Contribution ID: **102**   Type: **Presentation**

A single data ecosystem as framework for a joint infrastructure data world

Tuesday, 9 October 2018 16:15 (15 minutes)

There exists a need for better solutions for cross-domain data sharing and research collaboration, especially the need to process a multitude of different data types created in different contexts. To address this challenge the Research Data Alliance (RDE) together with GEDE (Group of European Data Experts) have recently suggested a Data Object (DO) architecture to create a network of DOs linked to persistent identifiers (PID) pointing to associated metadata descriptions connecting multitudes of data repositories. This DO architecture will enable a global data environment that may cause a fundamental changes in data practices for more efficient data processing, data sharing and simpler data process automation, and will certainly effect the area of infrastructure service provision.

Inspired by biology, the concept of the ecosystem provides a framework to help to comprehend the inter-wined and highly interdependent nature of increasingly complex data infrastructures involved in cross-domain research and Open Science. To maximize the impact of DO architecture on the data environment, we need to overcome the fragmentation in data ecosystem concepts to work with only a single data ecosystem for the emerging global data infrastructure. Though, different kinds of data ecosystems have already been described, like ecosystems for big data, climate data, biomedical data, open data and personal data, we propose to apply the concept of a single data ecosystem for the whole data world covering all different data types used in research with the DO architecture as its main structural component.

The advantage is that such a concept of a single, global data ecosystem may allow unique access to novel ideas and analysis methods that may be useful to further the development and evolution of European research infrastructures and e-infrastructures and improve their operations in the global context. Because research questions have become bigger and more complex the need for cross-domain data sharing has to be addressed. Seeing each infrastructure as a separate ecosystem is too restrictive; one should use a single, extended data ecosystem that covers open and protected, big data and micro data, as well as data from all different research domains. Ecosystems are no fixed structures, but are changing and evolving, going through cycles of growth and reorganization. In this way, they can represent a framework to generate ideas to be applied to data ecosystems and their infrastructure components to study their functioning, further development and sustainability. To support the analysis of the data ecosystem we adapt concepts from ecological analysis, like using as basic components data generators, data consumers, data flows, data re-users, and feedback loops.

In summary, the consequent exploitation of the concept of a single data ecosystem may produce novel solutions for the depicted limitations in existing data sharing processes, because the interoperation of research infrastructures and e-infrastructures is a prerequisite for streamlining cross-disciplinary processes and global cooperation that will become easier and more transparent by the proper application of the DO architecture as part of a global data infrastructure.

Type of abstract

Presentation

Summary

The concept of the data ecosystem is not exploited adequately to maximize the impact of a comprehensive application of the Data Object (DO) architecture on the emerging global data environment. Applying a single, global data ecosystem covering open and protected data, big data and micro data, as well as data from all different research domains, may allow access to novel ideas and analysis methods and thus produce novel solutions for the existing limitations in cross-domain data sharing. Cross-disciplinary processes will become easier by the proper application of the DO architecture as part of a global data infrastructure.

Primary author: Dr KUCHINKE, Wolfgang (Heinrich-Heine University Duesseldorf)

Presenter: Dr KUCHINKE, Wolfgang (Heinrich-Heine University Duesseldorf)

Session Classification: Data Management Services

Track Classification: Area 1. Cross-Domain challenges / Data exchange across domains: researchers, technologist and policy makers perspectives