

DATA LIFE CYCLE LABS IN LSDMA

CONTACT DETAILS:

PROJECT NAME: LSDMA – LARGE SCALE DATA MANAGEMENT AND ANALYSIS

NAME: RAINER STOTZKA

ORGANISATION: KARLSRUHE INSTITUTE OF TECHNOLOGY

EMAIL: RAINER.STOTZKA [AT] KIT.EDU

Objectives:

“Large Scale Data Management and Analysis” (LSDMA) is a portfolio extension of the Helmholtz Association of research centers in Germany. It extends the data services with community specific Data Life Cycle Laboratories (DLCL). The DLCLs work in close cooperation with domain scientists which need to process, manage and analyze research data during its whole life cycle. The joint research and development activities in the DLCLs lead to community-specific tools and mechanisms. The DLCLs are complemented with a Data Services Integration Team (DSIT). Its generic R&D focuses on requirements and components of importance to more than one DLCL. It also communicates closely with the providers of storage resources.

On-going activities:

The community specific Data Life Cycle Labs are the core of the project. Each lab has one or more initial topics:

- Key Technologies: synchrotron radiation, nanoscopy, high throughput microscopes, eHumanities
- Energy: smart grids, battery research, fusion research
- Earth and Environment: climate models, environmental satellite data
- Health: virtual human brain map
- Matter: large instruments (ESFRI items Petra 3, X-Ray Free Electron Laser (XFEL), Facility for Antiprotons and Ions Research (FAIR)), heavy ion research, elementary particle physics

The work in the DLCLs is work in progress. In some projects data management and data analysis are already well developed: standardized data formats are used, analysis tools are able to process these formats, and the collection of metadata is recognized as an important task. Missing are well defined processes to describe and to plan the data life cycles, to extract meta data automatically from proprietary device-dependent formats, data organization strategies for long-term preservation and data sharing.

Results:

To demonstrate the successful concept of the Data Life Cycle Labs the work in the DLCL “Key Technologies” is described in more detail. It cooperates with several scientific groups at Karlsruhe Institute of Technology, Technical University of Dresden, the Max-Planck-Institute of Molecular Cell Biology and Genetics in Dresden, the University of Heidelberg, and the University of Mannheim. The partners work on a variety of projects with a common need for sophisticated and data-intensive imaging, federated management, and analysis systems.

As an example a data ingest client has been developed. It ingests digital data objects into a repository system of the Karlsruhe Large Scale Data Facility with an average data rate of more than 400 MB/s. Until now 250 TB which could not be stored locally at the experiment site have been transferred.

The work within the DLCLs and the close cooperation between domain scientists and data scientists enables the realization of new data intensive scientific experiments. Domain scientist can concentrate on the production of scientific results by utilizing modern data organization, handling, and analysis technologies.

URL: <http://www.helmholtz-lsdma.de/>