An OAI Framework for biodiversity and contextual content: „PlanktonNet“ as pilot study

Anita Macario and Bastian Onken
Data and Computing Centre of Alfred Wegener Institute for Polar and Marine Research, Germany

Abstract

Digital objects in the field of earth and biological sciences are known to be often compound and complex. If one takes biodiversity as an example, an exhaustively long list of information systems can be found on-line. These systems often contain valuable and historically relevant information gathered over several decades using distinct archival repositories, data models and transport protocols. In order to assure long term preservation of this distributed and not yet networked mass of resources, data models and transport protocols need to be presented as metadata portals.

The proposed information network was built upon existing PlanktonNet data providers, repositories and other ecosystems and will serve to provide a sustainable and long term solution to currently harvest OAI-PMH compliant PlanktonNet data and associated information. Data providers are planned (using XSLT) to expose relevant metadata using FOXML (using OAI-PMH) and other repositories (e.g. publications, geological and environmental data repositories, etc.). The OAI framework is designed to harvest (for re-use and exchange purposes) will be made available as result of the GBIF initiative.

Why FEDORA?

FEDORA offers a scalable open access repository framework compliant with international standards (ONI, Storage, OA-PMH, SDV/REST web services, etc). The flexible and extensible digital object model and associated services include: ORIG, relationships, curation metadata and licensing (schema validation).

FEDORA offers preservation through content versioning, and control access to object and collection metadata. FEDORA architecture includes a generic RDF-based relationship model that represents relationships among objects and their components.

FEDORA's ability to distribute and load object storage among several OAI-PMH compliant repositories is tested in a federated environment together with metadata semantic services as key factors for a successful network of biological information systems.

References:

1. OAI http://www.openarchives.org
2. GBIF http://www.gbif.org
4. Fedora Commons http://fedora-commons.org
5. Pangaea http://www.pangaea.de
6. panFMP (PANGAEA framework for metadata portal) is generic and flexible harvester powered by Apache Lucene indexing and searching engine.
7. PlanktonNet data portal http://data.planktonnet.eu
8. PlanktonNet site http://www.planktonnet.eu
9. RDF-based relationship ontology
10. RDF – Resource Description Framework

Interoperability lessons from „Pathways“

A long tail of analogous information systems based on heterogeneous data models/interfaces and using distinct metadata description and transport protocols are currently available. In order to offer valuable services-oriented gateways targeted at specific projects (e.g., PlanktonNet data portal), simultaneous access to as many as possible repositories is wished. True repository interoperability can only be accomplished by agreeing in an interoperability layer in which the data model (including granularity) and services are commonly shared across repositories. Further details can be read on https://www.pathways-eu.org

PanFMP and the OAI framework as pilot study

The authors from Carl Luggren and Sandy Posada from FEDORA team for digital curation in content models and Low Schildt for producing the panFMP and the OAI framework as pilot study. This work is funded by a grant from the Sixth EU Framework Programme (PANs) and Alfred Wegener Institute for Polar and Marine Research, Germany.