



Parallel Track IV - New Services Overview

PARALLEL TRACK 4: NEW SERVICES OVERVIEW

DOWNLOAD THE DETAILED POST [CONFERENCE REPORT ON THE NEW SERVICES TRACK](#)

During the preparation phase of the EUDAT proposal a number of scientific communities were engaged with and were involved in discussing the first EUDAT core service set. From this the well-known short list of 4 core services was extracted, the requirements were defined and basically much focus was put to get these 4 services running. Due to the ongoing interactions it became obvious that one of these services has 4 different flavors for different types of data providers. In parallel EUDAT defined two more services which are currently being worked on also a result of the EUDAT interactions (see the note on community engagement on the web: <http://www.eudat.eu/published-articles>).

Now it is time to begin further intensive discussions about what the candidates for future common services within a Collaborative Data Infrastructure could be. Already back in March 2013 at the EUDAT user forum in London some candidates were proposed. Currently EUDAT has launched a questionnaire covering many communities and individuals to understand their preferences and ideas. Furthermore, for some areas (dynamic/realtime data, workflow support, semantic services, access issues) EUDAT is organising special workshops in Mid-September to understand better what services could be useful as solutions for these areas. The EUDAT 2nd conference will be a further step, the presentation of an evaluation of all suggestions received to date and an opportunity to openly discuss the outcome with all community experts and individuals that are interested. This will help EUDAT to find proper service cases for the coming phases that are based on the opinions of multiple communities together with the 5 core ones.

SESSION 4.1 - SEMANTIC ANNOTATION

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

CHAIR: HERBERT SCHENTZ, ECOSYSTEM RESEARCH & MONITORING, UMWELTBUNDESAMT GMBH

Semantic services are of common interest, although it is not a straightforward process to identify exactly which services in the semantic domain could be seen as common services shared by some or many communities. Currently EUDAT is discussing one concrete service which we call "Semantic Annotation". Semantic annotation can be applied to derived and typical long tail data, rather than to regular raw data created by machines. A typical example of the use of semantic annotation is a scenario where data is produced by humans and will therefore contain errors. Consequently scientists will want to annotate the errors and create references to accepted ontologies. This paradigm is becoming important to an increasing number of disciplines. At a certain level of abstraction, semantic annotation can be seen as a common service that can be applied to processes of data enrichment in many scientific disciplines. Such an annotation module could be used as plug-in for EUDAT core services, and also as a plug-in for community services. EUDAT will approach this work on two strands: on the one hand we want to implement a semantic annotation service as soon as possible, and, on the other hand, we want to start an elaborative discussion on other possible common semantic services.

AGENDA:

Morris Riedel: General Overview New Services

Former suggestions, current survey, etc.

First short Q+A

Herbert Schentz: Workshop Results

Start of EUON - European Ontology Network

Michael Mirtl, Environment Agency Austria & **David Vicente** Barcelona Supercomputing Center

Semantic Annotation Work in EUDAT

Other contributions and General discussion on Semantics in EUDAT

SESSION TITLE 4.2 - DYNAMIC AND REAL-TIME DATA

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

CHAIR: ALBERTO MICHELINI, DIRECTOR OF THE NATIONAL EARTHQUAKE CENTER, ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA (INGV)

Some dynamic data is generated by sensors which produce data streams that may be temporarily incomplete (owing to latencies or temporary interruptions of the transmission lines between the field sensors and the data acquisition centres) and that may consequently fill up over time (automatically or after manual intervention). Dynamic data can also be generated by massive crowd sourcing where, for example, experimental collections of data can be filled up at random moments. The nature of dynamic data makes it difficult to handle for various reasons: a) establishing valid policies that guide early replication for data preservation and access optimization is not trivial, b) identifying versions of such data - thus making it possible to check their integrity - and referencing the versions is also a challenging task, and c) performance issues are extremely important since all these activities must be performed fast enough to keep up with the incoming data stream. There is no doubt that both applications areas (namely data from sensors and crowdsourcing) are growing in their relevance for science, and that appropriate infrastructure support (by initiatives such as EUDAT) is vital to handle these challenges.

AGENDA:

• **Alberto Michelini:** Dynamic Data Workshop Results

• **Herman Stehouwer:** Crowd Sourcing Use case

• **Alberto Michelini:** Sensor Data Use Case

• Other contributions and General discussion on Dynamic Data in EUDAT

• **Alberto Michelini:** Update New Services

SESSION TITLE: 4.3 - WORKFLOWS

DATE & TIME: WEDNESDAY 30TH OCTOBER 2013, 09:00 - 10:30

CHAIR: CHRISTIAN PAGÉ, CERFACS

Well-described and documented scientific workflows that can be executed to achieve new results are becoming more and more important in all scientific disciplines to cope with the increasing amount of data in appropriate ways and to increase the reproducibility of scientific results. This is true both for raw data generated by sensors and software systems and processed in regular ways, and also for many areas of derived data - the long-tail data. As we move towards "data fabric solutions", workflow support for manipulating data will be essential.

Data infrastructure initiatives such as EUDAT and DataONE are already working on workflow

systems and building up expertise, while large institutions such as LANL and SDSC are

also looking into such workflow systems to offer services for scientists. It is not yet fully clear which environment will be offered in these cases or

exactly what types of services is data infrastructures should offer. EUDAT will continue to work with community experts to test service concepts that allow users to execute workflows on data stored in the EUDAT data domain.

AGENDA:

• **Christian Pagé:** Workflows

• **Christian Pagé:** Climate Use Case

• **Erhard Hinrichs:** Linguistics Use Case

• Other contributions and General discussion on Workflow Issues in EUDAT

• Update New Services

• Final Round on New Services

Abstract

Attachment	Size
 EUDAT_2nd_conference_Track4.1.pdf	435.96 KB
 EUDAT_2nd_conference_Track4.2.pdf	401.23 KB
 EUDAT_2nd_conference_Track4.3.pdf	423.8 KB
 EUDAT_Second_Conference_New_Services_Summary.pdf	858.26 KB

[Read more](#)