e-IRG workshop Dublin 22-23 May 2013

Track 1: Coordination of e-Infrastructures

Rossend Llurba

e-IRGSP3

e-IRG
e-Infrastructure
Reflection Group

Track 1 – 2 sessions

Session 1 (Chair: Lajos Balint)

4 presentations

- Bob Jones
- Stephen Moffat
- Sandra Collins
- Bob Jones

Session 2 (Chair: Sverker Holmgren)

3 presentations

- Stefano Cozzini
- Jon Ison
- Sverker Holmgren

Track 1 – Session 1

Lajos Balint introduced the track:

- Complexity; we are talking about ERA

- Questioned if coordination is enough

- Is e-IRG well positioned to do such a coordination activity?

Bob Jones (CERN)

Interoperable e-infrastructures as seen by Helix Nebula

- Helix Nebula - European cloud computing partnership: big science teams up with big business

- Explained history, partners and evolution of Helix Nebula

- Long term goal: multi-tenant '**Open Market Place for Science**', where data, scientists, funding bodies, SMEs and downstream industry meet to work towards common interests. Creating an ecosystem to transform data into valuable information

- Explained the timeline (2011 – 2014)

- Initial flagships use cases (from CERN, EMBL and ESA) and their results

Bob Jones (CERN)

- Changes from situation before Helix Nebula to situation in 2012
  - Basically connection Helix Nebula's commercial partners to GEANT
- Identified gaps
  - A standard set of APIs for suppliers and consumers
  - A federated identity management system with single sign-on facility to access cloud services across multiple-suppliers
  - Automation of cloud management processes to provide on-demand services
- Federation via Blue Box
- Helix Nebula Blue Box services currently under test and future tests
  - Interaction between public infrastructure (EGI) and Helix Nebula
- New flagship use cases
- Study how to work with national structures and funding agencies:
  - Identify potential commercial suppliers and research community users
  - Compare business models in Helix Nebula with those used on a national basis
  - Understand how national engagement would impact the Helix Nebula governance model
- Interoperability aspects

Bob Jones (CERN)

In the discussion that followed it was noted that:

- The connection to NRENs is in the proposal phase. It was discussed with DANTE and will be only used for public generated data for science
- The API discussion is a difficult one
- Bob Jones sees potential for an e-Infrastructure commons and would like Helix Nebula to be part of it.

Stephen Moffatt (IBM)

IBM Research….from Innovation to Impact

- Is working with IBM teams to bring technologies and experience to the Irish Marketplace

- Shared a perspective from the IBM Research team on:
  - What kinds of application and services users will value over the coming 5 years.
  - What kinds of framework are useful for assuring that appropriate infrastructure can be put in place
  - What kinds of challenges face IT infrastructure professionals

Stephen Moffatt (IBM)

- Shared some examples of the kinds of applications for which IBM sees demand in public and private enterprise
  - a method of relating lots data to their customers (data obtained e.g. when buying online, using loyalty cards, visiting websites etc.).
  - The unfairness of local property tax in Ireland and a model how to map the costs of local services to property owners
  - Computing electricity grid complex problems.
  - Finding out which bus routes in Galway are performing and which not.

Stephen Moffatt (IBM)

- Some observations from e-Infrastructure professionals
  - Technologies (databases, ETL tools, warehouses) are becoming easier to use in terms of the infrastructure needed to support them.
  - Openness is becoming important and is good because it will keep price pressure on vendors.
  - Standards for interoperability are fast improving which will avoid lock-in.
  - We need to think about the kinds of skills and experience we need to profitably put these analytical technologies and principles to work. We need to invest in mathematics, engineering and research skills.

- Movie: A boy and his atom.

Open Access and sharing: the Irish position in Open Access

- Why open access is important
  - *High-level principled argument*
    - Results of publicly funded research should be publicly available
    - Enables research findings to be shared with the wider public
    - Enhances knowledge transfer
  - *Practice-based pragmatic argument*
    - Improves research efficiency
    - Enables reuse of research outputs
    - Provides the basis for better research monitoring and evaluation
    - Preservation of research outputs ensures our cultural heritage is protected and curated
    - Scientific outputs are kept in formats that ensures they are permanently usable and accessible

Sandra Collins (Digital Repository of Ireland)

- **National steering committee on Open Access Policy**
  - "Coordinate activities and combine expertise at a national level to promote unrestricted, online access to outputs which result from research that is wholly or partially funded by the State"

- **National membership: all national funding agencies, researchers, libraries and repositories**

- **Phased approach**
  - Broaden infrastructure coverage beyond universities
  - Increase number of publicly funded publications made available
  - Use OA to develop a national picture – funding acknowledgements & value-added metrics
  - Promote OA to underlying research data and materials
  - Develop sustainable long-term solutions

Sandra Collins (Digital Repository of Ireland)

- **General principles**
  - All researchers to have deposit rights for in an OA repository
  - Release policies

- **High-level principles**
  - Infrastructure and sustainability
  - Advocacy and coordination
  - Exploiting Open Access

Sandra Collins (Digital Repository of Ireland)

- Evolving agenda
  - Leadership –engagement, raise awareness and compliance
  - Infrastructure – access for all and national picture
  - Clarity - costs of implementation, copyright
  - Use – recognise intrinsic value of research, impacts and metrics
  - Research data
- Reasons for not sharing data
- Change practice
  - Peer reviewed data plans in grants
  - Allowable costs
  - Data citation, PIDs
  - Metrics for career progression, funding
  - Training
  - Sustained e-infrastructure
  - Copyright, IPR, licensing, data protection, embargoes

In the discussion it was noted that:

- Training people in sharing data in general is OK but you will need to have also the specific disciplinary knowledge. Therefore we will have to work with communities in the way they work and use base practices.

Bob Jones (CERN)

Science, Strategy and Sustainable Solutions, a collaboration on the directions of e-infrastructure

- The business of research
  - To justify investments in e-infrastructure they must shown a clear **impact** for the research communities
  - To gauge the impact, this market of end-users must be well understood by funding agencies and e-infrastructure services providers
- User forum
  - A pan-European forum for organisations and projects that operate at an international level
  - Present to the policy makers and the infrastructure providers where there are common needs and opinions and where there is divergence
  - Independent of any supplier and engage across research domains

Bob Jones (CERN)

- Existing e-infrastructure user engagement: e-IRG, EGI, GEANT, TERENA, PRACE.

- need for *"a single organisation with a central role for user communities with a particular emphasis on involving large, advanced and well-organised user communities at a European level and beyond."* (e-IRG Roadmap 2012)

- and that *"On the strategic level user communities will have to organise themselves to drive the long-term strategy."*

Bob Jones (CERN)

- Membership
  - ESFRI cluster projects (BioMedBridges, CRISP, DASISH, ENVRI)
  - EIROforum members
  - Flagship Projects: Human Brain Project and Graphene
  - ERF (European Association of National Research Facilities) that will provide representation of national research infrastructures across Europe

- A number of open User Forum events will be organised to present the work and obtain input from a wider community

Bob Jones (CERN)

- Topics to be addressed (not limited to)
  - The growth of demand and expectations of infrastructure and services
  - The value of a proposed service or infrastructure to the community
  - Interoperability and Sustainability
  - Identifying inhibitors to use (regulatory, procurement, legal, technical)
  - Sharing of best practices and successful approaches
  - Collaboration and creation of common services
  - Creation of common understanding between service providers and users
  - "user aggregation" of needs with respect to working with industry

Bob Jones (CERN)

- Candidate common service needs
  - Single sign-on: consistent access to resources
  - Virtual organisations (collaboration)
  - Persistent storage: long-term preservation of data and its access
  - Data Management services
  - Standards – web services
  - Workflows – support of access to HPC/grid/network resources (compute and data) across Europe
  - Training
  - Global scope: beyond Europe

Bob Jones (CERN)

- **Deliverables**
  - Prioritise and publish issues facing the scientific communities in the areas of e-infrastructures
  - Maintain a database of contact information
  - Provide information on the potential for a service in terms of market size and likely adoption
  - Organise representative input from the scientific communities through workshops and polls
  - Participate in strategic discussions with e-infrastructure providers
  - Participate and provide input on strategic directions from the scientific community for the EC and national funding agencies

Bob Jones (CERN)

- Next steps
  - Decide on an initial coordination team
  - Approach relevant individuals
  - Host an initial meeting to refine the ToR
  - Appoint a chairman and organising panel
  - Disseminate the ToR and present the activity to relevant bodies

Bob Jones (CERN)

In the discussion the following issues were mentioned:

- Sustainability depends on impact on research, society, economics etc.
- How to measure impact. A problem is that impact of e-Infrastructures is indirect and if we cannot measure the final direct impact how can be then measure the indirect impact?
- Are we happy with expanding the community to a large set. How to get broad input but keep it manageable? Are there other mechanisms that can be used besides F2F meetings?
- Mechanism is missing to involve the users in designing the service. Too much top-down?
- Include e-Infrastructure providers in deliverables slide
- Should focus more on which aspects have to be centralised. There is a large potential for collaboration. Find the common services and keep services that are for only one scientific domain in that domain.

Track 1 – Session 2

Stefano Cozzini (CNR/IOM)

e-infrastructure Requirements and Needs for Nano foundries and Fine Analysis (NFFA) scientific community

- Introduced the NFFA project
  - The NFFA Design Study supports the construction and operation of an ERIC consisting of Nanoscale Science Research Centers at European sites that already host Large Scale Facilities for Fine Analysis of matter.

Stefano Cozzini (CNR/IOM)

First data repository in nanoscience

- Goal:
  - Store all the data produced in the NFFA centres
  - make them accessible by means of appropriate search tools

- Includes the relevant information
  - for data analysis
  - for the full reproducibility of preparations and experiments

- Enables exploitation of truly complementary data as obtained by different methods on true replica samples and environment conditions

Stefano Cozzini (CNR/IOM)

Data repository requirements

- Less invasive as possible for users
- Useful for external references: smart and easy access, keyword management, data readability, view sorting and priorities
- Useful for NFFA management: user access, technical liaison, reviewing, proposal support, IPR issues
- Useful for the comparing data and protocols and in calibrating instruments
- Easy to connect to other e-infrastructures

Stefano Cozzini (CNR/IOM)

Conclusions

- Data repository needs for NFFA identified
- Prototype built with the scope of validating the approach by user communities
- The prototype deployment and early users usage identified strong and weak features
- a new enhanced version of the prototype is under development
- Coordination with other communities and e-infrastructure services (data and others) is to take into account

Jon Ison (EMBL)

e-infrastructure - a hackers perspective

- Introduced the Bioinformatics infrastructures
  - ELIXIR, the European life sciences infrastructure for biological information, which is part of ESFRI.
  - BioMedBridges, which develops technical bridges for interoperable data and services in the biomedical science infrastructures
- Definitions of infrastructure
  - "the basic facilities, services and installations needed for the functioning of a community". ➔ hardware, software and data
  - Another definition is "the underlying base or foundation for an organisation or system". The real foundation is the people involved, their motivation and skill; we have to consider education and training.

Jon Ison (EMBL)

- e-infrastructure must deliver:
  - Collaboration: share ideas, resources, workload
  - Efficiency: achieve more for less
  - Productivity
  - Innovation: result of the previous three

- Data aspects
  - access
  - Open (taxpayer paid for it)
  - don't create barriers
  - keep personal data private

- Standards
  - don't invent new formats
  - You can't invent a standard (it emerges)

- .

Jon Ison (EMBL)

- Software tools
  - Well documented
  - Clean and stable interface
  - Work as advertised
  - Supported, versioned
  - Open source
- Common identifiers
- Common vocabularies
- Promote best practices
- Discussion
  - general problem is engaging people but funding is also important
  - danger of getting people in contradicting directions

Sverker Holmgren (White Paper 2013)

- Roadmap 2012: Outlines Europe's need for a single "e-Infrastructure Commons"

- White Paper 2013 (WP 2013): Further discussion and recommendations; draft now open for public consultation

- Aim of recommendations: a single "e-Infrastructure Commons" for knowledge, science and innovation
  - open and accessible continuously adapting to the changing requirements of research
  - open to new technological opportunities
  - meet the challenges of implementing the EU's 2020 Strategy

- Key elements of the e-Infrastructure commons:
  - high quality e-Infrastructure services that are well managed and seamlessly integrated from a users' point of view.
  - services that are flexible and can change dynamically, efficiently and in a future-proof manner.

Sverker Holmgren (White Paper 2013)

- Commons: Resources (including management system, governance etc) accessible to all members of a community
- Needed because in the current complex e-Infrastructure landscape
  - insufficient coordination, collaboration, and integration of e-Infrastructures services
  - legal issues
  - lack of "visibility" of European e-Infrastructure services
  - but high awareness by users of borders, interfaces, and technologies of the individual components
  - Lack of business models for sustainability
  - Lack of models for integration with commercial providers
  - lack of coherence from many user communities

Sverker Holmgren (White Paper 2013)

Proposed approach

- Establish the e-Infrastructure Commons through a joint and truly common strategic effort between users and primary strategic actors and suppliers.
  - a common strategic vision should not form a barrier to innovation in any of the individual (existing) services.
  - users need to become more directly involved in strategy, coordination and innovation activities
  - users need be prepared and empowered to pay for e-infrastructure services.
- Three distinct core functions:
  - Community building, high level strategy and coordination: a single organisation with a central role for user communities
  - Service provision: a flexible, open, and competitive approach to national, European, and global service provision; with advanced collaboration among the interested public and commercial service providers.
  - Innovation: Implementation of major innovation projects through the best consortia including e-Infrastructure suppliers, industry, users and academia.

Sverker Holmgren (White Paper 2013)

- Strengthening the role of users.
- Provide a platform for strategy and coordination bodies of the various e-infrastructure components to discuss:
  - Expanding the user base
  - Avoiding Digital Divide
  - Promoting sustainable business models
  - Promoting effective structures for international governance
  - Separating out business models for operations, support and innovation
  - Promoting effective governance models, giving user communities of all types proper roles

Sverker Holmgren (White Paper 2013)

- WP 2013 contains recommendations to:
  - International user communities
  - International organizations of eRIs
  - National governments
  - The EU

- Follow-up aspects in White Paper 2013
  - e-Infrastructures in support of Open Science
  - Data Policy Recommendations for Large-Scale Research Projects
  - Big Data Access and Standards
  - Cloud Computing
  - Legal Barriers to Commercial Use of e-Infrastructures

  Each topic comes with recommendations

Sverker Holmgren (White Paper 2013)

There was a short discussion on sustainability

- There are many business models but nobody likes them. Confront this issue.

- It is a main issue both for service providers as researchers.