FOR816-Ecuador Data Warehouse: An EML-based relational project database ...and more

Philipps University Marburg

Thomas Lotz, Maik Dobbermann, Jörg Bendix, Thomas Nauß, Dietrich Göttlicher
Department of Geography, Philipps University Marburg
DFG-RU816: Biodiversity and Sustainable Management of a Megadiverse Mountain Ecosystem in South Ecuador
www.tropicalmountainforest.org

EML - a metadata standard for the ecological science

Maintaining robust metadata is the critical task for scientific databases to support data exchange (make data searchable) and to keep data usable by future scientists. The Ecological Metadata Language (EML)[1] provides a sophisticated XML-based scheme to describe scientific ecological data sets. The FOR816dw has implemented a relational database structure containing all fields to be compliant to the EML standard.

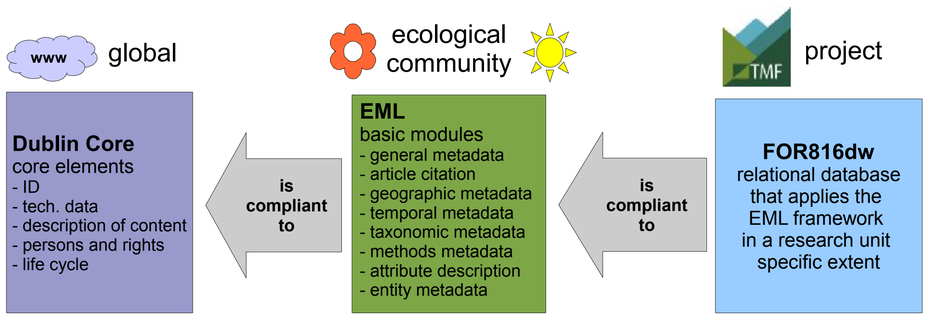


Fig. 1: The FOR816dw provides EML compliant metadata to the stored research data. The data sets are potentially searchable and useable by the further ecological community and are compliant to the general Dublin Core Standard

1) http://knb.ecoinformatics.org/software/eml/

Data in the FOR816 Data Warehouse

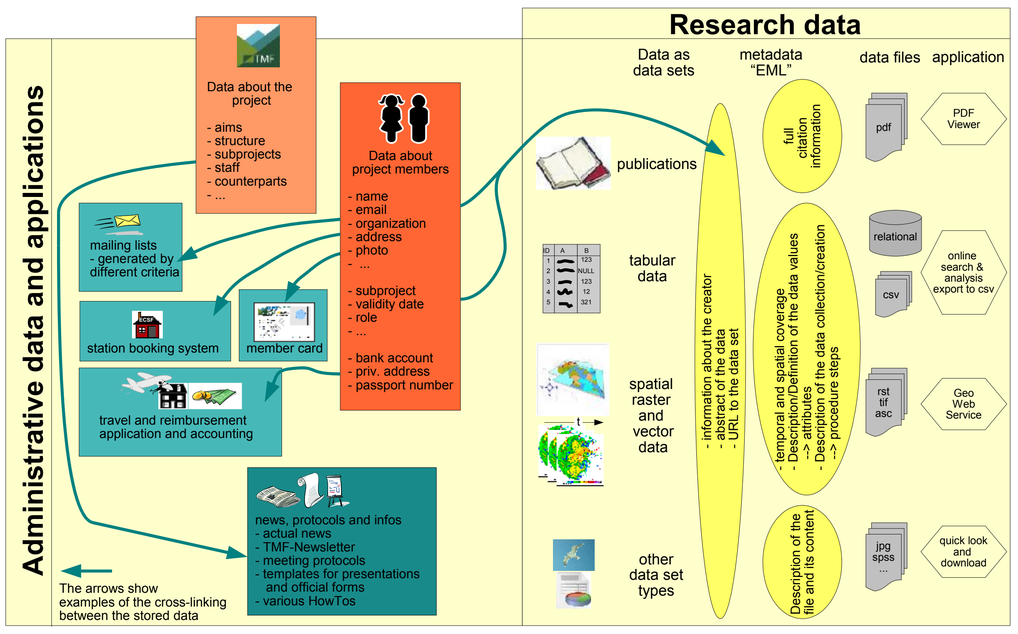
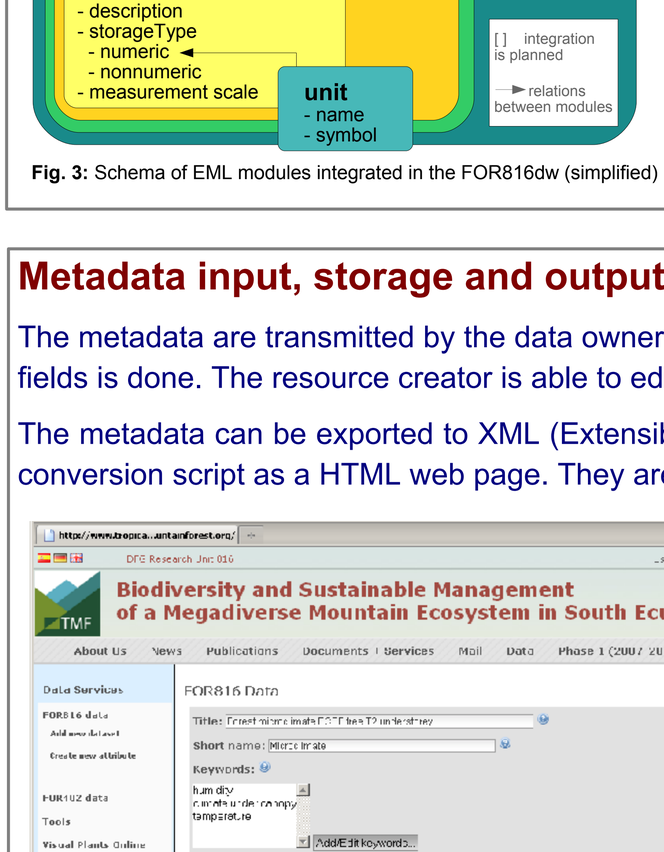


Fig. 2: The integrative handling of administrative and research data of the project within the FOR816 data warehouse.

EML modules used in the FOR816dw

The implementation of the EML framework into the relational structure of the FOR816dw considers the main modules of the specification. All metadata are based on the resource module, which holds the basic information. A resource can be a dataset or literature. The supporting modules help to give detailed description on those top-level resources.



A dataset contains one or more entities which might hold tabular data, raster data or data in an other format. Each entity is described by specific fields but all values of each entity are related to an attribute. The attributes are defined once and can be reused by other datasets. This makes values of different creators (possibly measured by different methods) comparable within the Data Warehouse.

All metadata are included into the extended data search and can be specified by the user to find datasets or literature of interest.

Fig. 3: Schema of EML modules integrated in the FOR816dw (simplified)

Technical implementation

The FOR816dw is based on the Model–View–Controller (MVC) design pattern.

Administrative data and metadata are stored in relational databases and are mapped to the model by Hibernate. The view (Java ServerPages) is dynamically generated by the controller (J2EE) based on user request and model data.

Research data are stored as single values in a database (in case of tabular data) or in the structured file storage.

The user has the possibility to search, download, analyse and visualize the data.

All used software and libraries are open source or free. Programming is object oriented and is extendable for further needs.

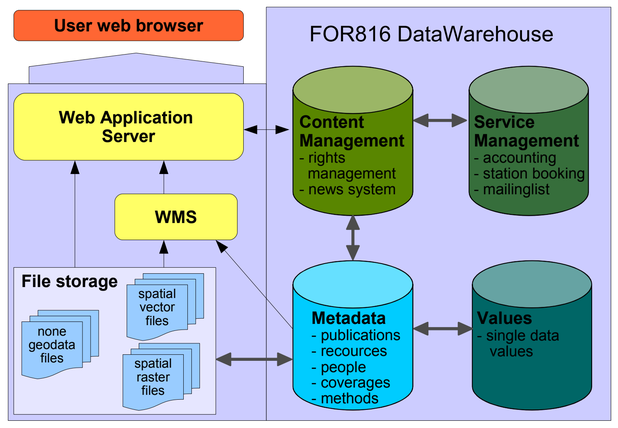


Fig. 4: Schematic system architecture of the FOR816dw

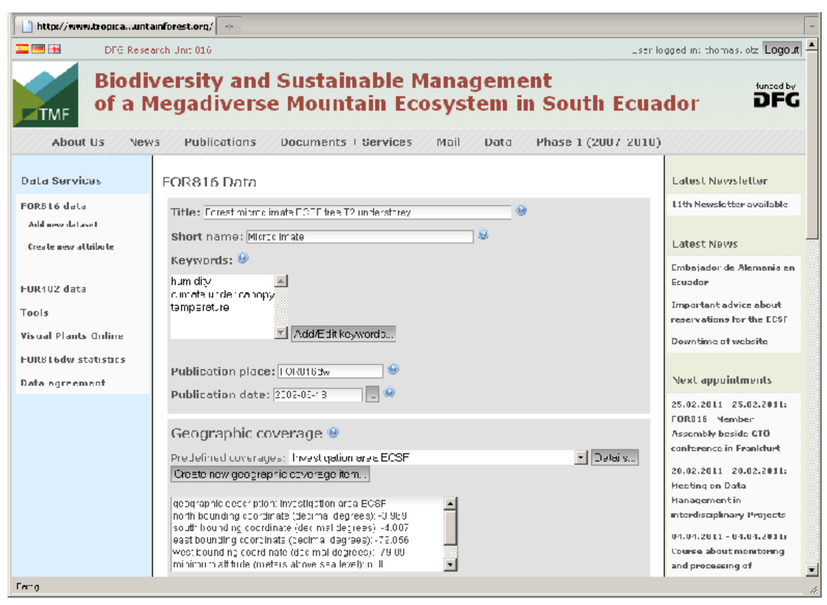
Assignment	Package	Internet information
Webserver	Apache	apache.org
Web application server	Tomcat	tomcat.apache.org
Relational database system	MySQL	mysql.com
Web application framework	Struts, J2EE	struts.apache.org java.sun.com/j2ee
Dynamic web content	Java Server Pages (JSP)	java.sun.com/products/jsp
Scripting language	PHP	php.net
Object-relational mapping	Hibernate	hibernate.org
Full text search engine	Lucene	lucene.apache.org
Metadata scheme	EML	knb.ecoinformatics.org/software/eml
OGC Web Map Service	GeoServer (WMS)	geoserver.org
Map Viewer	OpenLayers	openlayers.org
Data series visualization	Highcharts	highcharts.com

Fig. 5: Overview of applied software packages

Metadata input, storage and output

The metadata are transmitted by the data owner via a browser based wizard during the data upload. All mandatory fields have to be filled and a validity check for specific fields is done. The resource creator is able to edit the metadata afterwards. All fields are stored in the relational database „Metadata“ (see Fig.4).

The metadata can be exported to XML (Extensible Markup Language) and are made user-friendly visible by an XSLT (Extensible Stylesheet Language Transformation) conversion script as a HTML web page. They are automatically attached to a data package, when a dataset is downloaded.



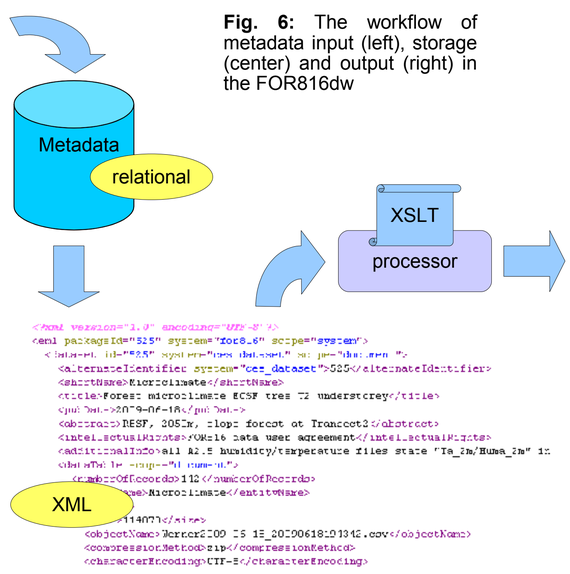
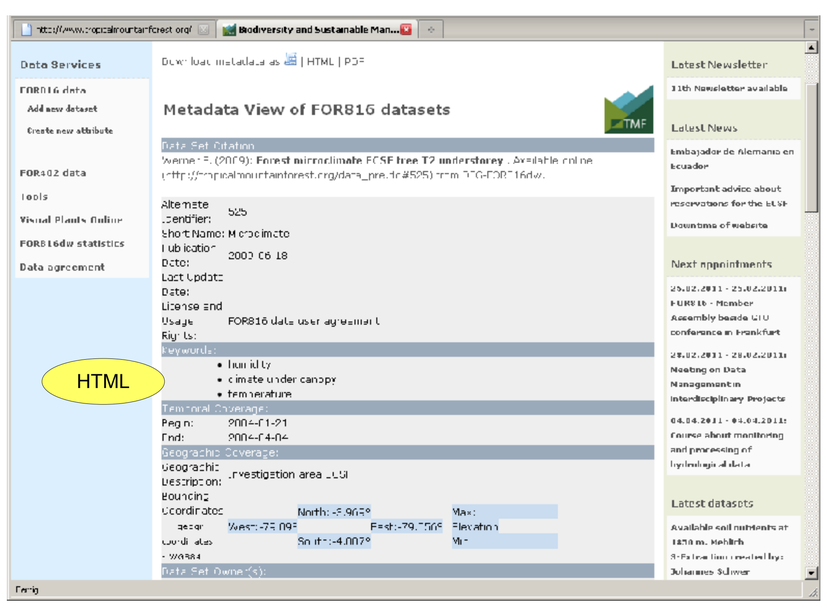


Fig. 6: The workflow of metadata input (left), storage (center) and output (right) in the FOR816dw



Acknowledgements This work has been carried out in the framework of the DFG Research Unit 816, subprogram „Central data services“ (BE 1780/21-2).