

EUDAT 2nd Conference Workshops

WORKSHOP TITLE: CUMULONIMBO: THE SKY IS NO LONGER THE LIMIT FOR BIG DATA!

MONDAY 28TH OCTOBER 09:00 - 13:00

Big Data: Only for analytical processing? These days that idea is ancient history. Now, with CumuloNimbo, Big Data can deal with any workload!

So far, Big Data has been a synonym for performing large analytical queries on massive amounts of data. However, in the CumuloNimbo workshop at the 2nd EUDAT conference, we will see how Big Data is getting even bigger, thanks to the new advances brought about by the CumuloNimbo project in ultra-scalable transactional processing.

Big Data is no longer constrained to just the analytical processing world. Unlike today, transactional workloads in the near future will scale out linearly without having to resort to sharding (which frequently requires that substantial modifications are made to applications). An extra benefit is that users will not have to resort to NoSQL (which currently implies sharding) in order to scale. If SQL is needed, it can be used and will scale out. If NoSQL is enough, it can be used and fewer resources will be consumed for query processing. If both SQL and NoSQL are needed, the two interfaces will be available to access the same data store in a coherent manner. Additionally, since updates will now scale as much as is needed, it will be possible to have Big Data combined with complex event processing (CEP) on massive event streams. This approach is based on enabling CEP queries to correlate these massive event streams over stored Big Data at network rates.

With these new advances, the Big Data world will be transformed to incorporate any workload from online analytical processing (OLAP) to online transaction processing (OLTP) and even to complex event processing. The new Big Data milieu will also support all the different data management paradigms available - from NoSQL technologies to CEP and SQL. Additionally, it will be possible to integrate specialized data stores (such as graph databases or document-oriented data stores) with SQL and thus applications will be able to exploit the best of the SQL and NoSQL worlds.

AGENDA:

- CumuloNimbo Overview.
- Ultra scalable transactional processing
- SQL support over NoSQL
- Beyond HBase
- CoherentPaaS and LeanBigData
- Discussion

To learn more about the exciting advances in Big Data that are being brought about by the CumuloNimbo project, please join us at the CumuloNimbo workshop at the 2nd EUDAT conference. For more information, see <http://cumulonimbo.eu>

Overview

WORKSHOP TITLE: DATA SHARING AND ENRICHMENT: IMARINE AND LIFEWATCH SOLUTIONS & EXPERIENCES

MONDAY 28TH OCTOBER 14:00 - 18:00

Overview

Data and Interoperability, in themselves challenging subjects, are also closely tied to research communities. Data sharing across the boundaries of communities and organizations, in the sense that data is usable in other contexts than those for which it has been generated, is a pressing need to guarantee richer, better quality and timely science-based knowledge creation. Addressing this challenge requires approaches that not only facilitate access to data produced by others but are also able to transform and enrich the shared data in a way that makes them suitable for being consumed in different contexts.

The goal of this workshop is to highlight these major interoperability issues as well as to showcase and compare the diverse yet complementary approaches and solutions developed by different stakeholders to promote data interoperability on a large scale. Experiences discussed will include:

- the ones gained by the iMarine (www.i-marine.eu) consortium, developing a cutting-edge e-infrastructure and fostering a collaborative approach to open data access and interoperability, by supporting marine and biodiversity specialist communities to unlock knowledge and support science policy decision making.

- Lifewatch's (www.lifewatch.eu) showcase "Patterns of ecosystem fragility to alien and invasive species in Europe", its tools and data services. The showcase has collected and shared data from existing research provided by some of the participating universities and institutions. The data includes about 11570 species from 314 different terrestrial and marine ecosystems.

14:00 - 14:20 Welcome and Workshop Objective - D. Castelli - ISTI-CNR

14:20 - 15:10 LifeWatch: Standards required/used by the community - H. Schentz - Umweltbundesamt GmbH, A. Oggioni - LTER

15:10 - 16:00 iMarine: Accessing and managing Biodiversity data - P. Pagano - ISTI-CNR

16:00 - 16:15 Coffee Break

16:15 - 17:00 LIFEWATCH ICT CORE bricks: Integration of preparatory phase projects developments for the coordination and management of the distributed e-Infrastructure construction : Antonio José Sáenz Albanés - LW ICT-Core, Daniel Fuentes - LW ICT-Core

17:00 - 17:45 iMarine: Analyzing and processing Biodiversity data: A. Manzi - CERN, G. Coro- ISTI-CNR, P. Pagano- ISTI-CNR

17:45 - 18:00 Conclusions : N. Fiore - LW Service Centre

WORKSHOP TITLE: SIM4RDM - CROSS-STAKEHOLDER RESEARCH DATA MANAGEMENT

MONDAY 28TH OCTOBER 14:00 - 18:00

Overview

The SIM4RDM project aims to improve current policies in the area of managing research data. The project has surveyed a selection of research data management (RDM) stakeholders across the EU and is currently building a self-assessment tool to help them evaluate and improve their RDM maturity.

This tool is based on an online questionnaire and produces overall maturity scores and detailed results measured against five scales, thus allowing users to decide what RDM areas they wish to improve first. The tool will suggest sets of concrete activities designed to increase the user's RDM maturity, and users are also pointed to online resources and case studies to help them complete these activities.

The process of developing the self-assessment tool revealed that there are overlaps in the activities suggested to funding bodies, infrastructure providers, research institutions, publishers and researchers. However RDM policies and recommendations are usually developed in cultural and sectorial isolation: senior executives in these stakeholder groups have very few opportunities to discuss RDM issues of common interest with each other.

SIM4RDM has therefore planned a workshop at the 2nd EUDAT Conference to offer senior executives (from funding bodies, infrastructure providers, research institutions, publishers and research centres across Europe) a unique opportunity to provide feedback on the functionality of the self-assessment tool, and to explore opportunities for cross-stakeholder collaboration on RDM issues of common interest.

The workshop is structured in two parts. In the first part, the participants will highlight RDM issues that are crucial to their activities and which they would like to explore through cross-stakeholder collaboration. In the second part of the workshop, the participants will engage in small-group exercises to provide feedback on the functionality of the self-assessment tool and to explore some of the opportunities for cross-stakeholder action highlighted earlier in the workshop.

AGENDA:

13.00 Networking lunch

14.00 Project overview and event outline - Matthew Dovey and Gabriel Hanganu

14.15 RDM collaboration pitches

Simon Hodson

Ingrid Dillo

Jonathan Tedds

Rebecca Lawrence

Leif Laaksonen

15.30 Cross-stakeholder collaboration discussion

16.00 Coffee break

16.20 Interactive hands-on session

(attendees divided in small groups work on the pitched ideas and their potential integration in the sim4rdm framework)

17.50 Reports from groups

18.00 Wrap-up and close

Further information about SIM4RDM is available on the [project website](#).

WORKSHOP TITLE: SOCIAL SCIENCES AND HUMANITIES (SSH) TACKLE THE BIG DATA CHALLENGE
MONDAY 28TH OCTOBER 11:00 - 17:00

Overview

Social Sciences and Humanities (SSH) are in general not known for Big Data challenges since the term often is reduced to processing large volumes of data. However, there are more aspects such as data to be worked on being highly distributed or having complex relationships with each other that need to be exploited etc. Partly, it has also to do with the fact that new trends in SSH are completely ignored such as studying the human brain based on brain imaging methodologies and genetic insights, the human language capacity and society based on massive crowdsourcing, the dynamics of diversities in cultures and languages based on extensive observations and phylogenetic methods, understanding the challenges of aging societies by collecting a wide range of markers including physiological ones, and many more. These questions address, amongst others, the interest to better understand the principles of maintaining stable minds and societies - thus grand challenges in the area of SSH. The workshop will see contributions from experts addressing these new challenges in SSH that combine large and complex data with new computational challenges. One aspect that is similar to all these initiatives is that the investigators often do not have the facilities to develop large software packages, to manage appropriate storage and computer systems and to centralize the required facilities. We want to see what kind of directions are currently being worked on, what their challenges are and what kind of infrastructures are required to allow SSH researchers to carry out this kind of new research. The workshop targets experts from humanities departments who tackle data issues of a certain scale and complexity that clearly crosses normal situations and can describe the challenges and the opportunities. It also targets technologists that have an interest in helping to find methodological and technological solutions for the described challenges.

Agenda :

11.30 Introduction
11.35 Peter Doorn -- *How to tackle the challenge of Long-term Access to Big Data in the Humanities and Social Sciences?*
12.15 Nicola Masini -- *Satellite digital data for Cultural heritage: new strategies to share and extract information. The Virtual laboratory Italy-China*
12.45 Riccardo Pozzo -- *From Data Science to Data Humanities*
13.15 Lunch Break
14.15 Nanna Floor Clausen -- *Dealing with Danish Cencuses - problems and opportunities*
14.45 Binyam Gebre -- *Massive Crowdsourcing: changing humanities.*
15.15 Luca Pezzati -- *Digitizing Cultural Heritage*
15.45 William C. Block -- *Freedom on the move*
16.15 Coffee Break
16.30 DH and H2020 - *General Discussion*
17.00 End

MEETING: PID INFORMATION TYPES

WEDNESDAY 30TH OCTOBER 14:30 - 18:00

Overview

The WG meeting will take place just a couple of weeks after the 2nd RDA Plenary. We will review the outcomes of the plenary meeting and continue potential discussion topics, review the use case documents, discuss possible PID types, mix and synchronize with other WGs, get newcomers up to speed and have a glance at the next steps. The meeting will be interactive and creative, we will try to have a minimum set of presentations and hopefully a maximum of ideas and pragmatism. If you are from the EU and did not have a chance to attend the 2nd RDA Plenary, you should join us in Rome as this is a good way for you to keep track of WG activities.

The meeting is open to anyone interested in WG activities, newcomers and known faces alike - if you are not a member of the WG yet, feel free to use the meeting to become involved!

MEETING: EPIC USER MEETING

WEDNESDAY 30TH OCTOBER 14:30 - 18:00

Overview

The European Persistent Identifier Consortium (EPIC) User Meeting 2013 takes place immediately after the close of EUDAT conference. The meeting is open to all but subject to pre-registration. For more information please contact Ulrich Schwarzmänn email: [uschwar1\[at\]gwdg.de](mailto:uschwar1[at]gwdg.de)
http://pidconsortium.eu/index.php?page=activities/2013_uf

WORKSHOP: DIGITAL PRESERVATION OF CULTURAL DATA

WEDNESDAY 30TH OCTOBER 14:30 - 18:00

Medicine and Natural sciences, including astronomy, biology, chemistry, earth sciences and physics, already make use of standardized formats and e-Infrastructures services to generate, curate, share and analyse research data. The need for novel more efficient and affordable solutions for digital preservation is now increasing also in the Social Science and Humanities, in particular the Digital Cultural Heritage (DCH) sector is producing a large volume of digital content that needs to be safely stored and curated, permanently accessed, and easily shared and re-used by researchers.

Each digitisation programme is currently addressing the issue of preservation in a separate manner, a shared implementation of common e-Infrastructure layers could be beneficial and cost effective.

Moreover, preservation models are often inspired by the ISO OAIS standard, where transfers and preservation are built on information packages containing both data and metadata. Even if the transferred files are in standard formats, the implementation of standards cannot be guaranteed and it is not in control neither by the institutions that produces the software for implementing them, nor by the memory institutions.

E-Infrastructures and DCH communities entered a dialogue in the last years and several data-infrastructure projects exist and look how to set up data infrastructures, including DCH use cases:

- DCH-RP: Digital Cultural Heritage Roadmap for Preservation
- SCIDIPES: SCience Data Infrastructure for Preservation - Earth Science
- APARSEN: Alliance Permanent Access to the Record of Science in Europe network
- EUDAT: Towards a European Collaborative Data Infrastructure
- CHAIN REDS: Coordination and Harmonization of Advanced e-Infrastructures for Research and Education Data Sharing
- DARIAH: Digital Research Infrastructure for Arts and Humanities
- DASISH: Data Service Infrastructure for the Social Science and Humanities
- CLARIN: Common Language Resources and Technology Infrastructure
- SCAPE: SCALable Preservation Environments

At the same time, new projects are about to start, such as a joint Pre-Commercial Procurement project – which is now under negotiation – whose main objective is the development of an open source software licensed reference implementation for different format standards as a tool to be used by memory institutions to check conformance with standard specifications.

Aim of the workshop is to bring together such kind of projects and initiatives working world-wide in the domain of DCH, e-infrastructures and digital preservation to share and present the advancements in the state of the art, find synergies and discuss opportunities for cooperation, starting from concrete use cases.

Target Users:

- Researchers in the humanities
- Teaching and learning actors (schools, training centers, university courses)
- Cultural and creative industry for the creative use and re-use of the digital cultural content
- Content providers (e.g. cultural managers of national institutions and libraries, small institutions, private and public publishers, etc.).
- Policy makers and programme owners
- E-infrastructure providers, technology providers and R&D institutions
- R&D projects and initiatives focusing on digital preservation

Workshop Agenda:

14:30 - 14:45 Welcome and introduction (Antonella Fresa, Promoter Srl)

First Part: DCH and the e-infrastructures

14:45 - 15:10: Using EUDAT services to replicate, store, share, and find cultural heritage data in Poznań? Supercomputing and Networking Center (Maciej Brzeźniak, Poznan Supercomputing and Networking Center – Damien Lecarpentier, CSC – IT Center for Science)

15:10 - 15:35 Authentication and Authorisation in the Cultural Heritage community (Roberto Barbera, Istituto Nazionale di Fisica Nucleare)

15:35 - 16:00 Scalability in preservation of cultural heritage data (Simon Lambert, Scientific Computing Department – STFC)

16:00 - 16:30 Break

Second Part: OAIS model, standards, provenance and authenticity

16:30 - 16:50 Standard models and formats for digital preservation (Börje Justrell, Swedish National Archives)

16:50 - 17:15 Implementation of authenticity evidence record model for supporting preservation scenarios (Luigi Briguglio, Engineering R&D Lab)

17:15 - 17:30 Coordination of digitisation, digital access and digital preservation in Sweden (Sanja Halling, Digisam)

17:30 Conclusions (Antonella Fresa, Promoter Srl)

Participation to this workshop is free but registration is required. To participate **only** to this workshop please register at <http://www.digitalmeetsculture.net/article/digital-preservation-of-cultural-data/>. If you wish to participate to the EUDAT Conference as well you should register to the full Conference (http://www.alfafcm.com/ita/eudat_registration_form) and choose the option of the “Digital Preservation of cultural data” workshop on October 30th in the registration form. If you already registered to the EUDAT Conference and now you want to add your participation to the workshop, please use the form available at <http://www.digitalmeetsculture.net/article/digital-preservation-of-cultural-data/>.

For more information please visit the workshop web page at <http://www.digitalmeetsculture.net/article/digital-preservation-of-cultural-data/> or contact Claudio Prandoni, Promoter Srl, prandoni@promoter.it



[Read more](#)