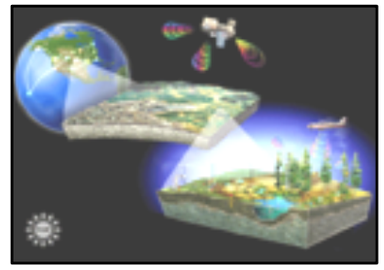
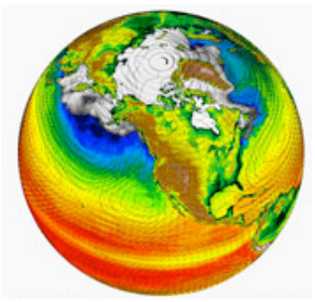
AT4: Capacity Building
Some Elements for Discussion

Jean-Pierre Vilotte, CNRS-INSU (France) and AT3 Champion EI&DM Belmont CRA





The Open Data Iceberg

partly FAIR, partly Cloudy

Technology ← The Technical Challenge

Processes & Organisation
- ← The Ecosystem Challenge
- ← The Funding Challenge
- ← The Support Challenge

People
- ← The Skills Challenge
- ← The Incentives Challenge
- ← The Mindset Challenge

motivation and ethos.

Developed from: Deetjen, U., E. T. Meyer and R. Schroeder (2015).

Some principles

Attention to data use must be increased to make the best use of data.

The positive-sum exchange of data must remain in the heart of scientific peer-review practices to verify and improve the quality of the data, the veracity of the information extracted from it. It would indeed be damaging for research trust if data are misused.

Education and training in improving data use must be increased to ***meet research needs, to maximise the overall value of generating, collecting, preserving, curating and archiving data and to deliver decision support.***

Democratise research by improving education and access.

Young researchers will gain from increased knowledge, skills and access to data and tools.

Education and training must seek to ***develop computational thinking and introduce data management and data analysis methods.*** Such changes in research practices require ***revision and knowledge of the legal and ethical frameworks within which researchers work***.

The costs of data stewardship and use should be visible to researchers.

Public funded research data is financed by taxes that everybody pays => ***data must be protected as open common knowledge, honouring intellectual property rights, that every one, every where can share and which is preserved.***

Research communities must organise to remain in control of global research data stewardship and governance, which must be ***agile and flexible enough to align with disruptive research practices*** and multidisciplinary data-use.

Researchers must be well-informed about the real costs of support for research data to get research done with less demand on resources through careful design and operation of federated transnational research facilities and ingenious strategies for finding the shortest path to the sought-for knowledge.

Some Elements

Stimulate new thinking and curriculum change in the next generation of researchers.

Incentives, programmes and events should be used to attract and prepare young scientists to ***data stewardship and data-use challenges***, and to facilitate their ability to harvest the new opportunities :

- establish a ***coordinated international framework*** for ***multidisciplinary immersive gatherings*** (summer/ winter schools, 'boot camps') where early career scientists and experts combine their insights with hands-on practicals;
- develop ***Belmont 'certification'*** of existing education programmes that stimulate ***productive combinations of communities, approaches and skills*** around Belmont research challenges;
- ***establish a knowledge-hub that collect and point to all kind of digital learning ressources*** for transnational and interdisciplinary data use; and offer them via user-oriented portal focussing expertise and ***promoting data stewardship, open-data access and data interoperability and reuse;***
- ***engage young scientists with current best research practices and organisations in data-driven global change research***, including ***data management and stewardship, data analytics methods, high-level languages, cost-effective federated infrastructures operations***; and
- engage with ***global research communities committed to pooling their efforts*** to accelerate global data stewardship and data-intensive research.

Some Elements

Increase data-use by building 'intellectual ramps' and providing education.

There are many kinds of '***intellectual ramp***', examples include:

- online best-practice guides, MOOCs, Jupyter-like notebooks for those planning data-intensive research;
- automated metadata and provenance generation tools (linked to standards and standards organisations);
- visual interfaces to pre-integrated bodies of data;
- aids for extracting selected data and delivering them in standard forms;
- augmented familiar research environments with data analytics methods and libraries (e.g. Python environments), together with streaming-data workflow tools, environment and repositories;
- reference data for an application domain and facilities for data archiving and publishing.
- specialised environments for safe play and incremental engagement.

Education and ramps must develop hand-in-hand.

- Once a ramp has successfully engaged a ***critical mass of communities***, the methods it supports become part of their culture.
- ***Teaching the methodological concepts and the judgement about when and how to employ a family of methods*** will reveal a strategy for building their ramps.
- Successful ramps allow teachers and students to focus on those concepts and judgements.