

Inaugural EUDAT Workshop on Big Data Services



Faced with an ever-increasing deluge of big data, adequate web services are a key prerequisite for ubiquitous, flexible and fast data access. Now, several large European initiatives have teamed up in a concerted effort to address the service challenge. On 12 and 13 November 2015, the inaugural EUDAT Workshop on Services for Big Data was held successfully in Barcelona, Spain. Representatives of three decisive big data projects – European Data Infrastructure (EUDAT), EarthServer and EPOS – came together to discuss innovative alternatives for value-adding services.

To consolidate activities around these specific themes the workshop was divided into several tracks focusing on the topics of big data semantics, federated data mining and multi-dimensional array databases for large time series. Discussions started by capturing

best practices and discussing the current state of development and activities in the respective areas. Questions like ‘How can data processing be orchestrated optimally?’ and ‘How can scientific workflows make use of EUDAT services?’ were discussed intensively in various working groups.

Peter Wittenburg, scientific coordinator of the EUDAT Data Infrastructure, convened a critical variety of expertise from Europe and the USA. The experts especially focused on the topic of multidimensional arrays because of the major role in scientific and engineering data. In a summary Mark van de Sanden, EUDAT workpackage leader, and Peter Baumann, workshop facilitator of the EUDAT ‘Array Database’ track, pointed out possible roles of EUDAT in the future:

- IaaS service provider: providing a cloud infrastructure to run array databases
- SaaS service provider: providing an array database as an domain-independent, horizontal service
- Providing tools for easy data movement between EUDAT DCI domain and user domain
- Providing domain services (e.g., geo, astro, life sciences) based on a common horizontal platform of array services, thereby leveraging cross-community effects.

Peter Baumann resumed his experiences of running large-scale infrastructures in his presentation, saying: “Of course multidimensional arrays do not stand alone, they are intertwined with other data types, but typically they constitute the big data part. Therefore, it makes sense to integrate arrays into common data management platforms.”

Critical advantages

The flexibility of querying data, achieving data independency, scalability and standards conformance are critical advantages of array database technologies. Among the challenges spotted were integration of heterogeneous data types, including arrays, into a single common information space for users. Array intensive domains like the Earth, Space and Life Sciences were considered as possible candidates of future EUDAT services.

The following presenters contributed their expertise to the ‘Array Database’ track:

- Peter Baumann (Workshop Facilitator, Array Database expert) - Jacobs University Bremen, Germany
- Kwo-Sen Kuo (Array Database expert) - NASA collaborator, US
- Stefan Pröll (Data Citation expert) - SBA Research, Austria
- Simone Mantovani (Atmospheric Analysis expert) - MEEO s.r.l., Italia
- Alessandro Spinuso (Seismology expert) - KNMI, Netherlands
- Luca Trani (Seismology expert) - KNMI, Netherlands
- Thomas Zastrow (expert for Data Analysis in the Humanities) - Max Planck Gesellschaft, Rechenzentrum Garching, Germany
- Mark van de Sanden (EUDAT Workpackage Leader) - SURFsara, Netherlands

The European Data Infrastructure (EUDAT) aims to contribute to the production of a Collaborative Data Infrastructure (CDI). The project’s target is to provide a pan-European solution to the challenge of data proliferation in Europe’s scientific and research communities. Increasing complexity and massive growth of data has outpaced the development of tools to deal with it.

Paradigm shift

Corresponding to this challenge the intercontinental initiative EarthServer aims for unleashing the potential of big data through a disruptive paradigm shift in service technology. EarthServer has established open ad-hoc analytics on massive Earth science data, based on and

extending leading-edge array database technology, rasdaman. Now the participating data centres are extending this to a Petabyte of 3D and 4D data cubes. Technology advances will allow real-time scaling of such Petabyte cubes and intercontinental fusion.

The European Plate Observing System (EPOS) contributes by planning a research infrastructure for European Solid Earth science, integrating existing research infrastructures to enable innovative multidisciplinary research, as recently prioritised by the European Strategy Forum on Research Infrastructures ESFRI for implementation.

<https://www.gim-international.com/content/news/inaugural-eudat-workshop-on-big-data-services>
