



## GEOSS Clearinghouse Quality Metadata Analysis

J. Masó (1), P. Díaz (1), M. Ninyerola (2), E. Sevillano (2), and X. Pons (2)

(1) Centre for Ecological Research and Forestry Applications, Spain, (2) Universitat Autònoma de Barcelona, Spain

The proliferation of similar Earth observation digital data products increases the relevance of data quality information of those datasets. GEOSS is investing important efforts in promoting the acknowledgment of the data quality in Earth observation. Activities, such as the regular meeting of QA4EO and projects as GeoViQua have the aim to make the data quality available and visible in the GEOSS Common Infrastructure (GCI). The clearinghouse is one of the main components of the GCI, which catalogues all the known Earth observation resources and provide it via the GEO Portal. Actually, after several initiatives to stimulate that (such as AIP4) most of the relevant international data providers referenced their data in the GEOSS Component and Service Registry, therefore, the GEOSS clearinghouse can be considered a global catalogue of the main Earth observation products. However, there are some important catalogues still in the process of being integrated.

We developed an exhaustive study of the data quality elements available on the metadata catalogue in the GEOSS clearinghouse, to elaborate a state-of-the-art report on data quality. The clearinghouse is harvested using the OGC CSW port. Metadata following the standard ISO 19115 is saved in XML-ISO 19139 files. The semi-automatic methodology, previously applied in regional SDIs studies, generates a big metadata database that can be further analyzed. The number of metadata records harvested was 97203 (October 2011).

The two main metadata nodes studied are directly related with data quality information package (DQ\_DataQuality) in ISO. These are the quality indicators (DQ\_Element) and the lineage information (LI\_Lineage). Moreover, we also considered the usage information (MD\_Usage).

The results reveal 19107 (19.66%) metadata records containing quality indicators; which include a total of 52187 quality indicators. The results show also a main representation of the positional accuracy, with 37.19% of the total. Nevertheless, one important thing to highlight is the diversity of quality indicators used, beyond the positional accuracy. In that sense, the completeness, the consistency and the temporal accuracy, represent 35.71%, 19.78% and 6.81% respectively on the quality indicators analysed. The indicators can be classified in quality measures, 25944 indicators (49.71%) have an associated measure. The measures are quantitative (22275 - 85.86%), conformance with some specifications, in our case mainly to INSPIRE (3669 - 14.14%) and coverage result, according to ISO19115-2 extension (5 - 0.02%), i.e. the quality is represented in a additional raster file, unfortunately, the link to the file is missing.

The lineage is composed by production processes (LI\_ProcessStep) and sources used in elaborating a dataset (LI\_Source), this information can be combined in several ways. (i) Direct list of the data sources (LI\_Lineage: LI\_Source); there are 3771 metadata records (3.88%) of this group. (ii) Direct list of the processes (LI\_Lineage: LI\_ProcessStep) involved in data production; there are 9261 metadata records (9.53%) of this group. (iii) Description of more complete provenance, consisting in describing the processes and the sources involved on each process (LI\_Lineage: LI\_ProcessStep with LI\_Source); there are 1226 metadata records (1.26%) of this group.

The MD\_Usage is an entity part of the identification information package, 1133 records (1.17%) contain usage information, but, only the mandatory specificUsage and userContactInfo elements are described.

These results demonstrate that the documentation of quality indicators and lineage is far from general in the Earth observation data but current status is enough to start developing tools for exposing and exploiting the quality data that is already present in catalogues. Also, usage information has to be extended and its use generalized.