



EUDAT 2nd Conference Parallel Tracks

PARALLEL TRACK 1: EUDAT SERVICES



SESSION TITLE: SAFE REPLICATION AND DATA STAGING

CHAIR: JOHANNES REETZ, RZG / MPG, GERMANY

DATE & TIME: TUESDAY 29TH OCTOBER - 14:00 - 16:00

OVERVIEW: Safe Replication (SR) and Data Staging (DS) are EUDAT services for moving data between sites and storage systems for two kinds of purposes. The purpose of SR is to keep the data from a repository safe by replicating it across different geographical and administrative zones according to a set of well-defined policies. It is also a way to store larger volumes of data permanently at those sites which are providing powerful on-demand data analysis facilities. In particular, SR operates on the domain of registered data where data objects are referable via persistent identifiers (PIDs). SR is more than just copying data because the PIDs must be carefully managed when data objects are moved or replicated.

The general purpose of DS is to move data between storage systems, specifically, DS transfers data from the domain of registered data into a temporary storage space, and vice versa, it moves data from any storage space into the domain of registered data, typically a repository.

The session will introduce the problem space of SR and DS, presents the achievements that have been made during the last year for enabling communities to make use of the SR service as well as DS, demonstrates a few use cases, outlines the commonalities and differences between the policies for SR, presents new developments towards a common service layer interface and a data policy management framework.

AGENDA (each presentation is followed by 5 min Q&A):

14:00 - 14:10 **B2SAFE and B2STAGE: Two Core Services of the EUDAT CDI**, Johannes Reetz, RZG, Germany

14:10 - 14:20 **B2SAFE adoption in the EPOS community**, Claudio Cacciari, CINECA, Italy

14:25 - 14:35 : **Utilisation of B2SAFE** by the MPI-TLA CLARIN Center, Willem Elbers, Max-Planck-Institute for Psycholinguistics, The Netherlands

14:40 - 14:50 **CLARIN-CUNI**, Pavel Stranek, Charles University, Czech Republic

14:55 - 15:10 **YPH and the Biomedical Scientific Case for EUDAT**, Peter Coveney, University College London, UK

15:10 - 15:30 **Community integrated B2STAGE tools**, Stefan Zasada, University College London, UK and Giuseppe Fiameni, CINECA, Italy

15:35 - 15:45 **B2SAFE Policies**, Willem Elbers, Max-Planck-Institute for Psycholinguistics, The Netherlands

15:45 - 15:55 **B2SAFE - Data Policy Manager Service Case**, Maria Francesca Iozzi, SIGMA/University of Oslo, Norway

SESSION TITLE: SIMPLE STORE

CHAIR: MARK VAN DE SANDEN, SURFsara, The Netherlands

DATE & TIME: TUESDAY 29TH OCTOBER - 16:30 - 18:30

OVERVIEW: Many researchers are collaborating across institutional boundaries and are facing the problem of finding an easy way of storing and sharing their data. But even single researchers should want to safeguard their own results and data in a safe storage place that will be kept open and accessible for other researchers in the future. They often have large numbers of small files, for example, files containing derived data in the form of spread sheets or analysis results. Although the information in these small files is important, they usually do not belong to large research collaborations with well-defined data storage plans. These kind of files are known as the "long tail data", which are often stored locally on laptops and departmental storage devices with the risk of losing valuable scientific data, either because other researchers do not have easy access to the data or because such storage systems are often not adequately secure. To solve this problem EUDAT is working on a service called Simple Store which combines a shared storage space with the possibility to add metadata and to assign persistent identifiers. This session will present the current work and plans on the Simple Store service, related work from other projects and there will be time for discussion on the subject of "long tail data". Options to provide cloud storage as a temporary low-barrier storage solution will be discussed.

AGENDA:

Welcome & Introduction, Mark Van De Sanden, SURFsara

Mark van de Sanden has a Bachelor degree in computer engineering from the technical college's Hertogenbosch and started work at the National Aerospace Laboratory (NLR) as system administrator. In 1997 he joined SURFsara as an UNIX system administrator of supercomputing environments. In 2002 he started working on the SURFsara mass storage infrastructures and is currently team leader of the Data Services group. He is involved in large scale data projects like the Large Hadron Collider (LHC) as part of the WLCG NL-T1 and Long Term Archiving (LTA) facilities of the LOFAR low frequency telescope. In EUDAT he is coordinating the service building work package.

Meeting DRIHM Citizen Scientist needs with SimpleStore, Alberto Parodi - CIMA Research Foundation

Expert in atmospheric modelling and statistical analysis of extreme events, in the development of simplified models of dry and moist convection and the study of the main sources of uncertainty in the high resolution numerical modelling of deep moist convective processes. Awarded with a CNR-MIT grant in 2002 in the framework of the bilateral USA-Italy investigations on climate change and hydrogeological disasters. Since 2003 has developed teaching activities at the University of Genova in the following fields: Hydraulics, Fluid Mechanics, Dynamics of Atmosphere and Computational methods in Environmental Engineering. Coordinator of FP7 project DRIHMS (Distributed Research Infrastructure for Hydro-Meteorology Study, www.drihms.eu, 2009-2011) and DRIHM (Distributed Research Infrastructure for Hydro-Meteorology Study, www.drihm.eu, 2011-2015). Coordinator of FP7 project DRIHMS (Distributed Research Infrastructure for Hydro-Meteorology Study, www.drihms.eu, 2009-2011), DRIHM (Distributed Research Infrastructure for Hydro-Meteorology Study, www.drihm.eu, 2011-2015) and DRIHM2US (Distributed Research Infrastructure for Hydro-Meteorology Study to United States of America, www.drihm2us.eu, 2012-2014). Antonio Parodi is author and co-author of 30 papers published in international peer-reviewed and referred journals.

SESSION TITLE: 1.3 METADATA

CHAIR: DAAN BROEDER, MPI for Psycholinguistics, The Netherlands

DATE & TIME: WEDNESDAY 30TH OCTOBER - 09:00 - 10:30

OVERVIEW: The data subject is hot; it is considered the new oil of the digital area. Data is produced at staggering rates. The main questions are "Where store it?", "How to find it?" and "How to make the most of it?". In this context, metadata is the key to improving and ensuring the quality of data for current and future usage, and metadata catalogues are the place to find the interesting relevant and most valuable data sets. EUDAT is developing a joint metadata catalogue that combines the metadata, gathered from various repositories and sources, and hence bridges key information from research across science domains. This session will present current work and plans on the joint metadata service, the need for harvestable repositories and the challenge to understand and to bridge domain specific ontologies. It will also present related work to the dispersed metadata challenge. The session foresees ample discussion time to hear views and contributions from the participants.

AGENDA:

09:00 - 09:30 Data interoperability in cultural heritage: the Europeana approach - Nuno Freire, The European Library, Europeana Foundation
Nuno Freire is a Senior Researcher at The European Library. He holds a PhD in Informatics and Computer Engineering from the Instituto Superior Técnico of the Technical University of Lisbon. During his entire career he has been involved in research projects in the area of digital libraries. His areas of interest include information systems, information retrieval, information extraction, data quality, and knowledge representation, particularly in their application to digital libraries and bibliographic data.

09:30 - 10:00 Data Federation via Metadata - Bill Michener, University of New Mexico's University Libraries

Bill Michener is Professor and Director of e-Science Initiatives at the University of New Mexico's University Libraries. He serves as Project Director for two large National Science Foundation supported projects: (1) Data Observation Network for Earth (DataONE)—a large DataNet project that supports cyberinfrastructure development and community engagement for the biological, environmental, and Earth sciences; and (2) the New Mexico Experimental Program to Stimulate Competitive Research. He is actively involved in research related to creating information technologies supporting data-intensive science, development of federated data systems, and community engagement and education. He has a PhD in Biological Oceanography from the University of South Carolina and has published extensively in marine science, as well as the ecological and information sciences.

Bill has authored five books and more than 100 journal articles and book chapters. He has a strong background and training in organizational sustainability and governance, project management, and meeting facilitation. Presently, he serves on the Board of Directors (or Administrative Board) for Dryad, Inc., the Organization for Tropical Studies, and the Cornell Lab of Ornithology, as well as the Governance Committee for an emerging international organization that seeks to nurture a network of organizations that are involved in Public Participation in Scientific Research (i.e., citizen science). He serves as editor of the Ecological Society of America's Ecological Archives, Associate Editor for Ecological Informatics, and as a member of the Ecology Editorial Board, and was recently appointed to the Technical Advisory Board for the Research Data Alliance.

10:00 - 10:30 B2FIND: The EUDAT Metadata Service - Daan Broeder, The Language Archive - MPI for Psycholinguistics

Daan Broeder, for many years senior developer for archive and infrastructure solutions and now deputy director of TLA unit of the MPI for Psycholinguistics, is technologist (electronics and IT) and has a long record in leading development tasks in international projects. Currently he is a member of the executive board of the Dutch CLARIN project and participating several EU projects concerned with research infrastructure development as DASISH and EUDAT. In addition he is convener of new ISO standards in the linguistic domain (TC37/SC4: Component Metadata Infrastructure "CMD" and persistent identification "PISA").

Abstract



Attachment	Size
BillMichener_dataone_metadata.pdf	1.43 MB
Claudio Cacciari.pdf	2.16 MB
DaanBroeder.pdf	1.68 MB
EUDAT_2nd conference_Track1.1_0.pdf	393.06 KB
EUDAT_2nd conference_Track1.2_0.pdf	453.03 KB
EUDAT_2nd conference_Track1.3_0.pdf	471.04 KB
JohannesReetz.pdf	987.57 KB
MariaFrancescalozzi.pdf	960.71 KB
NunoFreire.pdf	3.82 MB
PavelStranak.pdf	3.14 MB
PeterCoveney_0.pdf	3.09 KB
StefanZasada.pdf	1.08 MB
Willem Elbers_clarin_demo_.pdf	1.06 MB
Willem Elbers_practical.policies.pdf	1 MB

[Read more](#)