

ADVANCES IN BIODIVERSITY INFORMATION MANAGEMENT: AN INITIATIVE FROM THE BIOTA/FAPESP PROJECT FUNCTIONAL GRADIENT

Theme: Biodiversity Information Systems

Debora Pignatari Drucker, Jose Augusto Salim, Carlos Alfredo JOLY

Biodiversity is a data-intense science, based on a large number of disciplines in order to advance our understanding on the extent and trajectory of life on earth. Huge volumes of primary data, ranging from genomics to environmental, are currently being produced and archived in numerous databases, bringing together challenges for its maintenance, curation, interoperability, analysis and interpretation. New computational technologies are developed by the information scientists continuously, and the field of biodiversity informatics needs to take profit of such advancements to improve the ability to make sense of this huge amount of data. We describe the biodiversity information system developed for the Project Functional Gradient (Process Number Thematic 03/12595-7), within BIOTA/FAPESP Program; an effort to contribute to the endeavor of enhancing the long-term value of existing data by making it available for further high quality research. Our system accommodates data collected in biodiversity surveys: ecological, environmental, and specimen data not necessarily related to physical testimonies registered in scientific collections. The main characteristics of our system is that it is generic: designed in such a way that it accommodates data of any kind of survey, even future ones for which the survey design still has to be established; and is spatially enabled, which means survey data of whatever kind is geographically referenced. Moreover, all information is documented in Ecological Metadata Language, a worldwide adopted metadata specification used to describe field surveys datasets in a structured format written in XML (Extensible Markup Language). Metadata is crucial to provide information about people and institutions involved with the survey, methods adopted, taxonomic, temporal and spatial frameworks related to the dataset. The data packages (datasets accompanied by metadata documentation) are stored in a Metacat instance installed in the Institute of Biology, UNICAMP, which is a metadata catalog that allows storage, query and retrieval of XML documents. Our approach links a metadata catalog with a data repository, making information reachable by numeric, geographic and key-word based queries. Metadata from twenty-nine different biodiversity assessments at the Brazilian Atlantic Forest were documented and data and metadata retrieval were successfully tested at the system. Our effort is a contribution to advance on biodiversity information integration, including interfaces with modeling tools and climate change scenarios; as well as a contribution to improve our capacity to understand the future consequences of present choices.

Key-words: biodiversity information, Atlantic Forest, data management

This research was supported by BIOTA/FAPESP Process Numbers 07/59409-4 and 03/12595-7