The NextGEOSS CKAN-based Data Hub: Faster Discovery And Easier Access To Earth Observations Through Smart Use Of Open Web-technology.

Bente Bye (1), Wolfgang Ksoll (2), Nuno Catarino (3), Marie-Francoise Voidrot (4), Erwin Goor (5), Julian Meyer-Arnek (6), Pedro Goncalves (7), and Nuno Grosso (3)

(1) BLB, Hønefoss, Norway (bentelil@hotmail.com), (2) Viderum, Berlin, Germany (ksoll@viderum.com), (3) DEIMOS, Lisbon, Portugal (nuno.catarino@deimos.com.pt), (4) OGC, Brussels, Belgium, (5) VITO, Mol, Belgium, (6) DLR, Berlin, Germany, (7) Terradue, Rome, Italy

With data from COPERNICUS and its Sentinel missions as well as from in situ sources both provided by commercial and public sectors, a tremendous amount of Earth observation data is available today. To find the right data for government, businesses and civil society, a CKAN based data hub is set up to connect data sources with post-processing facilities (e.g. numerical intensive time series computed in cloud operations) and/or end consumer. Several pilot projects addressing different topics (e.g. agriculture, smart cities, energy, etc.) are used for proof of concept and to demonstrate ways to improve benefits from the already existing data. Individually and in sum they support the UN sustainability goals (SDG) improving our lives here on Earth.

To reach these goals openness and standards are necessary attributes of the strategy. In Europe it begins with open data provided by the nations and the European Union. Through our engagement in GEO and contributions to GEOSS they are aligned globally. The software CKAN (Comprehensive Knowledge Archive Network) is globally spread and open source maintained by Open Knowledge International. CKAN allows harvesting of the metadata from sources with different communication protocols and different metadata formats (e.g. Dublin Corr or with extensions DCAT or GeoDCAT). By harvesting the metadata the terrabytes of data can remain stored at the data provider or owner while only the description of the data (contributor, license, spatial information and so on) is collected. Once harvested, the metadata can be searched by users in several ways. A GUI (Graphical User Interface) can be used for access with a normal browser. Several interfaces are provided for API-access. Currently there is an OpenSearch interface available and for semantic web access RDF-access following the DCAT-standards will be made available later. In addition, this NextGEOSS data hub have a user feedback mechanism integrated which proves the CKAN's ability to use external programs for internal purposes. We will present the experiences from developing a beta version of a European data hub and cloud platform emphasizing how open source software support discoverability and access to open data and how user communities benefit especially when this data hub is seamlessly integrated with value adding software for number crunching in clouds. These results are important contributions to increase value and expand the use of existing (open) data beneficial to both science, businesses and society at large.