

# Open Science goes Geo – ESSD and RDA

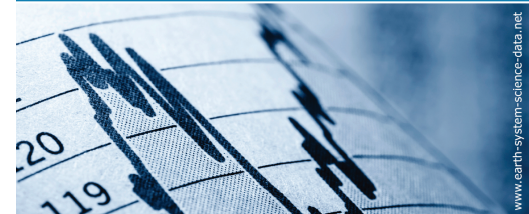
Hans Pfeiffenberger, Dave Carlson

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

EGU2015, SC23 Open Science goes Geo - Part I: Research Data  
2015-01-13, Vienna

Earth System Science

**Data** The Data Publishing Journal



Volume 1 • Number 1 • 2008

 Copernicus Publications  
The Innovative Open Access Publisher



AWI 



## Agenda

- **This will be a List of Required Reading**
- **And some Concrete Examples**
- **Recommendations for Young Researchers**



## 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**





## 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
- **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**.



Policy paper

# Open Data Charter

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

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



[G8 Open Data Charter and Technical Annex](#)



## The „economic“ case: Making primary data available **doubles the amount of knowledge gained**

- Hubble Space Telescope data
- More examples, mostly from “survey” data

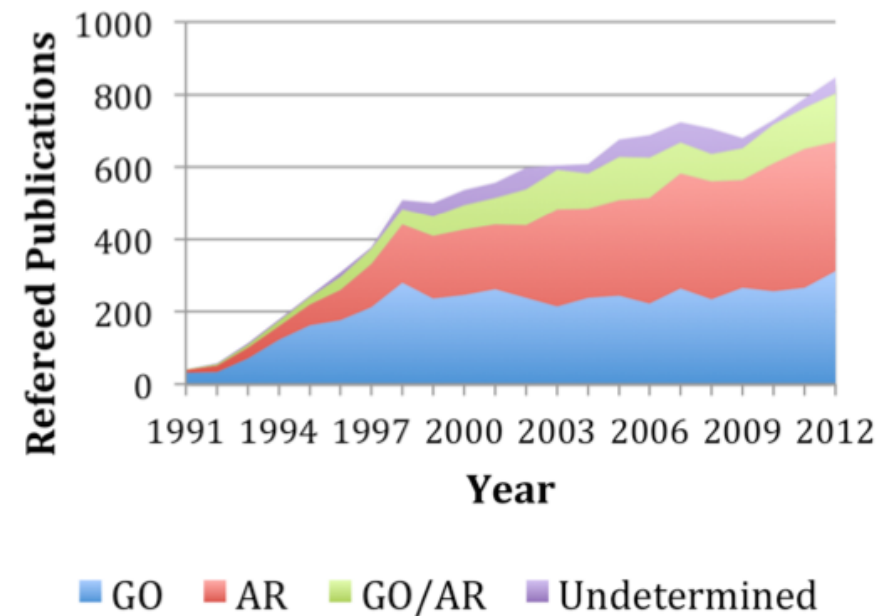
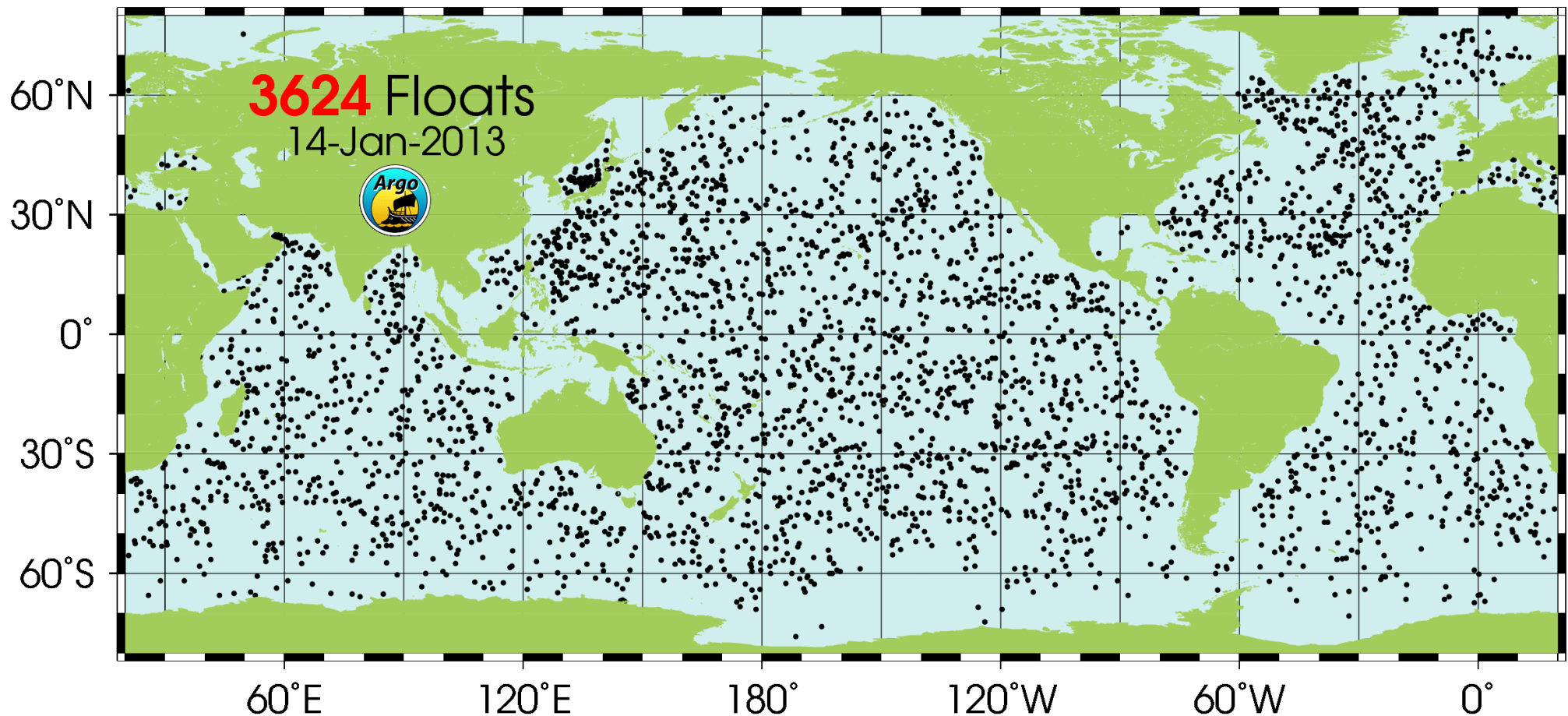


Figure 1. Number of refereed publications based on Hubble Space Telescope data in the Multi-mission Archive at the Space Telescope Science Institute. GO = guest observer programs (papers published by the principal investigator and immediate collaborators), AR = archival research (papers published by 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.



## ARGO, the biggest experiment in the world





## **ARGO : sharing data openly and immediately [7]**

**ARGO is really fascinating:**

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

- **Founded 2008**, to address
  - **quality** (through peer review)
  - **and rewards** (through *unquestionable* cite-ability)
- Concept originally for the “**long tail**”, but now many **huge data aggregation projects/products**
- Has an **Open Only** policy
- Now **Indexed by Scopus, and SCI/ISI/WoS**
  - **Authors can find citations of their data through Web of Knowledge and Scopus**



## 2013: CO above Troll Station, Original Data

BAS microwave radiometer CO 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 systematic  
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 correct 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  $\sqrt{\text{error1}^2 + \text{error2}^2}$

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	0.234
0.562341	53.021	0.065	-20.151	0.002	0.056	0.319
0.421697	55.337	0.072	-27.600	0.002	0.061	0.349
0.316228	57.609	0.080	-29.442	0.004	0.067	0.298

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
- Constraints

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<sup>1</sup>, S. C. Painter<sup>1</sup>, A. P. Martin<sup>1</sup>,  
R. Sanders<sup>1</sup>, and J. Felden<sup>2</sup>

<sup>1</sup>Ocean Biogeochemistry and Ecosystems Research Group,  
Southampton, SO14 3ZH, UK

<sup>2</sup>Center for Marine Environmental Sciences, Universität Bremen,  
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 pages of text 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<sub>2</sub>, CaCO<sub>3</sub>, 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

© 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. Aristegui, L. Beaufort, M. Estrada, D. Karentz, E. Koczyńska, 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!**

## 2012: Nature Climate Change, ESSD and CDIAC - interlinked

	A	B	C	D	E	F	G
1		<b>Terrestrial CO<sub>2</sub> sink (positive values represent a flux from the atmosphere to the land)</b>					
2		All values in petagrams of carbon per year (PgC/yr), for the globe. For values in carbon dioxide (CO <sub>2</sub> ), multi					
3		1PgC = 1 petagram of carbon = 1 billion tonnes C = 1 gigatonne C = 3.67 billion tonnes of CO <sub>2</sub>					
4		<b>Cite as:</b>					
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 CO <sub>2</sub> variability.				
14							
15		Terrestrial CO <sub>2</sub> sink as a residual		<b>Models</b>			
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

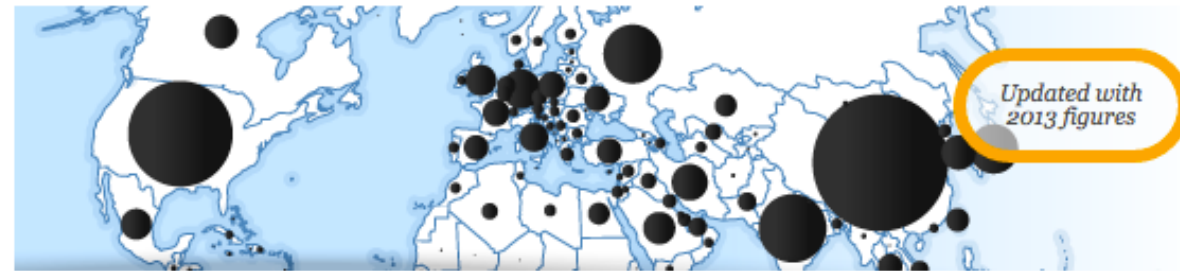
*The Global Carbon Atlas is a platform to explore and visualize the most up-to-date data on carbon fluxes resulting from human activities and natural processes. Human impacts on the carbon cycle are the most important cause of climate change.*

[7]

## Outreach

*Take a journey through the history and future of human development and carbon*

Go



## Emissions

*Explore and download global and country level carbon emissions from human activity*

Go

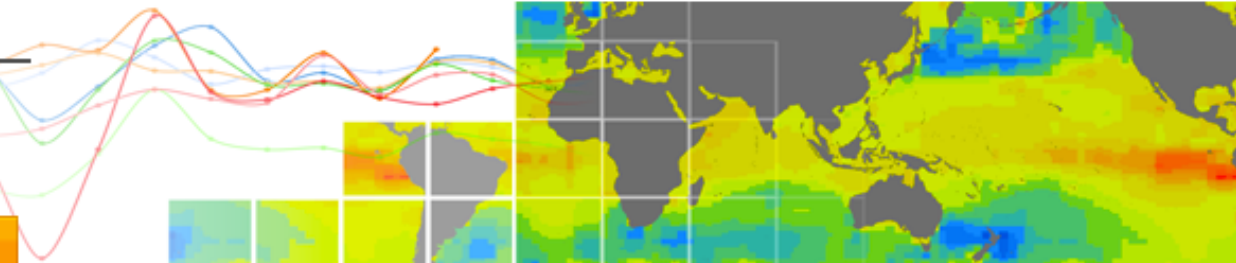
Funded by  
BNP Paribas

Implemented  
by WeDoData

## Research

*Explore and visualize research carbon data, and get access through data providers*

Go



(„data  
journalism“)



## Now to the Research Data Alliance (RDA) ...

- **Depending on your ambitions, take this as**
  - **Reassurance, that things will become easier over time**
    - **As a disciplinary scientist**
  - **An indication that this is the group to work with**
    - **As a data scientist or developer in ESSI**
- **Next plenary: Paris, September 2015**

# RDA/WDS Publishing Data Services WG



## Group details

**Workflow status:** Approved

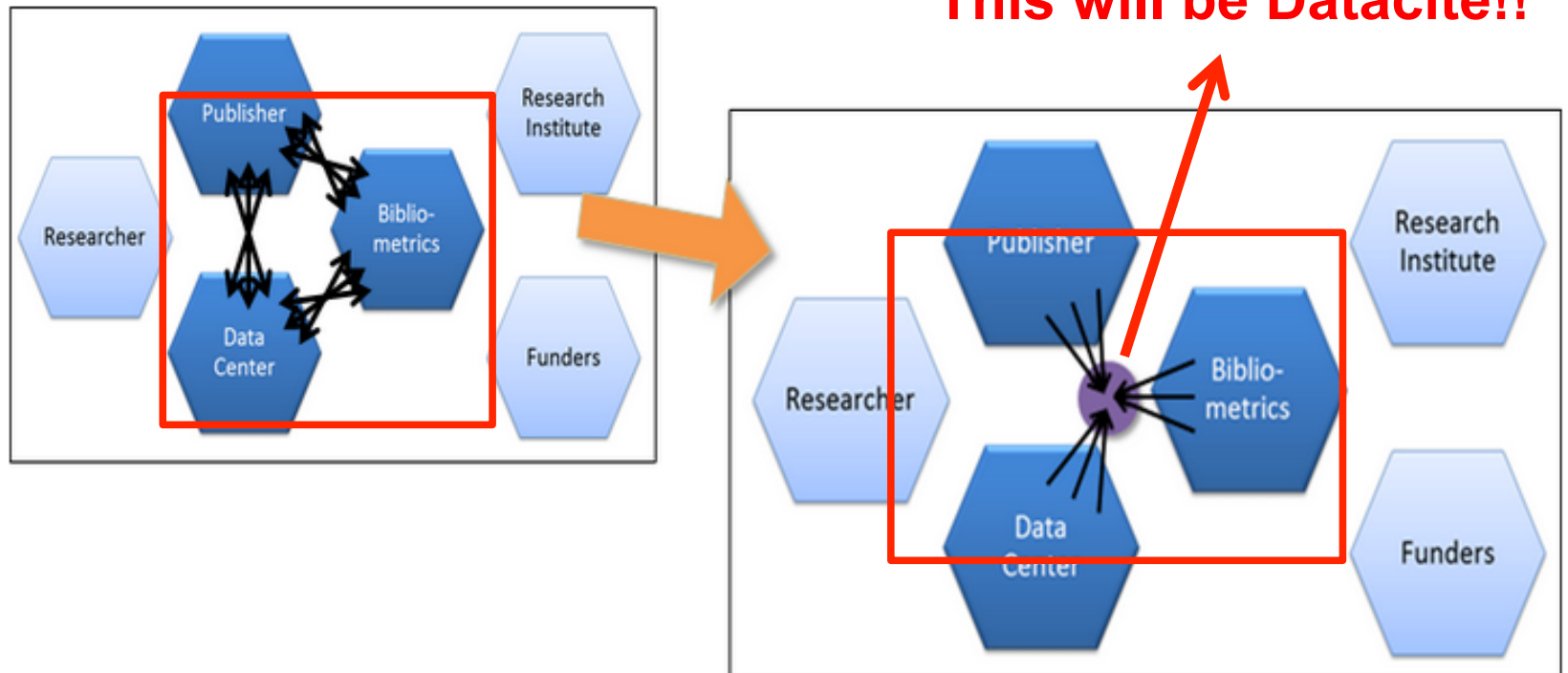
**Status:** Recognised & Endorsed

**Chair(s):** Hylke Koers, Adrian Burton

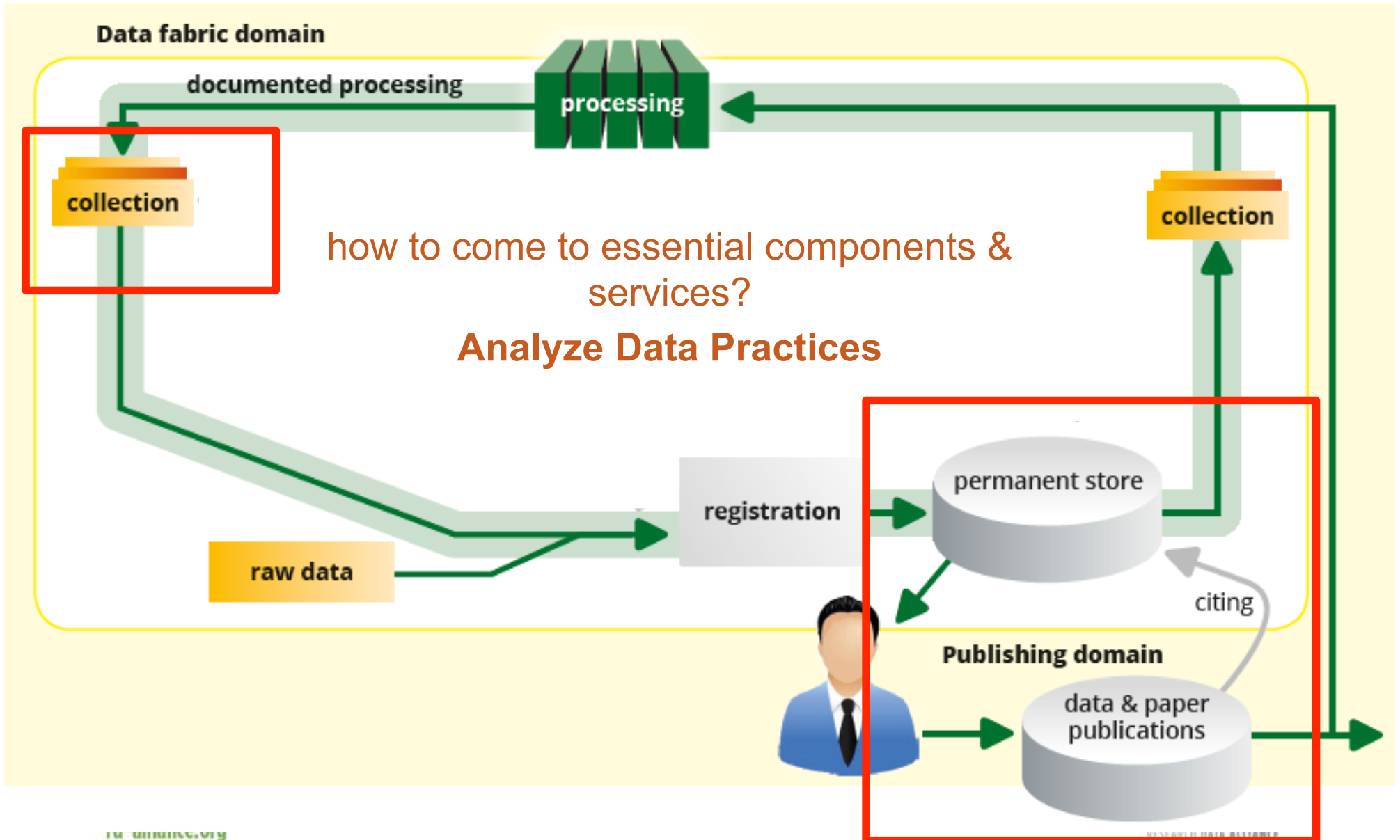
**Case Statement:** Download

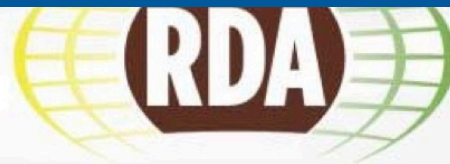
Building on pre-existing components and international initiatives, the WG is focusing on a one-for-all cross-referencing service for the links between data and publications. The challenge for the working group is "How to move from a plethora of (mostly) bilateral arrangements to a **one-for-all service model infrastructure** for the research data publication landscape?"

**This will be Datacite!!**









RESEARCH DATA ALLIANCE

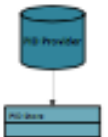
## Data Citation WG Making Dynamic Data Citable

<https://www.rd-alliance.org/working-groups/data-citation-wg.html>

Andreas Rauber, Ari Asmi, Dieter van Uytvanck

### ■ Goal:

- Ensure cite-ability of data at arbitrary levels of granularity, particularly when data is large-volume and dynamic
- Machine-actionable, variety of data types





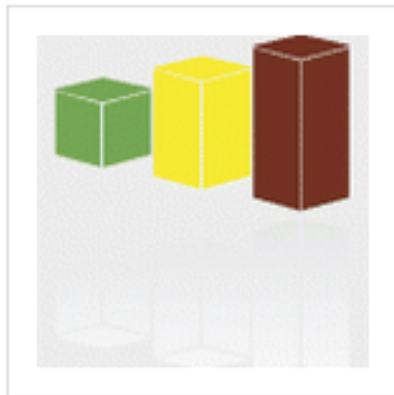
# RDA Interest Groups

Marine Data Harmonization IG

Geospatial IG

Biodiversity Data Integration IG

Big Data Analytics IG



## Group details

**Status:** Recognised & Endorsed

**Chair(s):** Morris Riedel, Peter Baumann, Kwo-Sen Kuo

**Case Statement:** [Download](#)

The Big Data Analytics (BDA) Interest Group seeks to develop community based recommendations on feasible data analytics approaches to address scientific community needs of utilizing la



## Recommendations

- To the **Young Researcher**
  - **Make sure your data are well documented!**
    - **Could you understand it 2 years after your project?**
    - **Could somebody else understand it now?**
  - **Put it in a reliable repository**
  - **Publish it**
  - **... Become part of data integration effort**
- To the **potential Data Scientist**
  - **Join RDA**
  - **... Become chair of a RDA WG**

Thank you!



## References

[http://www.earth-syst-sci-data.net/recent\\_papers.html](http://www.earth-syst-sci-data.net/recent_papers.html)

<https://royalsociety.org/policy/projects/science-public-enterprise/Report/>

<https://www.gov.uk/government/publications/open-data-charter>

<http://www.argodatamgt.org/Access-to-data/Argo-DOI-Digital-Object-Identifier>

<http://www.globalcarbonatlas.org/>

<https://rd-alliance.org/groups/rdawds-publishing-data-services-wg.html>

[https://rd-alliance.org/system/files/documents/140923\\_rda\\_wg\\_dc.pdf](https://rd-alliance.org/system/files/documents/140923_rda_wg_dc.pdf)

<https://rd-alliance.org/groups/working-groups>

<https://rd-alliance.org/groups/interest-groups>