Chapter 6
Semantic Search Engine for Data Management and Sustainable Development:
Marine Planning Service Platform

Giuseppe Manzella
ETT SpA, Italy

Francesca Frontini
CNR ILC, Italy

Roberto Bartolini
CNR ILC, Italy

Maurizio Maltese
Delta Progetti, Italy

Franco Bustaffa
Delta Progetti, Italy

Daniele Medone
Delta Progetti, Italy

Paolo D'Angelo
ETT SpA, Italy

Monica Monachini
CNR ILC, Italy

Maurizio De Mattei
Delta Progetti, Italy

Antonio Novellino
ETT SpA, Italy

Andrea Spada
SyO srl, Italy

ABSTRACT

This chapter presents a computer platform supporting a Marine Information and Knowledge System based on a repository that gathers, classify and structures marine scientific literature and data, guaranteeing their accessibility by means of standard protocols. This requires the access to quality controlled data and to information that is provided in grey literature and/or in relevant scientific literature. There exist efforts to develop search engines to find author's contributions to scientific literature or publications. This implies the use of persistent identifiers. However very few efforts are dedicated to link publications to data that was used, or cited in them or that can be of importance for the published studies. Full-text technologies are often unsuccessful since they assume the presence of specific keywords in the text; to fix this problem, it is suggested to use different semantic technologies for retrieving the text and data and thus getting much more complying results.

INTRODUCTION

The value of marine environmental data is very high for the cost inherent to data collection, as well as for the knowledge that can be extracted from them. It is estimated that the EU Member States spend approximately € 1.85 billion a year on marine research (European Commission, 2010). About half is on infrastructures for facilitating observation. This includes ships, underwater observatories, floating buoys, drifting devices, remotely operated or autonomous underwater vehicles, as well as many other platforms, all equipped with a range of sensors and analytical capabilities. Unfortunately, a significant amount of data has been lost and is being lost for many reasons.

The NOAA National Data Buoy Centre reported that about 15-20% of data are lost for vandalism to buoys. However, the major quantity of data have been lost for the changes in recording technologies, that have affected the data integrity and rescue. Before electronic computers came into general use, oceanographic data were recorded in manuscripts, data reports, and card index files. With the advent of electronic data storage, oceanographic observations were increasingly recorded on magnetic media such as tapes and disks. Unfortunately, all these media are subject to degradation over time with subsequent loss of unique data. This has occurred in some cases, but unfortunately technology turnover is not the only reason for data loss. There are still a lot of researchers that are not 'publishing' their data in data centre systems, and with the result that data is lost when the researcher retires. Unfortunately large amounts of research funds are spent every year, while already existing data remain underutilised.

It has been underlined that data from the marine environment is a valuable asset; use and re-use can address threats to the marine environment, and can be used for the development of policies and legislation to protect vulnerable areas of our coasts and oceans, in understanding trends and in forecasting future changes. More in general, better quality and more easily accessible marine data can support the 'blue growth' or, in other words, the further sustainable economic development.

Data is analysed, synthesised, interpreted and transformed into information, and, as a final step, can produce knowledge. The outcome of this process is published as a scientific article. The intangible value of the data has pushed public authorities and organizations to encourage free and open access to data. In 2003 the 'Berlin Declaration on Access to Knowledge in the Sciences and Humanities' was published in order to "promote the Internet as a functional instrument for a global scientific knowledge base and human reflection and to specify measures which research policy makers, research institutions, funding agencies, libraries, archives and museums need to consider." (UNESCO, 2013). To make data usable in a tangible way it is necessary to accompany data with documentation, i.e. protocols, reports, grey literature published papers.

There are many reasons limiting the open and free access to data and documents, among which the Intellectual Property Rights and Copyrights. The idea of universal access to research, education, and culture is made possible by the Internet, but existing legal and social systems don't always allow that idea to be realized. To achieve the vision of universal access the Creative Commons (Clarke, 2001) is trying to create a balance between the reality of the Internet and the reality of copyright laws (Clarke, 2005).

The Marine Planning and Service Platform (MAPS) project started from a schematic depiction of the flow from research to library resources that is interlinking documentation and their underlying data. Within the project 'documentation' intend protocols and reports, as well as grey literature and papers published in conventional scientific journals. MAPS has developed a web search engine where the information retrieval is obtained from metadata and full text indexing and the information allow to select

Related Content

Reverse Engineering from an XML Document into an Extended DTD Graph
Herbert Shiuand Joseph Fong (2009). *Journal of Database Management (pp. 38-57).*
www.irma-international.org/article/reverse-engineering-xml-document-into/3403

The Effects of Atmospherics and Influencers on Purchase Intention in Social Commerce
Pranay Sindhuand Kumkum Bharti (2023). *Journal of Database Management (pp. 1-23).*
www.irma-international.org/article/the-effects-of-atmospherics-and-influencers-on-purchase-intention-in-social-commerce/317222

Context-Aware Query Processing in Ad-Hoc Environments of Peers
Nikolaos Folinas, Panos Vassiliadis, Evaggelia Pitoura, Evangelos Papapetrouand Apostolos Zarras (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications (pp. 1844-1866).*
www.irma-international.org/chapter/context-aware-query-processing-hoc/8008

Transactional Support for Mobile Databases
Hong Va Leong (2005). *Encyclopedia of Database Technologies and Applications (pp. 701-706).*
www.irma-international.org/chapter/transactional-support-mobile-databases/11227

IAIS: A Methodology to Enable Inter-Agency Information Sharing in eGovernment
Akhilesh Bajajand Sudha Ram (2003). *Journal of Database Management (pp. 59-80).*
www.irma-international.org/article/iais-methodology-enable-inter-agency/3303