

e-IRG and Data

Parallel Session “Data Infrastructures”

Rossend Llurba (e-IRGSP2)
Matti Heikkurinen (e-IRGSP2)

- Data Management Task Force (DMTF)
- e-IRG and Data after DMTF

DMTF design goals and constraints

- Planning started early 2008
 - e-IRG board decision 19th March 2008
- Numerous data-related initiatives identified
 - Focus on surveying the landscape
- Goal setting needed to be realistic:
 - Capture key issues
 - Support cooperation between initiatives

Composition and management

- Membership
 - Core group from e-IRG delegates (Chaired by Dany Vandromme)
 - Large experts group
 - e-IRGSP2 as support organisation
- Subgroups
 - Survey of existing initiatives (Rudolf Dimper, Peter Wittenburg)
 - 18 social sciences, 12 health sciences, 33 natural sciences
 - Metadata and quality (Peter Wittenburg, Peter Doorn)
 - Requirements of metadata architecture, quality in specific contexts
 - Interoperability (Brian Coghlan, Peter Wittenburg)
 - Cross-disciplinary aspects, opportunities and challenges
- Liaison with ESFRI
 - Task force supported by ESFRI from the start
 - Final report endorsed by e-IRG and ESFRI

Task Force report

- “Heavyweight” document
 - 110 pages
 - Numerous contributors (23 for survey alone!)
- Content by necessity a “snapshot” of the situation
 - Challenging to keep up to date
- Review and consultation processes
 - Open call for experts
 - Public consultation version
 - Review by e-IRG and ESFRI; endorsement December 2009
- Distribution
 - e-IRG member states, ESFRI, e-IRG website

- Metadata is a key enabling tool for quality and interoperability. Goals:
 - Semantically interoperable
 - Available throughout the resource lifetime
 - Standardised, interdisciplinary, usable in aggregations
- Quality of data
 - Scope: discipline, universal standards are not seen as feasible
 - Individual researcher in a key role
 - Challenge interoperability!
- Interoperability
 - Programmes supporting cross-disciplinary access needed
 - Communities key driver, to be supported
 - Open standards and organisational guidelines needed

Survey of the Data initiatives

- Large number of domain-specific initiatives and databases
- Long-term sustainability is a major issue
 - Hardware and software infrastructure
- Focus on curating for reuse rather than long-term preservation
- Emerging trends
 - Move towards distributed/federated data repositories
 - Open access
 - New, ambitious projects changing the landscape
- Future
 - More detailed information about project-specific requirements needed
 - Cross-disciplinary research increases demand for interoperability
 - Common data analysis tools emerging (e.g. GIS)
 - Communication and cooperation between data initiatives should be stimulated

Metadata and Quality of Data

- Focuses on scoping and refining the question
 - What is metadata?
 - How it is used?
 - What are the relevant standards?
 - What are the quality issues related to data (and metadata)?
- Highlighting difficult questions
 - Responsibility for creating & maintaining metadata and data
 - Open access makes peer-review of data technically feasible
- Future directions
 - Data Seal of Approval
 - 16 guidelines for assuring and verifying quality of data
 - Open access: OECD and Berlin Declaration

Interoperability issues

- Key issue: cross-disciplinary use of data
 - Interoperability usually tackled within a discipline
- Resource-level and semantic interoperability
 - Resource levels:
 - Device
 - Communications
 - Middleware
 - Deployment strategies
 - Semantic interoperability
 - Data integration
 - Ontologies
- Several recommendations
 - From standardisation to organisational structures

e-IRG and Data after DMTF

DMTF follow-up activity

- A follow-up action is being prepared
 - Led by Norbert Meyer
 - Identify key issues for follow-up
 - Analyse post-DMTF developments
- Not a DMTF-II!
 - Lightweight organisation
 - Focus on activating community and liaise with other actors in data domain
 - Expedited approval process
- Disseminating the results
 - A stand-alone document or input for other e-IRG documents
 - Possible foundation for other data-related activities

- Roadmap recommendation 3.5 proposes
 - Prepare a blueprint for enabling data-intensive research
 - e-IRG committed its support
 - Concrete steps being discussed
- The e-IRG Workshop in Brussels
 - Data Infrastructure session!
- Next White Paper
 - e-infrastructure services for scientific data is a topic

Thank you!