RDA US Science workshop
Arlington VA, Aug 2014

Cees de Laat
with many slides from
Ed Seidel/Rob Pennington

Arlington, VA, USA Meeting

- Aug 12-13, 2013, next to NSF
  - Specifically to attract science groups!
  - Not easy to get them to attend RDA meetings!
- Organized by NCSA
- Built on RDA-Europe meeting sponsored by MPG
  - Similar topics discussed
  - RDA US and EU participants present
- Numerous science communities represented
  - Astrophysics, physics, chemistry, materials, scientific computing, earth sciences, civil engineering, National Labs, HPC centers, etc

Selected Recommendations

Persistence

- Explore options for innovative and cost effective solutions for deploying a network of persistent data archives that would serve the needs for a range of broad communities.

Persistent archives include *storage, data collections*, and *domain-specific curation* and *data services*

- Investigate the time value of data for different communities of use, and use that to inform potential solutions/offerings.

Sustainability

- Encourage funding for realistic data sustainability in all domains of scientific research, both public and private.
- Develop value propositions for all levels of funding organizations and users.
- Involve domain experts in sustainability discussions and decisions.

Tools

- Define a basic tool kit for data analysis and services, leveraging web tools where available.
- Connect science communities and informatics communities to develop user-requirement-driven tools.
- Develop or enhance linking service tools to connect to publishers.
- Promote the idea that software is data, and therefore discoverability, repurposing, and so on, apply to software as well as to data.

Selected Recommendations

Discovery

- Develop curation requirements (including metadata) that enable short and long-term discovery across publication, domain repository, institutional repository.
- Investigate metadata-free search and discovery mechanisms.
- Investigate workflows that connect discovery to use.

Selected Recommendations

Metadata & WorkFlow

- Metadata
  - Develop interactive and automated metadata validation and extraction tools.
  - Develop protocols to support automated and interactive metadata validation and extraction tools.
- Workflow
  - Gather, understand, and disseminate best practices in workflow packages, and in how various domains use workflows, and treat workflows as data.
  - Investigate workflows that connect discovery to use.

Selected Recommendations

Provenance & Education

- Provenance
  - Articulate the connection between provenance tracking and reproducible science.
  - Encourage the implementation of automated provenance collection
- Education
  - Work with educational institutions and science organizations to develop curricula and educational resources for data-intensive science
  - Develop Data Scientist and Data Manager as disciplines.

Selected Recommendations

Technology Trends & Data Infrastructure

- Technology Trends
  - Develop technology roadmaps that inform needed resources for code and infrastructure revision, including new or disruptive technologies like cloud.
- Data Infrastructure
  - Develop a new reference model to account for federated environments and their implications in all areas, and taking existing models into account.
  - Facilitate convergence between compute (HPC, cloud), data (Big Data, long tail), and networking communities.

Summary

- Critical---but difficult---to engage science communities in activities such as RDA
  - Meetings such as Munich and Arlington help; need more
- Draft report developed; posted for further comment
- Participants agreed this was useful; just a first step
- Anticipate a series of such workshops in coming year with multiple venues
  - Suggest coordination with EU, US and others
  - Recommendations to build and become more detailed
  - Useful not only for RDA but others
    - NDS, EUDAT, other infrastructure organizations
    - MGI, NEON, other science domain activities

Questions?

Supported by: