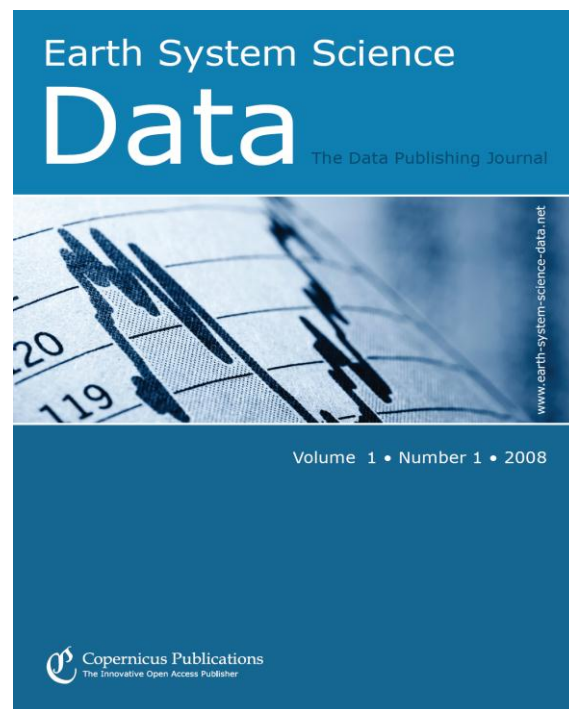


Data, Big Data and Publications

Hans Pfeiffenberger

Alfred-Wegener-Institute for Polar and Marine Research,
Helmholtz Association - Germany

STM Innovations Seminar, 2012-12-07, London



AWI 

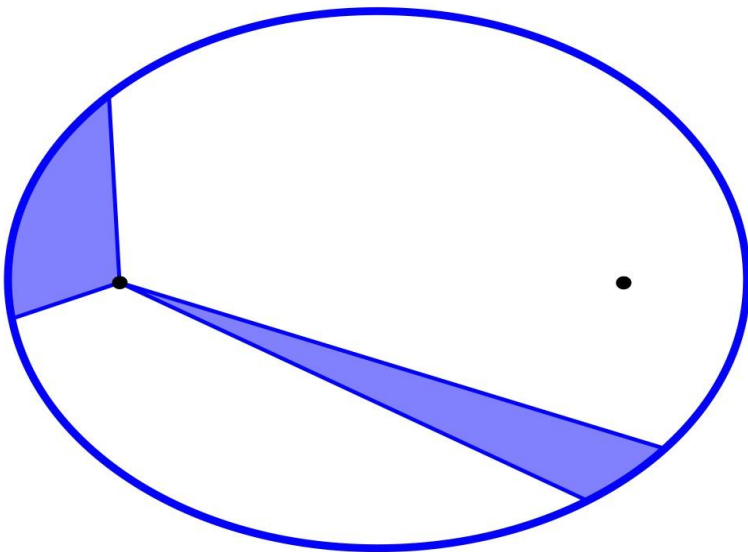
Data have “always” been the basis of science

- **2000 BC.**, Ur, Mesopotamia:
First known **record**
about eclipse of moon
- **700 BC**: Babylonians
predict eclipse of moon
- **585 BC**: Thales
predicts eclipse of sun
- **1300 years to find the pattern**
- **BIG DATA??**



1606 - 1618: Kepler's Laws (using Tycho Brahe's data!)

- Describe motion of planets
 - 12 years from second to third law



Planet	T	d	T^2	d^3	T^2/d^3
Merkur	0,241	0,387	0,058081	0,057960603	1,002077221
Venus	0,615	0,723	0,378225	0,377933067	1,000772446
Erde	1	1	1	1	1
Mars	1,881	1,524	3,538161	3,539605824	0,999591812
Jupiter	11,863	5,203	140,730769	140,8515004	0,999142846
Saturn	29,458	9,555	867,773764	872,3526289	0,994751131

T = siderische Umlaufzeit in trop. Jahren d = große Halbachse in astronomischen Einheiten (Abstand Erde–Sonne)

- 1684 – 1687 Newton De Motu – Principia
 - Explained (!) Kepler's Laws (not the primary data!)

PHILOSOPHICAL TRANSACTIONS: GIVING SOME ACCOMPT

OF THE PRESENT
Undertakings, Studies, and Labours

OF THE
INGENIOUS

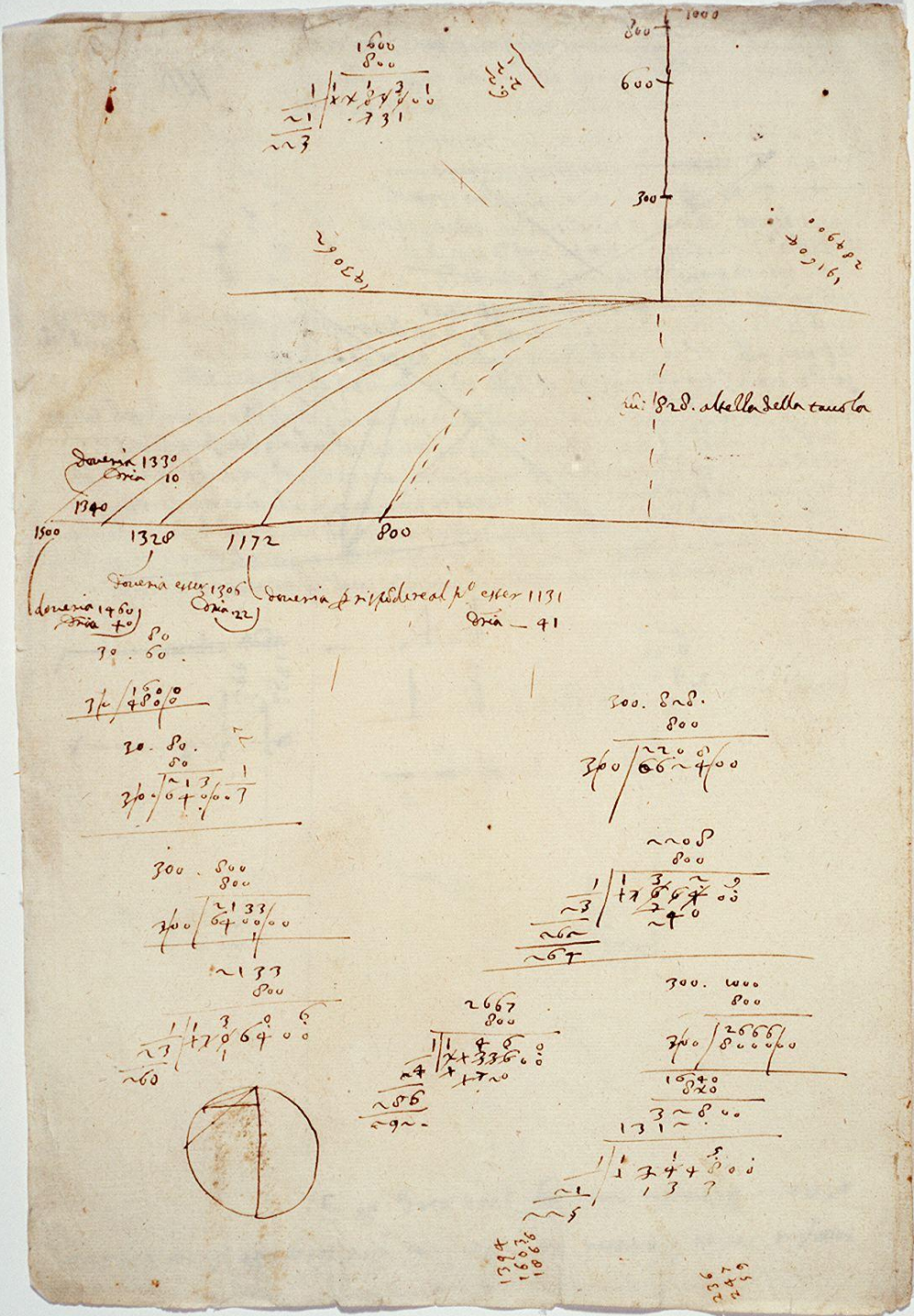
IN MANY
CONSIDERABLE PARTS

OF THE
WORLD.

Vol I.

For Anno 1665, and 1666.

In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little with-
out Temple-Bar, and James Allestry in Duck-Lane,
Printers to the Royal Society.

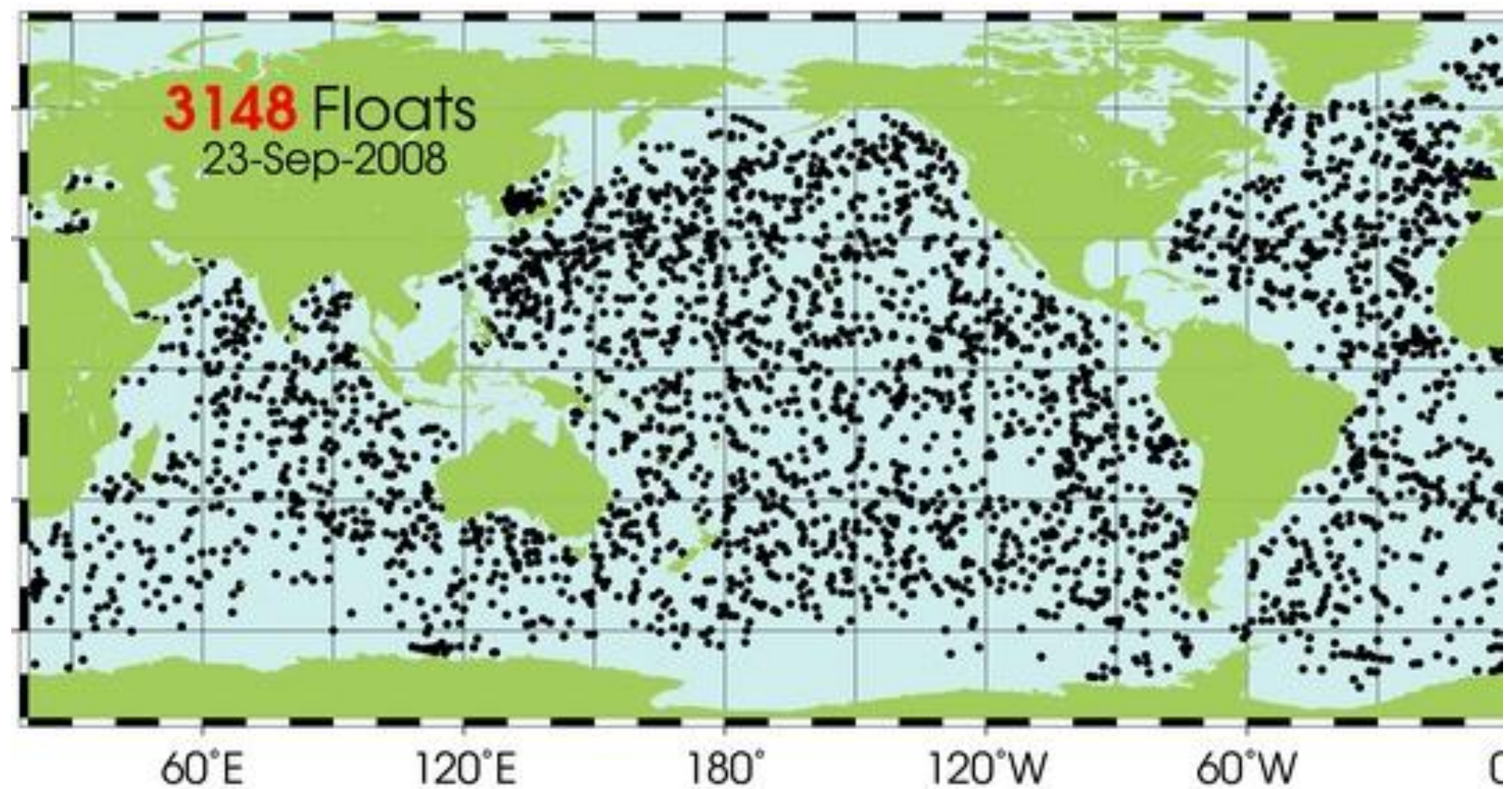


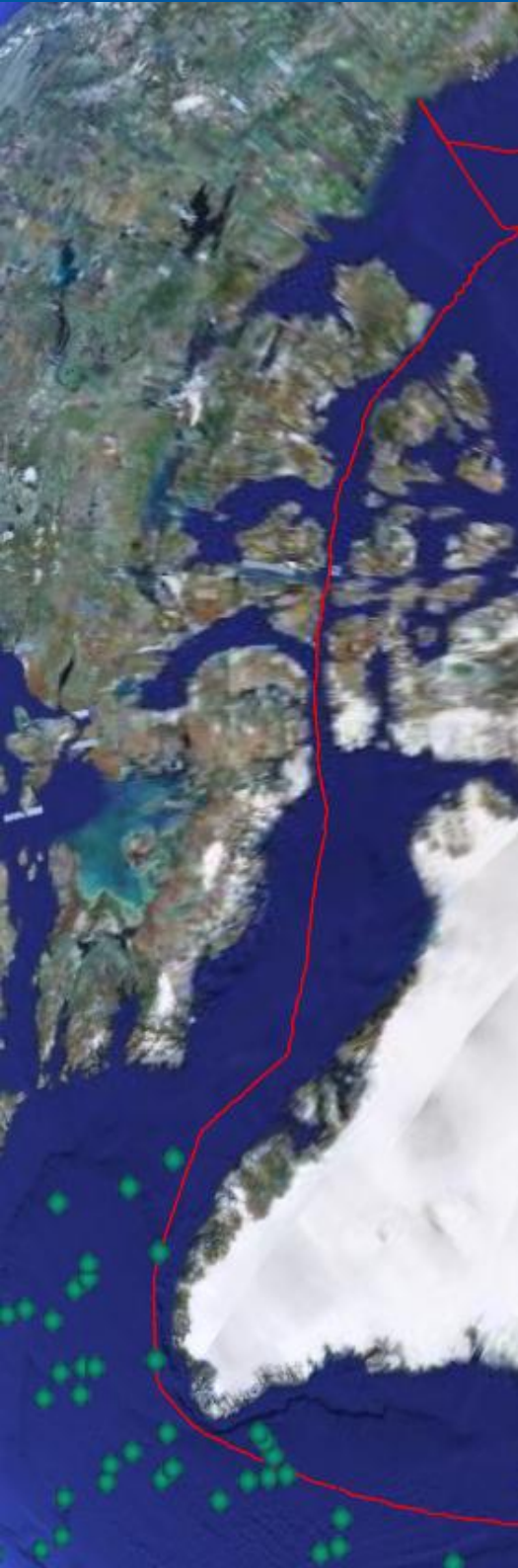
1938: Meitner-Hahn-Strassmann Uran-Experiment, Berlin



The last big breakthrough to be done with a **lab-notebook**?

The biggest experiment, worldwide (not CERN!)





6900499

NORWAY (Argo NORWAY)

Deployment
Latest Location

Web Products

880 Days
95 profiles at GDACs (origin Coriolis) including 0 DM profiles
Date: 13/04/2006 Lat : 64.6500 Lon : -.0216
Date: 09/09/2008 Lat : 67.0903 Lon : -9.0152
[AIC Coriolis JMA](#)
[CSIRO MEDS](#)

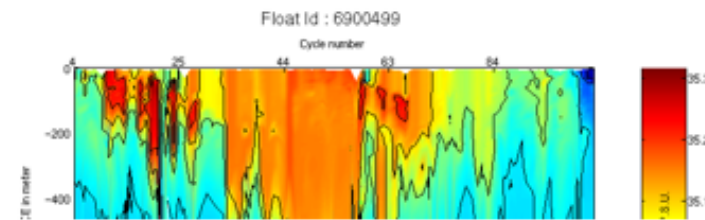
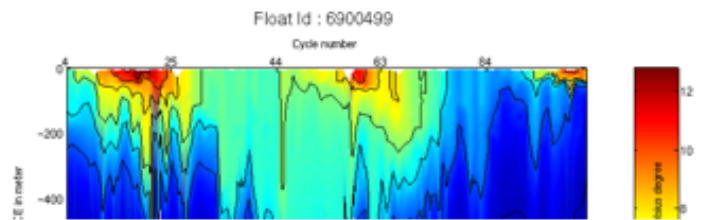
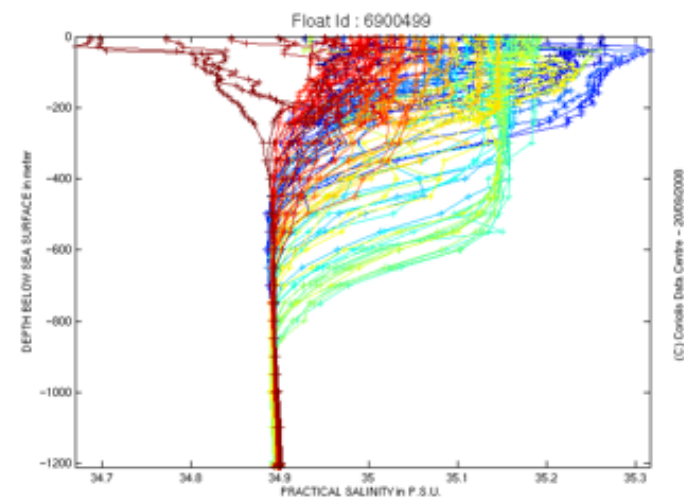
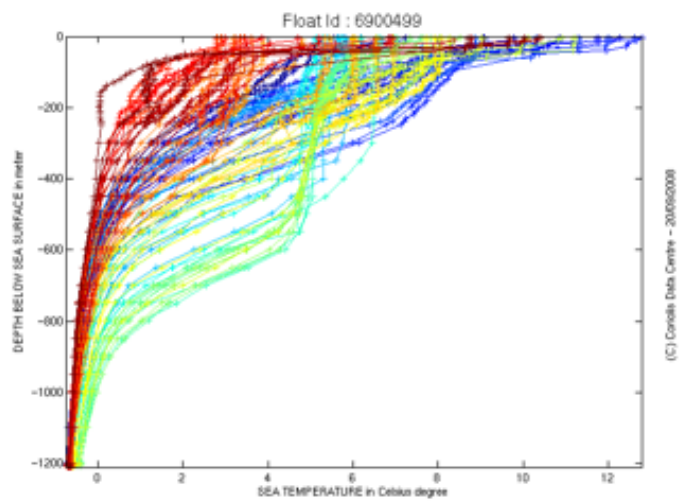
Data (netCDF)

[Profiles](#) [Metadata](#) [Trajectory](#) [Technical](#)

QC

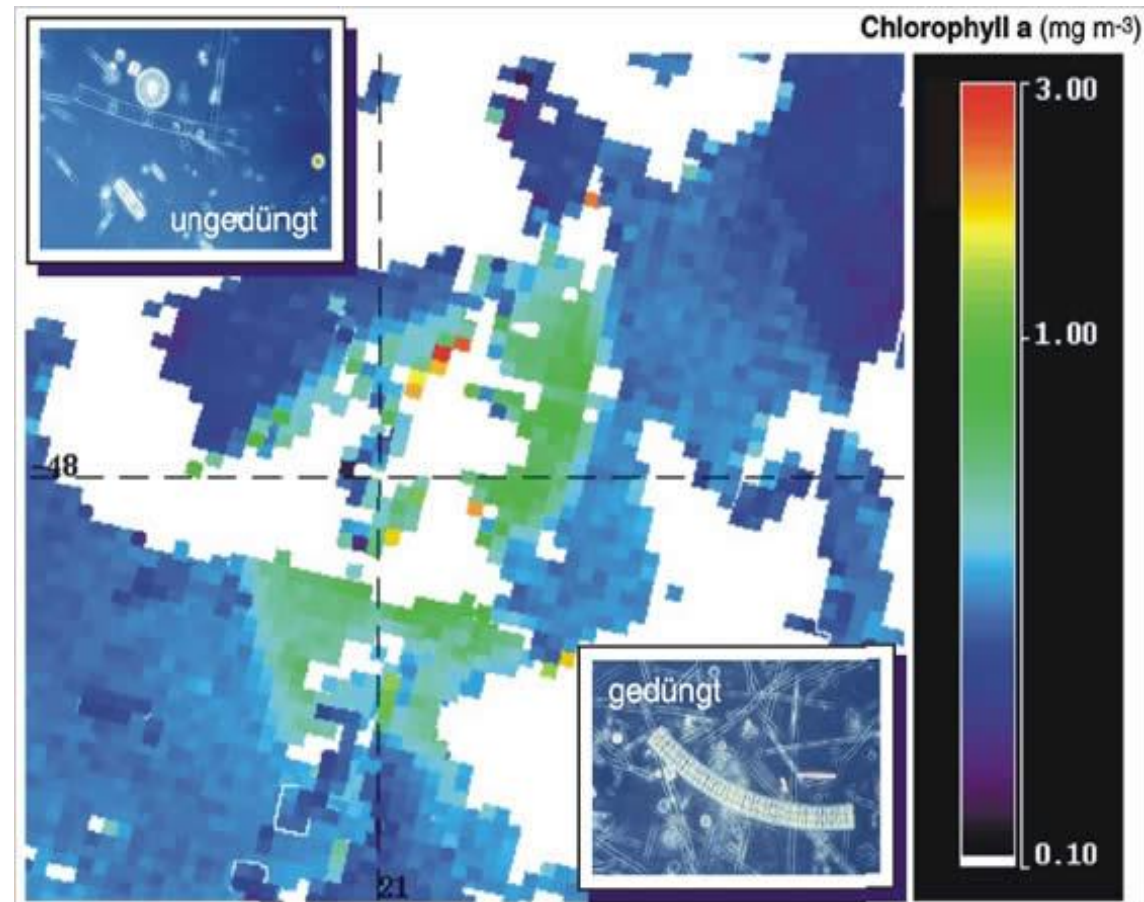
[Altimetry QC](#)

Subsurface Temperature - Subsurface Salinity (source [IFREMER/Coriolis](#))

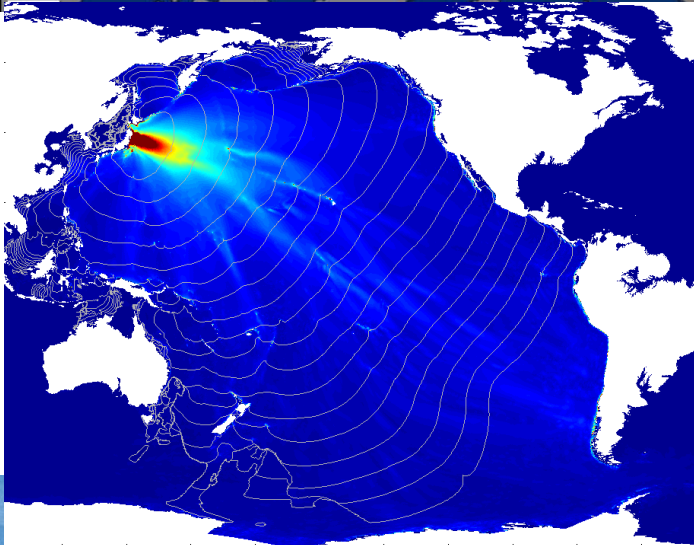
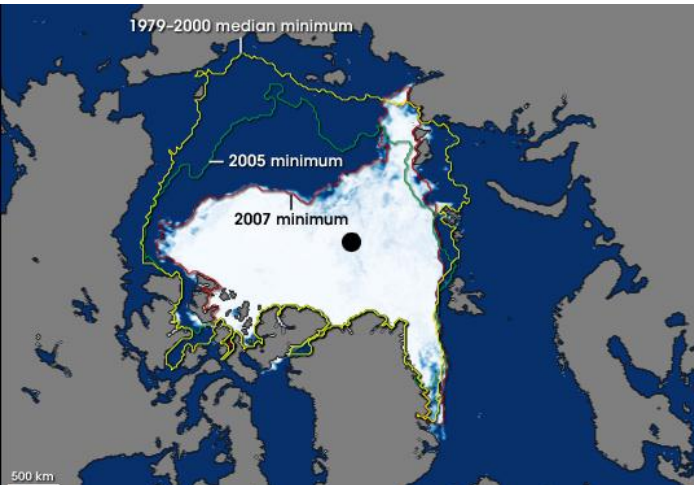


An important, typical Experiment

- EISENEX / EIFEX : Two expeditions of “**Polarstern**” :
With a few tons of iron fertilizer, south of Capetown
- EIFEX (2004):
 - 54 scientists and students from
 - 14 institutes and 3 companies from
 - 7 EU countries and South Africa
 - Oceanographers
 - Biologists
 - Chemists.....
- “**Biogeochemistry**”
- **+ Satellite observations !**



MaNIDA – Enabling Data-Intensive Marine Science



Global Change

- Assessing, understanding, and predicting environmental changes
- Human environmental impact

Hazards

- Risk analysis and support for disaster management
- Understanding environmental factors affecting human health

Resources

- Sustainable ecosystem management
- Energy from the ocean

The Big Challenge(s)

- **Global Change, Ageing Society ... „Theory Of Everything“**
- All are **Big Data** problems (by at least one definition)
- All are **multi-disciplinary** (except TOE)
- Most need **aggregation of globally distributed data**
- **Most are Heterogeneous and Complex**

Down to Earth !

- **What does an individual scientist want / need**
- **What is she prepared to do?**
- **And where are publications, after all ??**

2011: BGI („Beijing“ Genomics Institute)

Spiegel Online, 03.06.2011 (after EHEC identification)

<http://www.spiegel.de/wissenschaft/medizin/0,1518,766481,00.html>

- Das **Großunternehmen** beschäftigt rund **4000 Menschen**.
- Allein **180 Apparate** zur Entschlüsselung von DNA-Material stehen in Shenzhen bereit, dies macht das BGI nach eigenen Angaben zu einer der weltweit größten Einrichtungen für Genom-Entschlüsselung.
- "**300 Forscher sind nur für die Gen-Decodierung zuständig**", sagt Yang Bicheng, **Marketingleiterin** des BGI.

What „Spiegel“ did not mention:

- BGI has a private „**Cloud**“ and (half) a journal: „**Gigascience**“

One PICK of a TALE (I)



“[Researchers would prefer] just **one point of access** to all data, which would be **simple to use** and ‘**fool proof**’.”

But she **suspects it is wishful thinking** to ask for Google-like simplicity when one looks for

Looks simple! (Isn't)

“**chlorophyll data in the Atlantic at 200 meters depth**”

Karin Lochte

(Alfred Wegener Institute for Polar and Marine Research)

www.nature.com/nature

Data's sham

Research cannot flourish if data

More and more often these days, a measured not just by the publica the data it makes available to the ing archives such as GenBank have demc such legacy data sets can be for generati cially when data are combined from man in ways that the original researchers coul

All but a handful of disciplines still lack and cultural frameworks required to sup (see pages 168 and 171) — leading to a sharing of data by researchers (see page 16 needs to be addressed by funders, unive themselves.

Research funding agencies need to rec and access to digital data are central to be supported accordingly. Organization for instance, have made a good start. The

NATURE INSIGHT TRANSCRIBING THE GENOME

10 September 2009 | www.nature.com/nature | £10

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

nature



NATUREJOBS
Philadelphia
calling

DATA—WHAT DATA?

Learning to share your results

LUNAR EXPLORATION

Highland games

THE HUMAN BRAIN

Procrastination pathways

VERTEBRATE EVOLUTION

What jawless fish say about us



ature

no. 7261 | 10 September 2009

d must act accordingly.

tigators to do this. One impor- software: tools that streamline a with a description of what the d them, which algorithms have — information that is essential a effectively.

a when data can be mixed and software that can keep track of m. Such systems are essential if re ever to give credit — as they d of

“Data management should be woven into every course in science.”

cies her ent take responsibility for preserving digital data and making them accessible

One of ODE's HYPOTHESES



“Without the **infrastructure** that **helps scientists** manage their data in a **convenient and efficient way**, no culture of data sharing will evolve.”

Stefan Winkler-Nees
Deutsche Forschungs-Gemeinschaft
(DFG)

How do we manage data - so that

- Recognition / Reward become possible
- It can be found and aggregated
 - through complex questions
- Level of quality becomes apparent
 - **provenance**
 - **review** / **endorsement**
- => By linking data to **people** and **publications!**

PANGAEA - Elsevier

 [Purchase PDF \(743 K\)](#) |  [Export citation](#)

[Abstract](#) | [Article](#) | [Figures/Tables](#) | [References](#)

Marine Micropaleontology

Volume 66, Issues 3-4, 20 February 2008, Pages 192-207

[doi:10.1016/j.marmicro.2007.09.002](https://doi.org/10.1016/j.marmicro.2007.09.002) | [How to Cite or Link Using DOI](#)
 Copyright © 2007 Elsevier B.V. All rights reserved.

 Cited By in Scopus (2)

[Permissions & Reprints](#)

Organic matter rain rates, oxygen availability, and vital effects from benthic foraminiferal $\delta^{13}\text{C}$ in the historic Skagerrak, North Sea

Sylvia Brückner  ^a,  and Andreas Mackensen ^a, 

^aAlfred Wegener Institute for Polar and Marine Research, Columbusstr., D-27568 Bremerhaven, Germany

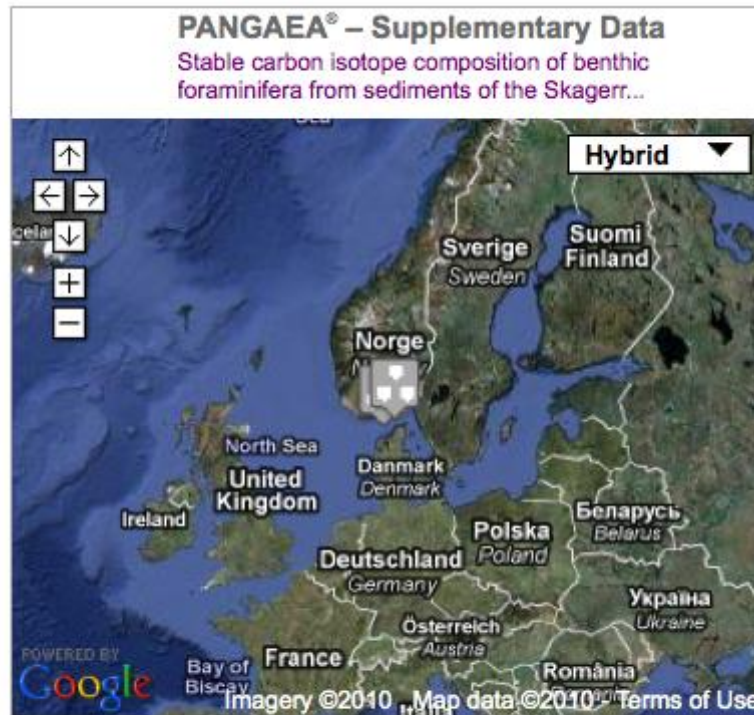
Received 27 March 2007; revised 21 September 2007; accepted 24 September 2007. Available online 4 October 2007.

Abstract

The sediment cores 225514 and 225510 were recovered from 420 and 285 m water depth, respectively. They were investigated for their benthic foraminiferal $\delta^{13}\text{C}$ during the last 500 years.

Purchase the full-text article 

-  PDF and HTML
-  All references
-  All images
-  All tables



Related Articles

-  [The tropical rainbelt and productivity changes off north...
Marine Micropaleontology](#)
-  [Temporal variability in living deep-sea benthic foramin...
Earth-Science Reviews](#)
-  [Early Maastrichtian benthic foraminiferal assemblages f...
Marine Micropaleontology](#)

2012: Nature Climate Change & ESSD

Earth Syst. Sci. Data Discuss., 5, 1107–1157, 2012
www.earth-syst-sci-data-discuss.net/5/1107/2012/
 doi:10.5194/essdd-5-1107-2012
 © Author(s) 2012. CC Attribution 3.0 License.

Open Access
 Earth System
**Science
 Data**
 Discussions

This discussion paper is/has been under review for the journal Earth System Science Data (ESSD). Please refer to the corresponding final paper in ESSD if available.

The global carbon budget 1959–2011

C. Le Quéré¹, R. J. Andres², T. Boden², T. Conway³, R. A. Houghton⁴,
 J. I. House⁵, G. Marland⁶, G. P. Peters⁷, G. van der Werf⁸, A. Ahlström⁹,
 R. M. Andrew⁷, L. Bopp¹⁰, J. G. Canadell¹¹, P. Ciais¹⁰, S. C. Doney¹², C. Enright¹,
 P. Friedlingstein¹³, C. Huntingford¹⁴, A. K. Jain¹⁵, C. Jourdain^{1,*}, E. Kato¹⁶,
 R. F. Keeling¹⁷, K. Klein Goldewijk²⁵, S. Levis¹⁸, P. Levy¹⁴, M. Lomas¹⁹,
 B. Poulter¹⁰, M. R. Raupach¹¹, J. Schwinger²⁰, S. Sitch²¹, B. D. Stocker²²,
 N. Viovy¹⁰, S. Zaehle²³, and N. Zeng²⁴

Discussion Paper | Discussion Paper | Discussion Paper

ESSD
 5, 1107–1157, 2012

The global carbon budget 1959–2011
 C. Le Quéré et al.

Title Page
 Abstract Instruments
 Data Provenance & Structure
 Tables Figures

nature
 climate
 Home | Opinion &

NATURE C
 The c

Glen P. Peters, Robbie M. Andrew, Tom Boden, Josep G. Canadell, Philippe Ciais, Corinne Le Quéré, Gregg Marland, Michael R. Raupach & Charlie Wilson

[Affiliations](#) | [Contributions](#) | [Corresponding author](#)

Nature Climate Change (2012) | doi:10.1038/nclimate1783
 Published online 02 December 2012

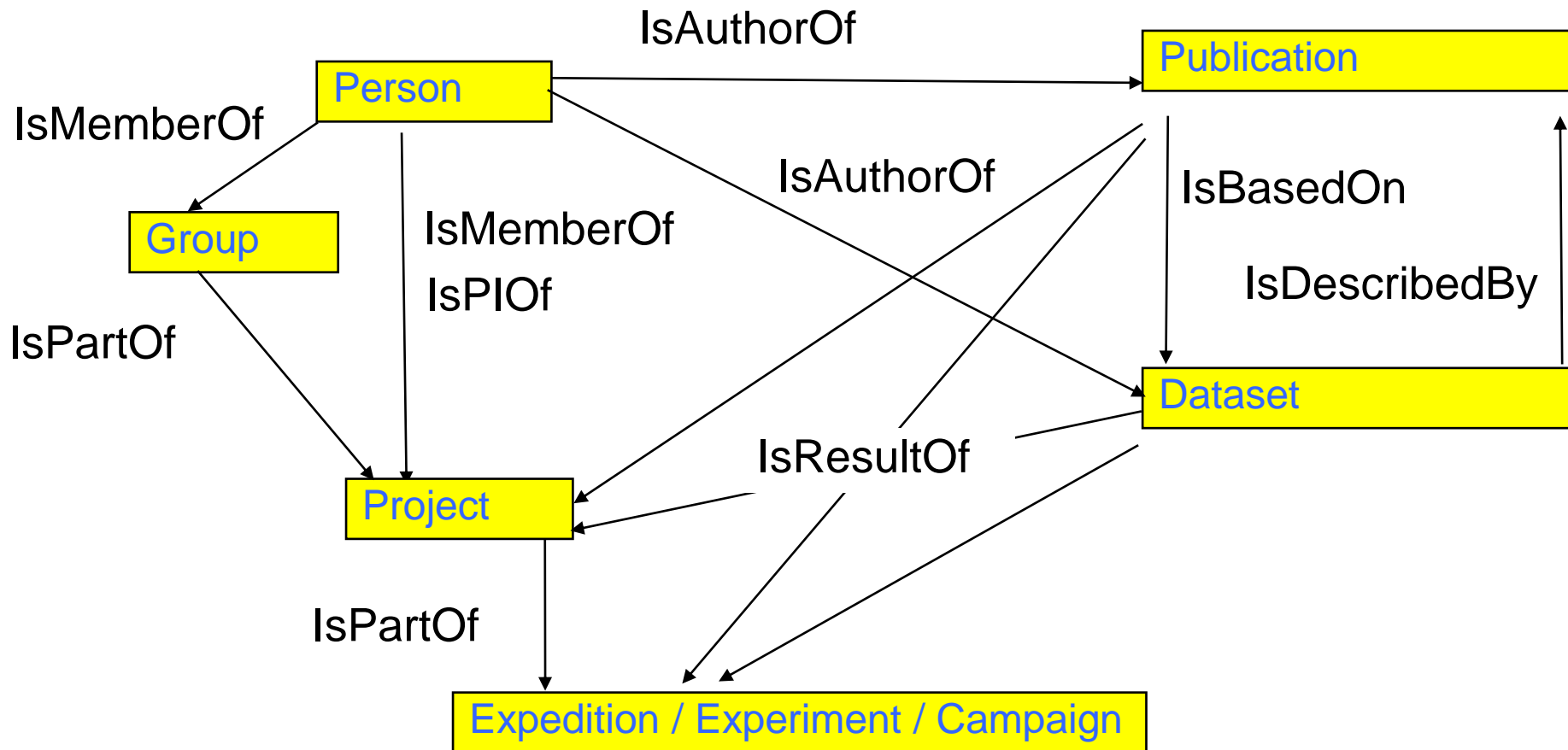
► Subscribe
 ► Recommend to library



✉ E-alert  RSS  Facebook  Twitter

Science jobs from **naturejobs**

Pfeiffenberger, Macario, Text, Data and People, OAI4, CERN 2005



eXpedition (in production since 2005)

Related Information: ["Reports on Polar and Marine Research"](#) (1982 to date)

[Primary data](#) (all polarstern datasets in PANGAEA)

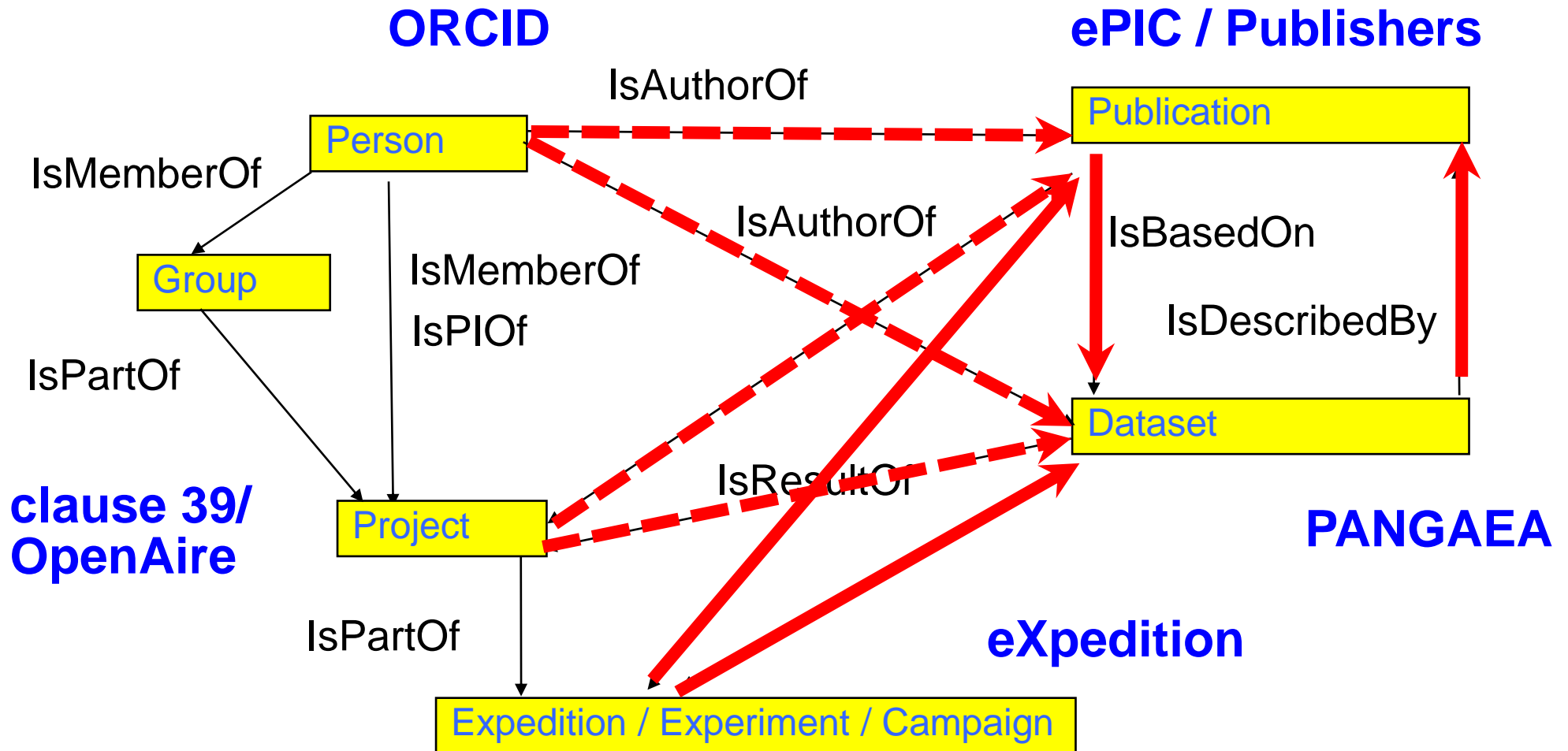
[Handbook and scientific device documentation](#)(in deutsch)

[DSHIP](#) (Polarstern Data Acquisition System)

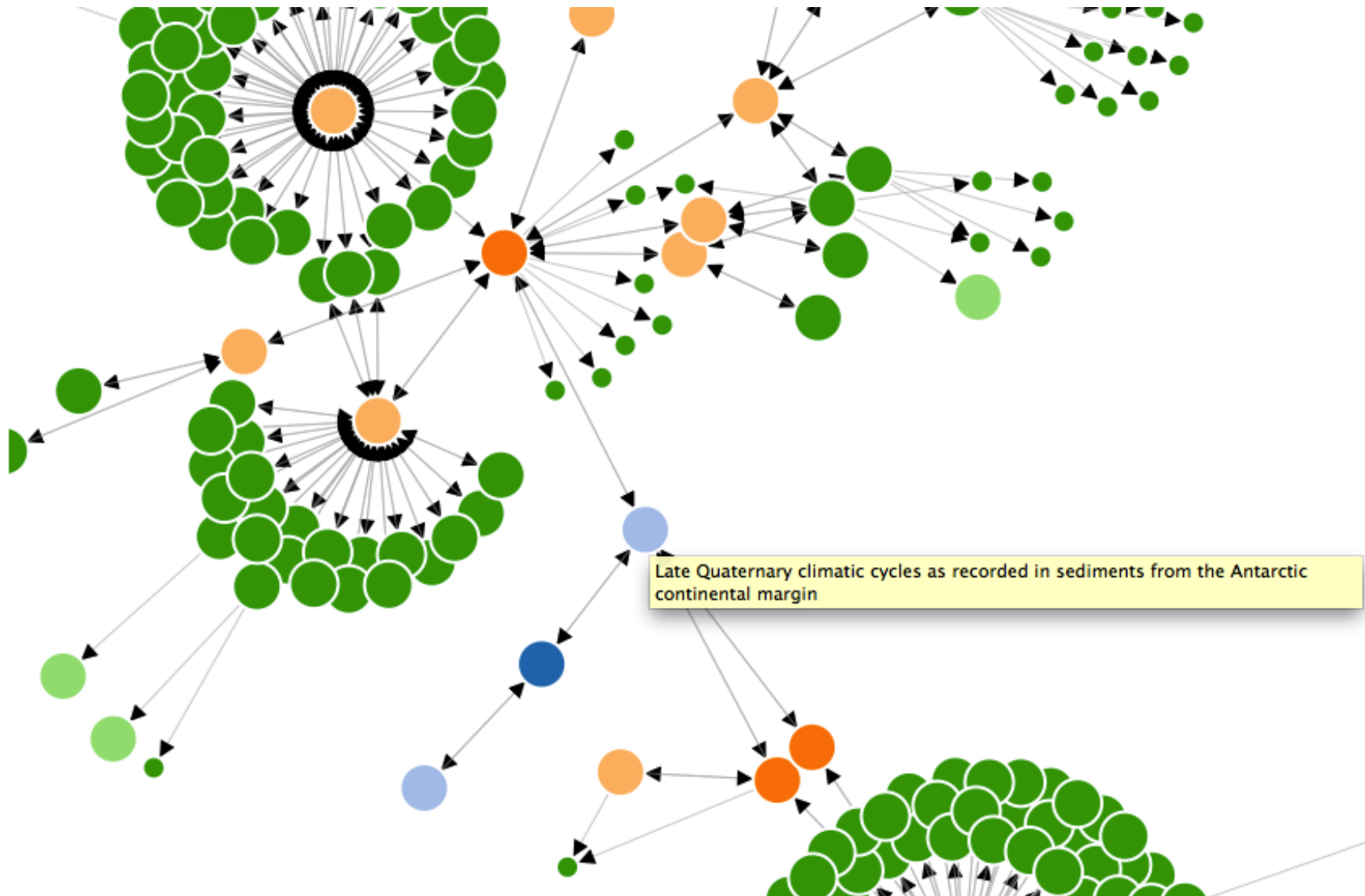
[VirtualPS: Virtual Polarstern Tour](#)

Expedition	Date Port	Region Research	Publications & Primary Data	Details
ANT-XXI/3 Coordinator: Pörtner, H. Chief scientist: Smetacek, V.	21.01.2004 - 25.03.2004 Capetown - Capetown [Map(png)]	Atlantic/Indian Ocean, Polar frontal zone Biology, EIFEX	ePIC: Publications ePIC: Reports on Polar and Marine Research ePIC: Weekly reports PANGAEA: Stations PANGAEA: Datasets <i>[Note: Publications and datasets for recent cruises may not yet be available]</i> Meteorology	▶
ANT-XXI/4	27.03.2004 - 06.05.2004	Lazarev Sea Biology, Krill, GLOBEC	ePIC: Publications ePIC: Reports on Polar and Marine Research	▶

Pfeiffenberger, Macario, Text, Data and People, OAI4, CERN 2005



Manida – Publications and Data network – A Big Data problem?



Conclusions

- There are **Huge Data** problems (such as genetics)
 - (relatively) homogeneous and not too complex
 - though costly and technologically challenging
- There are „**Big Data**“ problems (such as „Earth Science“)
 - involve **finding and exploiting patterns** in metadata and data
 - but heterogeneous and distributed (unlike Amazon,...)
- **Both need publications** linked to them
 - **Quality** assurance
 - The best „**metadata**“ one can have
 - Provide the **linking hubs** in the **digital assets ecosystem**

Thank you!

oa.helmholtz.de