Geophysical Research Abstracts
Vol. 18, EGU2016-12621, 2016
EGU General Assembly 2016

An Open and Holistic Approachfor Geo and Space Sciences

Bernd Ritschel (1), Christoph Seelus (1), Günther Neher (2), Iyemori Toshihiko (3), Akiyo Yatagai (4), Yukinobu Koyama (5), Yasuhiro Murayama (6), Todd King (7), Steve Hughes (8), Shing Fung (9), Ivan Galkin (10), Mike Hapgood (11), and Anna Belehaki (12)

(1) Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, Potsdam, Germany, (2) University of Applied Sciences FHP Potsdam, Potsdam, Germany, (3) Kyoto University, Kyoto, Japan, (4) Tokyo University, Tokyo, Japan, (5) NII, Tokyo, Japan, (6) NICT, Tokyo, Japan, (7) UCLA, Los Angeles, USA, (8) JPL, Pasadena, USA, (9) NASA, Goddard, USA, (10) UML, Lowell, USA, (11) RAL, Oxfordshire, UK, (12) NOA, Athens, Greece

Geo and space sciences thus far have been very successful, even often an open, cross-domain and holistic approach did not play an essential role. But this situation is changing rapidly. The research focus is shifting into more complex, non-linear and multi-domain specified phenomena, such as e.g. climate change or space environment. This kind of phenomena only can be understood step by step using the holistic idea. So, what is necessary for a successful cross-domain and holistic approach in geo and space sciences? Research and science in general become more and more dependent from a rich fundus of multi-domain data sources, related context information and the use of highly advanced technologies in data processing. Such buzzword phrases as Big Data and Deep Learning are reflecting this development. Big Data also addresses the real exponential growing of data and information produced by measurements or simulations. Deep Learning technology may help to detect new patterns and relationships in data describing high sophisticated natural phenomena. And further on, we should not forget science and humanities are only two sides of the same medal in the continuing human process of knowledge discovery.

The concept of Open Data or in particular the open access to scientific data is addressing the free and open availability of -at least publicly founded and generated- data. The open availability of data covers the free use, reuse and redistribution of data which have been established with the formation of World Data Centers already more than 50 years ago. So, we should not forget, the foundation for open data is the responsibility of the individual scientist up until the big science institutions and organizations for a sustainable management of data. Other challenges are discovering and collecting the appropriate data, and preferably all of them or at least the majority of the right data. Therefore a network of individual or even better institutional catalog-based and at least domain-specific data servers is necessary. In times of the WWW or nowadays Semantic Web, context enriched and mashed-up open data catalogs pointing to the appropriate data sources, step-by-step will help to overcome the burden of the users to find the right data. Further on, the Semantic Web provides an interoperable and universal format for data and metadata. The Resource Description Formation (RDF) inherently enables a domain and cross-domain mashup of data, e.g. realized in the Linked Open Data project.

Scientific work and appropriate papers in the geo and space domain often are based on data, physical models and previous publications, which again have been dependent on data, models and publications. So, in order to guarantee a high quality of scientific work, the complete verification process of the results is necessary. This is nothing new, but in times of Big Data a real challenge. So, what do we need for a complete verification of presented results? Yes, especially we need all the original data which has been used. But it is also necessary to get complete information about the context of the research objectives and the resulting constraints in the preparation of the raw data. Further on we need knowledge about the methods and the appropriate processing software, which has been used to generate the results. The Open Data approach enriched by the Open Archive idea is providing the concept for sustainable and verifiable scientific work. Open Archive of course stands for the free availability of scientific papers. But furthermore it focuses on mechanisms and methods within the realm of scientific publications for referencing and providing the underlying data, methods and software. Such reference mechanism are the use of Digital Object Identifier (DOI) or Uniform Resource Identifier (URI) within the Semantic Web -in our case for geo and space science data- but also methods and software code. Nowadays, more and more open and private publishers are demanding such kind of references in preparation of the publishing process. In addition, references to well documented earth and space science data are available via an increasing amount of data publications. This approach serves both, the institutional geo and space data centers which increase their awareness and importance, but also the scientists, which will find the right and already DOI-referenced data in the appropriate data journals.

The Open Data and Open Archive approach finally merges in the concept of Open Science. Open Science emphasizes an open sharing of knowledge of all kind, based on a transparent multi-disciplinary and cross-domain scientific work. But Open Science is not just an idea, it also stands for a variety of projects which following the rules of Open Science, such as open methodology, open source, open data, open access, open peer review and open educational resources. Open Science also demands a new culture of scientific collaboration based on social media, and the use of shared cloud technology for data storage and computing. But, we should not forget, the WWW is not a one way road. As more data, methods and software for science research become freely available at the Internet, as more chances for a commercial or even destructive use of scientific data are opened. Already now, the giant search engine provider, such as Google or Microsoft and others are collecting, storing and analyzing all data which is available at the net. The usage of Deep Learning for the detection of semantical coherence of data for e.g. the creation of personalized on time and on location predictions using neuronal networks and artificial intelligence methods should not be reserved for them but also used within Open Science for the creation of new scientific knowledge. Open Science does not mean just to dump our scientific data, information and knowledge into the Web. Far from it, we are still responsible for a sustainable handling of our data for the benefit of humankind.

The usage of the principles of Open Science is demonstrated on the scientific and software engineering activities for the mashup of the Japanese IUGONET, European Union ESPAS and GFZ ISDC related data server covering different geo and space science domains.