Research Data Sharing
without barriers

RESEARCH DATA ALLIANCE

RDA-Europe Science Workshop

10-11 February 2014, Munich

Bernard Schutz
Data Innovation Institute, Cardiff University
& Max Planck Institute for Gravitational Physics
(Albert Einstein Institute)
and

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

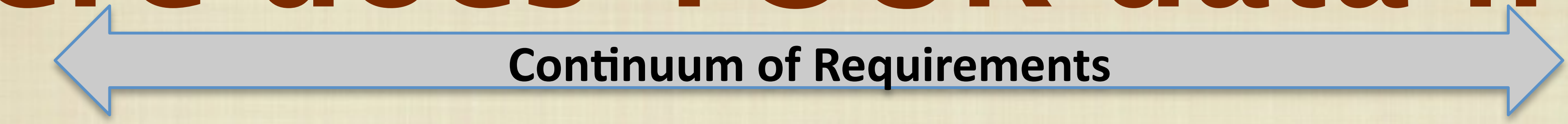
Science Data Workshop

- Hosted by Max Planck Society at their Munich Headquarters.
- Goal: Discover what contribution RDA can make to solving problems that working scientists have with their data?
- 12 prominent scientists from outside RDA spoke about their disciplines and their data challenges. Another 14 participants from the RDA TAB and other areas of data management.
- Disciplines ranged from astronomy to history, from meteorology to psychology.
- Data types included observational data, numerical simulations, text, and person-specific information.
- Analysis requirements included visualisation, signal analysis, text mining, pattern finding. Some were volume-intensive, some compute-intensive. Many were very heterogeneous.

Fascinating Research

- Just a few examples:
  - Dik Dee (ECMWF), Weather Forecasting. Data structures formed from heterogeneous unstructured data. Significance not easy to control.
  - Jan Bjaalie (Oslo), Neuroscience. Heterogeneous data from cellular level to MRIs.
  - Mark Hahnel (Cambridge), Stem Cell Biology. Serious social issues about publishing individual genomes. Data publishing and sharing rare. Might change with new publishing venues.
  - Jochem Marotzke (MPI Meteorology), Climate Modelling. Immense data and compute challenges. Heterogeneous data — even windshield wipers! Political dimension, reliability, reproducibility, publishing data for the public to access.

Multidimensional challenge: Where does YOUR data lie?


Continuum of Requirements

Well structured data	Heterogeneous data sets
Automatically generated metadata	Complex metadata issues
Static data	Dynamically changing data
Data acquired under controlled conditions	Crowd-sourced data
Centrally managed databases	Distributed data, no clear curation
Computationally simple	Data needing massive computing
Data that are used “raw”	Data that are understandable only after processing
Numerical data	Text data
Knowledgeable data communities	Communities scared of data
Communities with trust	Communities with no tradition of sharing, or even with distrust
Open data	Proprietary/embargoed data, data with copyright issues
Impersonal data	Data with privacy issues
Privately generated data	Data with publicly funded stakeholders

Issues probed

- Sharing and re-use of data: not common in most fields. How can communities benefit from the extensive experience and stable tools in areas like astronomy and climate research?
- Publishing and citing data: frequently referred to as a big issue. How to scientists get academic “credit” for producing and sharing data? Does it help their careers?
- Infrastructure and repositories: Some fields, like astronomy, are well organised. Other fields have no infrastructure, no experience, and not much motivation. What are the incentives?

RDA and Scientists

- A very frequent question from the scientists: *"What is the RDA trying to accomplish?"*
- The scientists frequently expressed the worry that *RDA is too top-down*, not driven strongly enough by needs of working scientists.
- Scientists also expressed concern whether RDA will make an impact, or will be overtaken and overshadowed by big companies like Google and Microsoft, whose activities create de facto standards.

Recommendations

- RDA certainly should invest in training young generations.
- RDA should push demo projects, act as a clearing house, and should be able to give advice on data management/access/re-use to everyone in research.
- RDA should have experts who could go to institutes and help them to implement solutions.
- The meeting looked ahead to September 2104, when the first RDA results were expected, and hoped for a meaningful quality assessment of the results. It cautioned that RDA should take care to not fall into the trap of overselling!