

# **SPATIAL DATA DISCOVERY AND INDEXING TOOLS: AN APPROACH BASED ON METADATA AND FITNESS FOR USE**

PEDRO CASTRO

Arc4DigiT - Applied Research Centre for Digital Transformation.  
CIBIO-InBIO - Research Centre in Biodiversity and Genetic Resources  
Instituto Politécnico de Viana do Castelo  
pmc@estg.ipvc.pt

JOAQUIM MAMEDE ALONSO

ProMetheus - Research Unit in Materials, Energy and Environment for Sustainability  
CIBIO-InBIO - Research Centre in Biodiversity and Genetic Resources  
Instituto Politécnico de Viana do Castelo  
malonso@esa.ipvc.pt

IVONE MARTINS

CIBIO-InBIO - Research Centre in Biodiversity and Genetic Resources  
Instituto Politécnico de Viana do Castelo  
ivonemartins@esa.ipvc.pt

JOÃO HONRADO

CIBIO-InBIO - Research Centre in Biodiversity and Genetic Resources  
Faculdade de Ciências, Universidade do Porto  
jhonrado@fc.up.pt

JOHANNES PETERSEIL

Umweltbundesamt Gmbh · Ecosystem Research and Environmental Information  
Management  
johannes.peterseil@umweltbundesamt.at

## **ABSTRACT:**

The provision of tools to support the searching and selection of relevant data is a critical issue in big, open and linked data development frameworks. This raises many challenges considering that spatial data is used in very different application contexts, and often used for purposes other than the producer's intended ones. There is a risk of misusing and/or misinterpretation of data by users that can lead to misleading results. Therefore, user-oriented data quality assessment is a valuable procedure to consider in the identification of datasets that satisfy the requirements of a particular application for a specific user.

In recent years, efforts have been made in metadata development and in meta-evaluation of external and (in)direct quality by the end-user(s), taking advantage of metadata documentation possibilities and quality communication. WebGIS metadata catalogues offer opportunities for the implementation and improvement of spatial data quality evaluation/assessment tools related to knowledge discovery, searching and indexing. Those opportunities are grounded on standard metadata profiles containing a description of attributes about dataset/database content, access and use conditions, thus allowing the assessment of data quality components and elements (ISO 19157) as well as data quality management (ISO 19158).

In this context, we describe the specification and development of a Web platform with tools to perform external (meta)data quality evaluation to support the quality-driven searching and selection of relevant datasets. The process is based on requirements

specified by the user, allowing to identify the data needs (or detect data gaps), e.g. for environmental/ecological modelling in the on-going ECOPOTENTIAL project (<http://www.ecopotential-project.eu/>). The approach is based on the evaluation of the fitness for use of spatial datasets centered on users' requirements, resulting from the analysis of the matching level (fitness-for-use) between the characteristics of the data according to users' requirements (specified through the definition of expected values for quality indicators) and data characteristics as detailed/documented by metadata (often described at the producer level). Additionally, the platform performs a metadata (content) quality evaluation regarding the compliance with specific metadata standard profile(s).

Socio-environmental spatial data are frequently rated as factual data. Therefore, metadata including data quality elements are crucial to assess their fitness for use in different application contexts. The metadata-based and fitness for use spatial data quality evaluation proposed here will contribute to disseminate spatial data quality evaluation principles, methods and tools, and to mainstream data quality management in open, distributed and collaborative environments.

#### **ACKNOWLEDGEMENT:**

This work has been carried out within the H2020 project 'ECOPOTENTIAL: Improving Future Ecosystem Benefits Through Earth Observations' (<http://www.ecopotential-project.eu/>). The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 641762.

**KEYWORDS:** spatial data quality, metadata, quality evaluation, fitness for use, metadata standards, environmental spatial data