

[3-1]

OPEN DATA, OPEN PUBLICATION AND OPEN SCIENCE APPROACH FOR GEO AND SPACE SCIENCE DOMAIN

B Ritschel^{1}, T Iyemori², G Neher³, Ch Seelus⁴, Y Koyama⁵, Y Murayama⁶, T King⁷, S. Hughes⁸, S. Fung⁹, M Hapgood¹⁰ and A Belehaki¹¹*

**¹Kyoto University, Data Analysis Center for Geomagnetism/GFZ German Research Centre for Geosciences*

Email: ritschel@kugi.kyoto-u.ac.jp/rit@gfz-potsdam.de

²Kyoto University, Data Analysis Center for Geomagnetism and Space Magnetism, Graduate School of Science

Email: iyemori@kugi.kyoto-u.ac.jp

³University of Applied Sciences, FHP Potsdam, Faculty of Information Sciences

Email: g.neher@fh-potsdam.de

⁴Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, Potsdam, Germany

Email: sceelus@gfz-potsdam.de

⁵National Institute of Informatics, Tokyo, Japan

Email: y-koyama@nii.ac.jp

⁶Institute of Information and Communications Technology, Tokyo, Japan

Email: murayama@nict.go.jp

⁷University of California, Los Angeles, California, United States

Email: tking@igpp.ucla.edu

⁸NASA, Jet Propulsion Laboratory, Pasadena, California, United States

Email: john.s.hughes@jpl.nasa.gov

⁹NASA Goddard Space Flight Center, Greenbelt, Maryland, United States

Email: shing.f.fung@nasa.gov

¹⁰STFC Rutherford Appleton Laboratory, Didcot, England, Great Britain

Email: mike.hapgood@stfc.ac.uk

¹¹National Observatory of Athens, Athens, Greece

Email: abeleha@otenet.gr

Geo and Space sciences thus far have been very successful even often an open and cross-domain 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 climate change or space environment, which only can be understood step by step using a holistic concept. So, what is necessary for a successful cross-domain and holistic approach in geo and space sciences? Research and science in general becomes more and more dependent from a fundus of multi-domain data sources. The buzzword phrase Big Data is reflecting this development. But Big Data also addresses the real exponential growing of data and information produced by measurements or simulations. 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 basis 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 finding or 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 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 these obstacles. Further on, the Semantic Web provides an interoperable and universal format for data and metadata, the Resource Description Formation (RDF), which inherently enables a domain and cross-domain mashup, 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 not a new discovery, 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 have been used to generate the results. The Open Data approach enriched by the Open Publication idea is providing the concept for sustainable and verifiable scientific work. Open Publication 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 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 science data centers and the scientists which are looking for data. The data centers increase their awareness and importance, whereas the scientists, will find the right and already DOI-referenced data in the appropriate data journals.

The Open Data and Open Publication approach finally opens out in the concept of Open Science. Open Science emphasizes an open sharing of knowledge of all kinds, 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 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 (DL) for the detection of semantical coherence of data for e.g. natural language processing or 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.

ACKNOWLEDGEMENTS

Many thanks to the students of the faculty of information science at the University of Applied Sciences Potsdam (FHP) who carried out different domain and terminological ontology mapping and merging experiments during project-oriented classes in the last 3 years. Particular thanks go to Alina Höhler and Marcel Thies -also students of the FHP- who not only installed the D2R framework at the ISDC server but also transferred context data from a Sybase RDMS to MySQL. They also designed PHP code for SPARQL queries and result output in Drupal, the CMS portal for the ISDC Semantic Web data server.

REFERENCES

Conference presentation

Ritschel, B., Seelus, C., Neher, G., Iyemori, T., Koyama, Y., Murayama, Y., King, T., Hughes, J., Fung, S., Galkin, I., Hapgood, M., Belehaki, A., (2015) Open Science via the Mashup of Heterogenous Data Server and Systems in the Geo and Space Science Domain, 12th Annual Meeting Asia Oceania Geosciences Society 2015, Singapore, Singapore.

Ritschel, B., Seelus, C., Neher, G., Iyemori, T., Koyama, Y., Murayama, Y., King, T., Hughes, J., Fung, S., Galkin, I., Hapgood, M., Belehaki, A., (2015) Experiments for the mashup of heterogenous data systems using semantic web concepts, 26th IUGG General Assembly 2015, Praha, Czech Republic.

Ritschel, B., Seelus, C., Neher, G., Iyemori, T., Koyama, Y., Murayama, Y., King, T., Hughes, J., Fung, S., Galkin, I., Hapgood, M., Belehaki, A., (2015) Semantic Web based Mashup of Data Systems for Open Data and Open Science, Japan Geoscience Union Meeting 2015, Tokyo, Japan.

Ritschel, B., Seelus, C., Neher, G., Iyemori, T., Koyama, Y., Murayama, Y., King, T., Hughes, J., Fung, S., Galkin, I., Hapgood, M., Belehaki, A., (2015) Opportunities for the Mashup of Heterogenous Data Server via Semantic Web Technology, European Geosciences Union General Assembly 2015, Vienna, Austria.

Ritschel, B., Seelus, C., Neher, G., Iyemori, T., Koyama, Y., Murayama, Y., King, T., Hughes, J., Fung, S., Galkin, I., Hapgood, M., Belehaki, A., (2015) Mashup of Geo and Space Science Data Provided via Relational Databases in the Semantic Web, American Geophysical Union Fall Meeting 2014, San Francisco, U.S.A.