Facing future users - the challenge of transforming a traditional online database into a Web service

Eva Tolonen
Helsinki University of Technology Library

FACING FUTURE USERS – THE CHALLENGE OF TRANSFORMING A TRADITIONAL ONLINE DATABASE INTO A WEB SERVICE

Eva Tolonen
Helsinki University of Technology Library, Finland

E-mail: eva.tolonen@hut.fi

Our users in the next century will be a generation grown up with mobile phones, Internet connections and a hectic lifestyle. They will call for immediate actions and solutions to their problems. Improvements are expected in high speed networking from a new generation of communications satellites which will make the Web more accessible and provide data at a greater speed to the end-users. The impact of the Web on information management is a transfer from building physical collection to providing electronic access to distributed documents regardless of the physical location of the user or the information. The success, however, will not only depend on high capacity networks, but it will also depend upon the quality of the service provided to the users, i.e. how well the information is organised and managed.

Background

In 1987 eleven countries had a vision: to develop an single channel through which they could share their nationally published energy research and technology information. The Energy Technology Data Exchange (ETDE) [1] agreement was the result of this vision and today 19 member countries span four continents. In Asia: Japan and the Republic of Korea; in Europe: Belgium, Denmark, Finland, France, Germany, Italy, The Netherlands, Norway, Poland, Spain, Sweden, Switzerland and the United Kingdom; in North America: Canada, Mexico and the United States; and in South America: Brazil. The participating countries conduct their information exchange as an agreement under the International Energy Agency (IEA) and the concrete result of this agreement is the ETDE- Energy Database.

To build the database the ETDE-members collect the energy-related research and technology information published within their borders. Members are responsible for submitting the information in a standardised form using agreed-upon guidelines and subject analysis requirements. In Finland, the Helsinki University of Technology Library functions as the Finnish input centre and takes care of the practical work. The funding organisation is the Technical Development Centre of Finland which therefor acts as the Finnish member organisation. The ETDE co-operation as a whole is coordinated by a central operating agent, the U.S. Department of Energy's Office of Scientific and Technical Information. Because of the decentralised input preparation only a minimal work is required by the Operating Agent to consolidate member input for the database. Today the ETDE database which contains over 3.8 million abstracted and indexed records is the world's largest and most comprehensive bibliographic

database in the field of energy research and technology Another important aspect of the ETDE agreement is the exchange of full-text of documents which are not commercially available and therefore not easy to acquire elsewhere. These full-text documents are available on microfiche and since 1996 in electronic form. Together with the bibliographic database the storage of full-text documents forms a very valuable information source and enables development of new products.

Although the vision from 1987 has remained unchanged the mechanisms are evolving. There are also financial constrains. After ten years of uninterrupted growth in membership, the ETDE Agreement lost one valuable partner when Australia withdrew, hopefully only temporary, its membership. Many members are facing budgetary problems. The value of the database is not easily demonstrated to the funding organisations, which all are governmental bodies. Although the development of information services and information technology are considered important, the cost to compile and provide information is not always recognised.

The database is designed to provide valuable information not only for governmental bodies but also for the industrial and academic sectors. The access is available to organisations, libraries and institutions in the ETDE member countries. The users consist of a diverse community: scientists, engineers, policymakers, intermediaries and students. The database has been offered to member countries through the large online-hosts Dialog and STN International and on CD ROM. In addition to these some countries have developed their own products usually in the form of organisational in-house systems.

User surveys

The results from user statistics indicate that during the 1990's there has been a overall decline in both the hours of use and the total amount of retrieved information. Another observation is the change in the user profile. There has been a continuous drop in usage by the researchers in the academic world. In the other sectors the number of users has slightly increased or remained the same. The decrease among the academic users has occurred along with the introduction of the Internet graphical user interface and browsers. The trend seems to be quite evident: users are likely to turn to the most easily accessible sources of information, even if it is not the most appropriate one. The aim of the ETDE Agreement was to provide a database as a research tool to the entire community of energy science and technology researchers within the various member countries. In this respect ETDE is now faced with the need to make the information in ETDE database more attractive. The new technologies for information exchange must be exploited. At the same time the need to show value for money is evident in all member countries. Thus the ETDE is looking for cost effective solutions by developing a new product which should be user friendly, owned by the ETDE and accessible over the network.

The new web service

To meet the challenge, a subgroup was established in 1998 to prepare a request for tender concerning design, operation and maintenance of a web-based information service for end-users in the energy field, " to pursue a virtual library with the ETDE Energy Database as the core and linking any number of suitable distributed Web sites containing energy research and technology information in order to provide energy-related scientific and technical information as well as full-text". The working schedule

for defining the technical and the administrative requirements was very tight. The decision making process involved both the Technical Working Group and the Executive Committee. It is rather difficult to establish a competitive information product based on joint agreements because the pace for the markets is far more rapid than that of an international community.

The call for tender resulted in three bids which were submitted by Germany, the Netherlands and the United States. These were evaluated by the Technical Working Group and by an ad hoc group set up only for evaluation purposes. The member community was approached to present their views concerning a possible Web based ETDE product and their attitudes towards the costs involved in developing such a product. The three formal bids were all regarded as too expensive solutions and hence they were all rejected. It was noted that the original specifications indeed may have been too superfluous. At the end the ETDE community agreed to build its future service on an existing service, the U. S. Department of Energy's Information Bridge. It will provide the ETDE members a possibility to clearly identify the product on the Web, not just another database among databases. The new Web product will also be a clear compromise between the functional properties and cost factors. It will be accessible within the member countries in October 1999 and it will provide both bibliographic information as well as links to the full-text of non-conventional literature cited in the database. The value of the ETDE database will remain in its content because it contains energy research and development information not published anywhere else. [2, 3]

The evaluation of metadata formats

Along with the process of introducing the database on Internet as a Web product the database record format was under evaluation with the ultimate goal of developing a new web-compliant data model for the ETDE and INIS databases. INIS (International Nuclear Information System) is the International Atomic Energy's bibliographic database using the same record structure. Of these two systems INIS is one of the oldest online databases and ETDE adopted the INIS record structure to facilitate records exchange between the two systems. With the purpose that the record structure should reflect the realities of a web-based scientific culture for search and retrieval of information, ETDE and INIS did jointly fund a study to look at new metadata structures. A contractor outside the ETDE and INIS communities was hired to collect, review and evaluate document-related metadata formats, both formal and de facto standards as possible models for conversion of the present format. [4]

The work was conducted under the guidance of an expert group with delegates from Australia, Canada, Finland, Germany, The Netherlands, United States and the INIS Secretariat. There were also assistance by the Operating Agent with the overall aim to simplify the current record format and reduce costs.

Data used in the evaluation was obtained from traditional and Web-based sources. Furthermore, interviews with publishers, database producers and metadata experts were carried out as well as input from the Expert group. The selected metadata formats for evaluation were Dublin Core, Scientific and Technical Attribute Set (STAS), Global/Government Information Locator Service (GILS) Core Elements and finally Text Encoding Initiative (TEI) Electronic Title Page Elements.

The Dublin Core Element Set [5] has been defined to support information discovery in the networked environment and the work had begun under the auspices of OCLC at a meeting held in Dublin Ohio. The Dublin Core element set consists of 15 elements, which are considered to be the minimal element set into which a description of an electronic resource can fit. These base elements can be extended by qualifiers, which can be used to define a content of a element and by subelements. Dublin Core is protocol independent although it is primarily accessed via HTTP. The Warwick Framework and RDF or Resource Description Framework (http//www.w3.org/TR/REC-rdf-syntax/, W3C Recommendation 22 Feb 1999) are architectures associated with Dublin Core and other formats. Since they are intended to provide a framework for many metadata initiatives, they have also been considered in the evaluation.

STAS [6] was originally developed by Chemical Abstracts Service (CAS) using the Z39.50 protocol. The STAS elements were defined in 1994 by a joint group including Fiz Karlsruhe, Chemical Abstracts, Dialog and CNIDR - Center for Networked Information Discovery and Retrieval. The development of STAS is a practical one and it is built on the elements that needed to be mapped from existing databases. There are 63 elements defined in the BIB-1 set developed for search and retrieval of bibliographic scientific information. If a mapped element does not match the element definition in the BIB-1 exactly, a new element is created within STAS.

The TEI [7] is an SGML-based DTD (document type definition) for markup of full text documents and it has been in use for several years primarily in the humanities. The focus of TEI in the evaluation was not on the SGML structure, but on elements identified for the description of an electronic title page. In practical usage the TEI Electronic Title Page has incorporated other elements than might appear on the title page itself. For example availability information has been added. The number of the core element exceeds 50 depending on the implementation.

The Government/Global Information Locator Service (GILS) [8] was developed by the U.S. Federal Government using the Z39.50 protocol for implementation. GILS is a standard for cataloguing, searching and retrieving information about resources. The cataloguing on the resource level includes 30 core elements, although this is varying by implementation.

These four metadata formats were evaluated against criteria established by the contractor and the ETDE Expert group with the objective to reduce complexity, support interoperability between formats and facilitate data import and export. Additional important factors were: continuity meaning degree of standardisation, use by similar communities and the extensibility of the format. In the evaluation Dublin Core and STAS were ranked highest against the established criteria. In order to provide more information about their applicability for the ETDE/INIS databases, the present ETDE/INIS elements were mapped to the Dublin Core and STAS formats.

The conclusions of the study made by the consultants were to "adopt the Dublin Core with its 15 elements and its extensible qualifiers and subelements as the document level minimal element standard". The Dublin Core is relatively stable with development and implementation projects on an international scale. Although it is not a standard, the organisations involved are likely to continue to move the convention

forward to standardisation for use in broader Internet environment with new technologies such as XML and RDF. The second recommendation was to extend the Dublin Core by using STAS element definition for special scientific needs, if necessary in the future. The third recommendation was to use an SGML-application to link to full text using the Dublin Core names as the header element tags in order to provide continuity. This would also promote the automatic creation of Dublin Core elements from the original electronic text. One of the original objectives was to reduce the cost of record creation. It was not showed, however, that the use of the Dublin Core would in fact result in the reduction of the database production costs. Furthermore the ETDE Expert Group was not able to form a general recommendation on what metadata format to choose or the timing for the change. A final decision was regarded too difficult to make at the time when the practices are still under development. The U.S. Department of Energy had in between decided to move into the direction of Dublin Core. It was agreed to wait and see the results of the U.S. experiment using the Dublin Core metadata format and its impact on ETDE Database before taking an international decision. The task of redesigning the present ETDE/INIS record format is still ongoing and active work is carried out on the new format structures and how these can be realised.

Conclusions

Will the new ETDE Web product find its users from the new generation demanding immediate solutions, remains to be seen. It is clear, however, that a pure bibliographic database is no longer sufficient. Added value produced via linking full-text documents and other relevant Web-based information will be the key to success of the ETDE Energy Database.

References

Energy Technology Data Exchange (ETDE) and the Energy Database. Homepage at