Automatically generating ISO 19115-1:2014 metadata from relational databases

Helmholtz-Zentrum Geesthacht
Centre for Materials and Coastal Research

Linda Baldewein, Dietmar Sauer, Ulrike Kleeberg, Housam Dibeh, Lars Möller

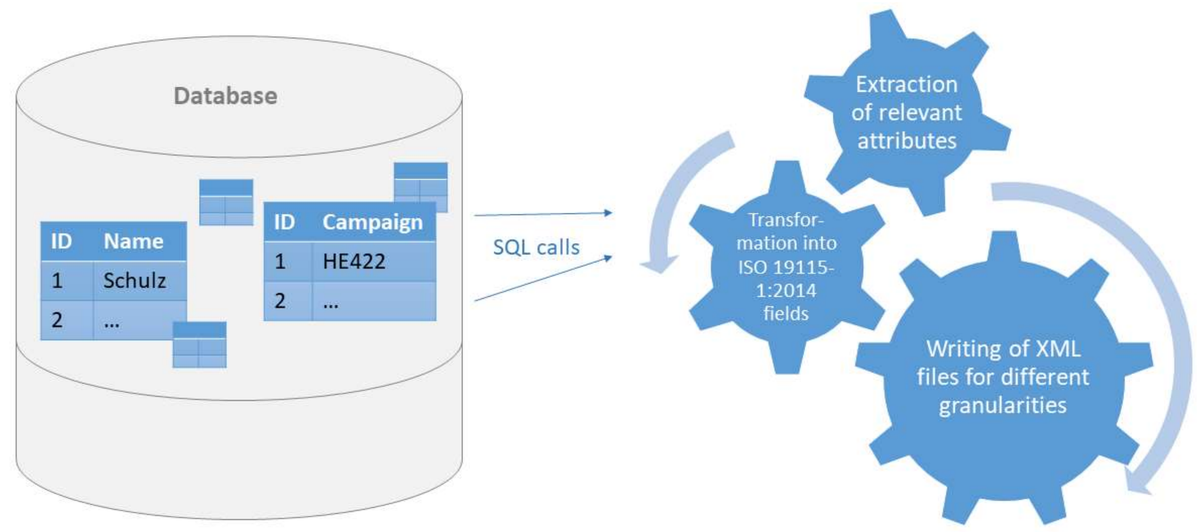
The approach

Generating metadata files from relational databases

The ISO 19115-1:2014 is a standard to describe geographic information in form of metadata. It is used to share and publish metadata within the Earth Sciences. At the Helmholtz Coastal Data Center (HCDC), a method was developed to automatically generate XML metadata files in the ISO 19115-1:2014 standard by extracting the relevant information from metadata stored in relational databases.

The relational databases contain measurement values, of for example biogeochemical samples. In addition, a lot of metadata information is stored across the various database tables, such as contact information, measurement parameters, units and methods and general information on the dataset, such as the campaign. Every table is checked for relevant metadata information. Each of the found attributes will be used to fill an ISO 19115-1:2014 element according to a mapping table.

The information is gathered from the database through SQL calls and aggregated, transformed and mapped to the relevant metadata fields in the ISO 19115-1:2014 standard. For example, contact information is mapped to the field



ETL process: Generating XML files from database tables

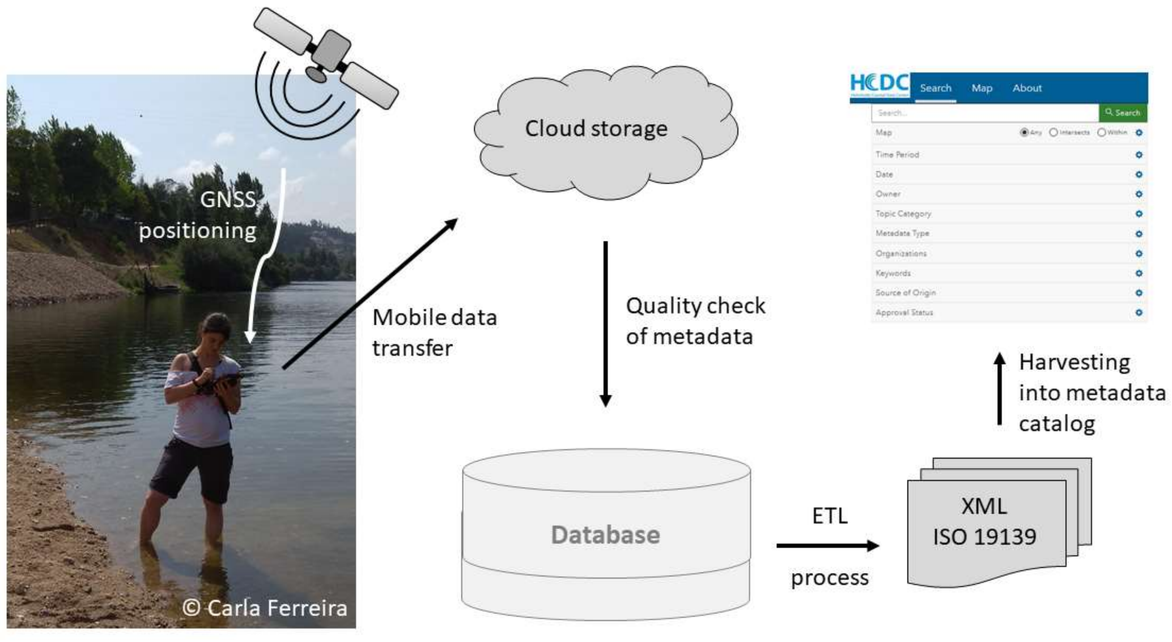
CI_ResponsibleParty and descriptions of parameters as well as standardized parameter names from controlled vocabularies, such as the ones provided by SeaDataNet, are mapped to MD_Keywords. The most evolved part of the process is the automated generation of the abstract of the dataset, which is part of the field CI_Citation. A generic template for all datasets is developed, that summarizes all known information of the campaigns, such as their purpose, and the involved projects. It also includes information on Digital Object Identifiers of the dataset.

Metadata generation

Metadata workflow from the field to the metadata catalog

To generate rich metadata files, the information needs to be available in the relational databases. Since 2018 we have been generating metadata of campaigns already in the field, while the scientists are still on the research vessel. All campaigns have been equipped with high accuracy GNSS receivers and water proof tablets to electronically gather metadata.

Using an app, the scientist fills in a questionnaire asking for example for the name of the station, the time of the sampling and the coordinates. The latter is directly received from the GNSS receiver, which is connected to the app to get the location of each sampling site with approximately one meter accuracy. The results of the questionnaire are then transferred via the mobile network from the tablet to a cloud storage. From there it is imported into the relational database, after carrying out quality checks.



Metadata workflow: From the field to the metadata catalog

The implementation

Helmholtz Coastal Data Center

Once all metadata elements are mapped to the appropriate ISO 19115-1:2014 fields, an ISO 19139 XML file is generated. This process is repeated for different granularities, such as each campaign and each project. The output XML files are stored in a folder, which is regularly harvested by the metadata catalogue system.

Helmholtz Coastal Data Center hcdc.hzg.de

The future

Reusability

The automated process described above has been adapted to different relational databases, storing different data types, for example biogeochemical data and real time observational data. It has proven to be easily adjustable to new situations. After an initial setup phase, it automatically generates ISO 19115-1:2014 conform metadata files.

In the future it might be possible to develop a user friendly generic tool from the described process that generates metadata files from any coastal research database. As the first step, the mapping process from generic database entries to ISO 19115-1:2014 fields can be published.

Helmholtz-Zentrum Geesthacht • Max-Planck-Straße 1 • 21502 Geesthacht / Germany • Phone +49 (0)4152 87-0 • Fax +49 (0)4152 87-1403 • contact@hzg.de • www.hereon.de
Contact: Linda Baldewein • Phone +49 (0)4152 87-1801 • Fax +49 (0)4152 87-2818 • linda.baldewein@hzg.de