Development of Extended Content Standards for Biodiversity Data

Wim Hugo (1), Jochen Schmidt (2), and Hannu Saarenmaa (3)

(1) South African Environmental Observation Network, (2) National Institute for Water and Atmospheric Research, New Zealand, (3) University of Eastern Finland

Interoperability in the field of Biodiversity observation has been strongly driven by the development of a number of global initiatives (GEO, GBIF, OGC, TDWG, GenBank, . . .) and its supporting standards (OGC-WxS, OGC-SOS, Darwin Core (DwC), NetCDF, . . .). To a large extent, these initiatives have focused on discoverability and standardization of syntactic and schematic interoperability. Semantic interoperability is more complex, requiring development of domain-dependent conceptual data models, and extension of these models with appropriate ontologies (typically manifested as controlled vocabularies).

Biodiversity content has been standardized partly, for example through Darwin Core for occurrence data and associated taxonomy, and through Genbank for genetic data, but other contexts of biodiversity observation have lagged behind – making it difficult to achieve semantic interoperability between distributed data sources. With this in mind, WG8 of GEO BON (charged with data and systems interoperability) has started a work programme to address a number of concerns, one of which is the gap in content standards required to make Biodiversity data truly interoperable.

The paper reports on the framework developed by WG8 for the classification of Biodiversity observation data into 'families' of use cases and its supporting data schema, where gaps, if any, in the availability if content standards have been identified, and how these are to be addressed by way of an abstract data model and the development of associated content standards. It is proposed that a minimum set of standards (1) will be required to address the scope of Biodiversity content, aligned with levels and dimensions of observation, and based on the 'Essential Biodiversity Variables' (2) being developed by GEO BON .

The content standards are envisaged as loosely separated from the syntactic and schematic standards used for the base data exchange: typically, services would offer an existing data standard (DwC, WFS, SOS, NetCDF), with a use-case dependent 'payload' embedded into the data stream. This enables the re-use of the abstract schema, and sometimes the implementation specification (for example XML, JSON, or NetCDF conventions) across services. An explicit aim will be to make the XML implementation specification re-usable as a DwC and a GML (SOS end WFS) extension.

(1) Olga Lyashevska, Keith D. Farnsworth, How many dimensions of biodiversity do we need?, Ecological Indicators, Volume 18, July 2012, Pages 485-492, ISSN 1470-160X, 10.1016/j.ecolind.2011.12.016.
(2) GEO BON: Workshop on Essential Biodiversity Variables (27-29 February 2012, Frascati, Italy). (http://www.earthobservations.org/geobon_docs_20120227.shtml)