Data, Big Data and Publications

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

STM Innovations Seminar, 2012-12-07, London

Data have “always” been the basis of science

- **2000 BC.**, Ur, Mesopotamia:
  - First known *record* about eclipse of moon

- **700 BC**: Babylonians
  - *predict* eclipse of moon

- **585 BC**: Thales
  - *predicts* eclipse of sun

- **1300 years to find the pattern**
- **BIG DATA??**
1606 - 1618: Kepler's Laws (using Tycho Brahe's data!)

- Describe motion of planets
  - 12 years from second to third law

1684 – 1687 Newton De Motu – Principia
- Explained (!) Kepler's Laws (not the primary data!)
1938: Meitner-Hahn-Strassmann Uran-Experiment, Berlin

The last big breakthrough to be done with a lab-notebook?

The biggest experiment, worldwide (not CERN!)
An important, typical Experiment

- **EISENEX / EIFEX**: Two expeditions of “Polarstern”:
  With a few tons of iron fertilizer, south of Capetown …
- **EIFEX (2004)**:
  - 54 scientists and students from
  - 14 institutes and 3 companies from
  - 7 EU countries and South Africa
  - Oceanographers
  - Biologists
  - Chemists…..
- “Biogeochemistry”
- + Satellite observations!
**Global Change**
- Assessing, understanding, and predicting environmental changes
- Human environmental impact

**Hazards**
- Risk analysis and support for disaster management
- Understanding environmental factors affecting human health

**Resources**
- Sustainable ecosystem management
- Energy from the ocean

---

**The Big Challenge(s)**

- **Global Change**, **Ageing Society** ... „Theory Of Everything“
- All are **Big Data** problems (by at least one definition)
- All are **multi-disciplinary** (except TOE)
- Most need **aggregation of globally distributed** data

- **Most are Heterogeneous and Complex**
Down to Earth!

- What does an individual scientist want / need?
- What is she prepared to do?
- And where are publications, after all??

2011: BGI („Beijing“ Genomics Institute)

Spiegel Online, 03.06.2011 (after EHEC identification)
http://www.spiegel.de/wissenschaft/medizin/0,1518,766481,00.html

- Das Großunternehmen beschäftigt rund 4000 Menschen.
- "300 Forscher sind nur für die Gen-Decodierung zuständig", sagt Yang Bicheng, Marketingleiterin des BGI.

What „Spiegel“ did not mention:
- BGI has a private „Cloud“ and (half) a journal: „Gigascience“
“[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 (Alfred Wegener Institute for Polar and Marine Research)
One of ODE’s HYPOTHESES

“Without the infrastructure that helps scientists manage their data in a convenient and efficient way, no culture of data sharing will evolve.”

Stefan Winkler-Nees
Deutsche Forschungs-Gemeinschaft (DFG)

How do we manage data - so that

- Recognition / Reward become possible
- It can be found and aggregated
  - through complex questions
- Level of quality becomes apparent
  - provenance
  - review / endorsement
- => By linking data to people and publications!
Organic matter rain rates, oxygen availability, and vital effects from benthic foraminiferal $\delta^{13}$C in the historic Skagerrak, North Sea

Sylviu Bröcker and Andreas Mackensen

Abstract

The sediment cores 225514 and 225610 were recovered from 420 and 265 m water depth, respectively. They were investigated for their benthic foraminiferal $\delta^{13}$C during the last 500 years.

H. Pfeifferberger, STM Innovations 2012-12-07, London

2012: Nature Climate Change & ESSD

The global carbon budget 1959–2011


This discussion paper has been under review for the journal Earth System Science Data (ESSDD). Please refer to the corresponding final paper in ESSDD if available.

H. Pfeifferberger, STM Innovations 2012-12-07, London
eXpedition (in production since 2005)

Related Information: "Reports on Polar and Marine Research" (1982 to date)
Primary data (all polarstern datasets in PANGAEA)
Handbook and scientific device documentation (in deutsch)
DShipt (Polarstern Data Acquisition System)
VirtualPSP: Virtual Polarstern Tour

<table>
<thead>
<tr>
<th>Expedition</th>
<th>Date</th>
<th>Port</th>
<th>Region</th>
<th>Research</th>
<th>Publications &amp; Primary Data</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANT-XXI/3</td>
<td>21.01.2004 - 25.03.2004</td>
<td>Capetown - Capetown</td>
<td>Atlantic/Indian Ocean, Polar frontal zone</td>
<td>Biology, EiPEx</td>
<td>ePIC: Publications</td>
<td><img src="image" alt="Map(png)" /></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ePIC: Reports on Polar and Marine Research</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ePIC: Weekly reports</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PANGAEA: Stations</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PANGAEA: Datasets</td>
<td></td>
</tr>
<tr>
<td>ANT-XXI/4</td>
<td>27.03.2004 - 06.05.2004</td>
<td>Lazard Sea</td>
<td>Atlantic/Indian Ocean</td>
<td>Biology, Krill, GLOBEC</td>
<td>ePIC: Publications</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ePIC: Reports on Polar and Marine Research</td>
<td></td>
</tr>
</tbody>
</table>
Manida – Publications and Data network – A Big Data problem?

Late Quaternary climatic cycles as recorded in sediments from the Antarctic continental margin.
Conclusions

- There are **Huge Data** problems (such as genetics)
  - (relatively) homogeneous and not too complex
  - though costly and technologically challenging
- There are „**Big Data““ problems (such as „Earth Science“)
  - involve **finding and exploiting patterns** in metadata and data
  - but heterogeneous and distributed (unlike Amazon,…)
- **Both need publications** linked to them
  - **Quality** assurance
  - The best „**metadata““ one can have
  - Provide the **linking hubs** in the **digital assets ecosystem**

Thank you!

oa.helmholtz.de