

Notes from “Data Management Workshop” of EGI Technical Forum 2011

Workshop page:

<https://www.egi.eu/indico/sessionDisplay.py?sessionId=14&confId=452#20110919>

Notes written by:

Gergely Sipos, workshop chair (gergely.sipos@egi.eu)

The “Data Management Workshop” was a 3x90 minutes long track within the “EGI Technical Forum 2011” event. The required outputs from the workshop were best practices, latest developments and requirements from the field of data management and each of these can contribute to the [EGI requirement tracking process](#). Several requirements that had reached the EGI.eu User Community Support Team and are stored within the EGI Requirement Tracker system relate to data management. Therefore the workshop brought together three groups from EGI:

- Software developers who produce applications for data management in distributed computing environments.
- Providers of data services in EGI: both software based services such as storage elements and file catalogues, and human services such as consultancy and support for sites and users.
- Scientific communities who require and/or already use data management software and/or services from EGI.

The first and second sessions of the workshop were filled with short presentations (10-12 minutes each). The third session included structured discussions about topics that were chosen by EGI.eu in advance (based on the existing user requirements about data management) and by the workshop participants during the event.

The workshop was attended by around 60 participants. Most of the attendees had committed themselves to attend all three sessions and actively participated in the discussion. All the presentation and discussion slides can be downloaded from the workshop agenda page: <https://www.egi.eu/indico/sessionDisplay.py?sessionId=14&confId=452#20110919>

In the third session the main points of discussion were:

1. Communication:

The first two sessions presented the results, foreseen results and needs of several projects, initiatives and research collaborations that deal with data in distributed computing environments. The attendees found these presentations very useful but expressed concerns about having a “communication gap” within the EGI community as they had not heard about many of these services, initiatives and requirements earlier. More frequent communication is needed between potential users and software projects that aim to produce solutions that are reusable in multiple scientific disciplines. The community expects EGI.eu (or EGI-InSPIRE) to play a more active role as communication facilitator between these parties. To achieve this EGI.eu/EGI-InSPIRE should put better communication channels and mechanisms in place that enable the various data service providers to inform the potential/existing users about their

services and future plans. Similar mechanisms should exist for the users to inform the providers about their needs and requirements.

EGI (EGI.eu and the NGIs) have been working on such communication mechanisms since the start of EGI-InSPIRE. The collaboration should evaluate whether the existing communication tools – the [EGI Helpdesk](#), [EGI blog](#), [EGI Inspired newsletters](#), Wikis (especially [EGI Wiki](#) and [EMI Wiki](#)), Deliverables & Milestones (from EMI, EGI-InSPIRE, etc), topical workshops – are good enough and are used heavily enough to fulfil their roles. The communication gap should be filled by the improved usage of these services and/or by implementing additional communication mechanisms within EGI.

2. Documentation and guides:

Understanding what software and what services are available for data management in EGI is a complex task. The participants of the workshop tried to identify the reasons of the complexity. The attendees agreed that, generally speaking, sufficient documentation exists for all data management software. These documentations are detailed enough for application developers to understand the capabilities of the particular software these describe and to integrate high level applications with these data management components. However, the number of software is high therefore the number of individual documentations that one need to find, read and understand is also high. “Meta-documentation” that puts the various individual solutions into context, that compares these solutions, that would help the application developers find the relevant services for his/her use case are hard to find or are not available at all. Because different communities have different focus areas within data management, meta-documentations should be specific for disciplines.

EGI.eu is expected to coordinate the development of such additional types of documentation. For example white papers, success stories written by communities who use EGI could be a good start. Such stories could describe the software components and the roles they play in delivering a high level application to users.

The possibilities of having “online discussion and help forums” were also discussed. The conclusion was: online help forums are useful for communities where the focus of work is relatively narrow and where exact solution can be identified for a problem. “Grid problems” cover broad areas and very often there can be several good or acceptable solutions for a single specific “grid problem”. Consequently, a question posted in any EGI/grid discussion forum would attract large number of answers. Identifying the valuable responses from these answers would still be a complex and time consuming investigation for the person who posted the question. The conclusion was that having an EGI online user forum (besides or instead of the existing EGI helpdesk and EGI Requirement Tracker) would not reduce complexities.

3. Integrated services:

The participants discussed whether the EGI data management services should be better integrated, so application developers and/or end users of these services could use and work with higher level interfaces and components.

More integrated APIs on top of individual data management service APIs could provide such an integrated interface for application developers.

High level, Web portal based services (such as Globus online for file transfer) or Web gadget based service (such as EGI AppDB, Training, RT gadgets) – are such integrated components at the end users' level.

- The attendees agreed that integrated APIs are not the priority. The priority is to have well documented and robust service specific APIs for the existing data management software. Using the well documented and robust APIs the application developers from user communities can build any custom, integrated function they wish.
- The representatives of EMI pointed out that there are ongoing developments within their project to have more integrated services for users. For example, the integration of LFC and Storage Elements behind WebDAV interface is one of these. This ongoing work was [presented](#) in the first session of the workshop.
- Endorsement mechanisms to recognise and flag useful, high quality community software within EGI would be useful. (What “high quality” means in this context should be defined while developing the mechanisms.) EGI should incubate software from user communities, should help software developers achieve the pre-defined “high quality” standards with their software, then promote the endorsed software throughout EGI channels. Mechanisms for the quality control and for the endorsement of EGI middleware ([UMD](#)) already exist. Similar mechanisms for community-level software should be defined.