Report from governance track

Chaired by Leif Laaksonen

e-IRG workshop on data issues

Amsterdam, December 3-4, 2012

Rossend Llurba

e-IRG
e-Infrastructure Reflection Group

Outline

- 4 presentations
- Discussion

Fulvio Marelli (ESA) "Harmonizing digital preservation policies for Earth Science data"

- Data is more valuable when combined together.
- Preservation of data is useless without preservation of the knowledge associated with the data.
- Ensure, enhance and facilitate archived data accessibility (allowing to combine data from different sources and to perform more complex analyses).
- Ensure coherency of approaches among different Earth Science providers.

Fulvio Marelli (ESA) "Harmonizing digital preservation policies for Earth Science data"

- ESA is coordinating the LTDP (Long Term Data Preservation) cooperation activities in the Earth Observation domain with European partners.
- LTDP workshops every two years to disseminate results within the EO/LTDP community.
- Earth Science can count 9 different data categories, each with its own data preservation policies, metadata and data formats, data description and semantics.
- Survey of earth science users to assess level of expertise w.r.t. long-term data preservation.

Kees Neggers (SURF)
“How to govern an ecosystem?

- Cyberinfrastructure ecosystem.
- Internet history: no grand design, no central management, evolutionary model, innovation driven by the advanced requirements of the science community.
- Lessons learnt: shared control plane required, not a centralistic model, create loose cooperation between domains, keep it simple, architecture based on openness and diversity, multi-domain connected via open standards, bottom-up development together with users (with opposition from incumbents), voluntary international cooperation.

Kees Neggers (SURF)
“How to govern an ecosystem?

- e-infrastructure innovation: will be driven through competition, co-operation and flexibility; needs openness, neutrality and diversity as guiding principles, must take account of the global context.
- distinguish three core functions: community building, high-level strategy and coordination; (competitive) service provisioning; innovation.
- Cooperation remains essential for the new internet and e-infrastructures.

Françoise Genova
"Lessons learnt from building the International Virtual Observatory Alliance"

- IVOA: mission, focus on development of standards and encourages their implementation, global endeavour from the beginning, "thin" interoperability layer, continuously adapting its organization and procedures to fulfill its mission at best.
- Goals IVOA similar to RDA for a single discipline; similarities and differences.
- IVOA membership and Executive Committee.
- IVOA structure: A formal procedure for acceptance of Recommendations (adapted from W3C); standards done by Working Groups; each REC has authors and Editors and is under the responsibility of one WG; interest Groups; standing and other Committees.

Françoise Genova
"Lessons learnt from building the International Virtual Observatory Alliance"

- Technical Coordination Group: 'technology aware' committee in support to the 'political' Executive Board; essential role to check the coherence of the global vision, manage interfaces, evaluate the WG proposed recommendations with respect to community comments, etc; TCG composition.
- IVOA stakeholders/participants: Constantly keep in mind several sub-communities: developers of standards and tools; those who implement them in archives and data centres; science users; essential to have both 'technologists' and 'data practitioners' (and scientists from the data centres) on board.
- Organisation of the work

Jamie Shiers (CERN)
"Harmonizing DP Policies for HEP Data"

- (Some) data related Issues – HEP: data preservation, data management, data access, storage management; but also databases, e-Infrastructures, software repositories.
- Desired outcomes: Adopted standards (within and across disciplines), deployed infrastructure, adopted policy, implemented best practices.
- Timeline: some issues need to be addressed rather urgently, others can (must) take longer to be addressed.

Jamie Shiers (CERN)
"Harmonizing DP Policies for HEP Data"

- Can we agree not only on common requirements but also on schedule?
- Summary
  - Data Preservation for long-term re-use is an important Use Case with clear links to other dimensions of the "Data Domain".
  - Strong motivation to address both technical and non-technical issues in an international / multi-disciplinary environment.
  - Let's profit from this motivation plus concrete experience to build something better, together, for the future.

Discussion

On the usability of the ‘Internet’ model for the data community:

- We cannot just copy what was done before because the world today is much more complex than when internet started. It will not be possible in only a few years and the challenge is to make it faster than was done for internet and with more different people.

On lessons learnt from the creation of internet:

- Internet is not invented. People worked in parallel and together based on sound principles. Top-down approach for investments was not always efficient.

Discussion

On governance:

- Don’t look now at the governance in detail but at the different features that have to be in place to create governance.
- Major impact will come from the working group level. This will create links with governance in different places.

On incentives for coordination of all the different data worlds:

- Progress should be made one step at the time
- There are things than one community knows better how to do than others. Discussion is good for acceptance and for building bottom-up.
- At this moment is not fully know what will result from connecting communities horizontally.

Discussion

On international aspects:

- e-IRG wants to be a coordination platform for discussions. Special attention to data area; has invited this semester two data projects as observer. e-IRG's ambition is to be more international.
- The main participants are now in the Northern hemisphere and Australia. It is not only a North-South problem. Also in parts of the South (e.g. Africa) many stakeholders are interested but not (yet) involved. The time is right now for strategic initiatives there (some have started).

Discussion

On impact and how to measure it:

- Small bridges as shown this morning will be essential. If they are in place there will be traffic, this traffic will grow and the bridges will be enlarged.
- The working groups will be enablers.
- Always tension trying to do some small or strive to a great structure.

Discussion

On how to bring the knowledge of other communities into RDA:

- Not useful **now** to have interoperability groups with communities that do this interoperability already.
- Organise liaison with these organised communities. They could be interested in participating in technical groups.
- Partnerships should be based on added value.
- Purpose of RDA is to create the horizontal, the connections between communities that create the added value.
- Involve communities in creation process.
- There should be some audit to see if best practices are used. RDA could help to adopt something that is common to everybody which will be beneficial.

Discussion

On recognition issues and motivation:

- Publications contribute to prestige scientists. A similar mechanism should be in the data area.
- Some collaborations and workshops being prepared to define jobs in data area.
- Metrics are important, evaluation of data, peer reviewing of data. We have to define goals to see where we are going.
- Use credits (like in movies) to recognize data contributions in research. Acknowledge the data provider. Specific data journals.
- It will take a long time before the academic community accepts recognition of data contributions.

Discussion

On role of private research and their meaning for governance structures:

- distinguish three cases: public, co-founded and private research
- RDA will have at some stage value that will interest companies but this has not been discussed in detail yet.

On standards in data community:

- Data community leadership has to accept that they will make mistakes and that there will be a say 5-year period before there are results. Doing things does not guarantee results.

Discussion

On the first steps to be taken and who should be doing what:

- Some communities are very advanced in creating a global network where exchange of data will be easier. They could have impact and it is a small step on the way.
- Work bottom-up except for things like looking at gaps and overlaps.
- At some stage there has to be some kind of planning that the output of the working groups have some common denominators (and are not diverging). Perhaps a task for the advisory committee.
- Brilliant ideas and solutions are generated by individuals. Does the construction that we have in working groups make it possible that ideas are developed?
- write about best practices in data.