HELMHOLTZ GEMEINSCHAFT

OPEN ACCESS

Who is afraid of Data Publishing – The ESSD Experience

Hans Pfeiffenberger, Dave Carlson

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

APE2014, 2014-01-29, Berlin

AWI

Who should be ~~Scared~~ of Data Publishing ?

Respectful (Jan Brase)

Everybody !!

- At least those who ~~don't like~~ a tough challenge
  recognize

Who should *really* be *Afraid* of Data Publishing ?

- Those who
  - Invented their data (Stapel),
  - Selected data with a bias (notorious: Clinical trials)
  - Read wrong or to much from their data (Reinhart/ Rogoff)

- Those who build business-models on a monopoly on knowledge or facts, e.g.
  - Non-OA Publishers
  - Institutes which consider data collections as "their" capital

Royal Society: Science as an Open Enterprise (2012)

- **Open enquiry has been at the heart of science** since the first scientific journals were printed in the **seventeenth century**. …

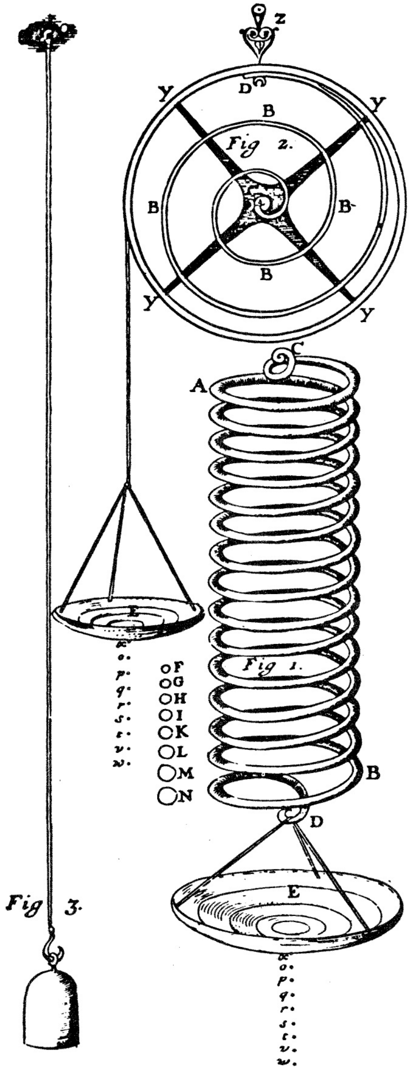
- Science's capacity for **self-correction** comes from this openness to scrutiny and challenge.

- **RS applied this to data:**

  Intelligent Openness

HELMHOLTZ | GEMEINSCHAFT OPEN ACCESS

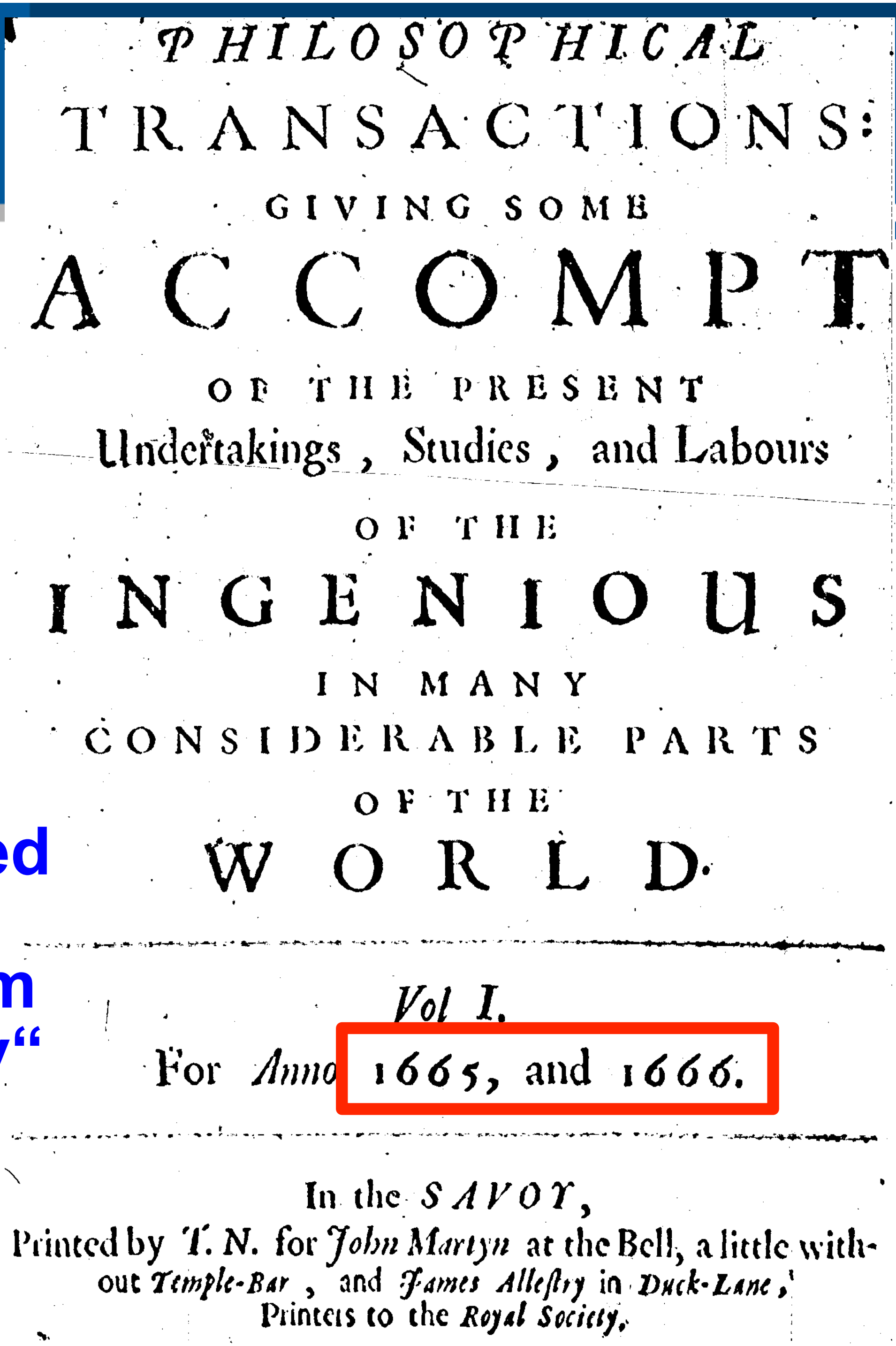
Scared in the 17th Century



Hooke, published

1676 by anagram „ceiiinosssttuv“

1678 in booklet



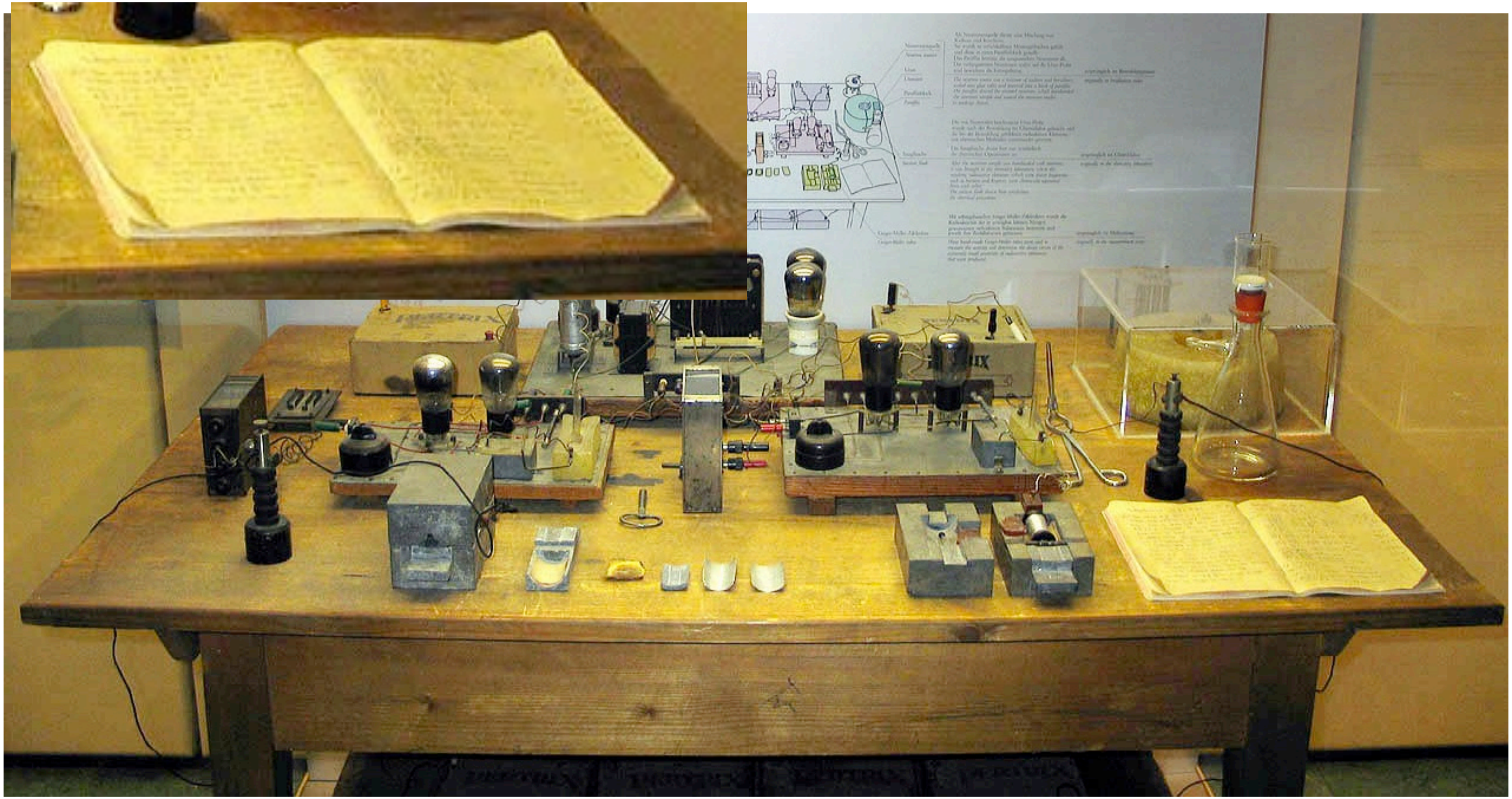
PHILOSOPHICAL
TRANSACTIONS:
GIVING SOME
ACCOMPT
OF THE PRESENT
Undertakings, Studies, and Labours
OF THE
INGENIOUS
IN MANY
CONSIDERABLE PARTS
OF THE
WORLD.

Vol I.

For Anno 1665, and 1666.

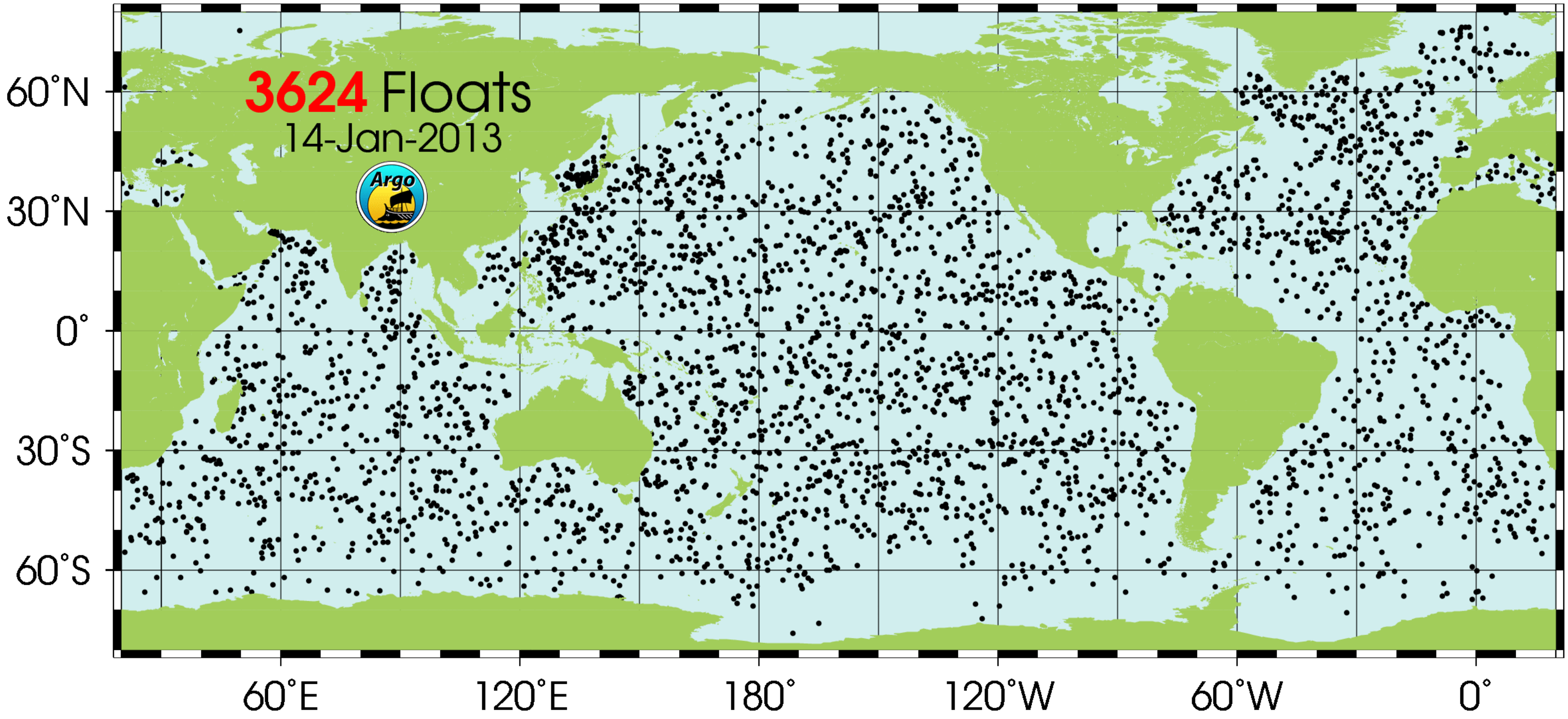
In the SAVOY,
Printed by T. N. for John Martyn at the Bell, a little without Temple-Bar, and James Allestry in Duck-Lane, Printers to the Royal Society.

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



The last big discovery by a small group with a lab notebook ?

What we do today:
ARGO, the biggest experiment in the world



ARGO is not Scared of „Data Publishing“ !

What is really fascinating: There are

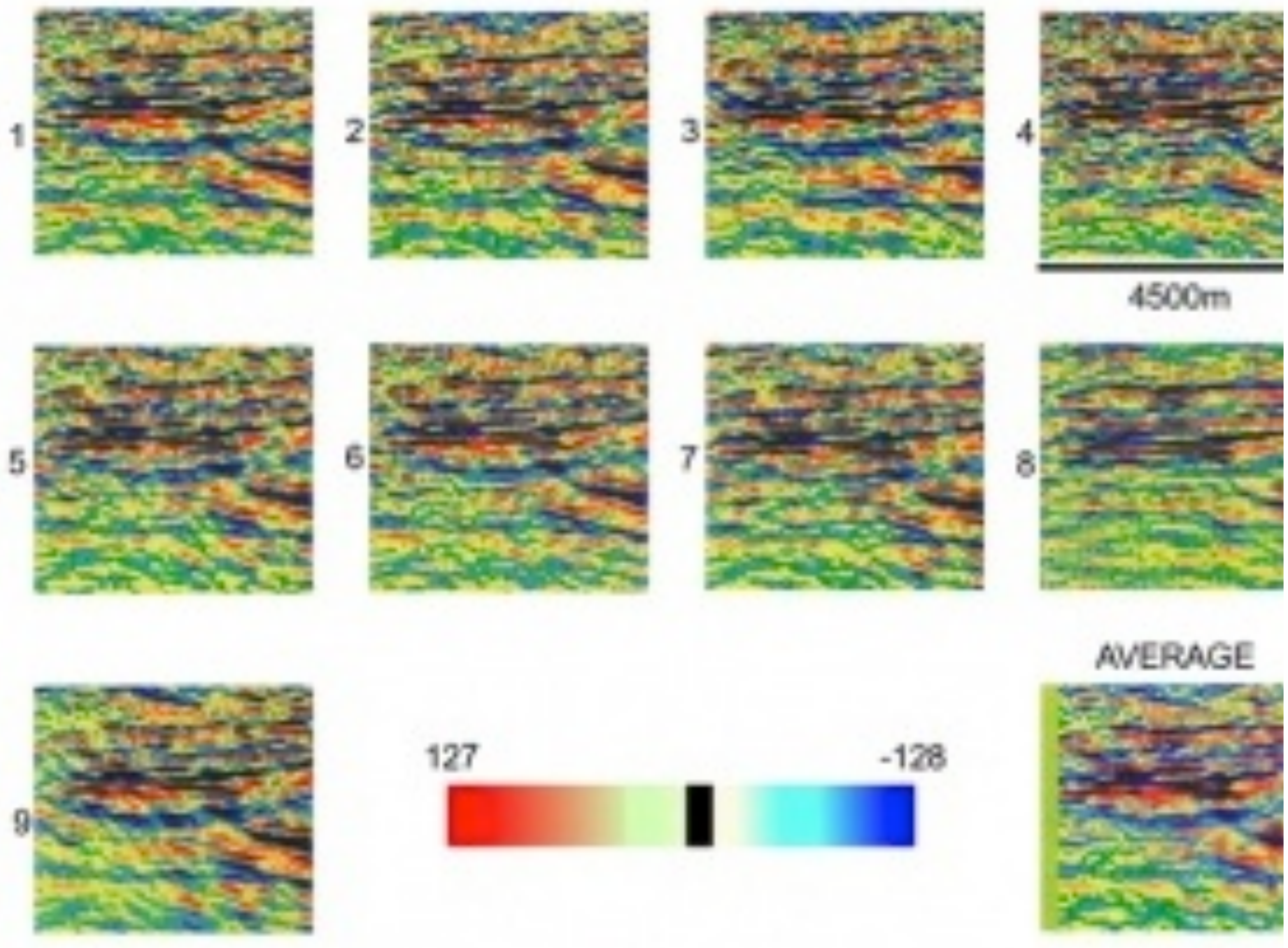
- **More than 3.000 buoys**
- **from more than 30 countries, lots of companies**

and yet there is:

- **Co-ordinated (quality) data management**
  - **One (“published”) standard for instruments**
  - **One (“published”) standard for formats**
  - **One (“published”?) standard for processing**
  - **Open access to data - (almost) no delay**

The Dangers of Working in Closed Silos –
„Does computation threaten the scientific method?“

- „using the same processed data from eight other companies, the **same algorithms** in the **same** programming **language**, using the **same input data**, just **coded independently**



- L.Hatton, A. Giordani ISGTW

Data Publishing Challenge #1

- **Quality of Data**
  - **Royal Soc. "intelligent Openness" (2012):**
    Data need to be "… assessable. Recipients need to be able to make some judgment or assessment of what is communicated.
  - **"Guidelines on Data Management in Horizon 2020" (2013):**
    "… are data provided in a way that judgments can be made about their reliability and the competence of those who created them)

Earth System Science Data (ESSD)

established 2008

<u>Advisory Board:</u>
Paul J. Crutzen
Sydney Levitus
Alexander Petrovich Lisitzin

<u>Editors in Chief:</u>
David Carlson
Hans Pfeiffenberger

<u>Publishing House</u>
Copernicus Publications
– OA Publisher, EGU

Require Estimate of Error and Data Provenance - No fancy interpretations!!

For balloon-borne ozone profile measurements a pump correction has to be applied in order to compensate the decreasing pump efficiency with increasing height and changing air temperature. Both, an inadequate pump correction and an erroneous estimate of residual ozone above the height of balloon burst may contribute to the overall measurement error of the ozone profile. Usually an independent column ozone observation X_D by spectrometer measurement is compared with the integrated column ozone X_S between the ground level and the height of balloon burst plus estimated residual ozone above that level to adjust the recorded profile values. The correction factor is

$$C = X_D / X_S.$$

Systematic differences and random errors of the electrochemical ozone sonde, type OSP, has been estimated by analysing 20 tandem ozone soundings at the Aerological Observatory Lindenberg in 1982 (Feister et al., 1985). Random errors are at their maximum of about 10 to 13% in the troposphere and above 32 km, and reach a minimum of 2 to 5% between 20 and 28 km. The mean random error is 11.5% in the troposphere, 7% in the stratosphere beneath the ozone maximum height (ca. 22 km), and 5.6% above that height.

2 Data Provenance and Structure

The first permanently operated German research base – later named Georg-Forster-Station – was established in 1976 in the Schirmacher Oasis at 70°46′ S, 11°41′ E. Since then the station was permanently used and operated as an annex to the Russian station Novolazarevskaya until 1987, and then as a German Antarctic station named after

4

ESSDD

1, 1–13, 2008

Antarctic ozonesonde profiles

G. König-Langlo and H. Gernandt

Title Page
Abstract | Instruments
Data Provenance & Structure
Tables | Figures
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

2013: CO above Troll Station, Original Data

Earth System Science Data
The Data Publishing Journal

- Home
- Online Library ESSD
- Online Library ESSDD
  - Papers in Open Discussion
  - Volumes and Issues
  - Special Issues
  - Most Commented Papers
  - Full Text Search
  - Title and Author Search
- Alerts & RSS Feeds
- General Information
- Submission
- Review

Volumes and Issues Contents of Issue 1

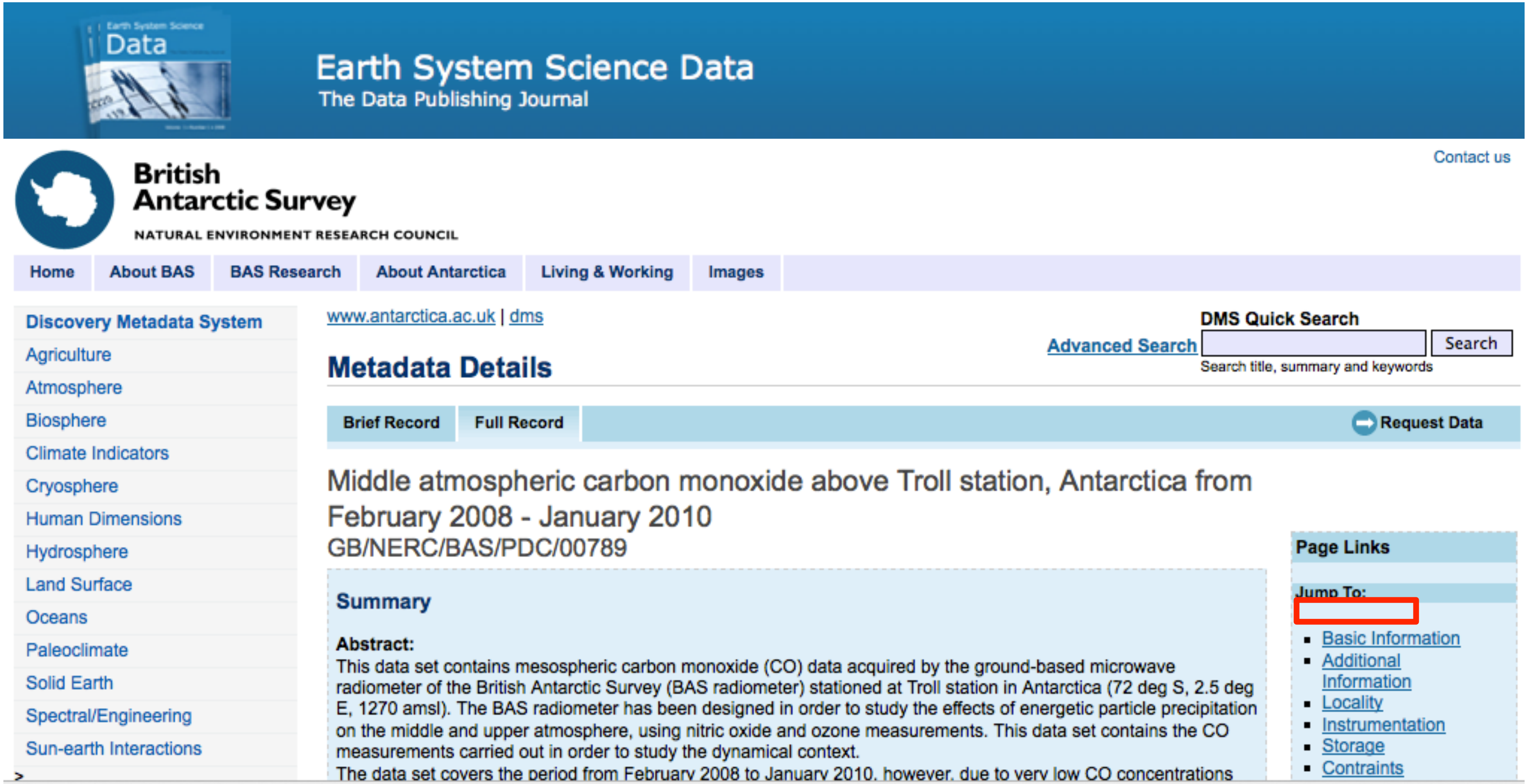
Earth Syst. Sci. Data Discuss., 6, 1-26, 2013
www.earth-syst-sci-data-discuss.net/6/1/2013/
doi:10.5194/essdd-6-1-2013

Mesospheric CO above Troll station, Antarctica observed by a ground based microwave radiometer

C. Straub[1], P. J. Espy[1], R. E. Hibbins[1], and D. A. Newnham[2]
[1]Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[2]British Antarctic Survey, Cambridge, UK

Abstract. This paper presents mesospheric carbon monoxide (CO) data acquired by the ground-based microwave radiometer of the British Antarctic Survey (BAS radiometer) stationed at Troll station in Antarctica (72° S, 2.5° E, 1270 a.m.s.l.). The data set covers the period from February 2008 to January 2010, however, due to very low CO

2013: CO above Troll Station, Original Data



2013: CO above Troll Station, Original Data

```
BAS microwave radiometerCO profiles acquired at Troll station, Antarctica between Feb 2008 and Jan 2010
Contact: Patrick Espy, tel:  +47 73 55 10 95, email: patrick.espy@ntnu.no
-------------------------------------------------------------------------------
date [UT]:                        2009-10-19 10:44:06
apriori contribution: The profile is most reliable where the contribution from the a priori profile is less than approx.
                    Negative values are a scaling artifact and should be regarded as close to 0.

The 2-sigma systematic errors provided have been determined using perturbation calculations:

temperature error:  error induced by the temperature profile (estimated error = 5K) needed as
                    additional information for the retrieval, mainly random
calibration error:  error induced by the calibration of the measured spectrum (estimated error = 10 percent), can be sys
spectroscopy error: we used lineintensity from HITRAN 2004 with an estimated error of 2 percent, systematic
channel shape error: uncertainty due to the use of a modified channel response function in the retrieval in order to cor
                    for an instability in one of the radiometers local oscillators after 2008-08-09, systematic
Error from measurement noise [K]:  0.1510, random
Smoothing error:    This error only needs to be considered if the profiles of the BAS radiometer are compared to
                    profiles with a significantly larger vertical resolution. For such a comparison the
                    better way would be to convolve the high-resolution profile with the AVK of the retrievals.

Sum of errors:      To build the sum of certain errors they are added up as follows sqrt(error1^2 + error2^2)
-------------------------------------------------------------------------------
pressure      altitude     vmr       apriori contribution   temperature error    calibration error    spectroscopy error
[hPa]         [km]         [ppmv]    [percent]              [ppmv]               [ppmv]               [ppmv]
    0.749894      50.679       0.060       -5.939       0.003       0.048       0.010       0.234       0.011
    0.562341      53.021       0.065      -20.151       0.002       0.056       0.011       0.319       0.012
    0.421697      55.337       0.072      -27.600       0.002       0.061       0.012       0.349       0.013
    0.316228      57.609       0.080      -29.442       0.004       0.067       0.013       0.298       0.015
```

Data Publishing Challenge #2

- Citability / Cite-worthy-ness / Reputation
  - NSF Proposal Preparation Instructions (2013)
    Proposals / PIs' CVs must contain:
    "A list of: (i) up to five products … Acceptable products must be citable and accessible including but not limited to publications, data sets, software, ..."
  - DFG "Rules of Good Scientific Practice" (2013):
    Recommendation 12 on authorship:
    contribution may be "preparation … of data"

HELMHOLTZ | GEMEINSCHAFT OPEN ACCESS

Fluxes of sedimenting material from sediment traps in the Atlantic Ocean

Review Status

This discussion paper is under review for the journal Earth System Science Data (ESSD).

S. Torres-Valdés[1], S. C. Painter[1], A. P. Martin[1], R. Sanders[1], and J. Felden[2]

[1]Ocean Biogeochemistry and Ecosystems Research Group, I
Southampton, SO14 3ZH, UK

[2]Center for Marine Environmental Sciences, Universität Bre
Bremen, Germany

A huge work to find, assess, collate (quality) data;

24 out of 43 text pages are source data references!

Abstract. We provide a data set assemblage of directly observed and derived fluxes of sedimenting material (total mass, POC , PON , $BSiO_2$, $CaCO_3$, PIC and lithogenic/terrigenous fluxes) obtained using sediment traps. This data assemblage contains over 5900 data points distributed across the Atlantic, from the Arctic Ocean to the Southern Ocean. Data from the Mediterranean Sea are also included. Data were compiled from a variety of sources: data repositories (e.g., BCO-DMO, PANGAEA), time series sites (e.g., BATS, CARIACO), published scientific papers and data provided by originating PI's. All sources are specified within the combined data set. Data from the World Ocean Atlas 2009 were extracted to coincide with flux

Does citation already work as an incentive?

- Home
- Online Library ESSD
- Online Library ESSDD
  - Papers in Open Discussion
  - Volumes and Issues
  - Special Issues
  - Most Commented Papers
  - Full Text Search
  - Title and Author Search
- Alerts & RSS Feeds
- General Information
- Submission
- Revi
- Prod
- Subs
- Com

Earth Syst. Sci. Data Discuss., 5, 491-520, 2012
www.earth-syst-sci-data-discuss.net/5/491/2012/
doi:10.5194/essdd-5-491-2012

Article | **Discussion** | Related Articles

Global marine plankton functional type biomass distributions: coccolithophores

C. J. O'Brien, J. A. Peloquin, M. Vogt, M. Heinle, N. Gruber, P. Ajani, H. Andruleit, J. Arístegui, L. Beaufort, M. Estrada, D. Karentz, E. Kopczyńska, R. Lee, T. Pritchard, and C. Widdicombe

Interactive Discussion **Status: Open (indefinitely extended)**

AC: Author Comment | **RC**: Referee Comment | **SC**: Short Comment | **EC**: Editor Comment

[Post a Comment] [Subscribe to Comment Alert] - Printer-friendly Version - Supplement

Reviewer: „no effort appears to have been made to engage the specialist scientists who have spent months or years at sea collecting such data. " - not knowing that:

Authors asked 164 potential contributors – got answer from 13!

Data Publishing Challenge #3

- **Linking text and data … and much more**
  - **The Lancet "Reducing waste from incomplete or unusable reports of biomedical research" (2014)**
  - **"… studies of published trial reports showed that … 40–89% were non-replicable"**
  - **Offered a long laundry list of "Components of study documentation" to be published**

Now this laundry list is *really scary!*

1 The protocol and related documents, such as details submitted for study registration

3 Supplementary materials, such as education materials for patients, clinician training resources, and videos

7 The primary data, data manuals, and statistical code for analyses

9 Reliable and stable bidirectional linkages between all these elements

2012: Nature CC & ESSD; Carbon data aggregation at global scale

nature climate change

Search Go

Advanced search

Home | Opinion & Analysis | Research | Current Issue | Archive | For Authors & Referees | About the journal

Opinion & Analysis > Commentaries > Article

NATURE CLIMATE CHANGE | COMMENTARY

The challenge to keep global warming below 2 °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

Current issue

December 2012 - Vol 2 Issue 12

- Current issue
- Subscribe
- Recommend to library

E-alert | RSS | Facebook | Twitter

2012: Nature CC & ESSD; Carbon data aggregation at global scale

nature climate
Home
Opinion &
NATURE C
The c
Glen P. Pe
Le Quéré,
Affiliation
Nature Cli
Published

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]

[1]Tyndall Centre for Climate Change Research, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK
[2]Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA
[3]National Oceanic & Atmosphere Administration, Earth System Research Laboratory (NOAA/ESRL), Boulder, Colorado 80305, USA
[4]Woods Hole Research Centre (WHRC), Falmouth, Massachusetts 02540, USA
[5]Cabot Institute, Dept of Geography, University of Bristol, UK

Discussion Paper | Discussion Paper | Discussion Paper | Discu

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
|◀ | ▶|
◀ | ▶
Back | Close
Full Screen / Esc
Printer-friendly Version

2012: Nature CC & ESSD; Carbon data aggregation at global scale

natu
clim
Home
Opinion &
NATURE C
The c
Glen P. Pe
Le Quéré,
Affiliation
Nature Cli
Published

Terrestrial CO_2 sink (positive values represent a flux from the atmosphere to the land)

All values in petagrams of carbon per year (PgC/yr), for the globe. For values in carbon dioxide (CO_2), multi

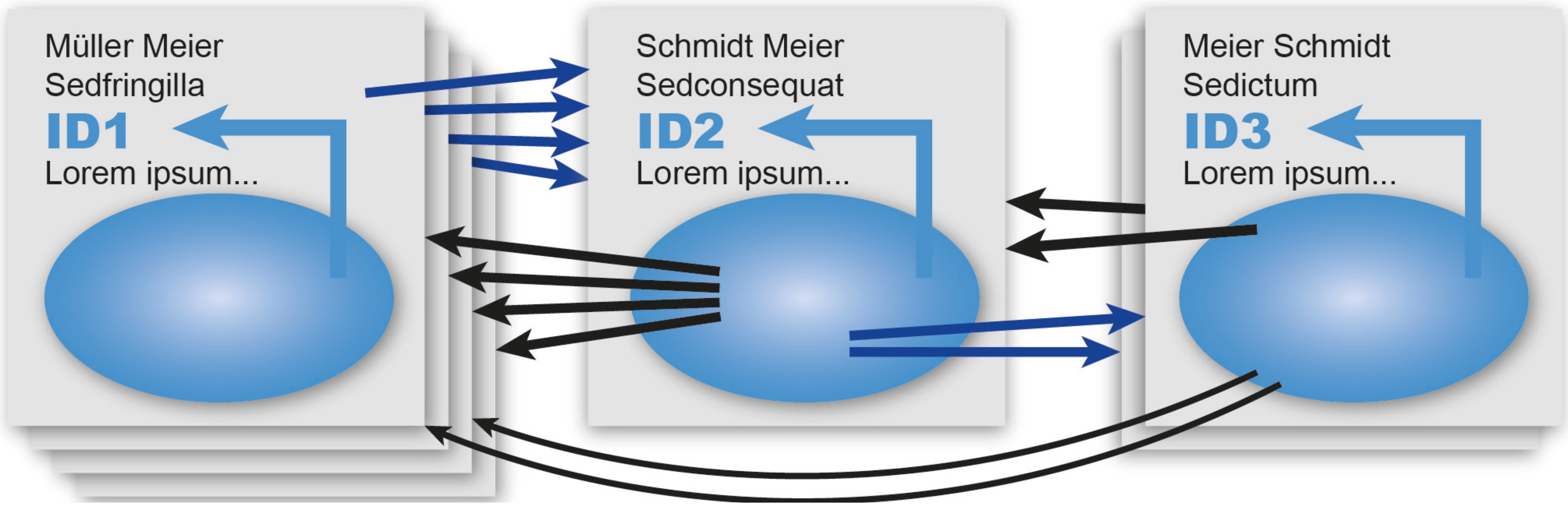
1PgC = 1 petagram of carbon = 1 billion tonnes C = 1 gigatonne C = 3.67 billion tonnes of CO_2

Cite as:

CLM4CN	Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence,
HYLAND	Levy, P. E., M. G. R. Cannell, et al. (2004). "Modelling the impact of future changes in clim
LPJ-GUESS	Smith, B., I. C. Prentice, et al. (2001). "Representation of vegetation dynamics in the mode
LPJ	Sitch, S., B. Smith, et al. (2003). "Evaluation of ecosystem dynamics, plant geography and
O-CN	Zaehle, S., P. Ciais, et al. (2011). "Carbon benefits of anthropogenic reactive nitrogen offs
ORCHIDEE	Krinner, G., N. Viovy, et al. (2005). "A dynamic global vegetation model for studies of the
SDGVM	Woodward, F. I. and M. R. Lomas (2004). "Vegetation dynamics - simulating responses to
JULES	Clark, D. B., L. M. Mercado, et al. (2011). "The Joint UK Land Environment Simulator (JULE
VEGAS	Zeng, N., A. Mariotti, et al. (2005). "Terrestrial mechanisms of interannual CO2 variability.

	Terrestrial CO2 sink as a residual		**Models**			
Year	of the global carbon budget		CLM4CN	HYLAND	LPJ-GUESS	LPJ
1959	0,42		0,79	2,02	0,42	-0,83
1960	1,14		0,75	1,53	1,16	0,81
1961	1,20		0,30	1,71	-0,07	-0,55
1962	1,76		0,79	2,37	1,25	0,57
1963	1,72		-1,20	1,81	0,26	-0,37

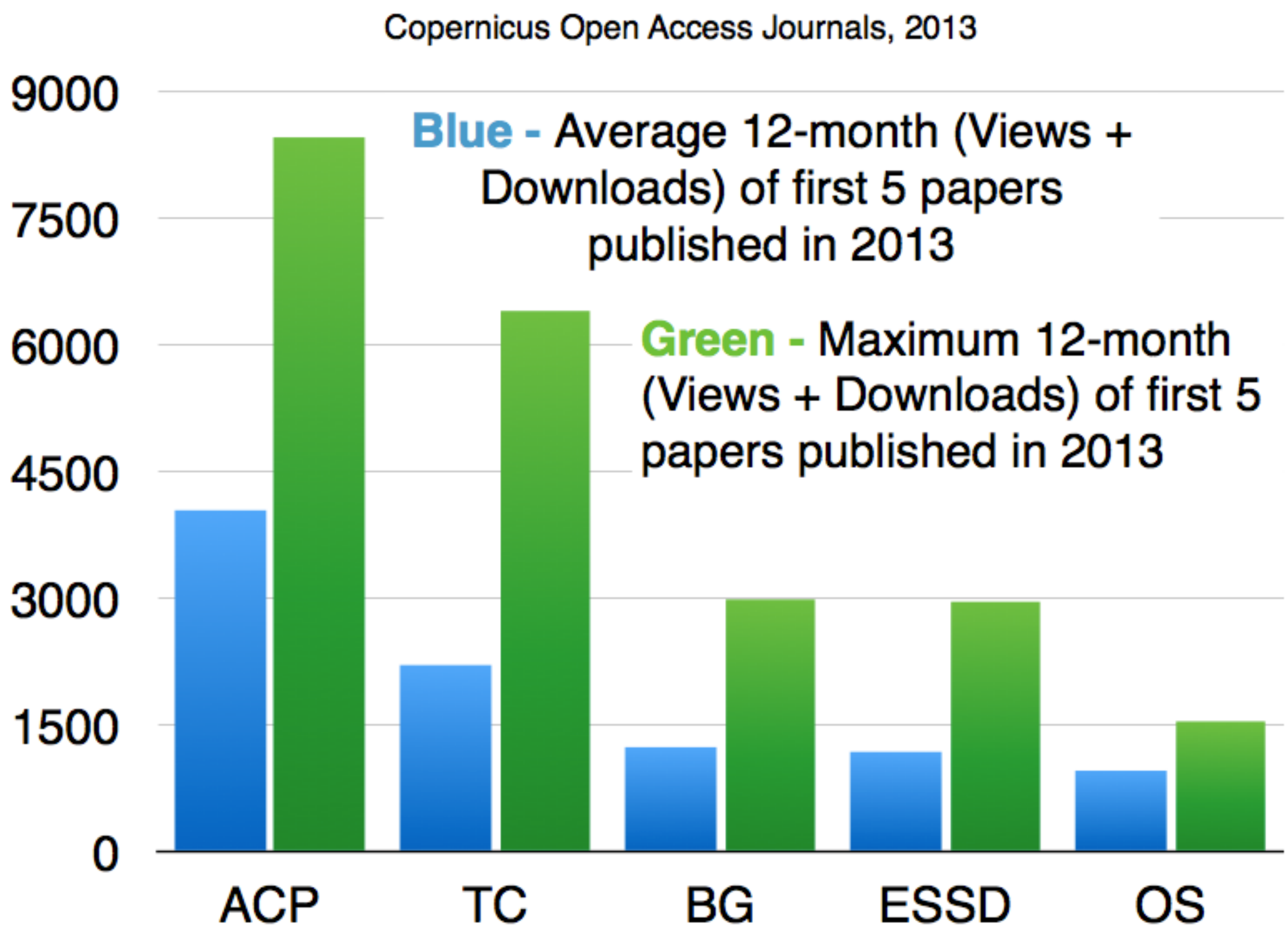
Linking Text and Data



Data (in repository)

Article in data journal

Article in „classical“ journal


Copernicus Open Access Journals, 2013
9000
7500
6000
4500
3000
1500
0
Blue - Average 12-month (Views + Downloads) of first 5 papers published in 2013
Green - Maximum 12-month (Views + Downloads) of first 5 papers published in 2013
ACP
TC
BG
ESSD
OS

Conclusions

- Socio-cultural change is on the way (may need just a few more decades)
  - Need for change/quality is recognized (Lancet)
  - NSF "5 products" rule offers the way out of the metrics dungeon
- "Technical" challenges remain, e.g.
  - Repositories for computer code etc.
  - Quality assessment for "protocols" etc.
  - bidirectional Linking of Everthing Open (b-LEO)
  - And, I did not even mention versioning …