A 10-year Vision for Global Research Data Infrastructures

Fotis Karagiannis
On behalf of GRDI2020 consortium

contact@grdi2020.eu www.grdi2020.eu

Where do we stand?

Data-driven research

Unifying fragments of data → Disruptive innovations

Unifying fragments of data → Research breakthroughs

Challenges!!!!!

Disconnected repositories with locally stored data

Coping with huge increases in data volume!
Sharing data across disciplines!!
Complexity of diverse data sets!!!

Curating & preserving data sets for current & future generations?

Terabytes | Petabytes | Exabytes

Where are we heading?

General consensus

The competitiveness of European research hinges on our ability to open up new research perspective by ensuring diverse data repositories are accessible & interoperable across geographical & disciplinary boundaries

Current landscape is fragmented/ compartmentalised

More challenges!!!!!

Costly!
Impedes collaboration!
Hinders sharing of best practices!
Grand global challenges harder to achieve!

Addressing the Challenges

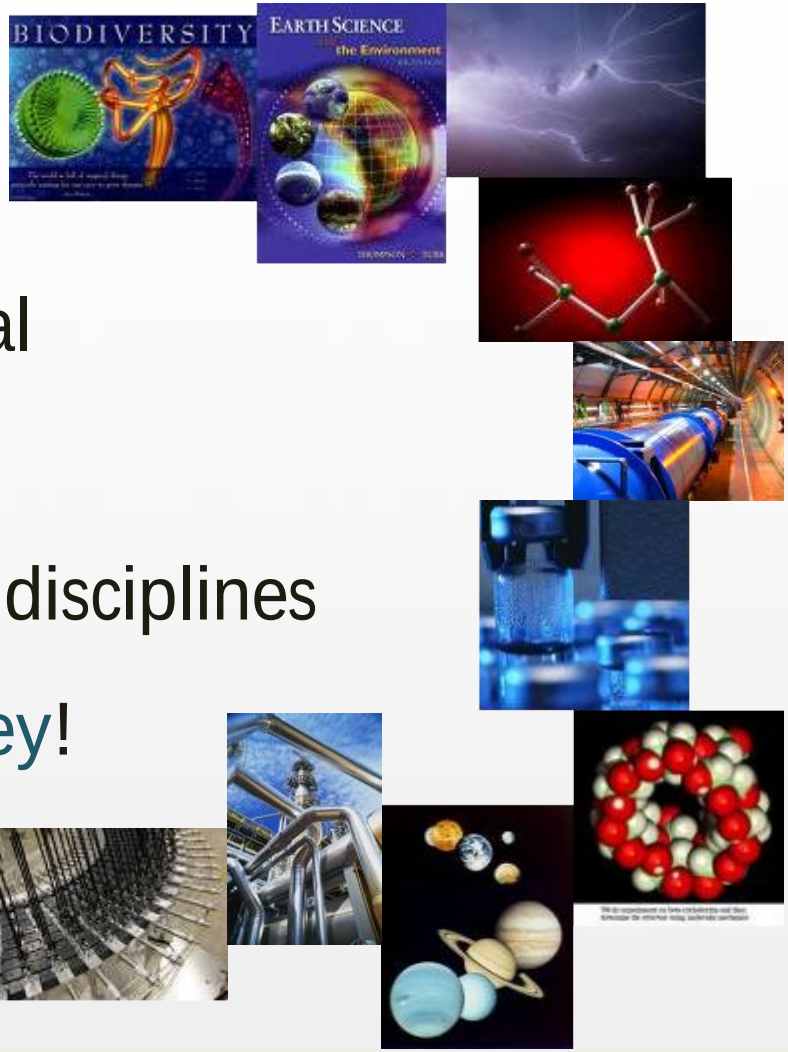
Tackling challenges that are:

- ❑ technological
- ❑ organisational - policy - social

Evaluating:

- ❑ Commonalities across diverse disciplines

→ Their requirements are key!



GRDI2020 Focus

GRDI2020 is about a Global Research Data Infrastructure or a network of interoperable GRDIs for year 2020

GRDI2020 will deliver a Roadmap to help shape the global research data infrastructure for the decades ahead.

It is our vision that GRDI2020 will be:

- A cost-effective, efficient collaborative research environment built on an interoperable & sustainable governance model fulfilling user needs across geographical borders & disciplines

Strengthening EU competitiveness - enhancing world-wide interoperability – increasing usability of data - reducing fragmentation & overlapping of effort

GRDI2020 approach

Community

- Bringing stakeholders close
- Two International Events

2 Working Groups:

Technological Aspects

Organisational /Policy Aspects

Insights

- Coordinating challenges & requirements
- Sharing best practices & knowledge
- Proposing initial solutions

Roadmap

- From state of the art to GRDI2020
- Mapping challenges & requirements to concrete steps
- Policy & sustainability Plan

Tech/Org/Policy/Social challenges (1/2)

- **Funding, Governance and Sustainability models**
  - Actors and stakeholders, business and cost-sharing model, the system functionality, the operational components, and user engagement. And above all, funding needs to be secured.
  - Decentralised or centralised approach? Top down vs. bottom up (user-oriented) ones, (multi-)disciplinary vs. interdisciplinary development strategies
  - **Can GRDIs be modelled after the other e-Infrastructure organisational structures?**
    - such as NRENs, NGIs or National Supercomputer centers?
  - **And/Or be integrated with any of them?**
- Impact of the Digital Agenda for GRDIs
  - http://ec.europa.eu/information_society/digital-agenda/index_en.htm may require adequate data management and open content for research projects in FP8

Tech/Org/Policy/Societal challenges (2/2)

- **Policy issues**: open access and free movement of knowledge, security related issues (privacy, protection, confidentiality, ownership)
  - The policy aspects of changes driven by technical, social or organisational developments: security, interoperability, discovery, provenance and trust, curation and preservation, use, etc.
- **Social issues**: Researchers are inclined to build their own (discipline-based) data “silos” to ensure fuller control
  - Even the use of a common repository does not automatically ensure any sharing of data, not to mention open access to other disciplines.
- **Technical issues** in assembling, securing, managing, preserving and making interoperable the huge amount of data that scientists are producing
- Priorities need to be set so that we can come up with a feasible plan for a GRDI2020

Building on GRL workshops

User focus
Open Access
Preservation
Trust & Quality
Data & Scalability
Interoperability
Sustainability
Services
Libraries
Unique Positioning

- Washington, US 2007
- Pisa, Europe 2008
- Taipei, Asia 2009

GRL2020 Think Tank

Building a Global Community

GRDI2020 workshops and sessions @ CODATA 22, Cape Town, S. Africa

http://www.codata2010.com/

GRDI2020 supporting the HLEG on Scientific Data

High-Level Expert Group on Scientific Data: Vision for data infrastructures 2030

http://cordis.europa.eu/fp7/ict/e-infrastructure/high-level-group_en.html

Chair: John Wood Rapporteur: David Giaretta

- Insights feeding into the GRDI2020 roadmap
- GRDI2020 Reports made available to HLEG (e.g. Data Analysis report)
- GRL2020 Reports made available
- Supporting 2 meetings for focused interaction

Linking with e-IRG

- GRDI2020 project uses the e-IRG DMTF report as an important input
- In contact with the leader of the follow-up activity
- Aligned with the e-IRG roadmap recommendation
  - Roadmap 2010 recommendation 3.5 "Sustainable data management infrastructure": e-IRG recommends that **sufficient EU and national resources are reserved for preparatory work to create a blueprint** for enabling data-intensive research. e-IRG also recommends that **established e-Infrastructure initiatives appoint a representative** to liaise with this new initiative. e-IRG will commit its expertise, contacts to policy makers, and data management experts to support this initiative.
- The project would be interested to see how it could support future activities in this domain in collaboration with e-IRG

GRDI2020
A Vision for Global Research Data Infrastructures

Thank you!

Questions?