Towards a preservation framework for Spatial Data Infrastructures: a metadata profile approach

Arif Shaon and Andrew Woolf

Science and Technology Facilities Council, UK, (arif.shaon@stfc.ac.uk)

With growing concerns about environmental problems, and an exponential increase in computing capabilities over the last decade, the geospatial community has been producing increasingly voluminous and diverse datasets. For example, some of the largest geospatial datasets, Earth observation (EO) data generated by Earth satellite sensor systems, grow at rates of up to terabytes per day. This phenomenal growth of geospatial data is expected to inflate even further over the next decade, with, for example, the large number of new EO missions upcoming in the next years. Interestingly, there is also evidence of a rise in demand for historical geospatial data, which is being seen as increasing in value, particularly for monitoring and analyzing social, environmental (e.g. global climate change) and economic changes that occur over time. However, without effective long-term preservation, both the current and historical geospatial data face the risk of being unusable over time. This may negatively impact the way complex and critical environment related issues, such as global climate change are dealt with in the future, as this crucially depends on accurate analysis of past related environmental data.

In general, geospatial data inherit the preservation challenges inherent to all digital information. These challenges are further complicated by some of the unique characteristics of geospatial datasets, such as diverse data formats, highly structured and the need for special knowledge for accurate interpretation. Moreover, the emergence of geospatial web services (e.g. Open Geospatial Consortium (OGC) standardised Web services) further add to these challenges, as static data are replaced by dynamic and changing services. For instance, state-of-the-art service-oriented infrastructures adopt exchange formats that reflect conceptual data models ('feature types') rather than the underlying database-specific storage schemas. On the positive side, it should be possible in principle, to apply existing widely adopted preservation mechanisms and standards, such as the Open Archival Information System (OAIS) reference model (a wide adopted ISO standard for digital preservation) to the long-term preservation of geospatial data. And Web services technologies enable harmonized and interoperable accessibility of data, which may be useful for developing effective long-term preservation approaches for geospatial data based on the existing approaches.

This talk discusses the applicability of the OAIS reference model to develop a preservation-aware Spatial Data Infrastructure (SDI) and presents a preservation-profile of ISO 19115 metadata profile for supporting such a SDI. Notably, the work presented may be seen as a preliminary effort for the European Space Agency (ESA) Long-term Digital Preservation (LTDP) initiative that aims to formulate a coordinated and coherent approach to the long-term preservation of the EO space data archives across different member states of the European Union.