used for iron analysis in order to understand the changes in iron distribution and size-fractionations, including soluble Fe (<200 kDa), colloidal Fe (200 kDa–0.2 μm) and labile particle Fe (>0.2 μm), throughout the development of the phytoplankton bloom.

Prior to the first infusion of iron, dissolved (<0.2 μm) iron concentrations in the ambient surface seawater were extremely low (0.06±0.015 nM) with colloidal iron being a minor fraction. For the iron addition, an acidified FeSO<sub>4</sub> solution was released three times over a 23-day period to the eddy. High levels of dissolved iron concentrations (2.0±1.1 nM) were measured in the surface water until 4 days after the first iron infusion. After every iron infusion, when high iron concentrations were observed before storm events, there was a significant correlation between colloidal and dissolved iron concentrations ([Colloidal Fe]=0.7627[Dissolved Fe]+0.0519, R<sup>2</sup>=0.9346). These results indicate that a roughly constant proportion of colloidal vs. dissolved iron was observed after iron infusion (~76%). Storm events caused a significant decrease in iron concentrations (<0.61 nM in dissolved iron) and changed the proportions of the three iron size-fractions (soluble, colloidal and labile particle). The changes in each iron size-fraction indicate that colloidal iron was eliminated from surface mixed layer more easily than particulate and soluble fractions. Therefore, particle and soluble iron efficiently remain in the mixed layer, probably due to the presence of suspended particles and naturally dissolved organic ligands. Our data suggest that iron removal through colloidal aggregation during phytoplankton bloom should be considered in the oceanic iron cycle.

### Article Toolbox

- Download PDF
- E-mail Article
- Cited By
- Save as Citation Alert
- Citation Feed
- Export Citation
- Add to my Quick Links
- Add to **collab**
- Permissions & Reprints
- Cited By in Scopus (16)

### Related Articles in ScienceDirect

- Major deviations of iron complexation during 22 days of...  
*Marine Chemistry*
- Iron cycling and nutrient-limitation patterns in surfac...  
*Deep Sea Research Part II: Topical Studies in Oceanogra...*
- Net and gross O<sub>2</sub> production in the southern ocean from ...  
*Deep Sea Research Part I: Oceanographic Research Papers*
- Modelling ice-edge phytoplankton bloom in the Scotia-We...  
*Journal of Marine Systems*
- Insights into Southern Ocean carbon export from the [de...  
*Deep Sea Research Part II: Topical Studies in Oceanogra...*

[View More Related Articles](#)

Supplementary Data

[View Record in Scopus](#)

### The research collaboration tool

- No user rating
- No user tags yet



Nishioka, J et al. (2008): Profiles of iron concentration from GoFlow bottles during the CARUSO-EISENEX experiment - M... Oost Ende is logged in  
Logout

Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten ...

 **PANGAEA®**  
Publishing Network for Geoscientific & Environmental Data Always quote citation when using data!

You are not logged in (LOG IN)

Advanced Search Font Size: - +

**Data Description** RIS BIBTEX

**Citation:** Nishioka, J et al. (2008): Profiles of iron concentration from GoFlow bottles during the CARUSO-EISENEX experiment, doi:10.1594/PANGAEA.701305,  
*Supplement to: Nishioka, Jun; Takeda, Shigenobu; de Baar, Hein JW; Croot, Peter L; Boyé, Marie; Laan, Patrick; Timmermans, Klaas R (2005): Changes in the concentration of iron in different size fractions during an iron enrichment experiment in the open Southern Ocean, Marine Chemistry, 95(1-2), 51-63, doi:10.1016/j.marchem.2004.06.040*

**Reference(s):** Boyé, Marie; Nishioka, Jun; Croot, Peter L; Laan, Patrick; Timmermans, Klaas R; de Baar, Hein JW (2005): Major deviations of iron complexation during 22 days of a mesoscale iron enrichment in the open Southern Ocean, *Marine Chemistry*, 96(3-4), 257-271, doi:10.1016/j.marchem.2005.02.002  
Croot, Peter L; Laan, Patrick; Nishioka, Jun; Strass, Volker; Cisewski, Boris; Boyé, Marie; Timmermans, Klaas R; Bellerby, Richard G J; Goldson, Laura; Nightingale, Philip D; de Baar, Hein JW (2005): Spatial and temporal distribution of Fe(II) and H2O2 during EisenEx, an open ocean mesoscale iron enrichment. *Marine Chemistry* 95(1-2)

Fertig

**Abstract**

An in situ iron enrichment experiment was carried out in the Southern Ocean Polar Frontal Zone and fertilized a patch of water within an eddy of the Antarctic Circumpolar Current (EisenEx, Nov. 2000). During the experiment, a physical speciation technique was used for iron analysis in order to understand the changes in iron distribution and size-fractionations, including soluble Fe (<200 kDa), colloidal Fe (200 kDa–0.2 µm) and labile particle Fe (>0.2 µm), throughout the development of the phytoplankton bloom.

Prior to the first infusion of iron, dissolved (<0.2 µm) iron concentrations in the ambient surface seawater were extremely low (0.06±0.015 nM) with colloidal iron being a minor fraction. For the iron addition, an acidified FeSO<sub>4</sub> solution was released three times over a 23-day period to the eddy. High levels of dissolved iron concentrations (2.0±1.1 nM) were measured in the surface water until 4 days after the first iron infusion. After every iron infusion, when high iron concentrations were observed before storm events, there was a significant correlation between colloidal and dissolved iron concentrations ([Colloidal Fe]=0.7627[Dissolved Fe]+0.0519, R<sup>2</sup>=0.9346). These results indicate that a roughly constant proportion of colloidal vs. dissolved iron was observed after iron infusion (~76%). Storm events caused a significant decrease in iron concentrations (<0.61 nM in dissolved iron) and changed the proportions of the three iron size-fractions (soluble, colloidal and labile particle). The changes in each iron size-fraction indicate that colloidal iron was eliminated from surface mixed layer more easily than particulate and soluble fractions. Therefore, particle and soluble iron efficiently remain in the mixed layer, probably due to the presence of suspended particles and naturally dissolved organic ligands. Our data suggest that iron removal through colloidal aggregation during phytoplankton bloom should be considered in the oceanic iron cycle.

Article Toolbox

- Download PDF
- E-mail Article
- Cited By
- Save as Citation Alert
- Citation Feed
- Export Citation
- Add to my Quick Links
- Add to collab
- Permissions & Reprints
- Cited By in Scopus (16)

**Related Articles in ScienceDirect**

- Major deviations of iron complexation during 22 days of... *Marine Chemistry*
- Iron cycling and nutrient-limitation patterns in surfac... *Deep Sea Research Part II: Topical Studies in Oceanogra...*
- Net and gross O<sub>2</sub> production in the southern ocean from ... *Deep Sea Research Part I: Oceanographic Research Papers*
- Modelling ice-edge phytoplankton bloom in the Scotia-We... *Journal of Marine Systems*
- Insights into Southern Ocean carbon export from the [de... *Deep Sea Research Part II: Topical Studies in Oceanogra...*

View More Related Articles

Supplementary Data  
View Record in Scopus

**The research collaboration tool**

- No user rating
- No user tags yet

Thanks!

The British Library is working with national science libraries to support researchers by providing methods for them to **locate, identify, and cite research datasets with confidence**

- This is the start of an long and open dialogue
- There are many open issues to address

We welcome your comments, questions, and ideas!

Adam.Farquhar at BL.UK



**Memorandum of Understanding  
Paris, 2-Mar-2009**



## Issues

### Technical

- Metadata
- Requirements for landing pages
- Improvements to submission, update APIs
- Granularity – Identify an entire dataset, subset, row, or cell
- Updates – Identify a dataset as it was on 1-Jan-2009
- Corrections
- Queries and subsets of large datasets

### Organisational

- Governance process for JRA
- Encouraging adoption