Services Working Group

15:00   Prepared short interventions and discussions

16:00   Coffee break

16:30   Continued

18:00   Boat trip and dinner (for registered people only)

Rapporteur: Arjen van Rijn

Contributions and interventions

- Views on data services: EUDAT and the e-IRG Blue Paper
  Norbert Meyer (Poznan Supercomputing and Networking Center)

- Providing data services to researchers outside the core community that produced the data
  Andrew Treloar (Australian National Data Service)

- Facilitating Data Upload, Access and Deployment
  Rainer Stotzka, Marcus Hardt (Karlsruhe Institute of Technology – Inst. For data processing & Electronics)

- Data preservation: proposal for an RDA working group
  David Giaretta (Alliance for Permanent Access)

- Multi-dimensional Big data: From Data to Service Stewardship
  Peter Baumann (Jacobs University Bremen & rasdaman GmbH)

Impressions (1)

Disclaimer: these are my impressions, not necessarily shared by all attendants.

Much discussion about 'Persistent Identifiers'.

But what is meant by 'persistent': until tomorrow, next year, 5 years, 50 years?

And if you promise '50 years', will they believe you?

And: to what does a PID refer: the object, the actual dataset, the service?

Impressions (2)

What does the user want? Here's the unsorted list from eIRG's Blue Paper:

- Large datasets
- Restricted access
- Metadata structure
- Federated AAI
- Accounting
- Data provenance
- Integration
- Interoperation
- High trust and security
- Reliability

Access

Advanced search functionality

Data preservation

Interpretation of data

Unified access

High quality

Simplified access

Searching information

Important stakeholders

Impressions (3)

A useful approach! Four key functionalities:

1. Find

    (… and don't try to change a researcher's searching behaviour)

2. Assess

    (… here procedures for trusted repositories such as the DSA might come in handy; but is certification scalable if 'everybody' becomes a data producer?)

3. Access

    (… provide least restrictive access but that cannot not always be open;)

4. Reuse

    (… especially challenging if data are reused in other communities than the original one)

Impressions (4)

'Data producers (researchers) are no good at providing data services. For that you need separate, dedicated organizations. Bring your data directly to the databank.'

'A researcher wants to know everything about the data, so let him query freely. (But then: not all data is always inside the database).'

'There cannot be enough network bandwidth.'

Impressions (5)

Let's avoid a technology dominated perspective!

It's the users we're trying to serve!