

COVAX News

Inside this issue:

What users want	2
ENEA Profile	2
AIT demonstrator	2
TEXTML at BTH	5
Archives event	5
Events/Contacts	6

Covax aims:

- build a web service for European Contemporary Culture
- make accessible over the Internet existing document descriptions
- satisfy access provision needs of Memory Institutions
- achieve inter-operability between systems through standards

How do users want online databases to work?

Covax is designing online database systems which will ultimately be useful to end-users searching the web. Although our primary focus is to create systems to support interoperability, resource discovery and access, we conducted a survey in mid 2001 amongst a small but international sample of 'end users' of databases to find out what they feel they are looking for in terms of functions and features. We reached the following conclusions, from a questionnaire survey producing 86 responses from six partner sites in Austria, Italy, Spain, Sweden and the UK. Those responding were a mixture of information professionals, academic, clerical staff and others related to partner institutions

used for general information searching.

- Respondents mainly search for full documents or books (58%) and to a lesser extent for other materials, images and locations.
- The search fields usually used are subject, word in any field, author and title. The fields related to publishing (date, place, publisher) are seldom used.
- The operator mainly used for searching is AND (41%) and sometimes help is needed about its use. It is interesting to note that a large number of people do not use operators. There are two possible reasons for this: the users do not know

in some way. This selection was necessary to achieve rapid results in the available timescale.

- The educational level,



Contemporary Culture Virtual Archive in XML

familiarity with the use of computers and databases, and number of times per week respondents use on-line databases, is high. More than half of the end users use online databases more than 5 times a week.

- The main reasons for using and searching databases are academic, professional or personal research. They are also

(Continued on page 2)

PARTNER PROFILE: Ente per le Nuove tecnologie L'Energia e l'Ambiente (ENEA)

ENEA is a public corporation operating in the field of research and innovation for sustainable development, for the purpose of promoting together the objectives of development, competitiveness, and employment and that of environmental protection.

It also acts as an agency for public administrations by providing advanced services

in the sectors of energy, the environment, and technological innovation.

In particular, it:

- carries out, develops, and promotes research and innovation, through the creation of demo systems and pilot projects, with aims and objectives related to sustainable development, within the framework of the national research programme and in

conformity with the commitments arising from Italian participation in the European Union and other international organisations on the subjects of energy, environment, and technological innovation;

- supports the innovation processes of manufacturing, in particular of small and medium-sized enterprises, also promoting

(Continued on page 2)

PARTNER PROFILE: ENEA

By Luciana Bordoni, ENEA, Rome

(Continued from page 1)

demand for research and technology in conformity with the principles of sustainable development;

- aids the process of technology transfer and positive experiences in the energy and environmental field to enterprises, in particular small and medium-sized ones, and to public administrations within the

framework of national policy and that of the European Union;



- supplies, on request, in the sectors of its competence and within the framework of programme agreements with the Ministries of Industry, Com-

merce, and Handicraft, of Universities, and of Scientific and Technological Research and the Environment, as well as with other public administrations, specialised technical and organisational support to administrations responsible for public activities, at the national and international level, to regions, and to local government units.

How do users want online databases to work?

by Francisca Hernández, Residencia de Estudiantes, Madrid and partners

(Continued from page 1)

how to use operators or the appropriate operator is not available.

- This questionnaire shows that users wish to customise search results by changing the display format or result order.
- Ease of search, speed of response and clear results presentation is what users expect a database to offer. The help function is less important to them.
- The majority of users (71%) are prepared to wait up to one minute for the response, but a few are more impatient and they wouldn't wait for more than 30 seconds, whereas others are ready to wait 5 minutes if necessary.
- When they need help with

database searching, users look for a help button or search a different way (such as asking a librarian).

- They usually request full display, print or download when they have found information. Only 24% download information.
- If end users do not retrieve the expected records, they think this is because they do not exist in the database or they have not found appropriate search elements. Sometimes they think the results are too broad and they would need help to effectively narrow down the search.
- If they do not find an item in a catalogue they try another search or search another site. A minority of users use the help button or ask for outside help.

- 69% of the users find it useful or sometimes useful to store their query for future reference. A third of respondents never does.
- Two thirds of those polled think it is easy to find what they are looking for and 25% find it difficult. This result can be explained by the high level of familiarity with the use of on-line databases amongst our users.
- They are moderately interested in document reproductions and perhaps occasionally they would be willing to pay for a reproduction (61%), depending on the document.

The results of the study will be useful for our validation and usability assessment phases now underway as well as informing our exploitation plans.

"Ease of search, speed of response and clear results presentation is what users expect a database to offer."

COVAX metasearch engine and access to native XML-data bases

by Walter Koch, AIT (Angewandte Informationstechnik Forschungsgesellschaft mbH), Graz

The COVAX-Project was initially based on proprietary XML-Servers (TAMINO) holding

metadata formatted in XML. These metadata catalogues from museums, libraries and archives follow different domain specific standards such

as MARC21, AMICO, EAD and TEI. During the project lifetime other native (the database stores XML records

(Continued on page 3)

COVAX metasearch engine and access to native XML-data bases

(Continued from page 2)

directly) XML-Servers have become available. One of them was IXIASOFTs "TeXtML"-Server. This Server is implemented only for the Windows platform, providing an API (Application Programming Interface) to ASP (Active Server Page) programs only.

Figure 1 shows the deployment diagram with the different components included in the COVAX Architecture. To connect the XML-server directly to the COVAX-component in general the RMI (remote method invocation)- approach has been chosen. The queries which are distributed by the COVAX-metasearch-component (the left block in Figure 1) are XPATH-like (See Figure 2) and are routed via URLs to the WEB-Servers which are connected to the XML-Databases.

This technique cannot be applied directly to the integration of the TeXtML-Server. For that reason AIT - Angewandte Informationstechnik Forschungsgesellschaft mbH - has developed an interface which consists of two parts: one part (a Windows DLL) supports the TeXtML-API and is interfaced to an ASP script which resides on a WEB enabled directory at the WEB-Server's site. By

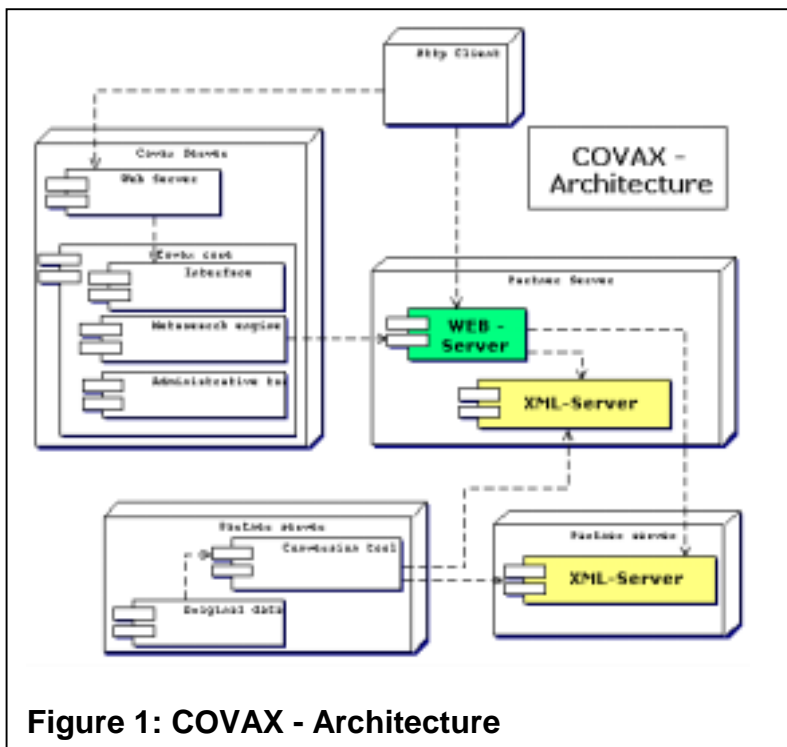


Figure 1: COVAX - Architecture

this means a functional integration of the XML-Server into the COVAX system has been achieved. Another problem to be solved was the mapping of XPATH- (XQL)- queries to the standard used by the TeXtML-server. The

most frequent query expressions and search functions have been mapped, so the end user is not aware of the types of XML-Server to which his query has been broadcast. The query in Figure 2 (the substring "Aus" in the

"the end user is not aware of the types of XML-Server to which his query has been broadcast"

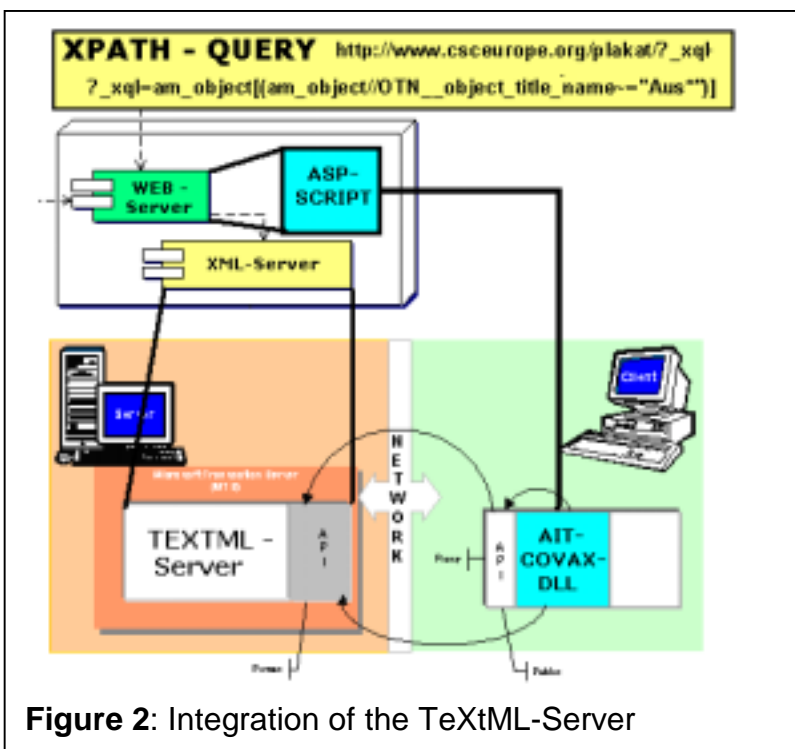


Figure 2: Integration of the TeXtML-Server

COVAX metasearch engine and access to native XML-data bases

(Continued from page 3)

title field) had to be transformed to the form as outlined in Figure 3 (<anystr/> means "wildcard"-sign):

The semantic mapping from the DublinCore (DC) data element: dc-title to the AMICO-dtd relevant element: OTN__object_title_name is done within the COVAX-system on the user interface side. Another option for this mapping is to do it at the index-level of the XML-Server (see Figure 4). "Key NAME" corresponds to "index NAME".sherw00d1

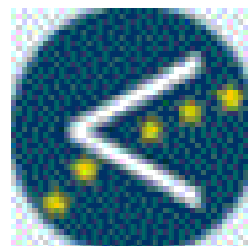
A hit delivered by the XML-Server (XML-formatted) is represented in Figure 5 (an example from the Basle Poster Collection).

The integration of the TeXtML-Server into the COVAX-system has - besides the fact that the potential user of the COVAX-System has more than one option for an XML-Server - another benefit. Many utilities are available supporting the TeXtML-Server, and the powerful API enables fast development of (ASP-based) applications. For example, AIT has developed a simple (meta-) data management and search facility for single data bases. Using this tool (Figure 6 and 7), the COVAX-user (who runs the data base) can manage his/her own data base, which is also accessed via the the metasearch engine without any additional effort.

AIT has in addition developed an installation tool which allows the setup and management of different data bases which can be searched in parallel via the COVAX Metasearch engine. This possibility is already used within the Culture-2000 project VECTOR. In this project different archives all over Europe are maintaining their different catalogues separately. The "Virtual Catalogue" again is achieved by the use of the COVAX-System.

```
<?xml version="1.0" encoding="UTF-16" ?>
<query VERSION="1.0"
  RESULTSPACE="R1">
  <key NAME="dc-title">
    <elem>
      Aus
      <anystr />
    </elem>
  </key>
</query>
```

Figure 3: TeXtML-Query representation



```
<index NAME="dc-title" TYPE="FullText">
  <elements>
    <element NAME="OTN__object_title_name">
      <content>
        <elemdata DEPTH="INFINITE" />
      </content>
    </element>
  </elements>
</index>
```

Figure 4: TeXtML-Index

```
<?xml version="1.0" encoding="UTF-8" ?>
<response>
  <am_object>
    <AID__amico_identifier>4</AID__amico_identifier>
    <am_titles>
      <OTG_object_title_name>
        <OTN__object_title_name>Lockvögel im Museum /
        Ausstellung im Naturhistorischen Museum
        Basel</OTN__object_title_name>
      </OTG_object_title_name>
    </am_titles>
```

Figure 5: XML-Record (result)

"Many utilities are available supporting the TeXtML-Server, and the powerful API enables fast development of (ASP-based) applications"

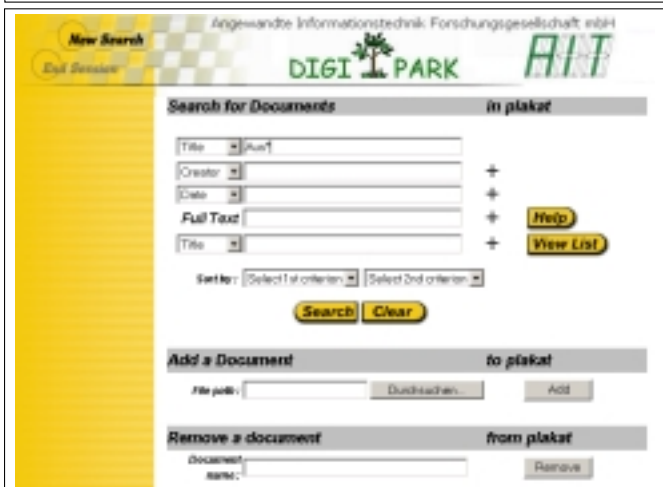


Figure 6: Meta Data Management based on a TeXtML-Demonstration Package



Figure 7: Search Result

TEXTML at Blekinge Tekniska Högskola

By Peter Linde, Blekinge Tekniska Högskola, Karlskrona

We decided to use IXIASOFT's TEXTML-server, version 1.6 (<http://www.ixiasoft.com>) this spring after receiving favourable reports about the software's performance from our COVAX partner AIT in Graz.

The main reason for installing and running the TEXTML-server package at BTH, except of course for doing what we wanted it to do, was the price. Unfortunately the price went up from 3.000 USD to 10.000 USD just before we decided to purchase, but even so it compares very favourably with TAMINO if you only look at the price.

It also has proven itself quite easy to install and start running. The only negative remark we make is regarding the administrative user interface, which is not very user friendly. For example, it is quite complicated to manipulate records in the different document bases - to show, delete, transfer etc.

All communication with the database is done via different indexes. This gives a lot of possibilities to do

searches on xml-records at even the lowest attribute level, which suits us fine since the main bulk of our records are full text documents, mostly fiction, tagged for research purposes in TEI.

But indexing also creates some problems that we have encountered during the COVAX project. To retrieve records via the Metasearch engine all the partners using TEXTML must have corresponding indexes that also match all the accesspoints for the TAMINO-partners. This has been shown to be crucial and has raised problems with the early prototypes of the COVAX system.

Our profile in the project has been to create and archive fulltext records. During our tests it has been proven quite timeconsuming to shuffle these big files back and forth from the partner repository to the main COVAX search site. This has resulted in long wait times. To solve this problem we have developed a search-and-display interface for our TEXTML-database (for display we used Sebastian Rahtz's XSL stylesheets - <http://www.tei->

[c.org/Stylesheets/](http://www.tei.org/Stylesheets/) - with minor modifications). Thus we can produce HTML versions of our fulltext documents.

Being able to do this, plus being able to do searching within the document bases will make it possible to instruct the COVAX search engine only to retrieve document metadata, not the full document. In the metadata there will be a pointer to the full document at our site. So when the user searches COVAX and finds any of our full text documents, the user will find a link to the full text document at our site. If he follows the link it will take him to the full document, which he now can investigate with further searches directly from our own web-interface based on the elements used for the full-text documents.

We are currently running the TEXTML software on a Dell PowerEdge 1300, under Windows 2000. As soon as the COVAX project allows us we intend to upgrade to version 2.0 which was released in September 2001.

"when the user searches COVAX and finds any of our full text documents, the user will find a link to the full text document at our site"

"Safeguarding our culture", Society of Archivists' Annual Conference, Aberystwyth, UK, 4-7 September 2001

The UK's Society of Archivists kindly invited Robin Yeates to present a paper about Covax to their main annual meeting, held in the normally very pleasant but sometimes wet, Welsh, holiday, university and seaside town of Aberystwyth.

The paper was entitled *'Using XML to make cultural resources accessible - the COVAX project'*. Its theme of integrating library, archive and museum resources obviously struck a chord with many of the audience.

However, the short time to explain the technical approach of the XML-based project meant that some people may have had difficulty understanding it all!

As a technology project primarily concerned with developing appropriate 'middleware' for networks, it is becoming increasingly difficult to introduce this kind of work to outsiders, but those trying to achieve similar goals were extremely interested to receive further information after the event.

If you are using XML technologies to exchange and make available cultural resource data, contact us to find out more, once you have read this newsletter of course...

The Society of Archivists has a website at

<http://www.archives.org.uk/>



COVAX-IST-1999-11820

LITC, South Bank University
103 Borough Road
London
SE1 0AA



Phone: +44 20 7815 6924
Fax: +44 20 7815 7050
Email: yeatesrb@sbu.ac.uk

We're on the Web!

www.covax.org



Contemporary Culture Virtual
 Archive in XML

A website has been established in Catalan, English, German, Italian, Spanish, and Swedish at <http://www.covax.org>, where you can find other issues of Covax News and other information about the project.

Carlos Wert, the project Coordinator and Francisca Hernández have published an introductory article on Covax in Cultivate Interactive Issue 3 January 2001 (<http://www.cultivate-int.org/issue3/covax/>).

Project coordinator

Carlos Wert
Residencia de Estudiantes
Pinar, 23
28006 Madrid
Spain.
Email: ile@interlink.es
URL: <http://www.covax.org>

Events

Online Information 2001 - London, UK
4-6 December, 2001

<http://www.online-information.co.uk/>

LITC is an exhibitor at this major event and Robin Yeates will be present to discuss Covax.

An overview article about Covax by Robin Yeates will appear in a special VINE issue to be published for this event and available at the stand.

Workshop on Artificial Intelligence for Cultural Heritage and Digital Libraries **AI*IA 2001 Conference, Bari, Italy, 25 September 2001**

This workshop represents an opportunity to analyse the contribution that the Artificial Intelligence and the Web could bring in this contest and, also, to directly compare ideas and results coming from academic and re-

search corporations and from companies/bureau that are involved in developing new technological solutions in the field of cultural heritage and of digital libraries. The goal is to create a community of institutions that could deal with this new

challenge. In this context Luciana Bordonni presented in the paper entitled "The Covax project" the main aims of Covax.

(http://www.di.uniba.it/~aiia/ws2/ws2_en/)



Rights and disclaimer

COVAX is part funded by the European Commission.

The authors are solely responsible for the content of this newsletter.

The site does not represent the opinion of the Community and the Community is not responsible for any use that might be made of data appearing herein.

Reproduction of the newsletter is authorised, except for commercial purposes, provided the source is acknowledged.

Some of the documents and multimedia sequences in this newsletter and on the COVAX website might contain references, or pointers, to information maintained by other organisations. Please note that we do not control and cannot guarantee the relevance, timeliness or accuracy of these outside materials.

© COVAX 2001