ENVIRONMENTAL DIGITAL LIBRARIES

I. Papadakis¹ and V. Chrissikopoulos²

¹Department of Informatics, University of Piraeus
²Department of Archive and Library Sciences, Ionian University
E-mail: jpap@unipi.gr¹, vchris@ionio.gr²

ABSTRACT

Although a vast amount of information about the environment can be found in various locations around the globe, difficulty has been expressed in gathering and organizing such information. The role of Web-based digital libraries in this case is to provide a common framework for the management of such information. A digital library of environment-specific data can be defined as a unified entity, which brings together collections of digital materials as well as the services needed to access these materials. In our case, the collections of materials are environmental data. There are a number of technical and non-technical issues that should be faced and dealt with when building or evaluating such a digital library. In this paper, we refer to existing environmental resources on the Web and discuss various organizational and technical issues about environmental digital libraries.

ΠΕΡΙΒΑΛΛΟΝΤΙΚΕΣ ΨΗΦΙΑΚΕΣ ΒΙΒΛΙΟΘΗΚΕΣ

I. Παπαδάκης¹ και Β. Χρυσικόπουλος²

¹Τμήμα Πληροφορικής, Πανεπιστήμιο Πειραιάς
²Τμήμα Αρχείων και Επιστημών Βιβλιοθήκης, Ιόνιο Πανεπιστήμιο
E-mail: jpap@unipi.gr¹, vchris@ionio.gr²

ΠΕΡΙΛΗΨΗ

Αν και υπάρχουν πολλές πηγές πληροφοριών για το περιβάλλον σε διάφορες τοποθεσίες στη Γη, είναι γνωστό ότι είναι δύσκολο αυτές οι πληροφορίες να συλλεχθούν και να οργανωθούν. Ο ρόλος των ψηφιακών βιβλιοθηκών που είναι βασισμένες στο Web στην περίπτωση, είναι να παρέχουν ένα κοινό εννοιολογικό πλαίσιο για τη διαχείριση αυτών των πληροφοριών. Μια ψηφιακή βιβλιοθήκη περιβαλλοντολογικού περιεχομένου ορίζεται ως μια αυτοτελή συντόμη που συστηγάζει συλλογές ψηφιακού υλικού όπως επίσης και τις υπηρεσίες εκείνες που χρειάζονται για την πρόσβαση στην περιβαλλοντολογικό υλικό. Υπάρχουν πολλά τεχνικά και μη θέματα που πρέπει να αντιμετωπιστούν κατά την κατασκευή και αξιολόγηση τέτοιων ψηφιακών βιβλιοθηκών. Στην περίπτωση, οι συλλογές ψηφιακού υλικού αναφέρονται σε περιβαλλοντολογικό υλικό. Υπάρχουν πολλά τεχνικά και μη θέματα που πρέπει να αντιμετωπιστούν κατά την κατασκευή και αξιολόγηση τέτοιων ψηφιακών βιβλιοθηκών. Στην περίπτωση, την εργασία αναφερόμαστε σε περιβαλλοντολογικό υλικό που υπάρχει στο Web και αναλύουμε διάφορα οργανωτικά και τεχνικά θέματα σχετικά με περιβαλλοντολογικές ψηφιακές βιβλιοθήκες.
1. INTRODUCTION

Environmental documents represent a resource of public information that should be efficiently managed and preserved. Before the Web, bulky environmental documents were made available to the public at only a few sites, such as libraries or schools. The public could purchase documents but only at substantial cost. Documents were difficult to copy due to bulk and complexity (color maps, photos, tables, and technical appendices). Fortunately, the Web’s importance as a means of disseminating such data was soon realized. Thus, since the emergence of the Web, the availability of environmental information has been greatly enhanced. Moreover, in a digital, online world, information can be viewed simultaneously by multiple parties. Some of Web’s features like flexibility, open architecture and wide usage are most applicable to the case of disseminating environmental data. With environmental information online, interested audience can search for technical as well as scientific environmental sources without even leaving their office.

On the other hand, the fact that groups of environmental information on the Web resides in many different locations isolated one from the other, requires the development of environmental systems that will be able to manipulate it more efficiently. Such systems should nevertheless remain simple to use, since they are addressed to inexperienced users. Indeed, the potential users may belong to two categories: End-users (e.g., general public, policy makers) and Data-providers (e.g., biologists, geologists, physicists, oceanographers, etc).

The diverse and complex nature of environmental information dictates that it should be organized in digital libraries. Thus, an environmental report may address waterways, water quality, endangered species, recreation, economic development, land use, agriculture, soil, transportation and utility infrastructures, shipping, flood control, flood insurance programs, political structure, legislation, ecosystem protection, and history. It will include text, maps, charts, graphs, tables, photos and even video. A digital library defined as a system that includes full text delivery of digitized documents as well as original digital documents of environmental information consisting of additional services and value-added features such as indexing and searching is capable of satisfying users expectations for Web-based applications of environmental data.

Another use of an environmental digital library is for environmental policymaking and planning [1]. The ability of digital libraries to handle large and at the same time diverse amounts of data makes them most applicable for information–intense applications such as environmental planning.

The organization of this paper is as follows: In the next section, we give an overview of existing projects of environmental data on the Web and in section 3 we point out some of the most commonly met problems. Section 4 presents organizational and technical issues of environmental digital libraries on the Web that are most commonly met and finally, in section 5 we conclude with a discussion about the future of environmental digital libraries and their role in disseminating environmental information.

2. ENVIRONMENTAL INFORMATION RESOURCES ON THE WEB

Many of today’s projects that deal with dissemination of environmental data on the Web are actually directories that consist of annotated lists of online resources [2, 3, 4]. They could be described as “metasites” that provide access to standalone sites containing specialized environmental information. Environmental data in such a case is categorized by topic. The most popular environmental topics we meet are: climate, pollution, water resources, ecosystems, threatened species, radiation, agriculture and environmental planning. They act as “starting points”
for users that wish to locate environment specific information by browsing topics or by searching equivalent indices. Such projects originate either from governmental institutions [5] or educational ones [6] and private companies [7, 8, 9, 10]. The common feature of such projects is that they provide links leading to external resources. Usually, in such sites there is an invitation for submitting an unlisted environmental server. The services provided resemble subsets of many well-known Web search engines like Yahoo! [11]. Some of them [12, 13] provide additional services like Web-publishing, online catalogs, advertising and evaluation of its content.

Another category of environmental projects on the Web is the one that consists of applications providing Web interfaces to databases of environmental data. Queries can be addressed to standalone databases or to a set of individual databases behaving as a unified entity. Unlike the previous category, the structured nature of data stored in databases provides more advanced options to users wishing to retrieve environmental information.

Environment Australia Online [14] constitutes Australia’s government effort to provide a central access point for environmental data to Australians. This project consists of a number of environmental databases organized by topics such as air, biodiversity, marine and water geographic, heritage, industry, parks and reserves.

The Environmental Data Registry (EDR) [15] provides among others a suite of integrated data tools operating as a dictionary, a directory and a road map for people who use environmental data in the U.S.

The HUD E-Maps [16] enables users to create maps depicting information from the U.S department of Housing and Urban Development (HUD) together with environmental data. Maps can be generated to show location, type and performance of HUD-funded activities in every neighborhood across the U.S. in combination with local environmental information.

The Right-To-Know Network (RTK.NET) [17] provides free access to numerous databases, text files and conferences on the environment. By combining information available on RTK.NET, users can identify specific factories and their environmental effects, analyze reinvestment by banks in their communities and assess people and communities affected.

The Canadian Environmental Directory [18] provides a Web interface to a database containing Canadian organizations that deal with the environment. Among other features, the user is able to add comments to a specific organization, which will be added to the database and displayed to the public in the “notes” section of the organization’s record.

A database of special functionality is Earth’s Environmental Experts Database (3E Database) [19]. This project was born out of a desire to be able to find environmental experts quickly. Users can search for environmental experts in the database according to their specialization, language skills and countries they have worked in. The prime aim of the 3E database is to improve the quality of environmental studies by helping people locate a suitable expert.

Another Web-based database of special environmental interest is the WWW-Server for Ecological Modeling [20]. It provides information about existing mathematical models in ecology as well as a front-end to a system for documentation of mathematical descriptions of ecological processes.

The Environmental Protection Agency (EPA) [21] developed by the U.S. government, maintains a number of environmental information databases of various topics that may be queried as a unified
entity from other applications. The central access point to the repository infrastructure is provided by Envirofacts [22], which allows users to retrieve environmental information from EPA databases on various topics. This strategy has promoted the development of diverse environmental Web applications that nevertheless rely on a common environmental terminology and use standard data elements and values. Thus, relying on a common repository framework, users have the ability to develop a better understanding on environmental data. It is also easier to exchange data between systems and therefore to compare data coming from different information systems.

The Central European Environmental Data Request (CEDAR) [23] project is Europe’s effort to support environmental information dissemination across its members. This project provides mainly organizational information like terminology and legislation about the environment and its ultimate goal is to encourage companies to develop projects that support and protect the environment in Europe.

There are also efforts on the Web that promote the use of innovative technologies for the protection of the environment. The Center for Environmental Initiatives in M.I.T. (CEI) [24] examines the role of science and technology in developing better environmental policy, both from technical and from institutional points of view.

The Global Network of Environment & Technology (GNET) [25] is another environmental technology, news and business center that promotes the use of innovative environmental technology. It serves as an information meeting place for companies wishing to develop projects that invent and utilize such technology.

RENEW America [26] is a non-profit organization comprising of community and environmental groups, businesses, U.S. government leaders and civic activists that exchange ideas and expertise for improving the environment. RENEW America’s Environmental Success Index chronicles more than 1,400 effective environmental programs nationwide that measurably protect, restore or enhance the environment.

Apart from the above projects that provide general information on the Web about various environmental topics, there are also projects that concentrate on storing and managing environmental information about a specific location. These are mainly organized as information systems that provide digital library facilities to its users.

Montana’s Natural Resource Information System [27] provides an index to interactive applications about Montana’s natural resources as well as several interactive maps of Montana’s region that automatically zoom in / out to any point based on latitude / longitude, township / range / section, place, name or quadrangle name.

The Mojave Desert Ecosystem Program (MEDP) [28] aims at examining and evaluating mission critical U.S. department of defense installations within the Mojave Desert ecosystem to maintain military preparedness and readiness while protecting the environment. The project is based on a queryable database of environmental data and on an integrated decision support system.

Another environmental information system that acknowledged the need to provide digital library services to its users is THETIS [29]. THETIS provides dissemination, visualization and decision support of Coastal Zone Management (CZM) information for the Mediterranean Sea that is geographically distributed over a number of heterogeneous data repositories. The objective of THETIS is to promote information management for the support of CZM for the Mediterranean Sea.
through an open federated architecture that facilitates resource sharing while maintaining operational autonomy for information contributors.

The U.S. Berkeley Environmental Library [1] is a full-fledged digital library of environmental documents of various formats supporting environmental planning. It consists of digital as well as digitized documents that are accompanied with adequate metadata describing the content of the documents to the project’s search engine. The search engine is based on the Dienst protocol [30].

Although most environmental journals are met in printed form, the electronic version of the Environmental Periodicals Bibliography [31] contains roughly 630,000 indexed citations of environmentally related articles dating from 1973. The fact that since 1997, this site includes citations to all relevant environmental articles, assures researchers a reliable overview of the published literature.

An exception to the above rule is the Electronic Green Journal [32], which is an online environmental journal that provides peer-reviewed articles, book reviews, news and information on current printed and electronic sources concerning international environmental topics.

Finally, an electronic bookstore can be found on the Web which is dedicated to the environment [33] and provides e-commerce services such as an online catalog and payment through credit cards to people wishing to purchase printed material about the environment.

3. PUBLISHING ENVIRONMENTAL DATA ON THE WEB: PROBLEMS MET

Publishing environmental information on the Web is not an easy task. The diverse nature of environmental data and the fact that part of such information is not digital but digitized, generates certain problems as far as the design of their corresponding Web-based information systems is concerned.

The fact that browsing an environmental information system most frequently depends on some kind of geographical map, makes it difficult to browse through various parts of the map to locate text data that correspond to an unknown location on the map. Presently, users must search a map visually for requested data, which is a rather tedious procedure.

The fact that environmental issues recur over time and across geographical areas poses a need for comparing data items that come from maps and text belonging to different applications. Thus, a researcher who wishes to compare data referring to the same subject but deriving from different sources, must execute this task manually by visually comparing the different data objects. This is also a discouraging factor for interested users wishing to acquire environmental information from today’s environmental information systems.

In the case of digitized material where pages are stored as individual files, there is actually no means to print or download a document consisting of multiple pages in one simple operation. Consequently, users are forced to manually accomplish this task.

Having the above thoughts in mind, the next section refers to a number of organizational as well as technical issues that have to be considered when building a Web-based application capable of handling large sets of environmental data.
4. ORGANIZATIONAL AND TECHNICAL ISSUES

Manipulating efficiently large and diverse sets of data and publishing them on the Web are not new issues to the scientific community. A number of other disciplines have encountered this issue (e.g., healthcare systems [34]) and came to the conclusion that it is best to employ digital libraries technology in order to provide services that will satisfy user needs and expectations from such systems. There are, however, a number of organizational and technical issues that must be taken under consideration when building such systems. Lalis et al. [35] address some of the most critical organizational and technical issues for the development of a digital library. Their work will be extended and presented in this section from an environmentalist’s point of view.

4.1 Organizational issues
Organizational issues for the development of a Web-based environmental digital library concern problems that are met when dealing with any kind of non-technical requirements. It is crucial that such issues should be discussed during the early stages of such projects and in any case prior to the technical ones. Short descriptions of the most common organizational issues for the development of an environmental digital library are outlined below.

The fact that environmental topics most commonly exist in geographical regions that may include more than one countries, dictates that it is important to apply the same rules and regulations for all data maintained at the repository of the digital library, irrespective of the physical location of each data set. For example, a distributed digital library with satellite images from two countries existing in two digital library nodes physically located one in each country, must apply the same rules dealing with the dissemination of its data irrespective of the physical location of the data itself. In order to achieve this, the designers of the digital library must overcome difficulties deriving from the fact that each country may have different laws for the dissemination of satellite images of its geographical domain. Despite the fact that such a digital library consists of many separate nodes, users of the digital library must face it as a unified entity.

Designers of environmental digital libraries consisting of information provided by third parties should be aware that such information may be subject to copyright and property right laws. These laws refer to the right to publish and generally disseminate such information. Although a lot of discussions have been made on this subject, still a robust solution has not been found. This issue refers mainly to digital libraries containing reports and articles on various environmental topics.

Considering an environmental digital library as an information provider that provides resources and services upon these resources, a pricing policy should be established for these services. In many cases, the viability of such systems heavily depends on a carefully designed pricing policy. Different user categories should correspond to different pricing models. Thus, some basic material should be offered free of charge, scholarly material should correspond to special pricing models when referring to environmental scientists and members of the research community and value added information should be subject to commercial pricing models when referring to commercial users and authorities.

4.2 Technical issues
Building an environmental digital library also involves a number of important technical issues that mainly affect its usability and functionality. Many of the issues presented below are met in a number of digital libraries from various disciplines.
Especially for digital libraries that handle environmental documents, metadata play a central role. The term “metadata” refers to data that is used to describe other data. Thus, an environmental report about, e.g., the Mediterranean Sea could be accompanied with metadata like the author of the report, region it refers to, etc. Metadata are mainly used to provide content description to various kinds of resources. Conventional libraries successfully use metadata (in this case called bibliographic data) for the support of their resources. The employment of metadata in Web applications that handle diverse content, like digital libraries, is not a new trend. Metadata standards like USMARC [36] and Dublin Core [37] are widely used for the description of Web resources of various kinds. Moreover, the emergence of the XML standard has facilitated their broader dissemination since it provides a common syntax for their generation.

Another major issue of technical nature about environmental digital libraries is the need for a single point of access. Despite the fact that digital libraries may comprise nodes that are distributed across different physical locations, end-users should be unaware of this distribution. An end-user should be presented a unified system with exactly the same behavior for all of its components. Browsing and searching the digital library should be made according to a number of predefined logical abstraction levels that access each node of the digital library transparently.

The distinguished nature of environmental data requires advanced user interfaces as well as integrated functionality. Although it is possible with existing Web technology to embed virtually any kind of data (maps, reports, images, video, audio) into an html file, manipulating such data is not a straightforward procedure due to the lack of content description. However, the new XML standard addresses the need for content description on the Web and provides for integrated procedures on data of diverse nature such as environmental data.

A very important issue in developing environmental digital libraries is language support. Under current technology, it is not possible to support multiple languages simultaneously. Furthermore, despite the fact that some digital library applications offer multilingual support, it is still not possible to address a query written in one language to resources that are expressed in a different one. The need for multilingual support in digital libraries is of imperative importance especially in Europe where a variety of languages can be met. The Unicode standard addresses this need but there is still a lot of work to be done before it can be employed in real-time applications.

The large amount of information that exists in an environmental digital library dictates the development of techniques that will provide adequately good performance to such systems. Designers of an environmental digital library should, therefore, develop intelligent search strategies, as well as dynamic load balancing algorithms that will result in fast executions of user tasks. However, the fact that a lot of data travel through the network between the various architectural components of a digital library, dictates that the performance of such systems highly depend on the bandwidth of the network. Although the availability of bandwidth today is rather disappointing, there is a lot of activity concerning the development of faster communication protocols on the Web.

5. DISCUSSION

Throughout this paper, an effort has been made to map the current situation of environmental resources on the Web. The conclusion deriving from this work is that the Web is a place where a lot of information about the environment can be found. Such information is manipulated from many kinds of applications, varying from simple lists of resources to more complex digital library applications. There are though many questions about such applications that still must be addressed. One deriving issue has to do with the dual use of environmental data. For example, high resolution
satellite images could be used for military purposes. Although not directly related to weapons, this data is directly applicable to intelligence gathering, targeting and other military applications. When observing the technical and political context of environmental information, some user groups may disagree in including some information. Moreover, the quality of some environmental data could be challenged. The decisions about what information to include and what to exclude gives rise to debates over ownership, control, censorship, and public participation in policymaking.

The role of digital libraries to the dissemination of environmental data on the Web has also been examined. Therefore, the digital library can be used as a technology that results in making environmental applications more open, dynamic, interactive, content-based and transparent. By extending the characteristics of its precursor (conventional libraries), an environmental digital library creates new and interesting problems of design and evaluation which require rethinking of existing methods and ultimate contribute to the development of more sophisticated tools in the area of information dissemination on the Web.

REFERENCES

15. Environmental Data Registry (EDR), available at: http://oaspub.epa.gov/edr/EDRQUERY$.STARTUP
27. Montana Natural Resource Information System, available at:
   http://nis.state.mt.us/interactive.html
29. Houstis C., S. Lalis and H. Tsalapata, THETIS: an Environmental Information System for the
    Support of Coastal Zone Management, FORTH-ICS Technical Report, available at:
    http://www.ics.forth.gr/pleiades/THETIS/thetis.html....
33. NHBS e-bookstore, available at: http://www.nhbs.co.uk/
    Library Architecture based on TTPs’ Proc. of MIE2000, pp. 610-616..
    communication and research in Europe, available at:
    http://academia.darmstadt.gmd.de/seeheim/nikolaou.html