HELMHOLTZ GEMEINSCHAFT OPEN ACCESS

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

HELMHOLTZ GEMEINSCHAFT OPEN ACCESS

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

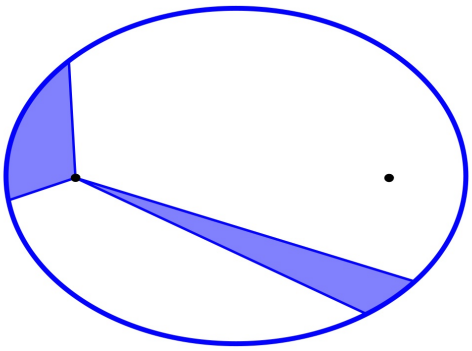
- **585 BC:** Thales

  predicts eclipse of sun

- **1300 years to find the pattern**
- **BIG DATA??**

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

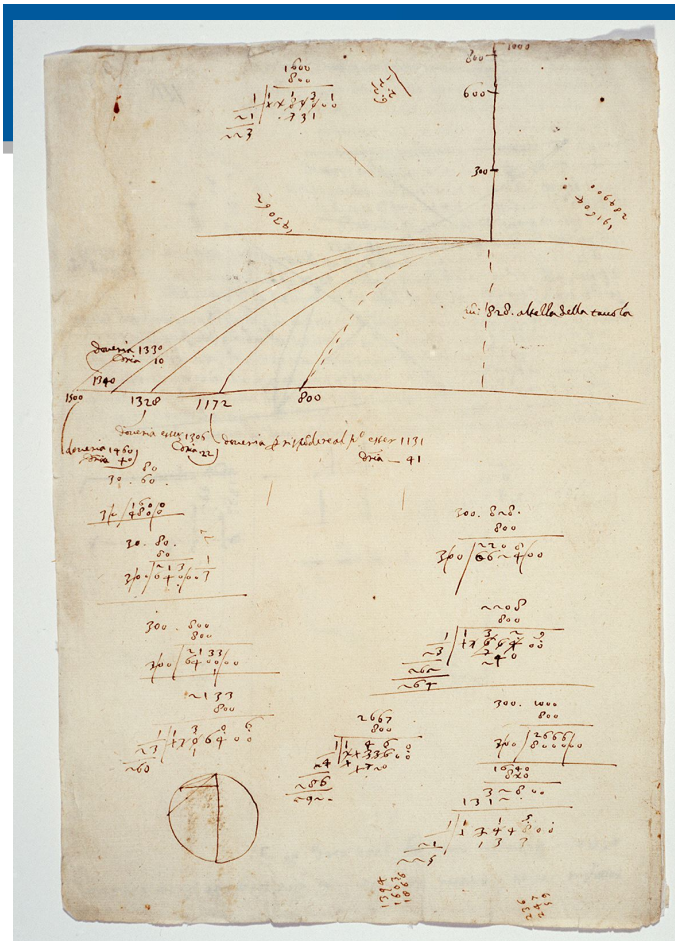
- **Describe motion of planets**
  – 12 years from second to third law



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!)**

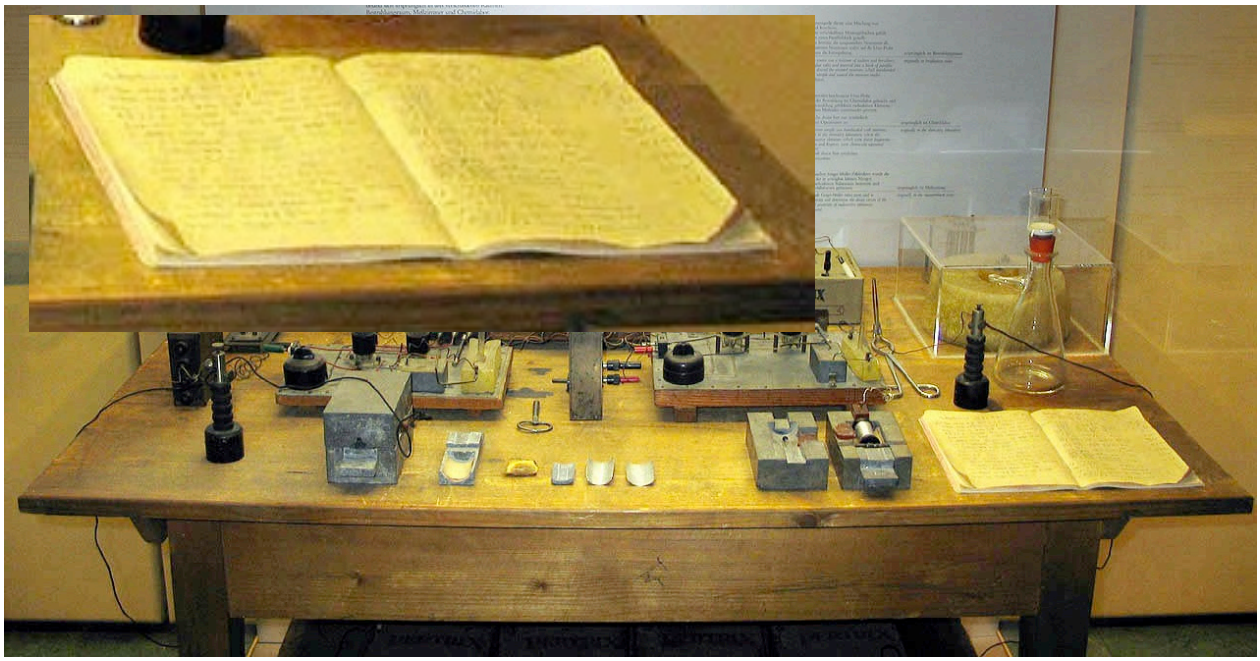


PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

Vol I.
For Anno 1665, and 1666.

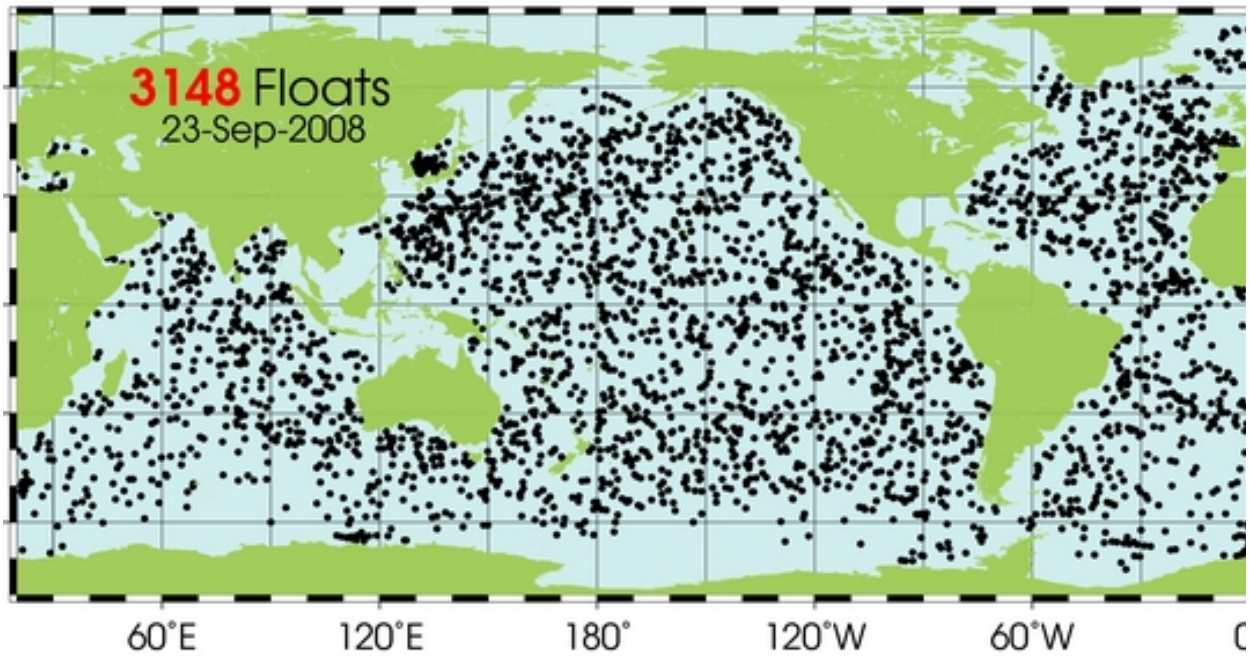
In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane, Printers to the Royal Society.

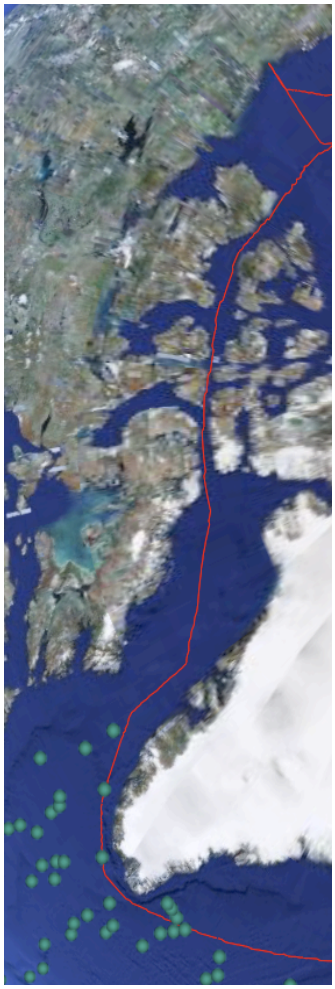
1938: Meitner-Hahn-Strassmann Uran-Experiment, Berlin



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

The biggest experiment, worldwide (not CERN!)

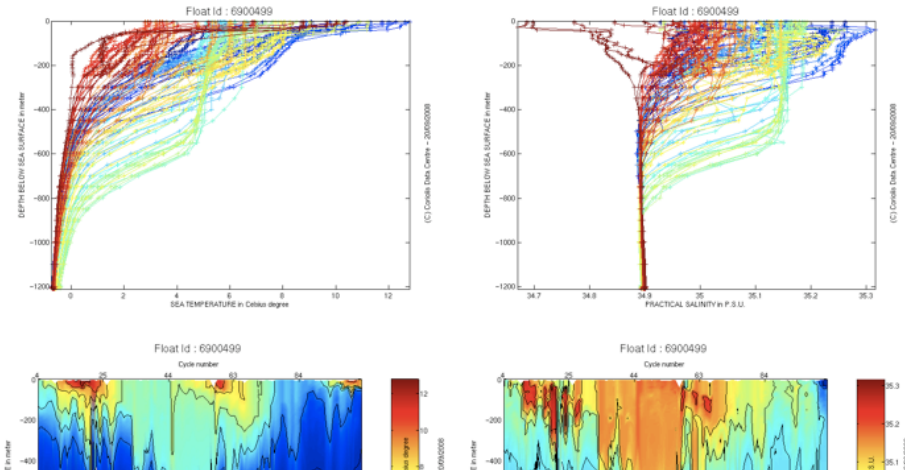




6900499

NORWAY (Argo NORWAY)	880 Days 95 profiles at GDACs (origin Coriolis) including 0 DM profiles
Deployment	Date: 13/04/2006 Lat : 64.6500 Lon: -.0216
Latest Location	Date: 09/09/2008 Lat: 67.0903 Lon: -9.0152
Web Products	AIC Coriolis JMA CSIRO MEDS
Data (netCDF)	Profiles Metadata Trajectory Technical
QC	Altimetry QC

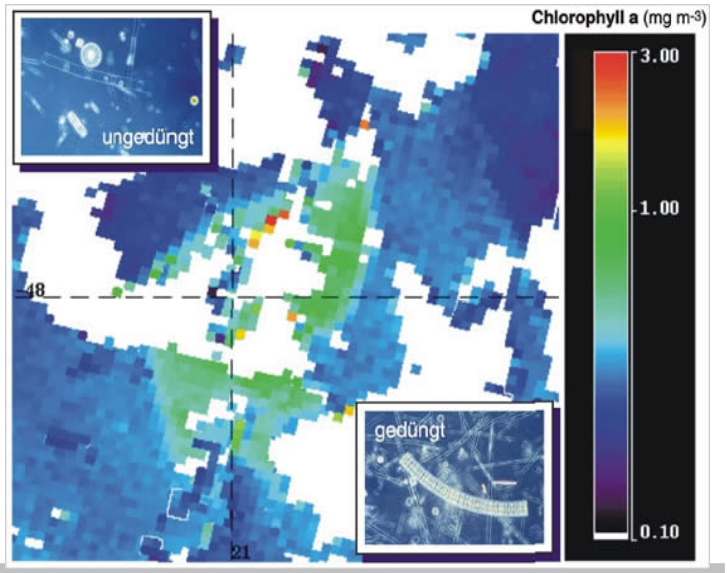
Subsurface Temperature - Subsurface Salinity (source IFREMER/Coriolis)



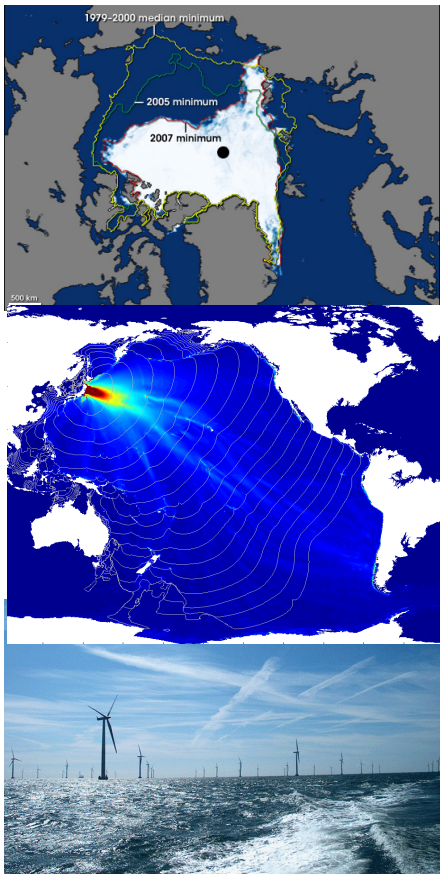
HELMHOLTZ GEMEINSCHAFT OPEN ACCESS

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

HELMHOLTZ GEMEINSCHAFT OPEN ACCESS

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 (1)

"[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 "chlorophyll data in the Atlantic at 200 meters depth"

Looks simple! (Isn't)

Karin Lochte
(Alfred Wegener Institute for Polar and Marine Research)

SEVENTH FRAMEWORK PROGRAMME

APA ALLIANCE FOR PERMANENT ACCESS

HELMHOLTZ | ASSOCIATION

DE Opportunities for Data Exchange

HELMHOLTZ | GEMEINSCHAFT OPEN ACCESS

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

www.nature.com/nature

Data's sham

Research cannot flourish if data

d must act accordingly.

ture

o. 7261 | 10 September 2009

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 demo 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 lac 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, univ themselves.

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

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 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

"Data management should be woven into every course in science."

ies her ent ake

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)

SEVENTH FRAMEWORK PROGRAMME

APA ALLIANCE FOR PERMANENT ACCESS

HELMHOLTZ | ASSOCIATION

Opportunities for Data Exchange

HELMHOLTZ | GEMEINSCHAFT OPEN ACCESS

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!**

HELMHOLTZ GEMEINSCHAFT OPEN ACCESS

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 | How to Cite or Link Using DOI

Permissions & Reprints

Cited By in Scopus (2)

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

Sylvia Brückner [a], and Andreas Mackensen [a]

[a]Alfred 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.

Purchase the full-text article
- PDF and HTML
- All references
- All images
- All tables

Abstract

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

PANGAEA® – Supplementary Data
Stable carbon isotope composition of benthic foraminifera from sediments of the Skagerr...

Related Articles
- The tropical rainbelt and productivity changes off nort... *Marine Micropaleontology*
- Temporal variability in living deep-sea benthic foramin... *Earth-Science Reviews*
- Early Maastrichtian benthic foraminiferal assemblages f... *Marine Micropaleontology*

HELMHOLTZ GEMEINSCHAFT OPEN ACCESS

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

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 Quéré[1], R. J. Andres[2], T. Boden[2], T. Conway[3], R. A. Houghton[4], J. I. House[5], G. Marland[6], G. P. Peters[7], G. van der Werf[8], A. Ahlström[9], R. M. Andrew[7], L. Bopp[10], J. G. Canadell[11], P. Ciais[10], S. C. Doney[12], C. Enright[1], P. Friedlingstein[13], C. Huntingford[14], A. K. Jain[15], C. Jourdain[1,*], E. Kato[16], R. F. Keeling[17], K. Klein Goldewijk[25], S. Levis[18], P. Levy[14], M. Lomas[19], B. Poulter[10], M. R. Raupach[11], J. Schwinger[20], S. Sitch[21], B. D. Stocker[22], N. Viovy[10], S. Zaehle[23], and N. Zeng[24]

ESSDD
5, 1107–1157, 2012
The global carbon budget 1959–2011
C. Le Quéré et al.
Title Page | Abstract | Instruments | Data Provenance & Structure | Tables | Figures

Discussion Paper | Discussion Paper | Discussion Paper | Discus

nature clim
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

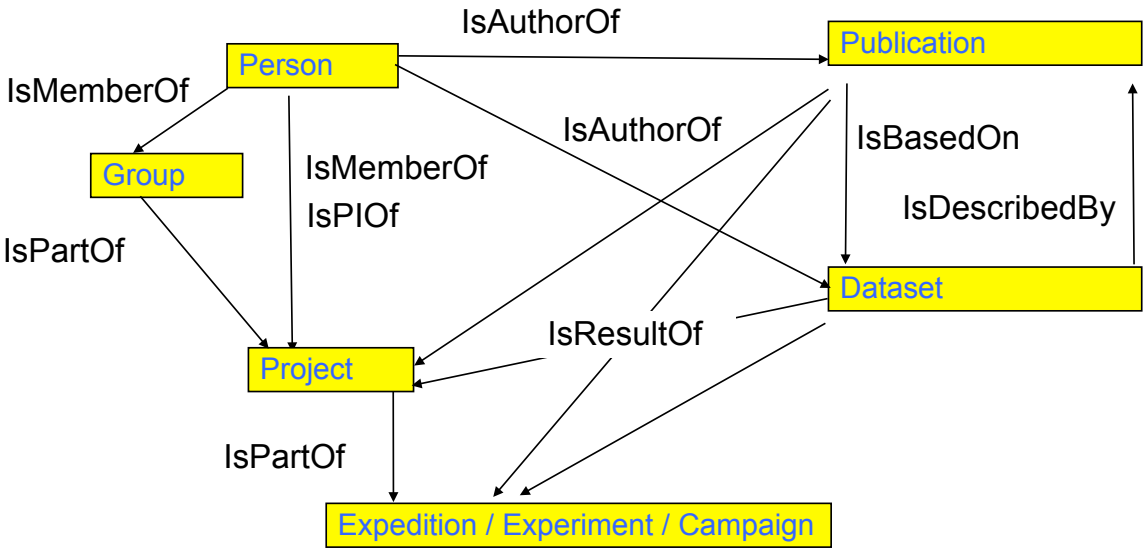
Affiliations | Contributions | Corresponding author

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

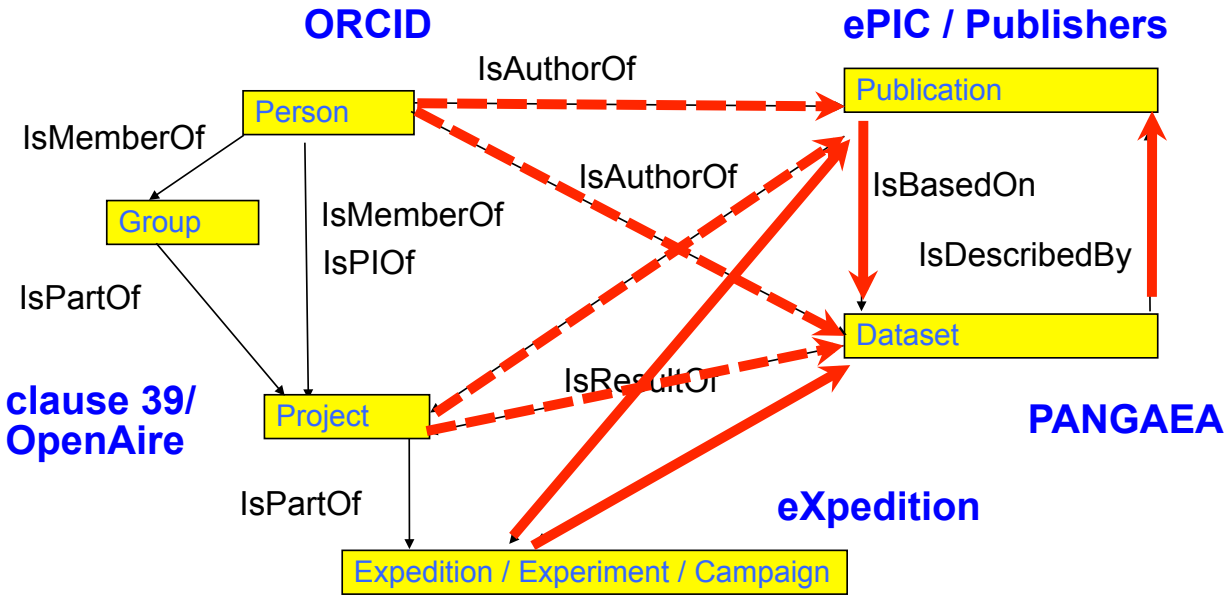


eXpedition (in production since 2005)

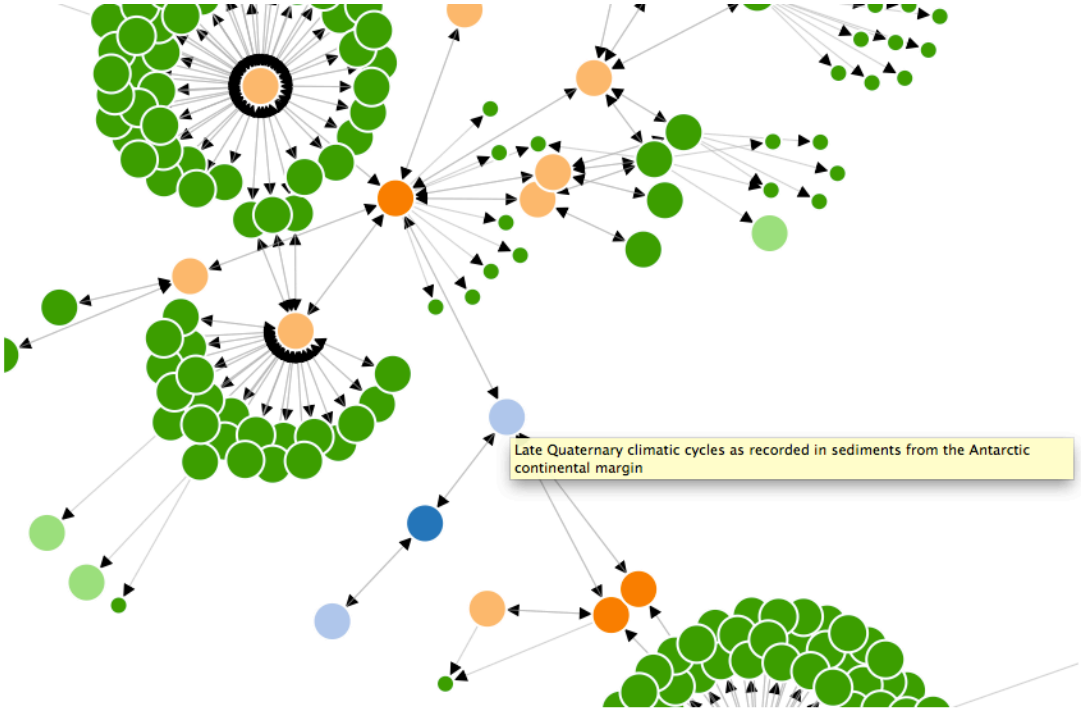
Related Information: "Reports on Polar and Marine Research" (1982 to date)
Primary data (all polarstern datasets in PANGAEA)
Handbook and scientific device documentation(in deutsch)
DShip (Polarstern Data Acquisition System)
VirtualPS: Virtual Polarstern Tour

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

HELMHOLTZ GEMEINSCHAFT OPEN ACCESS

Thank you!

oa.helmholtz.de

H. Pfeiffenberger, STM Innovations 2012-12-07, London