Research Infrastructures and Data

Chair: Michelle Barker

Rapporteurs: Barbara Sánchez Solís and Carthage Smith

Ensuring Trust and Quality

- Data, metadata, controlled vocabularies, standards, processes
  - Deep understanding of what a digital object is
- FAIR principles help to think about these issues
- Completeness of metadata
  - Time stamp, geographical scope, keywords ect.
- Open data fundamental part of scientific debate
  - Openness of hardware, algorithms etc.
- Provenance, provenance chains, preservation, privacy
- Sensitive data/anonymization
  - Bias: anonymization key for use of data vs maintenance of full information

Trust in both, container and content

- FAIR relating to broader ecosystem
- Quality of services (functionalities, operations)
- Cybersecurity
- Ability to access data
- Interoperability of data and tools
- Machine-actionable access and rights models
- Researcher-centric data ecosystem

Challenges

- Plethora of competing standards
- Small metadata standard for citizen access
- Gaps between HPC infrastructures and traditional repositories
- Data changes in real time – requires snapshots, versioning
- Ownership of data
- Multilinguality
- Abundance of data, journals, consequences of metrics

Is non-designed data/big data changing the game?

Outlook

- Change in publishing culture
  - Article of the future: deeper access to knowledge, interlinked with code, data, images etc.
- Bring together science and society expectations
- Trend of clustering data (linked data)
- Models on domain-agnostic data
- Share user experience
  - recommender systems etc.
- Adopt, adapt and develop
  - Tools, support, automated processes

RI ecosystem:cyber-infrastructure and data

- Trusted data and trusted cyber-infrastructure and trusted connections – need open PIDs
- Think in terms of the research data life-cycle
- Support development of open tools and software for connecting
- Must design data services/cyber-infrastructure around specific user needs
- Build on best practices: different domains at different levels of preparedness

Human infrastructure

- Need to build a workforce of data scientists and stewards
- Need new kinds of data specialists and stewards
- Need critical skills and education to deal with the data rich World of the future
- Recognition is a key issue that is not yet addressed but technology (AI) can help.
- Generational latency because of mentorship. Need cultural change

Google & Co.

- Google solns for storage storage and access to computing meet a need
- Google are solving discovery challenge and drving semantic web, AI etc.

But

- Google pays more and is more attractive vs academia
- Google de facto sets the standards
- Google threaten to drive out (public) competitors
- ?long-term sustainability and customer-dependency

Funders perspectives and competition

- On-line Poll: - ~70% of audience see incentives and mandates as a critical function
- Data sharing policies are not the same as ‘open data’.
- Balance national benefits and internat. cooperation?
- Communities need to triage and decide what must be kept and what is disposable
- Data life-cycle approach and importance of data management plans

So, it (science)is all about data

but

it is not just about data.