FP7 infra-2007 1.2.1 : Scientific Digital Repositories

Original Description of : FP7 infra-2007 1.2.1 : Scientific Digital Repositories

This topic fosters a coordinated approach to the deployment of digital repositories for the scientific communities by pooling existing resources at European level and supporting data storage, archiving, access, interpretation, interoperability, management and curation activities. This will enable scientists to effectively aggregate and combine information to generate and share knowledge, profiting from a transparent underlying data infrastructure across different communities, institutions and geographic boundaries. The contribution to common open standards and their widespread adoption is an essential element of this activity, to bridge heterogeneity and ensure long term preservation.

Expected Impact : The activity on Scientific Digital Repositories is expected to play a catalytic role in the way data repositories for the scientific communities and future generations of scientists are organized, preserved, accessed and support interoperability on the data level. This should optimize the way the e-Infrastructure is used to stock knowledge, add value to primary research data and information (make secondary research more effective), provide a valuable asset for industry and help bridging research and education. In this context, the data layers should emerge as a key aspect of the evolution towards a more advanced knowledge-based e-Science.

Proposed Partners :

Institution	***Contact***	***Reference***
DESY	Patrick Fuhrmann	www.desy.de
CNAF/INFN	Flavia Donno, Marco Paganoni	www.cnaf.infn.it
RAL	Jens Jensen	www.cse.clrc.ac.uk
Swiss Institute of Bio-Informatics	Heinz Stockinger	www.isb.sib.ch
Systems Biology	Adrian Honegger	www.systemsx.ch
CERIMED	Paul Lecoq	cerimed.web.cern.ch
Nordic Data Grid Facility	Lars Fischer, Michael Gronager	www.ndgf.org
European Synchrotron Radiation Facility	Volker Guelzow (temporary)	www.esrf.eu
Sun Microsystems	Axel Köhler, Philippe Trautmann	www.sun.org

Partner Details :

This table briefly sketches the possible contributions of the various partners. It should be seen as a basis for further discussion.

Institution	***Suggested Contribution***	***Status***
DESY	Coordination, Development	OK
CNAF/INFN	Development	OK
RAL	Development, Hep Community Application	
Swiss ISB/SIB	Development, BIO Community Application	OK
Systems Biology	BIO Community Application	
NDGF	Development, Hep Community Application, OTHER Applications	
ESRF	Synchrotron Radiation Community Application	

Institution	*Suggested Contribution*	*Status*
SUN	Vendor, Integration	

Our interpretation of the technical part of this bid :

We assume that the basic technical part of this bid is to enable different scientific communities to manage and share data. Being a European project it is obvious that this has to be done cross location and moreover cross country. Therefore we are proposing a set of organizations to become partners in the project, which are either themselves already distributed amount Europe or have experience in managing distributed data. We furthermore assume that we have to provide 'ready to use solutions' for the whole data chain described above.
The initiative for this project is basically coming from institutions historically involved in HEP experiments. Consequently it shouldn't be surprising that the proposed ideas on the basic structure of this proposal is very close to the building-blocks HEP is using to solve their mass data problems. Therefore it's the goal of one of the initial work packages to integrate the requirements of the different communities into a model suitable for all partners. Independent of the different ways communities are handling their mass data we have been identifying sections which are common to all of them.

- Storage Implementations
- Storage Control Protocols
- Data access protocols (local and wide area)
- Data discovery protocols
- Storage Location capability and discovery protocols
- Meta Data Storage and discovery
- High level data replication protocols
- Application interface layers
- Applications

To our understanding it is essential to make use of agreed standards wherever possible.

Types of contribution :

- Integration : *All partners* have to provide their knowledge on handling community specific data repositories and applications, in order to agree on common denominators of all the building blocks sketched above.
- Agreement on standards : *All partners* have to agree on standards in order to fulfil the requirements of this bid.
- Gap Discovery : *All partners* need to identify areas, requiring non-trivial development in order to provide a fully functional solution for this bid.
- Adaptation : *Community Specific Partners* need to suggest and implement an adaptation of community specific solutions to the agreed standard.
- Development : *Partners*, assigned for *development* need to implement solutions for gaps identified in the different areas.