Anforderungen an Persistenz und Identifier für Daten – aus der Sicht von Datenjournalalen

Hans Pfeiffenberger

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

Datenpublikationen mit DOIs - persistent und zitierbar, UFZ Leipzig, 2015-10-28
Late Quaternary climatic cycles as recorded in sediments from the Antarctic continental margin
Publishing - with a capital P

- A Marketing Meme – to emphasize a distinction
  - to publish: make publicly available, e.g. upload to website
  - to Publish - as in Scientific Publishing:
    - is a quite formal, “ritualised” process
    - requires systematic 3rd party scrutiny
    - QA supposedly yields higher quality
  - establishes priority

- quality + priority => reputation => willingness to share
- published items become part of “The Records of Science”
(Data) Publishing - with a capital P

- Standardized, well known process
  => TRUST => we can build on the work of others!!

- Apply these mechanisms to data
  - get all the benefits!! ??
    - thus: “Data Publishing with a capital P”

- In a broader sense, this is about:
  To make data a legitimate part of research culture
JOINT DECLARATION OF DATA CITATION PRINCIPLES - FINAL

When citing please use: Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 [https://www.force11.org/datacitation].

ENDORSEMENT LIST

PREAMBLE

Sound, reproducible scholarship rests upon a foundation of robust, accessible data. For this to be so in practice as well as theory, data must be accorded due importance in the practice of scholarship and in the enduring scholarly record. In other words, data should be considered legitimate, citable products of research.

Data citation, like the citation of other evidence and sources, is good research practice and
Journal Science Releases Guidelines for Publishing Scientific Studies

By BENEDICT CAREY  JUNE 25, 2015

In the midst of a debate over scientific misconduct, one of the world’s leading scientific journals on Thursday posted the most comprehensive guidelines for the publication of studies in basic science to date, calling for the adoption of clearly defined rules on the sharing of data and methods.

The guidelines, published in Science, come weeks after the

Dr. Marcia McNutt, the editor-in-chief of Science, co-authored new guidelines to improve transparency of data, methods and materials in published studies. Drew Angerer for The New York Times
Journals’ Transparency Criteria

- TOP: modular, agnostic to disciplines, low barrier to entry; categories are:
  - a) **data citation**
  - b) design transparency (**standard operating procedures**, protocols)
  - c) **materials** transparency
  - d) **data transparency**
  - e) **analytical methods (code)** transparency
  - f) preregistration of studies
  - g) **preregistration of analysis plans**
  - h) replication

- 3 levels
  - level 1: article states whether data are available and where to access
  - level 2: **data has to be in trusted repository**
  - level 3: data in TR, but reported analyses will be reproduced independently
Pfeifferberger, Macario, Text, Data and People, OAI4, CERN 2005

Diagram showing relationships between entities:
- Person
- Publication
- Group
- Project
- Dataset
- Expedition / Experiment / Campaign

Relationships:
- IsMemberOf
- IsPartOf
- IsAuthorOf
- IsBasedOn
- IsDescribedBy
- IsResultOf
Identifiers

- For data: persistent identifiers for data, ⇒ preferably DOIs, implying fixity, integrity!
- For authors, contributors: ORCID ⇒ all publishers to use ORCID, even the ORCID AuthN service
- For samples: IGSN

- For grants: FundRef (http://www.crossref.org/fundref/) ⇒ http://dx.doi.org/10.13039/501100000780, "European Commission",
  - "narrower": [{"resource": "http://dx.doi.org/10.13039/501100000889"},

ESSD – Data Publishing in practise
ESSD Principles and Criteria

- ESSD expects data to be at a repository and be
  - Open Access, static, with a DOI

- ESSD expects authors to describe in the article
  - provenance, methods, limitations, estimates of error

- ESSD expects reviewers to
  - actually look at the data
  - assess consistency of article and dataset
2013: CO above Troll Station, Original Data

---

**British Antarctic Survey**

**NATURAL ENVIRONMENT RESEARCH COUNCIL**

---

**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

---

<table>
<thead>
<tr>
<th>date [UT]:</th>
<th>2009-10-19 10:44:06</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
<td></td>
</tr>
</tbody>
</table>

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:** used line intensity 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 $\sqrt{(\text{error1}^2 + \text{error2}^2)}$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.749894</td>
<td>50.679</td>
<td>0.060</td>
<td>-5.939</td>
<td>0.003</td>
<td>0.048</td>
<td>0.010</td>
</tr>
<tr>
<td>0.562341</td>
<td>53.021</td>
<td>0.065</td>
<td>-20.151</td>
<td>0.002</td>
<td>0.056</td>
<td>0.011</td>
</tr>
<tr>
<td>0.421697</td>
<td>55.337</td>
<td>0.072</td>
<td>-27.600</td>
<td>0.002</td>
<td>0.061</td>
<td>0.012</td>
</tr>
<tr>
<td>0.316228</td>
<td>57.609</td>
<td>0.080</td>
<td>-29.442</td>
<td>0.004</td>
<td>0.067</td>
<td>0.013</td>
</tr>
</tbody>
</table>
Fluxes of sedimenting material from sediment traps in the Atlantic Ocean

S. Torres-Valdés¹, S. C. Painter¹, A. P. Martin¹, R. Sanders¹, and J. Felden²
¹Ocean Biogeochemistry and Ecosystems Research Group, Southampton, SO14 3ZH, UK
²Center for Marine Environmental Sciences, Universität Bremen, Germany

Abstract. We provide a data set assemblage of directly observed and derived fluxes of sedimenting material (total mass, POC, PON, BSiO₂, CaCO₃, 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, CARIBACO), 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...
2012: Nature Climate Change, ESSD and CDIAC

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Terrestrial CO₂ sink (positive values represent a flux from the atmosphere to the land)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>Terrestrial CO₂ sink as a residual of the global carbon budget</td>
<td>Models</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Year</td>
<td></td>
<td>CLM4CN</td>
<td>HYLAND</td>
<td>LPJ-GUESS</td>
</tr>
<tr>
<td>17</td>
<td>1959</td>
<td>0,42</td>
<td></td>
<td>0,79</td>
<td>2,02</td>
<td>0,42</td>
</tr>
<tr>
<td>18</td>
<td>1960</td>
<td>1,14</td>
<td></td>
<td>0,75</td>
<td>1,53</td>
<td>1,16</td>
</tr>
<tr>
<td>19</td>
<td>1961</td>
<td>1,20</td>
<td></td>
<td>0,30</td>
<td>1,71</td>
<td>-0,07</td>
</tr>
<tr>
<td>20</td>
<td>1962</td>
<td>1,76</td>
<td></td>
<td>0,79</td>
<td>2,37</td>
<td>1,25</td>
</tr>
<tr>
<td>21</td>
<td>1963</td>
<td>1,72</td>
<td></td>
<td>-1,20</td>
<td>1,81</td>
<td>0,26</td>
</tr>
</tbody>
</table>
Global Carbon Atlas - Visualization

- Note the two panels in www.globalcarbonatlas.org
  - data journalism ("emissions")
  - scientific visualization ("research")

- At the project website www.globalcarbonproject.org
  “real life” and linkage between publications and data
Thank you!

earth-syst-sci-data.net

expedition.awi.de

oa.helmholtz.de

copdess.org