INTEROPERABILITY LEVELS & OTHER DATA STANDARDISATION RELATED CONCEPTS

Alessandro Oggioni
WHY INTEROPERABILITY? FOR DATA REUSE

“data reuse” means taking a data asset and using more than once for the same purpose.

You do not want to create / modify repository for each new project.

reduce duplication and enhance collaboration.
INTEROPERABILITY

Ability of a system or a product to work with other systems or products without special effort on the part of the users or customer.

Interoperability is made possible by the implementation of standards.
INTEROPERABILITY: THREE LEVELS

▸ “Syntactic” interoperability

Machine readable aspects of data representation or data formats (e.g. XML, SQL, etc.).

▸ “Structural” interoperability

Defines the syntax of the data exchange. It is obtained through a data model sharing (standard).

▸ “Semantic” interoperability

How to understand the data. Terminological misunderstandings, but also conflicts between domains are resolved, in other words the data are described adequately to be recognized as being appropriate for a certain search purpose.
INTEROPERABILITY: THREE LEVELS

- “Syntactic” interoperability (data formats)
INTEROPERABILITY: THREE LEVELS

- “Structural” interoperability (data structure)

  Data owner
  Date of data collection
  Parameter 1
  Parameter 2
  Parameter 3

  ...
INTEROPERABILITY: THREE LEVELS

▸ “Semantic” interoperability (understand data)

My data are about …

BARK

dog

My data are about …

BARK

tree
INTEROPERABILITY IN OPEN SCIENCE PERSPECTIVE

How knowledge of informatics can be applied to support environmental science to reduce duplication and to enhance collaboration (be interoperable!)?

• use OpenData to embody the principles of being accessible, assessable, intelligible and usable data;

• create unique data encoding to allow analysis across multiple scale and to facilitate the integration of various data sets in a single analytical structure;

• standards and tools are needed to structure data into a linked format by using the potential of vocabularies and ontologies;

Hardisty, A. et al. 2013. BMC ecology
Standards are technical documents that detail interfaces or encodings. Software developers use these documents to build open interfaces and encodings into their products and services.

- INSPIRE (ISO 19115, 19139, etc.)
- Ecological Metadata Language (EMF)
- ABCD
- Darwin Core
- NetCDF
- ...
A **controlled vocabulary** is a list of terms according to a specific order.

A **thesaurus** is a formalized and shared collection of terms of a controlled vocabulary linked by broader / narrower / related.
NERC Vocabulary Server

The NERC Vocabulary Server gives data managers the means to access lists of controlled terms to describe data, thus saving the time and costs associated with unraveling the meaning of a given data set.

- **Introduction** — a brief summary of the value of implementing the Vocabulary Server
- **Connectivity** — consumer access options for the Vocabulary Server
- **Collection, concept and scheme URIs** — how to browse the content of the Server
- **An example of the ReSTful, SOAP API and SPARQL methods** — the GetCollection method
- Access the NERC Vocabulary Server version 2.0 (NVS2.0) [documentation](#)

**Introduction**

The NERC Vocabulary Server provides access to lists of standardised terms that cover a broad spectrum of disciplines of relevance to the oceanographic and wider community.

Using standardised sets of terms (otherwise known as "controlled vocabularies") in metadata and to label data solves the problem of ambiguities associated with data markup and also enables records to be interpreted by computers. This opens up data sets to a whole world of possibilities for computer aided manipulation, distribution and long term reuse.

An example of how computers may benefit from the use of controlled vocabularies is in the summing of values taken from different data sets. For instance, one data set may have a column labelled "Temperature of the water column" and another might have "water temperature" or even "temperature". To the human eye, the similarity is obvious but a computer would not be able to interpret these as the same thing unless all the possible options were hard coded into its software. If data are marked up with the same terms, this problem is resolved.
CNR-IIA-EKOLab Vocabularies web interface

Environmental Applications Reference Thesaurus - EARTH - standard interface - visual interface

Ice and Snow Thesaurus - SnowTerm - standard interface - visual interface

Earth Observation Systems Thesaurus - EOSterm - standard interface - visual interface

Geoterm Thesaurus - GeotermThes - standard interface - visual interface

Natural Hazards and Risks Thesaurus - NHWthes - standard interface - visual interface

Powered by Tematres
INTEROPERABILITY – THESAURUS

Phytoplankton Traits Thesaurus

Trait
  • Phytoplankton Trait
  • Demographic Trait
    • Density
  • Total Biomass
  • Functional Trait
    • Behavioural Trait
INTEROPERABILITY - THESAURUS
INTEROPERABILITY – CONCLUSIONS

Bridging the gap of expectations
INTEROPERABILITY AND FAIR PRINCIPLES

FAIR Principles

Make your data:
- **Findable**
- **Accessible**
- **Interoperable**
- **Reusable**

reduce duplication and enhance collaboration