PEG-BOARD
Bridging the gap between Research Group & Institutional Data Management Needs

Gregory J. L. Tourte

School of Geographical Science
The University of Bristol

g.j.l.tourte@bristol.ac.uk

DCC Roadshow — 2010–11–02

Outline

- Introduction
- Position within the wider debate
- Institutional and Research group solutions
- Conclusion

Who we are

- BRIDGE (Bristol Research Initiative for the Dynamic Global Environment) Research Group
- Based in the School of Geographical Sciences at the University of Bristol
- Very cross-disciplinary research with international collaboration
- Currently 8 academics, about 15 postdocs and 15 PhD students

What We Do

Our aim is to improve our understanding of natural climate/environmental variability and to use this knowledge to better predict future changes.

Major themes of the research include :

- Rigorous evaluation of climate models with accurate proxy palaeoclimate records
- Innovative Earth System modelling, and process studies for past, present and future change
- Impact of future climate change on spatial and temporal scales relevant to society, and including timescales from decadal to millennial.

The PEG-BOARD Project

- Palæoclimate & Environment data Generation — Building Open Access to Research Data
- JISC funded project within the Managing Research Data Programme
- In collaboration with UKOLN, The University of Leeds, The University of Southampton.

Where we come from. . .

- Ad-hoc development of technical solution
- Need to share data with research collaborators
- Need to share data with a variety of other disciplines
- Research grants usually cover data generation but not data maintenance
- Storage and development agility requirements were beyond capacity of the institutional IT team.

Data deluge

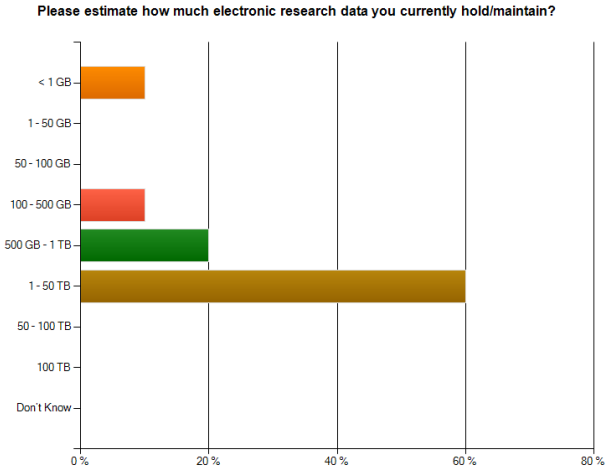
data sets that start at terabytes (thousands of gigabytes) in size [...] are now commonplace.

– Dr John A Taylor,
Finding new ways to deal with large datasets (CSIRO, 2007)

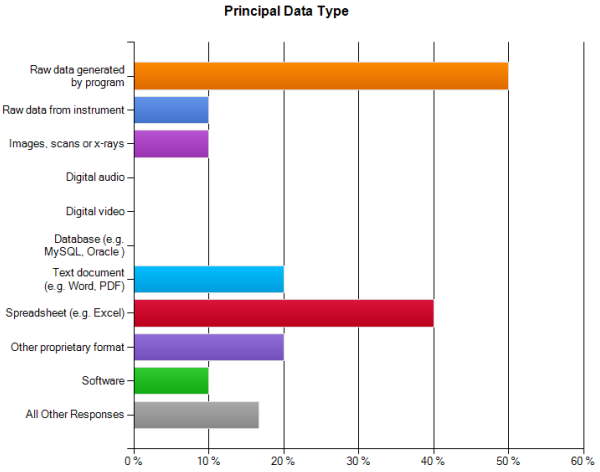
World's digital content equivalent to stack of books stretching from Earth to Pluto 10 times

– Richard Wray,
Internet data heads for 500bn gigabytes (The Guardian, 2009)

How much data do we work with?



What sorts of data do we work with?



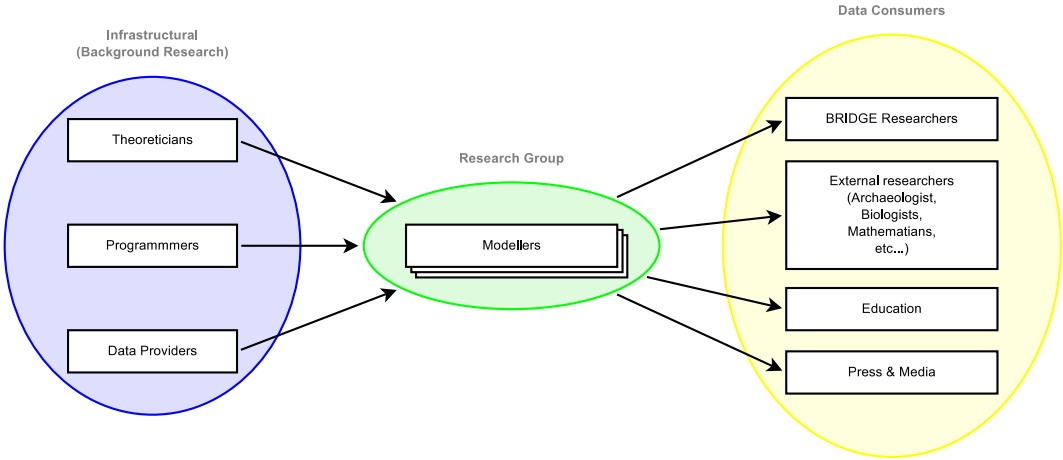
Science as a social activity

The understanding of science as a social activity has changed quite radically

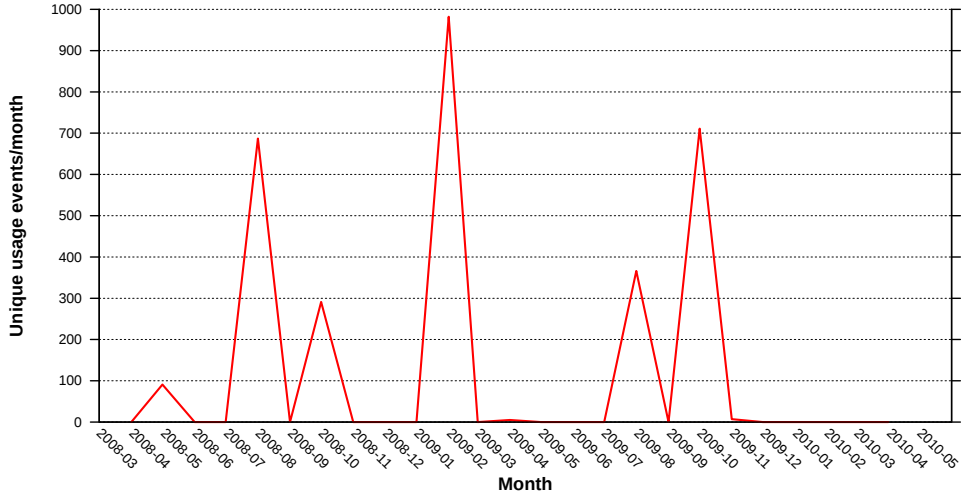
Mike Hulme and Jerome Ravetz
'Show Your Working': What 'ClimateGate' means (BBC, 2009)

- Always be a unique function for expert scientific review
- Should not exclude other interested and motivated parties
- Demands for openness in Science intensified by the Internet as a central element of social life
- Knowledge validation must also be scrutinised by extented community of citizens with legitimate stakes
- To be empowered for use in plublic debate and policy making, knowledge must be fully exposed to the proliferating new communication media used by community

Who works with our data?



The pulse of data in educational reuse



Political landscape

> *raw data for the climate models was not made available to the public. To try to restore public confidence [future development will be] more open to the public.*
>
> – Jonathan Gray, Climate Change, Climate Sceptics and Open Data (Open Knowledge Foundation Blog, 2009)

Where we are now. . .

BRIDGE :

- HPC infrastructure growing rapidly as well as performance
- We generate more and more data as we can do more complicated science with available technology
- Data production is outgrowing our capacity to afford an ad-hoc storage solution

Institution :

- Storage requirements for other departments is becoming critical, especially in less technical disciplines (arts and humanities)
- Need to make data available to wider community (other institutions and research groups, general public, press. . .)
- Need for cater for security and privacy of some potentially sensitive data

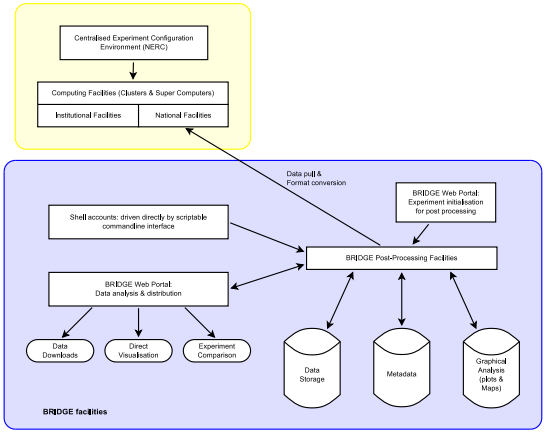
BRIDGE :

- HPC infrastructure growing rapidly as well as performance
- We generate more and more data as we can do more complicated science with available technology
- Data production is outgrowing our capacity to afford an ad-hoc storage solution

Institution :

- Storage requirements for other departments is becoming critical, especially in less technical disciplines (arts and humanities)
- Need to make data available to wider community (other institutions and research groups, general public, press. . .)
- Need for cater for security and privacy of some potentially sensitive data

Using PEG-BOARD as a case study for the Institutional Data Curation Solution



Any Questions?