



A vision for e-infrastructure in the 21st century

**eIRG workshop
Panel on interoperability
23 May 2013**



Background

These selective responses are set within the wider context of what e-infrastructures should offer to the European Research community in the future.

The eIRG white paper proposes to establish *e-infrastructure commons*, and these responses draw on a number of EIROforum initiated sources that represent cross-disciplinary initiatives engaging major public research organisations that are already “**big data factories**” and will continue to expand their needs and impact in the future

- **A Vision for a European e-Infrastructure for the 21st Century**
(<https://cds.cern.ch/record/1550136>)



Who are the main stakeholders to achieving global research data interoperability?

Future RIs currently in construction (FAIR, XFEL, ELIXIR, EPOS, ESS, SKA, ITER and upgrades to ILL and ESRF etc.), need to be convinced that the e-Infrastructure will exist and continue to evolve throughout their construction and operation phases if they are to take the risk and invest in its creation and exploitation

Need an infrastructure that supports all of the scientific and academic needs of the European community, including the *“long tail of science”*

Cannot be a one-size-fits-all solution

Broad but coherent set of services and tools which must be available to allow the specific needs of each community to be met

Essential that European industry engage with the scientific community in building and providing such services

But it is also important that the user community have a strong voice in the governance



Who are the main implementers of global research data interoperability?

- “pathfinder” initiatives have prototyped many aspects of what will be needed in the future
 - Includes much of the work in the existing e-Infrastructure projects (GEANT, EGI, PRACE), but also projects such as EUDAT, CRISP, Helix Nebula, OpenAIRE+, thematic data projects, such as Transplant and many others.
- Future service infrastructure and tools be fully based on open standards, open software, and provide open access to the data.



How can we create data infrastructures that overcome fragmentation?

- Fragmentation of users (big science vs. long tail)
- Fragmentation of infrastructure (not integrated services)
- Common platform (*e-infrastructure commons*) with 3 integrated areas
 - International network, authorization & authentication, persistent digital identifiers
 - small number of facilities to provide cloud and data services of general and widespread usage
 - Software services and tools to provide value-added abilities to the research communities, in a managed repository

These areas need to be supplemented by investment in application software in order to build and share expertise in ensuring that applications are capable of exploiting evolving computing architectures



How can we improve data re-use by improving interoperability at different layers?

- Many community specific research data infrastructures such as The Catalog of Life indexing the world's known species, iMARINE and GENESI-DEC, which have produced valuable data curation tools and expertise, along with data sharing policies
 - **But most of these are not integrated and not linked to compute services**
- Need a data continuum - linking the different stages of the data lifecycle, from raw data to publication, and cloud compute services necessary to process this data
- Take a look at Zenodo (www.zenodo.org)



Do we need a bottom-up or top-down process or do we need both?

- future e-infrastructure be driven by the scientific stakeholders
 - See proposal for a user forum