ESSD Experiences with Linking

E-IRG Prague, 2022-12-13

Hans Pfeiffenberger

Copernicus Gesellschaft e.V.

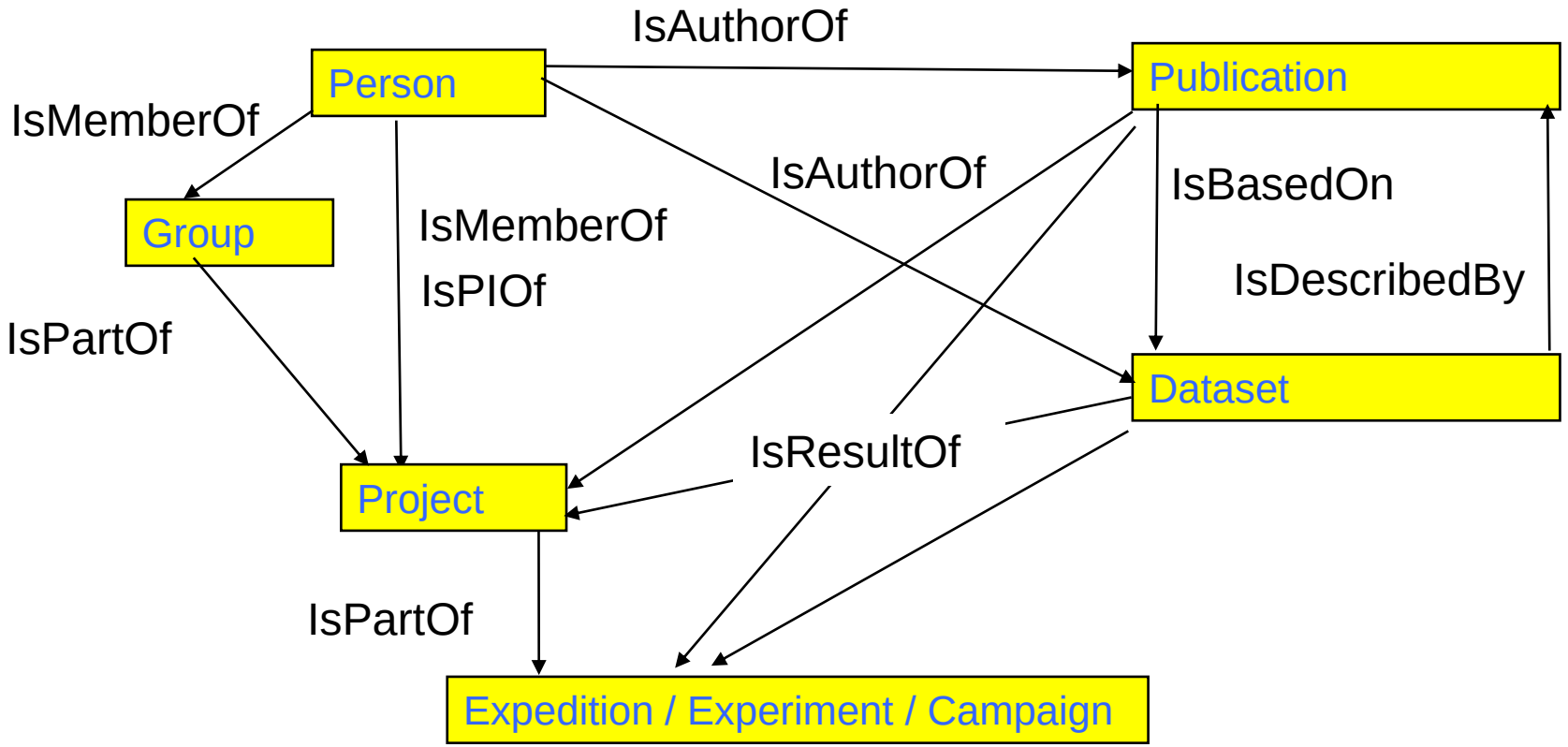
Agenda

1. Linking in 2005: link between articles, data => ESSD
2. Publishers' linking ecosystem
   => findability, availability, recognition/reward
3. Linking environment by 2014: identifiers for many kinds of objects
4. Linking example Global Carbon Budget
5. The future: executable articles / executable research compendia
6. Conclusions

"May all your problems be technical"

1. Linking in 2005: link between articles, data => ESSD

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



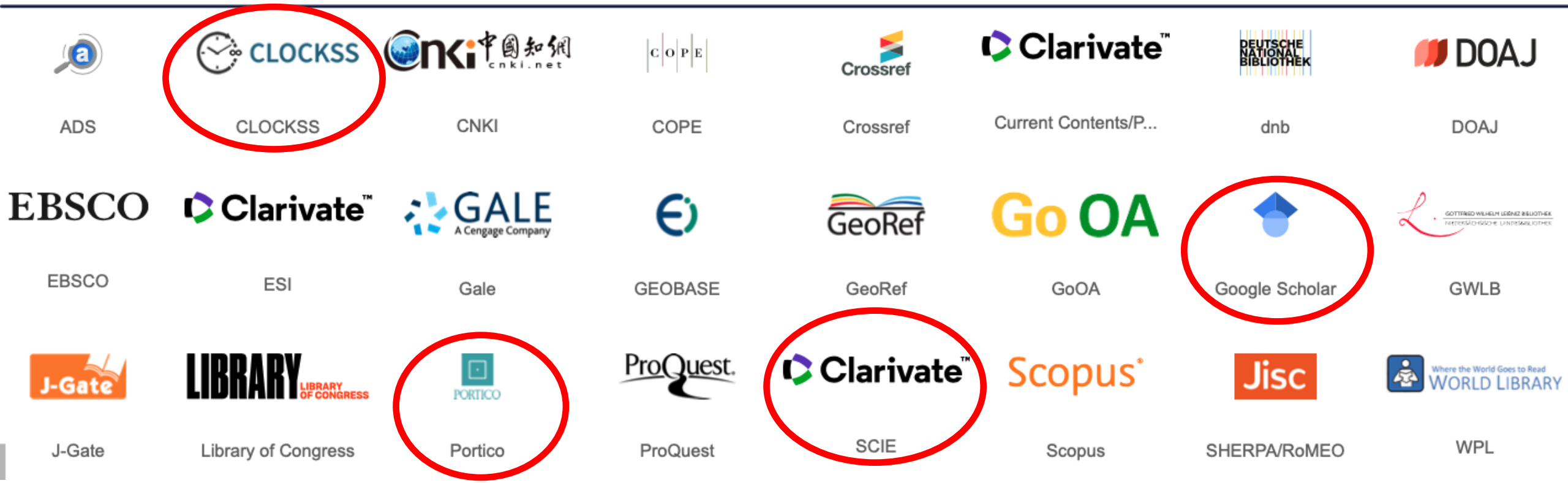
Data-DOIs

ESSD 2008, Motivation

- IPY 2007/2008: data legacy
- Provide peer review to data
- Provide recognition (reward) to data creators (“authors”)

2. Publishers' linking ecosystem
=> findability, availability, recognition/reward

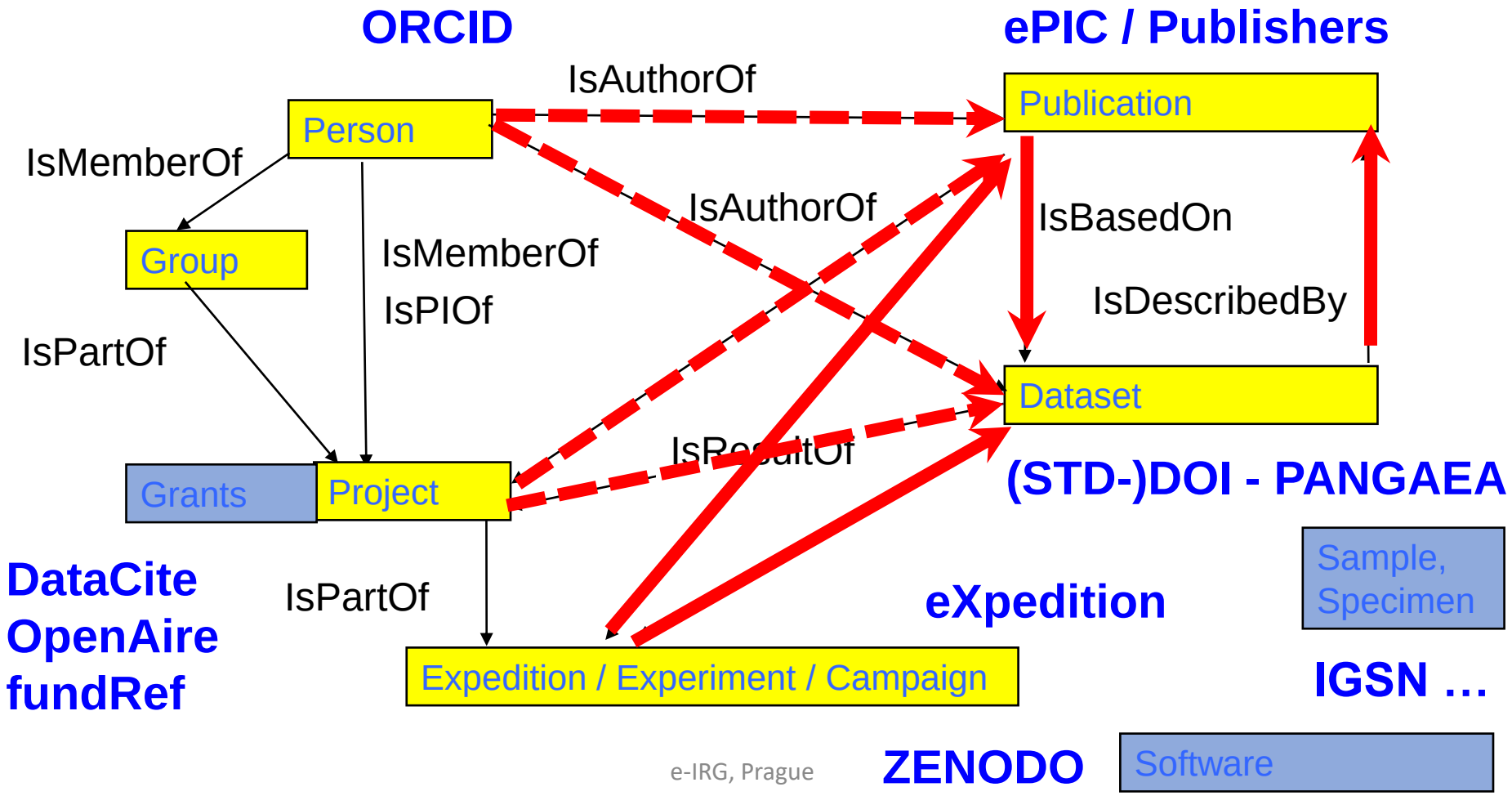
Earth System Science Data is indexed and archived in the following databases (as of May 2022):



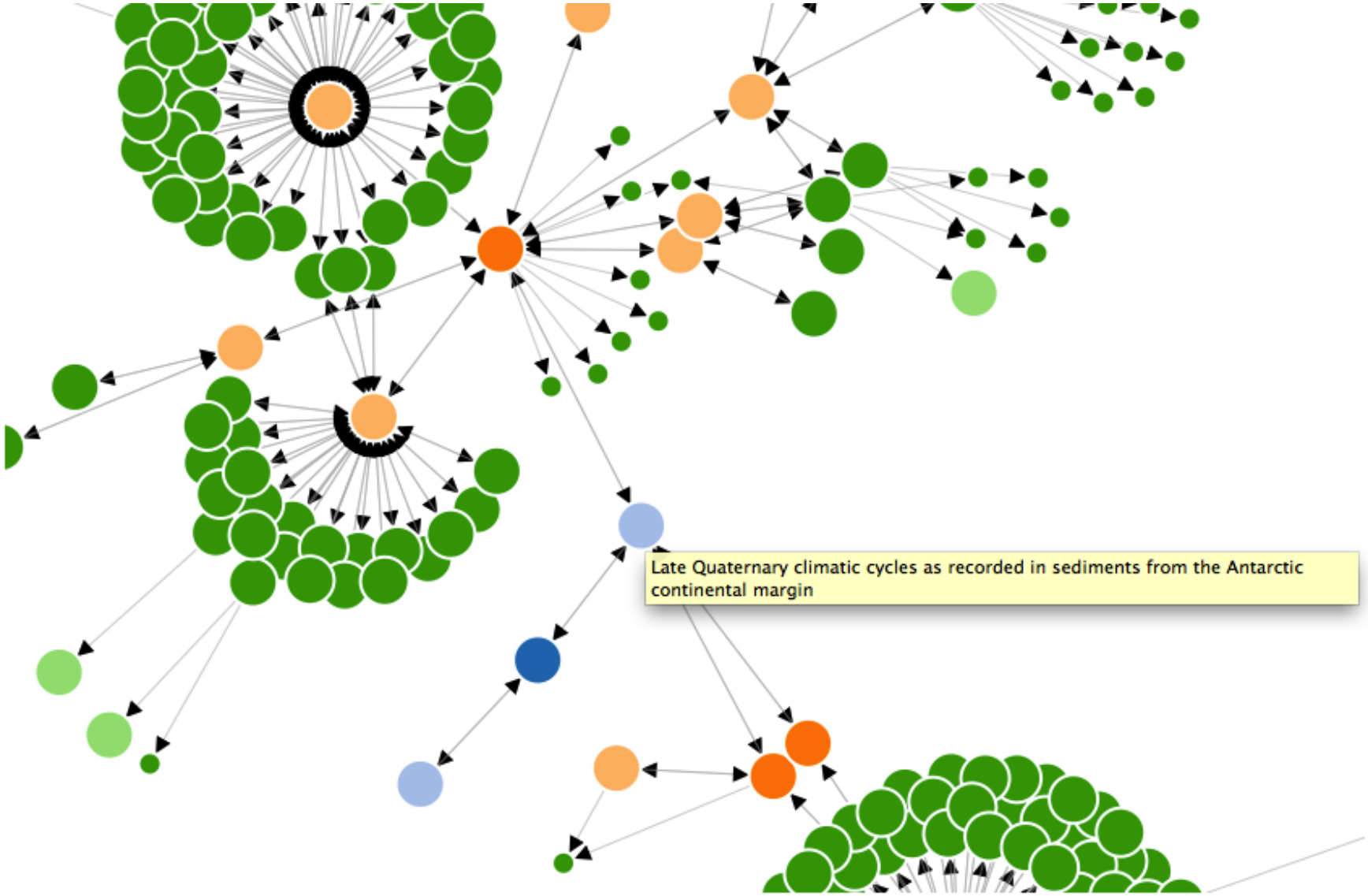
All articles (from all Copernicus journals) are preserved and indexed invoking a ton of services

3. Linking environment by 2014: identifiers abound

Pfeiffenberger, Carlson, TR32 DM WS Cologne 2014

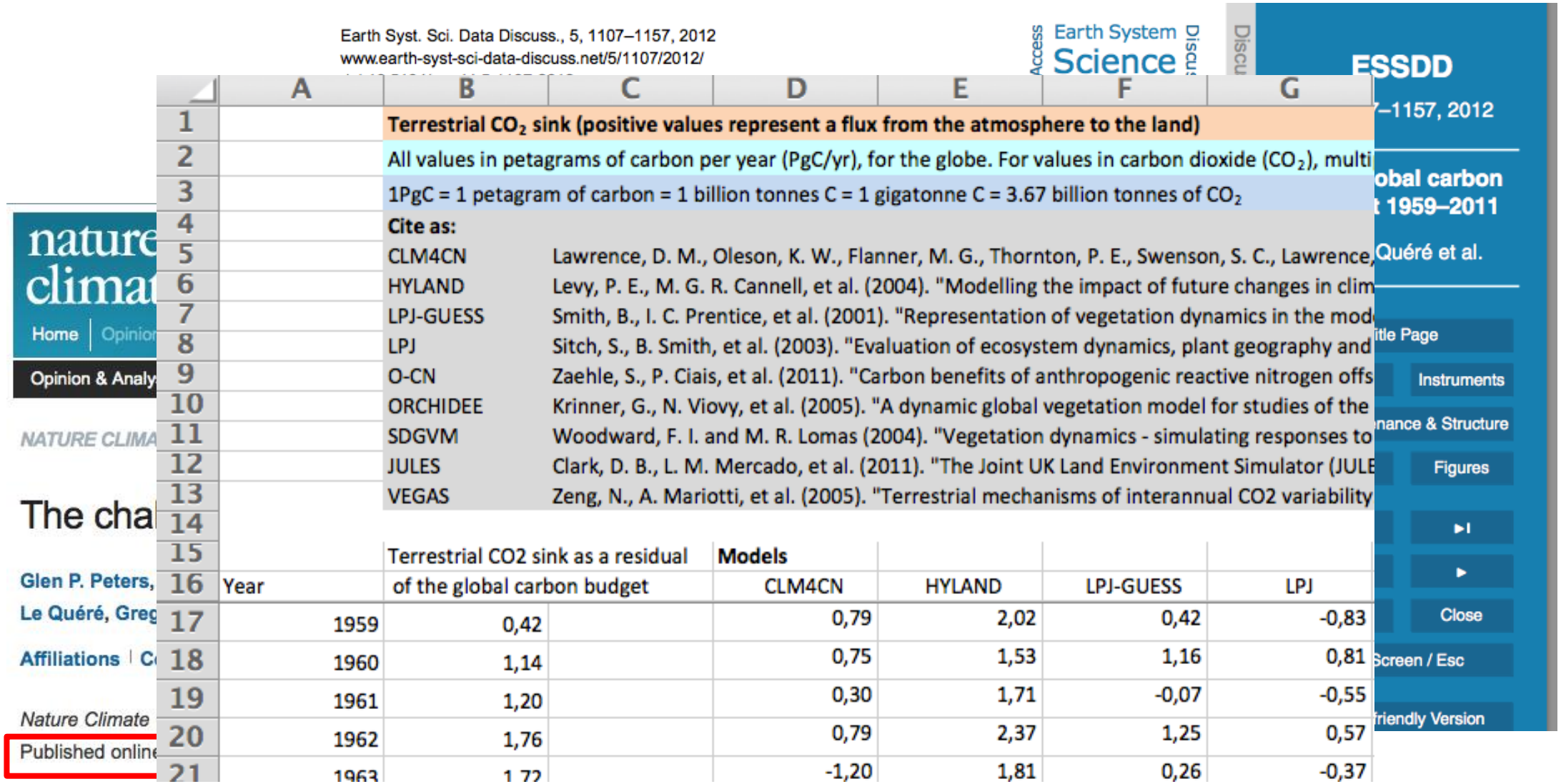


eXpedition – Publications and Data network



4. Linking example Global Carbon Budget

2012: Nature Climate Change, ESSD and CDIAC



Earth Syst. Sci. Data Discuss., 5, 1107–1157, 2012
www.earth-syst-sci-data-discuss.net/5/1107/2012/

Earth System Science

ESSDD

–1157, 2012

obal carbon
t 1959–2011

Quéré et al.

Title Page
Instruments
nance & Structure
Figures
Close
Screen / Esc
friendly Version

nature climat
Home | Opinion
Opinion & Analy

NATURE CLIMA

The cha

Glen P. Peters,
Le Quéré, Greg

Affiliations | C

Nature Climate
Published online

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 mod
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

Year	Terrestrial CO2 sink as a residual of the global carbon budget	Models 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

Global Carbon Budgets 2012-2022, DOIs

- In 2011, CDIAC had to be convinced about DOIs
- 2017: CDIAC shut down!
  - Good news: all DOIs still resolve! (ORNL => LBL)
  - Not so good news: Degraded metadata, most DOIs resolve to one page
    no authorship data at rescue-site

- From 2017 on: Data at GCOS, an ESFRI ERIC
  - (DOIs resolve)
  - Good referencing – machine readable?
  - Unfortunately, (still) no links to more/most recent versions

Global Carbon Budgets 2012-2022, Timing

- 2012: Nature Climate Change demanded an embargo of ESSD until they had published
  - Copernicus publishing engine needed to be halted, temporarily
  - Personal communication with Nature editor and Copernicus necessary to patch mutual citation

- Typical ESSD submission, today:
  - 20% without final data DOI (“review token”)
  - Many (50%?) on ZENODO (ease of use, immediate minting of DOI)
  - A major disciplinary data repository: One year backlog

5. The future: executable article/ “research compendia”

Assets defined in articles published in 2021:

100%
Any asset

100%
Data

1%
Video abstracts

17%
Code

1%
Video supplements

Crossref reports for all articles published in the last 2 years:

100%
Open abstracts

61%
Funding statements

100%
Open references

100%
Similarity reports

57%
Grant numbers

100%
License information

93%
Orcid IDs

What is demanded of e-infrastructures?

- Must read:
  - Lasser, J. Creating an executable paper is a journey through Open Science. Commun Phys 3, 143 (2020). https://doi.org/10.1038/s42005-020-00403-4

- The benchmark: from a research article, enable one-click execution of a Jupyter notebook, with all necessary data available
- "to reproduce each and every step taken to arrive at a publication's conclusions, from the raw data to the polished plot."
- "the possibility to extend the analysis beyond what is typically reported in a traditional publication"
  - E.g., the author "incorporated interactive elements as a sensitivity analysis"

6. Conclusions

- An easy-to-use socio-technical framework to publish interlinked research objects (on federated infrastructures) is still just a vision, years (or decades?) from becoming reality
- It faces competition from easy-to-use commercial platforms (Google Earth Engine, Microsoft Planetary Computer, Amazon)
- Sufficient, correct (not to speak of complete, rich) metadata is lacking on most platforms (or arrives late)
- Will current “manual” creation, chiseling of metadata be outrun by something like ChatGPT?
- Even to maintain the resolving of data DOIs is still fragile

THX! Q?

Bonus Material

References

- Le Quéré, C., Andres, R. J., Boden, T., Conway, T., Houghton, R. A., House, J. I., Marland, G., Peters, G. P., van der Werf, G. R., Ahlström, A., Andrew, R. M., Bopp, L., Canadell, J. G., Ciais, P., Doney, S. C., Enright, C., Friedlingstein, P., Huntingford, C., Jain, A. K., Jourdain, C., Kato, E., Keeling, R. F., Klein Goldewijk, K., Levis, S., Levy, P., Lomas, M., Poulter, B., Raupach, M. R., Schwinger, J., Sitch, S., Stocker, B. D., Viovy, N., Zaehle, S., and Zeng, N.: The global carbon budget 1959–2011, Earth Syst. Sci. Data, 5, 165–185 (2013) https://doi.org/10.5194/essd-5-165-2013
  - (446 citations as of 2022-12-11)
- Hans Pfeiffenberger, David Carlson, “Earth System Science Data" (ESSD) — A Peer Reviewed Journal for Publication of Data, D-Lib Magazine, Volume 17, Number 1/2 (2011) , https://doi.org/10.1045/january2011-pfeiffenberger
- Lasser, J. Creating an executable paper is a journey through Open Science. Commun Phys 3, 143 (2020). https://doi.org/10.1038/s42005-020-00403-4

ODE „Tales“, 2012

“[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”

Karin Lochte, 2012
(Alfred Wegener Institute for Polar and Marine Research)

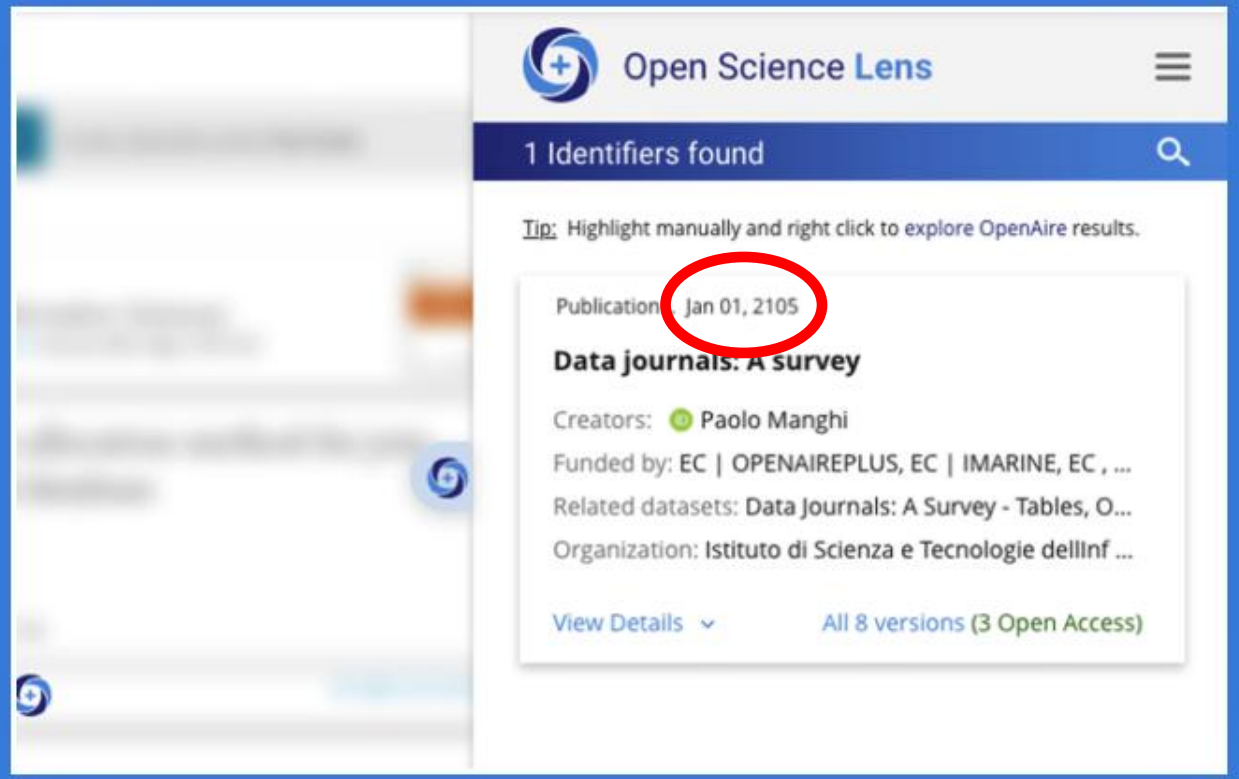
SEVENTH FRAMEWORK PROGRAMME

APA ALLIANCE FOR PERMANENT ACCESS

HELMHOLTZ | ASSOCIATION

Opportunities for Data Exchange

Fun



ABOUT

Open Science Lens delivers an innovative technological approach to **empower** Open Science and bring the information residing on OpenAIRE data infrastructure at the reach of scientists and citizens.