HELMHOLTZ GEMEINSCHAFT OPEN ACCESS

Open Access and Climate Knowledge in Theory and Practice

Hans Pfeiffenberger

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

JPI Climate Towards Open Climate Knowledge, 2015-01-13, Vienna

AWI

Agenda

- **The Theory:**
  BBB, RS, G8 – you name it
- **The Practise:**
  About Recommendations and Sledgehammers
- **Examples (Macro and Micro; Data and Software)**
- **Conclusions for JPI Climate A2K**

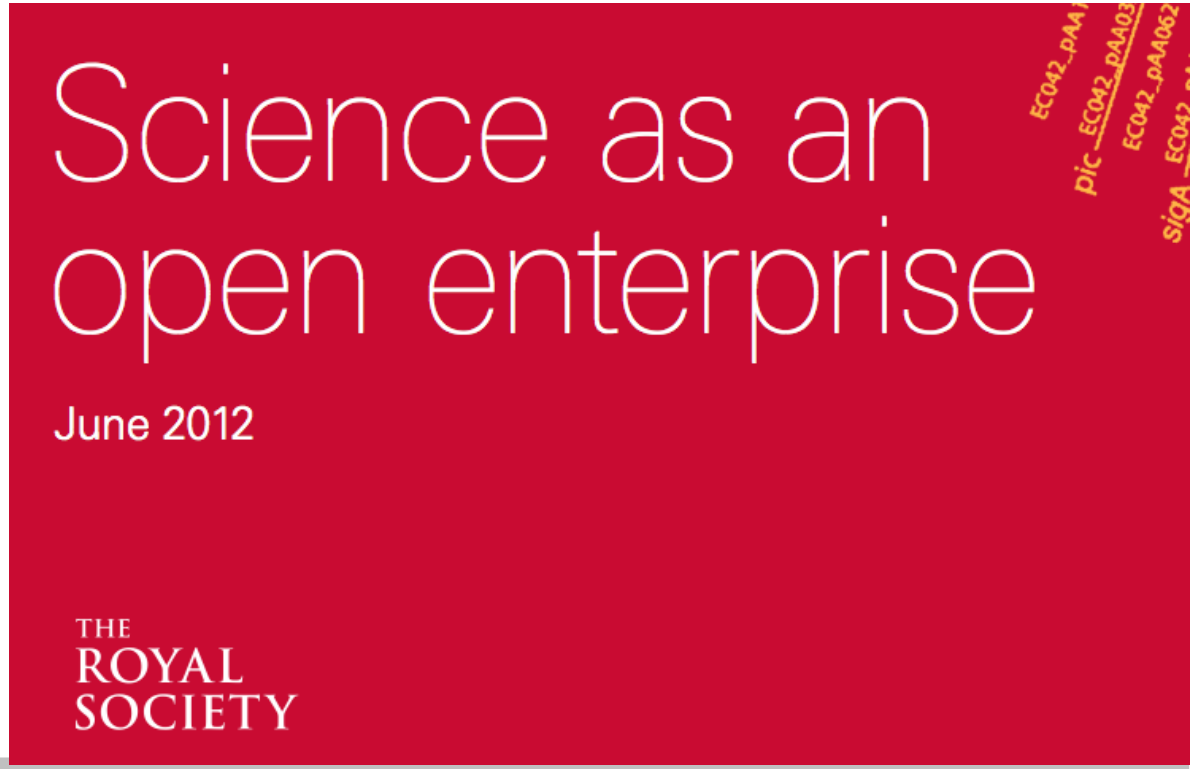
Royal Society: Science as an Open Enterprise (2012) [1]

- **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 take on data:**

  Intelligent Openness



HELMHOLTZ GEMEINSCHAFT OPEN ACCESS

GOV.UK

Search

Departments Worldwide How government works Get involved
Policies Publications Consultations Statistics Announcements

Policy paper

Open Data Charter

From: Cabinet Office
First published: 18 June 2013
Part of: G8 communiqué and documents, UK Government Partnership Summit 201 transparency and accountability of g

Published 18 June 2013

Contents

1. Principle 1: Open Data by Default
2. Principle 2: Quality and Quantity
3. Principle 3: Usable by All
4. Principle 4: Releasing Data for Improved Governance
5. Principle 5 Releasing Data for Innovation
6. Technical annex

Charter on open data signed by G8 leaders to promote transparency, innovation and accountability.

Documents

HTML G8 Open Data Charter and Technical Annex

The economic case: Making primary data available doubles the amount of knowledge gained

- **Hubble Space Telescope data**

- **ENCODE ("Human Genome 2.0")**
  - "clumsy etiquette-based restrictions" … "starting to show their age and a lack of clarity"
    Birney, The making of ENCODE, Nature 2012, doi:10.1038/489049a

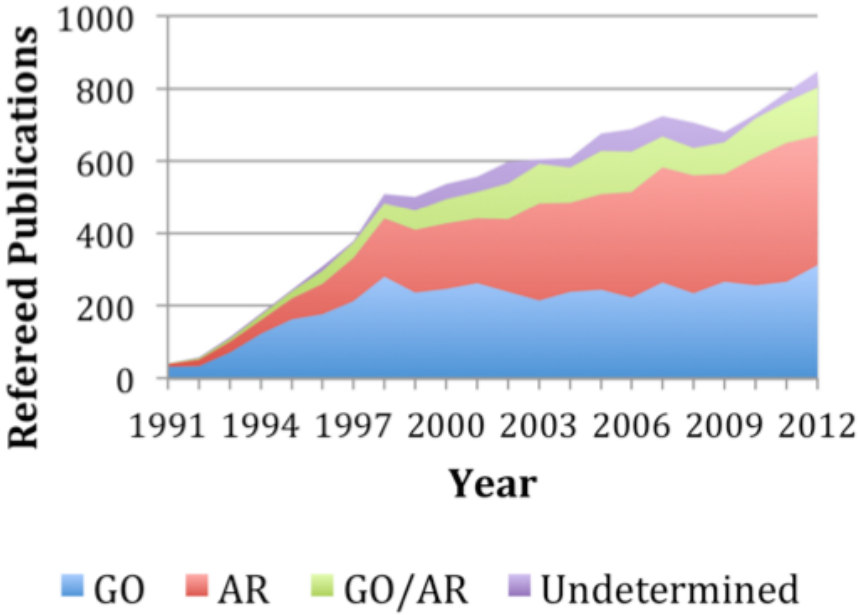


Figure 1. Number of refereed publications based on Hubble Space Telescope data in the Multi-mission Archive at the Space Telescope Science Institute. GO = guest observer programs (papers published by the principal investigator and immediate collaborators), AR = archival research (papers published by researchers not affiliated with the principal investigator), GO+AR = papers that include both GO and AR data, and Undetermined = papers for which the origin of the data is unclear.

It is so obvious that

- **Open Access, Open Data, Open Knowledge is a Good Thing! – is it not?**
- **Why do we even need to talk about it?**
- **Why doesn't it happen, just so?**

What about BBB? (1) (Budapest, Bethesda, Berlin)

- Budapest (2002) [3] was certainly the radical definition of "Open":
  - "free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. "
- However,
  - just by a number of individuals
  - just about articles

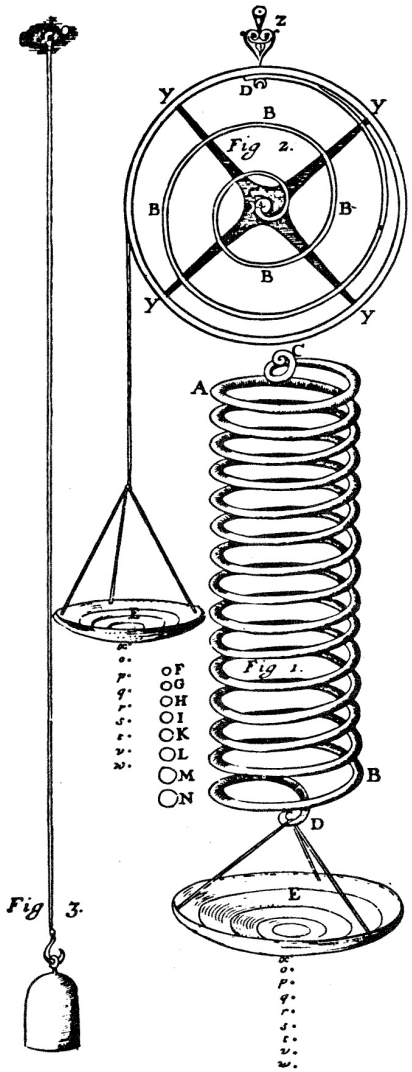
What about BBB? (2)

- Berlin (2003) [4] was certainly not radical anymore (thanks to Budapest)
- It was a declaration by notable institutions
- did include data and (implicitly) software
  - "Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material."
- But the commitment was weak
  - "… we intend to make progress by encouraging our researchers/grant recipients to publish their work according to the principles of the open access paradigm"

Status of BBB compliance

- We have (2015) ca.
  - 20-30% OA to articles
  - 1% to data (with disciplinary exceptions!)
- Why is appealing to researchers, citing the public good, not sufficient?
- As long as there is (perceived) risk and/or cost, but no rewards for compliance …
- Now, funders are getting out the sledgehammer
  - Netherlands: 60% by 2019 or else …

Openess in the 17th Century



Hooke, published his law

1676 by anagram „ceiiinosssssttuv“

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*.

The very first B: Bermuda (1996/97) [6]

- "Policies on Release of Human Genomic Sequence Data Bermuda-Quality Sequence"
  - Timely release, quality assurance

- Why did they care to write it?
  Nobody, no institution could have done it alone!
  (at that time)

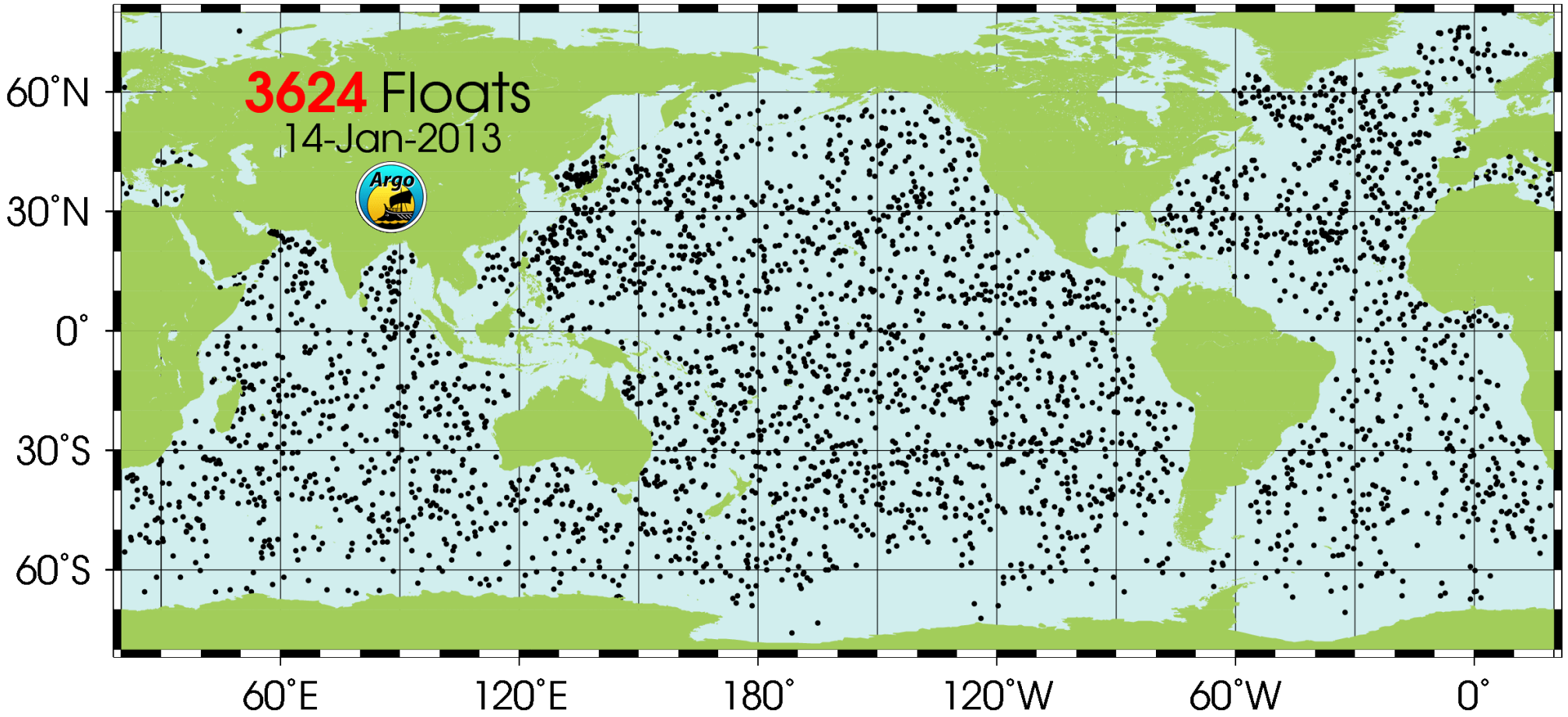
- Why does it still work, 100%?

  Meanwhile, journals refuse publication, otherwise!

AGU (2013) PLoS (2014)

- AGU reserves the right to refuse publication when authors are unwilling to make the underlying data available or otherwise refuse to comply with this Data Policy
- PLoS: Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection. PLOS journal editors encourage researchers to contact them if they encounter difficulties in obtaining data … . If restrictions on access to data come to light after publication, we reserve the right to post a correction, to contact the authors' institutions and funders, or in extreme cases to retract the publication.

ARGO, the biggest experiment in the world



ARGO : sharing data openly and immediately [7]

ARGO is really fascinating: There are

- **More than 3.000 buoys, built by lots of companies**
- **From / funded by more than 30 countries,**

- **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**

Earth System Science Data: The details in practise

- **Founded 2008, to address**
  - **quality (through peer review)**
  - **and rewards (through unquestionable cite-ability)**
- **concept for the "long tail",**
  but many huge data aggregation projects/products
- **Has an Open Only policy (but…)**

- **by the end of 2014: ca. 100 data articles published**
- **Indexed by Scopus**

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 [hPa]	altitude [km]	vmr [ppmv]	apriori contribution [percent]	temperature error [ppmv]		calibration error [ppmv]		spectroscopy error [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	

Sun-earth Interactions

measurements carried out in order to study the dynamical context.
The data set covers the period from February 2008 to January 2010, however, due to very low CO concentrations

- Storage
- Contraints

General Information

Submission

Review

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

Fluxes of sedimenting material from sediment traps in the Atlantic Ocean

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, Southampton, SO14 3ZH, UK

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

Review Status

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

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

HELMHOLTZ GEMEINSCHAFT OPEN ACCESS

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

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
© Author(s) 2012. This work is distributed
under the Creative Commons Attribution 3.0 License.

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

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!

2012: Nature Climate Change, ESSD and CDIAC - interlinked

	A	B	C	D	E	F	G
1		**Terrestrial CO_2 sink (positive values represent a flux from the atmosphere to the land)**					
2		All values in petagrams of carbon per year (PgC/yr), for the globe. For values in carbon dioxide (CO_2), multi					
3		1PgC = 1 petagram of carbon = 1 billion tonnes C = 1 gigatonne C = 3.67 billion tonnes of CO_2					
4		**Cite as:**					
5		CLM4CN	Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence,				
6		HYLAND	Levy, P. E., M. G. R. Cannell, et al. (2004). "Modelling the impact of future changes in clim				
7		LPJ-GUESS	Smith, B., I. C. Prentice, et al. (2001). "Representation of vegetation dynamics in the mod				
8		LPJ	Sitch, S., B. Smith, et al. (2003). "Evaluation of ecosystem dynamics, plant geography and				
9		O-CN	Zaehle, S., P. Ciais, et al. (2011). "Carbon benefits of anthropogenic reactive nitrogen offs				
10		ORCHIDEE	Krinner, G., N. Viovy, et al. (2005). "A dynamic global vegetation model for studies of the				
11		SDGVM	Woodward, F. I. and M. R. Lomas (2004). "Vegetation dynamics - simulating responses to				
12		JULES	Clark, D. B., L. M. Mercado, et al. (2011). "The Joint UK Land Environment Simulator (JULE				
13		VEGAS	Zeng, N., A. Mariotti, et al. (2005). "Terrestrial mechanisms of interannual CO2 variability.				
14							
15			Terrestrial CO2 sink as a residual	**Models**			
16	Year		of the global carbon budget	CLM4CN	HYLAND	LPJ-GUESS	LPJ
17	1959	0,42		0,79	2,02	0,42	-0,83
18	1960	1,14		0,75	1,53	1,16	0,81
19	1961	1,20		0,30	1,71	-0,07	-0,55
20	1962	1,76		0,79	2,37	1,25	0,57
21	1963	1,72		-1,20	1,81	0,26	-0,37

GLOBAL CARBON ATLAS

The Global Carbon Atlas is a platform to explore and visualize the most up-to-date data on carbon fluxes resulting from human activities and natural processes. Human impacts on the carbon cycle are the most important cause of climate change.

[7]

Funded by BNP Paribas

Implemented by WeDoData

(„data journalism“)

... And a recent story about software

- **Two licenses from climate modelling:**
  - **"amended" GPL, a no-go!**
  - **The opposite of GPL ?**
    - **"… This licence agreement is a personal licence, … Before using the Software, you have to ensure that your employer has accepted the terms of this license agreement. …**
    - **You are not allowed to distribute the Software or any part of it, neither in its original nor any modified form.**
    - **… Any modifications and improvements of the Software must be communicated to the coordinator of … You shall grant the licensors a non-exclusive, world-wide, irrevocable, perpetual, royalty-free license … "**

JPI Climate similar to IPY 2007-2008?

- **IPY Data Policy [8]**
  - **"IPY Joint Committee requires that IPY data, including operational data delivered in real time, are made available fully, freely, openly, and on the shortest feasible timescale."**
- **Had Zero Impact!**
  - **"A lesson in sharing", David Carlson, Nature 469, 293 (20 January 2011) doi:10.1038/469293a**
  - **"despite the best efforts …, we cannot say how users might discover or access IPY data five years hence."**

Conclusions

- Make it easy to comprehend and comply with policy
- Don't invent new policy, copy it if you can
- Don't invent new licenses, ever
  - choose the simplest; provide "legal interoperability"
- Determine and concentrate on priorities!
  - e.g.: "Open Science"; data and software ?
  - nobody knows ODT; possibly irrelevant (cloud!)
- Work with societies and publishers
  - most public funders are "conservative", can't police

Thank you!

This work is based on discussions etc. with
- **Dave Carlson, ESSD**,

and insights and input from all members of
- **Helmholtz Open Science Group** oa.helmholtz.de/en
- **Allianzinitiative** allianzinitiative.de/en
- **Science Europe WG on Research Data** scienceeurope.org/policy/working-groups/Research-Data

References

[1] https://royalsociety.org/policy/projects/science-public-enterprise/Report/

[2] https://www.gov.uk/government/publications/open-data-charter

[3] http://www.budapestopenaccessinitiative.org/read

[4] http://openaccess.mpg.de/Berliner-Erklaerung

[5] http://web.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml

[6] http://www.argodatamgt.org/Access-to-data/Argo-DOI-Digital-Object-Identifier

[7] http://www.globalcarbonatlas.org/

[8] http://classic.ipy.org/Subcommittees/final_ipy_data_policy.pdf