

IOWMETA - A modular metadata information system enabling the extraction of ISO-compliant metadata from heterogeneous sources

Susanne Jürgensmann, Leibniz Institute for Baltic Sea Research Warnemünde (Germany),
susanne.juergensmann@io-warnemuende.de

Susanne Feistel, Leibniz Institute for Baltic Sea Research Warnemünde (Germany),
susanne.feistel@io-warnemuende.de

Steffen Bock, Leibniz Institute for Baltic Sea Research Warnemünde (Germany), steffen.bock@io-warnemuende.de

The purpose of the metadata information system IOWMETA is to provide a comprehensive catalogue and a central infrastructure hub for all kinds of distributed research data stored at the Leibniz Institute of Baltic Sea Research Warnemünde (IOW). Metadata, describing heterogeneous data, can be standardized and thereby facilitate an international exchange of data and metadata. Comprehensive metadata and thorough documentation ensure long-term usability and maximum potential knowledge gain from already existing research datasets.

IOWMETA is based on the metadata standards of ISO 191xx and incorporates the open-source platform GeoNetwork, offering data access for scientists, federal agencies and the public. A set of more than 40 ISO-elements (including the mandatory ISO 19115 core) are put into practice as an ever-expanding relational database consisting of currently 100 tables. Complementary a software interface is in continuous development that reflects the underlying database schema in terms of its class structure and comprises basic import and export functionality. These two components build the technical core for subsequent modular extensions.

The interface facilitates the straightforward implementation of modules, mapping metadata from arbitrary data sources to ISO-compliant data structures and subsequently storing the results to IOWMETA. The following three examples serve as an illustration for the highly flexible extension capabilities of this modular approach.

The first module automatically extracts metadata from the oceanographic database of IOW, the so-called IOWDB. The IOWDB constitutes the most important long-term data source at IOW. The relational database contains not only research data but also associated metadata including information about scientists, platforms, methods and parameter descriptions. The scientific relevance of the data in IOWDB and its already well-structured content make it the ideal use case for automatic metadata extraction.

A maximum of descriptive information is extracted and transformed in order to meet the requirements of ISO 191xx. Additionally, the resulting metadata is enriched by a link back to the corresponding dataset stored in IOWDB and therefore provides a direct download option. This way, several ten thousands of datasets with the granularity of one dataset per CTD-cast are automatically described with standardized metadata.

Another module in development integrates metadata from a Zoobenthos database, containing taxonomy, abundances and biomasses. The extracted metadata i.e. referencing systems, data granularity and keywords differ from the previous example. However, the software interface allows an automatic integration of metadata into IOWMETA with little effort.

The third exemplary module uses as its data source a file system, rather than a relational database. A central data storage at IOW contains either raw, validated or processed data from scientific cruises. An XML file for indexing the content is placed in each folder of the file system, describing in a

predetermined manner the data stored in the folder. This XML can automatically be retrieved and parsed. The module in IOWMETA converts the information from the XML to the ISO elements.

Based on the same interface a different kind of extension exports metadata from the IOWMETA database. The metadata are recursively extracted and transformed into an XML structure, compliant with ISO 19139. Furthermore, IOWMETA includes a version-control system, allowing the later reconstruction of previous changes in data and metadata.