

Progress in Multi-Disciplinary Data Life Cycle Management

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 J. Phys.: Conf. Ser. 664 032018

(<http://iopscience.iop.org/1742-6596/664/3/032018>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 141.3.58.111

This content was downloaded on 22/03/2016 at 10:51

Please note that [terms and conditions apply](#).

Progress in Multi-Disciplinary Data Life Cycle Management

C Jung¹, M Gasthuber², A Giesler³, M Hardt¹, J Meyer¹, A Prabhune¹, F Rigoll¹, K Schwarz⁴, and A Streit¹

¹ Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

² Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany

³ Forschungszentrum Jülich, Jülich, Germany

⁴ GSI Helmholtz Centre for Heavy Ion Research, Darmstadt, Germany

E-mail: christopher.jung@kit.edu, martin.gasthuber@desy.de,
a.giesler@fz-juelich.de, marcus.hardt@kit.edu, joerg.meyer2@kit.edu,
ajinkya.prabhune@kit.edu, fabian.rigoll@kit.edu, k.schwarz@gsi.de,
achim.streit@kit.edu

Abstract. Modern science is most often driven by data. Improvements in state-of-the-art technologies and methods in many scientific disciplines lead not only to increasing data rates, but also to the need to improve or even completely overhaul their data life cycle management.

Communities usually face two kinds of challenges: generic ones like federated authorization and authentication infrastructures and data preservation, and ones that are specific to their community and their respective data life cycle. In practice, the specific requirements often hinder the use of generic tools and methods.

The German Helmholtz Association project "Large-Scale Data Management and Analysis" (LSDMA) addresses both challenges: its five Data Life Cycle Labs (DLCLs) closely collaborate with communities in joint research and development to optimize the communities data life cycle management, while its Data Services Integration Team (DSIT) provides generic data tools and services.

We present most recent developments and results from the DLCLs covering communities ranging from heavy ion physics and photon science to high-throughput microscopy, and from DSIT.

1. Introduction

The central role of data in science has been boosted in the past few years by the advance of Big Data¹. The sources of these data are experiments, observations and simulations. Policies like data privacy, data preservation and data curation directly affect researchers' handling of scientific data.

The project 'Large-Scale Data Management and Analysis' [2] of the German Helmholtz Association covers both generic and community-specific research and development for scientific data life cycles. Data experts in the Data Life Cycle Labs (DLCLs) perform joint R&D with selected domain scientists, while data experts in the Data Services Integration Team (DSIT) are responsible for generic data tools and services.

¹ Big Data is usually defined via the three Vs: volume, velocity and variety [1].



2. Selected results from Data Life Cycle Management

In this central section of the paper, highlights of the actual R&D performed by the DLCLs and the DSIT are presented. They show in an exemplary way the breadth and depth of the challenges and solutions in data life cycle management.

2.1. DLCL Key Technologies

For this paper, we focus on a novel imaging method based on Localization Microscopy (LM, see Figure 1). LM is a novel imaging technique which focuses on analysis of cellular nanostructures. For example, the chromatin nanostructures of eukaryotic cells has been difficult to analyze by light optical microscope techniques due to the limited physically resolution of 200 nm, the Abbe limit. For deep understanding of these subcellular nanostructures, it is necessary to have resolution ranging down to 20nm and less. Spectral Precision Distance Microscopy (SPDM), an embodiment of LM, allows capturing of high-resolution images at 20 nm range.

Presently, datasets produced during systematic research are in the range of several TB. There are three different kinds of datasets used: raw datasets, intermediate results and high-resolution images. In the near future, they will add up in size to 150-200 TB, which is about 100 times more than the data generated using a conventional fluorescence microscopes.

For managing the extremely large datasets, dealing with their descriptive metadata is very important. The metadata enable the comprehensive description of the data and their provenance allowing the datasets to be referenced and reused. The associated metadata of the dataset are partly embedded in the dataset itself and partly produced in an additional file during the experiment.

For producing valuable research results, several aspects of handling the datasets need to be fulfilled: data sharing, referencing, long-term storage, curation and performant data transfer.

The aforementioned prerequisites can be fulfilled using an Open Reference Data Repository [3]. Within LSDMA, the 'KIT Data Manager' [4] repository system has been developed. It provides a generic repository architecture that can be fully customized to build community specific data repositories. For sustainable and long-term data storage many data back-ends, e.g. the Large Scale Data Facility (LSDF) [5], can be integrated seamlessly. The repository system provides comprehensive high-level services for

- data management and staging,
- metadata management,
- authorization and sharing,
- data discovery based on metadata.

Currently the available services are extended to provide the seamless integration of various analysis workflows and image data annotation technologies.

These developments can be applied in all scientific fields in which novel measurement and imaging technologies are developed. Open Reference Data Repositories enable the results to be shared and discussed openly in the scientific communities.

2.2. DLCL Energy

The DLCL Energy has designed a concept for a user-oriented system for users' energy consumption data [6] and has started a prototype implementation. This modular system (see Figure 2) aims to tackle the technical challenges as well as the requirements posed by the privacy needs of the respective users.

Energy data often is faulty and incomplete. Moreover, different data sources need to be incorporated into one system. Thus, specialized input modules for different kinds of data are aggregated in the input handler which allows for error-tolerant import of those datasets.

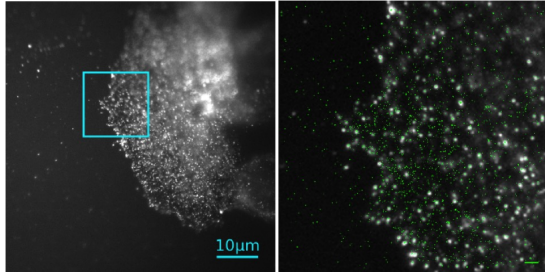


Figure 1: Left: widefield image of a whole cell with a marked ROI; right: enlarged ROI (white) with overlaid localization image (green); image courtesy: University of Heidelberg, Nadine Waltrich, Patrick Müller, Michael Hausmann

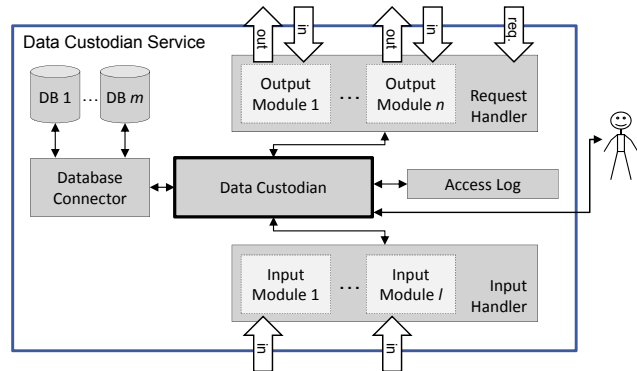


Figure 2: The modular system for a user-oriented energy management system.

Imported data is then processed by a central module of the system – the so-called data custodian – before being sent to the database connector for storage in one or more systems.

Request for data must be issued to the request handler as direct access to the data storage itself is not possible. The data custodian will decide whether processed data is released to the requesting third party or not. Any request for data and the subsequent decisions are logged in the access log. Decisions are based on the requesting party, the requested data quality, and the user-defined rules. The user can decide to allow for higher quality data to be shared, to reduce quality before releasing data, or to deny the release of data. Data quality is defined by temporal and spatial resolution as well as artificial noise. Temporal restrictions lead to reduced frequencies whereas spatial resolution can be lowered by aggregating different data source into one.

The user can define boundaries of data quality for different third parties or decide for each request manually. The Data Custodian Service provides decision support in order to reduce privacy threatening impacts of data distribution. Client-side visualization of the stored energy data helps users to understand the implications a release of their data might have. Thus, the user and third parties can negotiate data qualities which allow for third parties to conduct their analyses while at the same time protecting the users' privacy. The Data Custodian Service can be used not only on a local level but also as part of a larger system in a hierarchical architecture for the entire Smart Grid [7].

2.3. DLCL Earth and Environment

In climatology, a particular task is the comparison and calibration of observed data of remote sensing instruments mounted on the ground, on aircrafts, balloons and on satellites. This requires a matching of geolocations and time of a pair of two devices in given ranges. In [8] the used algorithm is described and the speedup of the geo-matching by using parallel processes to query a MongoDB [9] is explained. Meanwhile we imported geolocations and times of 22 devices and further improved the geo-matcher. In Figure 3 an example of the performance improvement due to parallelization is illustrated.

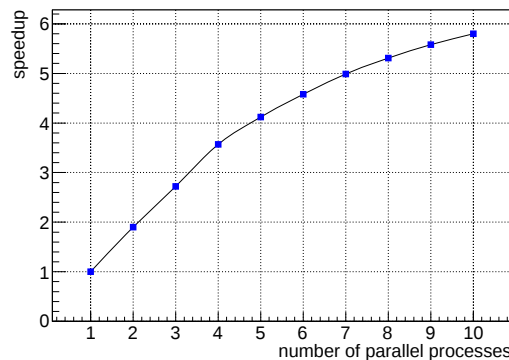


Figure 3: The speedup as a function of parallel processes is displayed for the matching of measurements of the Michelson Interferometer for Passive Atmospheric Sounding [10] and the High Resolution Dynamics Limb Sounder [11]. For higher numbers of client processes the speedup is limited by the I/O-rate of the MongoDB server. This bottleneck can be avoided by the usage of several server instances.

2.4. DLCL Health

The research of the anatomical structure of the human brain on the level of single nerve fibers is one of the most challenging tasks in neuroscience nowadays. In order to understand the connectivity of brain regions on the one hand and to study neurodegenerative diseases on the other hand, a detailed three-dimensional map of nerve fibers has to be created. One mapping technique is Three Dimensional Polarized Light Imaging (3D-PLI) [12] which allows the study of brain regions with a resolution at sub-millimeter scale. Therefore about 1,500 slices, each 70 micron thick, of the post-mortem brain are imaged with a microscopic device using polarized light.

The images of brain slices are processed with a chain of tools for calibration, independent component analysis, enhanced analysis, stitching and segmentation. These tools have been integrated in a UNICORE workflow [13], exploiting many of the workflow system features, such as control structures and human interaction. Prior to the introduction of the UNICORE workflow system, the tools involved were run manually by their respective developers. This approach led to delays in the entire process. The introduction of an fully automated UNICORE approach reduced the makespan of the entire workflow to hours rather than weeks and at the same time, the results are highly reproducible and scalable now.

Tailored solutions were worked out for some peculiarities of the workflow system. For example, in order to use results of one workflow job as input in the next job, the workflow system usually copies this data to the common workflow storage before copying it into the working directory of the next job. The amount of data for a single brain slice is on the order of magnitude of up to 1TB, with intermediate results at the same scale. Thus, the total amount of data easily adds up to several TB of data movement within the workflow, which can be avoided by working directly on a central workflow storage which is available on the file system on the machine running the single job. Additionally, configured storages can be used if there are shared file systems among multiple machines at a single site. Another task for processing large data sets in the 3D-PLI context is the workflow support for the iteration over arbitrary file sets of image data. A brain slice in the workflow is comprised of tiles. The number of tiles belonging to a single brain or their names are not known before the workflow execution. All tiles belonging to a slice are put in a directory, serving as input to the workflow. Thus, the workflow engine is configured to iterate efficiently over the tiles generating independent jobs and intermediate

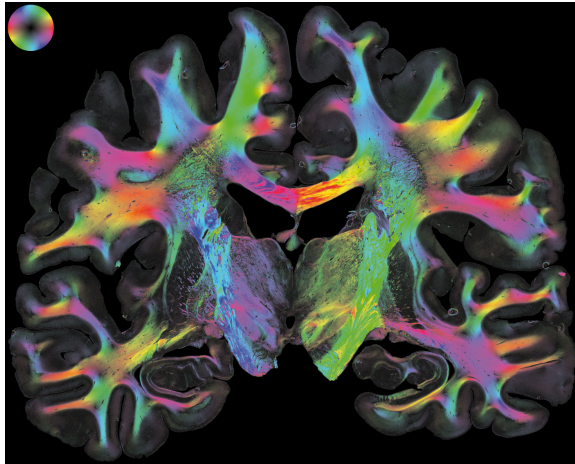


Figure 4: Final 3D-PLI representation of a human brain slice processed by the tailored UNICORE workflow.

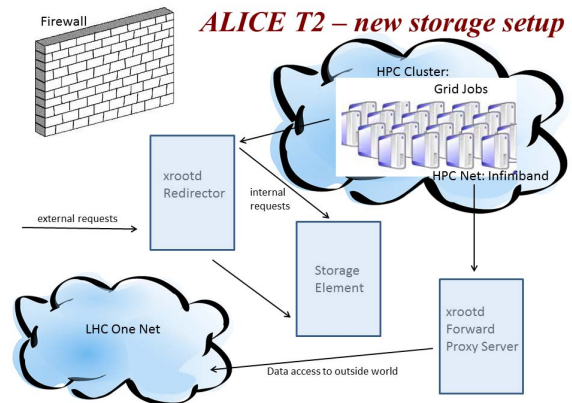


Figure 5: The GSI storage setup

results for each of them. Figure 4 shows the final result of workflow execution of one processed brain slice.

2.5. DLCL Structure of Matter

A photon science highlight of this DLCL was presented at this same conference ([14]); for this paper, we focus on heavy-ion physics.

The exact amount of computing, storage and archiving required for the Facility for Antiproton and Ion Research (FAIR, [15]) depends on many factors but is certainly beyond the capacity of a single computing centre. The required resources are dominated by the experiments Compressed Baryonic Matter (CBM, [16]) and Anti-Proton Annihilations at Darmstadt (PANDA, [17]). Current estimates for the sum of all experiments are 300,000 cores and 40 PB of disk space plus the same amount for archive during the first year of data taking. Especially in order to be able to meet peak demands for computing, it may be necessary to offload some of the computing tasks to public or community clouds, local HPC resources and super computers.

In this contribution an enabling technology is described which gives the possibility to include local HPC resources into a distributed computing environment for FAIR. A prototype has been implemented which will be operated in production mode for the ALICE Tier-2 centre at GSI [18] within the global Worldwide LHC Computing Grid [19] environment.

An xrootd [20] based storage infrastructure has been developed and implemented which can also be used by Grid jobs in the firewall protected environment of the GSI [21] HPC cluster.

The main elements are the xrootd redirector as well as the xrootd forward proxy server. The redirector is using the split directive of xrootd and redirects external clients to the external interface of the GSI storage element and internal clients to the internal interface which is directly connected to the local Infiniband [22] Cluster. The xrootd forward proxy server provides the possibility to Grid jobs running inside the protected HPC environment to read input data from external data sources using the proxy interface. Writing to external storage elements is possible via the same technique. The setup is shown in Figure 5.

2.6. DSIT

Most of the participating institutions in DSIT have a strong background in the X.509 certificate based federated identity management, which is (among others) used within the High Energy

Physics (HEP) community. Experience shows that a relevant share of users unable to use X.509 certificate protected resources. At the same time a significant increase of users with access to SAML [23] based authentication is observed. This can especially be observed in the education sector, where SAML infrastructures are rolled out and provide user accounts to all students and employees by default.

To allow a widespread usage of sophisticatede-infrastructure, a better co-existence between SAML and X.509 has to be established. One approach would be to modify the infrastructure services to support SAML natively. However, this is impractical due to the complexity and manifold operational processes established upon X.509.

DSIT is therefore working on concepts and methods to provide migration paths from X.509 certificates to other authentication infrastructures such as SAML or OpenID Connect [24]. The goal is to be transparent to the user, e.g. by translating credentials on behalf of the user. This can be applied at several levels, e.g. as described in [25–27]. It is important to note that information to be used for authorisation decisions should survive a translation process and furthermore, the establishment of trust relations between credential provider and consumer is a complex issue in its own.

In LSDMA's DSIT we are working on two approaches towards improving access to our resources via SAML, yet supporting X.509 where feasible. One involves the use of an online certification authority (CA), e.g. DFN SLCS [27], while the other builds on the replacement or extension of core authentication components in the infrastructure.

The first approach is technically feasible, as a user visits the SLCS portal to which he authenticates via SAML at his home institutions identity-provider (IdP). The DFN SLCS CA is accredited within the IGTF trust framework [28] and can issue certificates for use in LCG. However, organisational challenges have yet to be addressed. Firstly every IdP has to sign a contract with SLCS to ensure that all users given the entitlement required to obtain a certificate have undergone existing identity vetting procedures. Secondly, most IdPs do not currently support different levels of assertion for their users, left alone adhere to one common scheme for this. Currently, commandline access is not supported and to some extent, certificates still have to be handled by hand. LSDMA is working on an improved web client to replace the current java-webstart based approach. LSDMA also pushes forward a commandline client, which uses the ECP profile of SAML [29].

The second approach is to extend the Lightweight Directory Access Protocol (LDAP) as one of the core components that handle local authentication of users. We are building on top of the existing work of [25, 30], which was started at KIT. The initial status is that non-web logins via SAML/ECP are supported to any service which can authenticate against either Pluggable Authentication Modules (PAM) or LDAP. Both ECP modes are supported: the less secure proxy mode as well as the more secure enhanced client mode. Both methods uses the user provided password via LDAP to authenticate via SAML/ECP to a backend IdP. This approach is very versatile and can in principle be extended to support technologies like OpenID on the backend. We are currently working on an extension to support kerberos [31] and gridFTP [32] on the client side. Kerberos, however requires users once to generate a service password which has to be subsequently used. The support for gridFTP is implemented via the globus authorisation callout, in which the subject of the X.509 certificate is passed via LDAP. Trust is established by relying on the fact that gridFTP has verified the users certificate.

Additionally, we are conducting conceptional work to ensure that third party authorisation can be supported. For this we plan to follow the VOMS concept in that an external (web)service is offered for the administration of group membership. The membership information then can be retrieved by the Service Provider (SP) from the external (web)service and mapped into unix group IDs by the LDAP service.

3. Lessons Learned

From a distance, it might seem that scientific communities 'only' need large-scale storage, but practice shows that this is not the case. The challenges posed by the advance of scientific Big Data are diverse. Though the communities recognize these challenges, their main focus stays on the analysis of their scientific data. Most of the challenges are community-specific, as can be seen by the joint R&D performed by the DLCLs and their respective communities.

Tools and workflows for running experiments can be changed and replaced only gradually. Even if new technologies and approaches promise major advances, their implementation might not be feasible. This makes the collaboration of domain scientists and data experts in the planning phase of new experiments so valuable.

When LSDMA started in 2012, all its subprojects started simultaneously. Syncing developments between DLCLs and DSIT was a process that took time. The communities knew their immediate needs rather well, but carving out mid- and long-term requirements that were common within these communities required much communication and reflection. As the needs of the LSDMA communities differ substantially, tools and workflows developed for one particular community are rarely taken up by another community; yet the ideas and concepts of these tools are used when designing solutions for other communities.

Handling scientific data has been a very important topic and will become even more important.

Acknowledgements

The authors wish to thank all people and institutions involved in LSDMA as well as the German Helmholtz Association for funding the project.

References

- [1] Laney D 2001 3D data management: Controlling data volume, velocity, and variety Tech. rep. META Group
- [2] Jung C, Gasthuber M, Giesler A, Hardt M, Meyer J *et al.* 2014 *J.Phys.Conf.Ser.* **513** 032047
- [3] Prabhune A, Jejkal J, Hartmann V, Stotzka R, Bach M, Schmitt E, Hausmann M and Hesser J 2015 *IEEE BigDataService 2015 Conference*
- [4] 2015 KIT Data Manager URL <http://www.kitdatamanager.net/>
- [5] Stotzka R, Hartmann V, Jejkal T, Sutter M, van Wezel J, Hardt M, Garcia A, Kupsch R and Bourov S 2011 *Parallel, Distributed and Network-Based Processing (PDP), 2011 19th Euromicro International Conference on* (IEEE) pp 373–379
- [6] Rigoll F and Schmeck H 2014 *VDE-Kongress 2014* (VDE VERLAG GmbH)
- [7] Rigoll F, Hirsch C, Kochanneck S, Schmeck H and Mauser I 2014 *To be published in Proceedings of the 11th IEEE International Conference on Autonomic and Trusted Computing (ATC-2014)*
- [8] Ameri P, Grabowski U, Meyer J and Streit A 2014 *Trust, Security and Privacy in Computing and Communications (TrustCom), 2014 IEEE 13th International Conference on* pp 652–659
- [9] MongoDB homepage (last visited on 2013-10-08) URL <http://www.mongodb.org/>
- [10] Fischer H, Birk M, Blom C, Carli B, Carlotti M, von Clarmann T, Delbouille L, Dudhia A, Ehhalt D, Endemann M, Flaud J M, Gessner R, Kleinert A, Koopmann R, Langen J, López-Puertas M, Mosner P, Nett H, Oelhaf H, Perron G, Remedios J, Ridolfi M, Stiller G and Zander R 2008 *Atmos. Chem. Phys.* **8** 2151–2188

- [11] Gille J C, Barnett J J, Whitney J G, Dials M A, Woodard D, Rudolf W P, Lambert A and Mankin W 2003 The high-resolution dynamics limb sounder (hirdls) experiment on aura URL <http://dx.doi.org/10.1117/12.507657>
- [12] Axer H, Axer M, Krings T and v Keyserlingk D G 2001 *Journal of Neuroscience Methods* **105** (2) 121–131
- [13] UNICORE Workflow System Manual (last visited on 2014-07-28) URL <http://www.unicore.eu/documentation/manuals/unicore6/files/workflow/workflow-manual.html>
- [14] Gasthuber M *et al.* Architecture of a new data taking and analysis infrastructure and services for the next generation detectors of Petra3 at DESY to be published in Proceedings of CHEP 2015
- [15] FAIR webpage (last visited on 2015-05-06) URL <http://www.fair-center.eu/>
- [16] (last visited on 2015-05-06) URL <http://www.fair-center.de/index.php?id=184&L=0>
- [17] Protopopescu D and Schwarz K 2011 *Journal of Physics: Conference Series* **331** 072028
- [18] Schwarz K 2011 *J.Phys.Conf.Ser.* **331** 052018
- [19] WLCG webpage (last visited on 2015-05-06) URL <http://wlcg.web.cern.ch/>
- [20] Homepage of XRootD URL <http://xrootd.org/>
- [21] GSI webpage (last visited on 2015-05-06) URL <https://www.gsi.de/en/start/news.htm>
- [22] Infiniband Trade Association (last visited on 2015-10-22) URL <http://www.infinibandta.org>
- [23] Organization for the Advancement of Structured Information Standards 2005 Security assertion markup language (saml) v2.0
- [24] Oudot C The OpenID Connect Protocol (last visited on 2015-10-22) URL https://2015.rmll.info/IMG/pdf/rmll2015_openidconnect.pdf
- [25] Köhler J, Labitzke S, Simon M, Nussbaumer M and Hartenstein H 2012 *Proc. of the 11th IEEE Int. Conference on Trust, Security and Privacy in Computing and Communications, TrustCom-2012 - 11th IEEE Int. Conference on Ubiquitous Computing and Communications, IUCC-2012* 557–564
- [26] Hardt M, Hayrapetyan A, Memon A and Millar P 2014
- [27] DFN SLCS <https://slcs.pca.dfn.de/gridshib-ca> (last visited 18.2.2015)
- [28] IGTF: Interoperable Global Trust Federation (last visited on 2015-10-22) URL <https://www.igtf.net/>
- [29] Enhanced Client or Proxy (ECP) (last visited on 2015-10-22) URL <https://wiki.shibboleth.net/confluence/display/CONCEPT/ECP>
- [30] Köhler J, Simon M, Nussbaumer M and Hartenstein H 2013 *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **7905 LNCS** 462–473
- [31] Kohl J T, Neuman B C and Theodore Y 1991 *European Conference Proceedings* pp 295–313
- [32] Bresnahan J, Link M, Khanna G, Imani Z, Kettimuthu R and Foster I 2007 *Proceedings of the First International Conference on Networks for Grid Applications* GridNets '07 (ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering)) pp 19:1–19:5 ISBN 978-963-9799-02-8 URL <http://dl.acm.org/citation.cfm?id=1386610.1386636>