

IST-2001-33127

**SciX**

Open, self organising repository for scientific  
information exchange

## D13: Core pilot

Responsible author: Vlado Stankovski

Co-authors: Tomo Cerovšek, Žiga Turk, Janez Žmavc

Type: software/pilot

Access: public

Version: 1.0

Date: September 28, 2003

## EXECUTIVE SUMMARY:

As a demonstration project SciX plans to demonstrate a sustainable digital library infrastructure. The demonstrators were structured into these workpackages and deliverables:

- WP2 should create an e-journal infrastructure (D12)
- WP3 should create core digital library services (D13)
- WP4 should create advanced digital library services (D14)
- WP5 should create content syndication services for non academic, industrial users (D15)

The deliverables D12-D15 are not documents but running demonstrator software. This software will be documented in the D10 "Overall implementation report", which will include (1) the "as-built" architecture, (2) end user as well as (3) programmer's and integrator's documentation.

The purpose of this report is:

- to provide documentary statement, that the core pilot has been created;
- to provide addresses and URLs stating what services are using the pilot;
- to briefly enumerate what features have been implemented to date.

The purpose of this report is not:

- to provide documentation of the core pilot,
- instructions for use,
- release notes etc.

This report will be updated as the functionality of the core grows; this report will be superseded by the D14 "Intelligent maintenance and use pilot" when advanced features and services will become available.

## RELEASE HISTORY

date	changes
29.9.2003	report on D13v1

**TABLE OF CONTENTS:**

EXECUTIVE SUMMARY: .....	2
RELEASE HISTORY .....	2
TABLE OF CONTENTS: .....	3
<b>1. BUSINESS OBJECTS IMPLEMENTED .....</b>	<b>4</b>
<b>2. COLLABORATION AMONG THE SERVICES .....</b>	<b>7</b>
2.1 LOCALIZATION OF SERVICES .....	6
<b>3. APPLICATIONS .....</b>	<b>8</b>
3.1 DIGITAL LIBRARIES: .....	8
3.2 INTERNATIONAL DIGITAL LIBRARIES .....	10
3.3 ELECTRONIC JOURNALS .....	11
3.4 CONFERENCES .....	12

## 1. BUSINESS OBJECTS IMPLEMENTED

An overview of the SciX architecture is shown below. On the third layer from top, there is a business objects layer providing web services to applications. It is by different combinations of these business objects that different kinds of applications can be, and have been built.

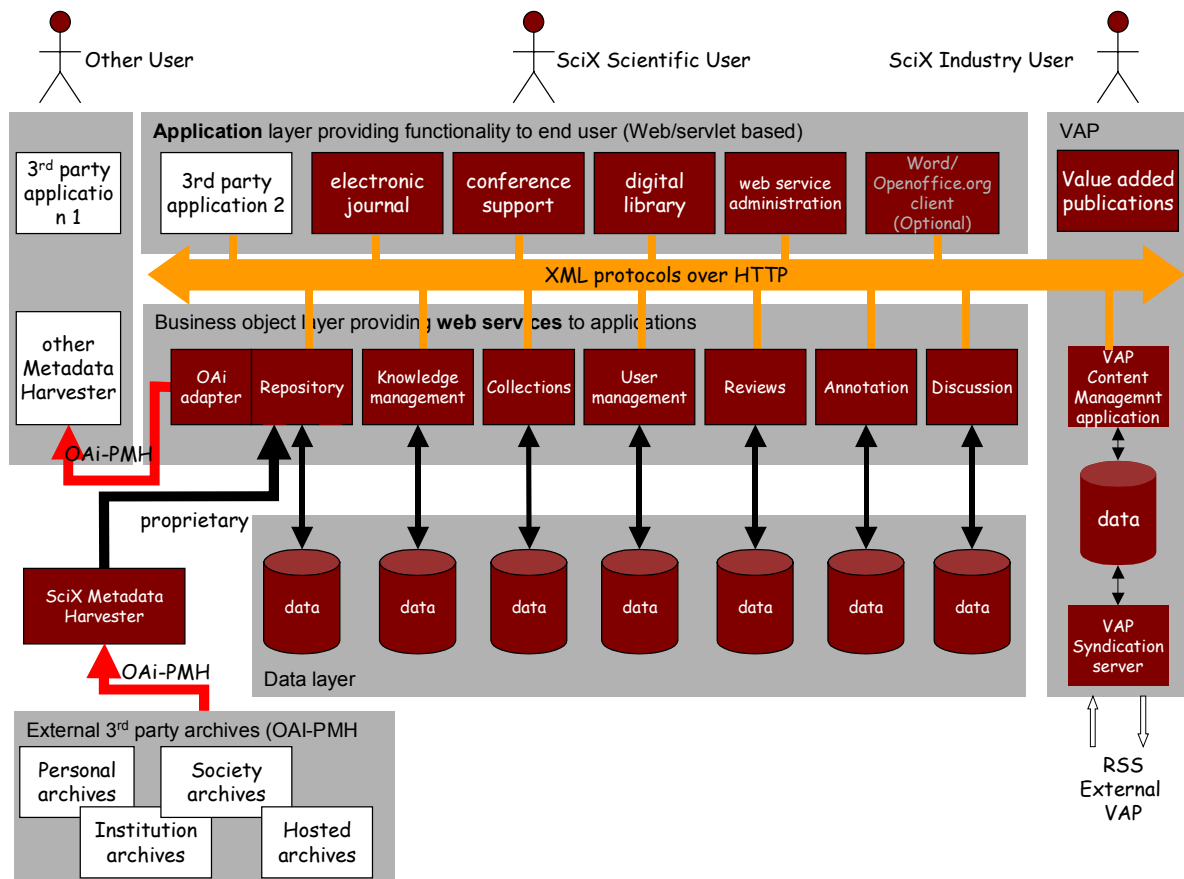


Figure 1: SciX architecture overview.

The services are at different stages of development. They will be listed as:

- P - operational, publicly available.
- B - under development, consortium.
- A - alpha, under development, developers.
- P - planned.

Services will be addressed from left to right, as they appear in the figure:

**OAI adapter (A).** SciX project is building its solution on top of open source solution "Open Archives Initiative Protocol for Metadata Harvesting Toolkit" from Virginia Polytechnic. It implements OAI-PMH v2.0.

**Repository (P).** Repository services are handling metadata and full text and allow for all basic functions identified in the requirements sections. The repository provides Dublin Core compatibility in the sense that any Web pages it creates provide proper Dublin Core head elements. The repository is also compatible with citation management systems by providing BibTex, Refer and Harvard formats for the data input. In terms of access rights, the repository service can be set up in various ways; in some cases anonymous users would get full rights, in others, only registered users would be allowed access.

**Knowledge management (A).** The service is based on the Bow toolkit developed at the University of Michigan for clustering and statistical text analysis. It allows for automatic classification and clustering of works in a repository, searching for similar papers etc. The services were tested with one application (itc.scix.net). Seamless integration with the rest of the system is subject to work in WP4 and D14.

**Collections (P).** Collections service allows end users to publish collections of papers that they have found interesting in a repository, together with an introductory commentary. The envisioned use is that for example a teacher collects a reading list for his students and publishes that. A collection is created in such a way, that in the repository, the end user collects 'favorite papers'. These papers may then be added a collection. The collection is then published. End users may, of course, at to it, or delete items from the collection. It goes without saying that ownership of collections and the rights to modify them are restricted to the end user setting up the collection.

**User management (P).** All services have a built with a notion of a user and of a group of users. They use this for access right management and personalization of the services. Both these features depend on the service's logic. The generic user management service takes care of adding new users, authentication of users, log-in, preferences management etc. The service allows for user log-in, display end editing of her profile, setting and changing of the password, lost password recovery etc. The maintenance of the user identity allows other services to remember user's search history, subscriptions personalized notifications in the repository (or collection, discussion or any other service). All running applications use the user management functions, however, they are choosing different strategies on how to restrict functionality for different classes of users.

**Reviews (B).** Works in the repository may be reviewed. The review is part of a workflow that is different in journals, conferences or reviews that take place after a paper has been published in a repository that is not part of a traditional peer-reviewed media. Generic and conference review service have been set up. Reviewing service was only implemented for the IAPS 2004 and although fully functional in that case is has not yet been fully integrated with the rest of the platform. It is expected, that reviews will become a seamless part of the platform in the D12 e-Journal Infrastructure due Oct. 31<sup>st</sup>.

**Annotations (A).** Annotation is a private or public comment about some information managed by the SciX platform. Theoretically any information can be commented, however, in the applications we set up, users can annotate repository items. The feature was not built into any application.

**Discussions (P).** Discussions are very similar to annotation with the following differences: they are public and they are structured in the sense that a discussion is followed by another discussion forming a tree-like structure.

**Ratings (P).** Rating service allows users to rate contents of other services. Typically, repository items are rated by the end users.

### 1.1 SCIX METADATA HARVESTER (A)

The harvester is based on The OAI Harvester Open Source project. It is a Java application providing an OAI-PMHv2.0 harvester framework. This framework can be customized to perform arbitrary operations on harvested data by implementing some Java interfaces. The harvester is using the XML API of the repository service to push harvested data into a repository.

### 1.2 LOCALIZATION OF SERVICES (P)

During the project the need has arisen to provide some services in languages other than English. A system for organizing services in other languages was therefore set up. It combines coding convention with software tools that allow for the separation of structure, logic and user interface. Only elements of the user interface need to be translated and this is done in a way that a non-programmer can do this.

## 2. COLLABORATION AMONG THE SERVICES

The Figure 1 charts that the services are talking to each other and the applications using the XML protocols. They do, if the services are located on different servers. If not, a faster proprietary internal communication is used. The choice of one over the other is entirely transparent to the logic of the application. In fact the services are self adaptive in this case and discover the most appropriate means of conversation.

For example, when displaying a paper from a digital library, repository service needs to know if there were any related discussions posted. This needs to be learned from the discussion service. If the discussion service is found to be on the same server, aligned with the repository service, internal query is performed. If not, an XML RPC request is triggered to the URL of the service defined to handle the discussions of papers in the repository.

Particularly the user management is a good candidate for remote service since several applications may want to share a single user database.

### 3. APPLICATIONS

These applications on the Web use the SciX platform.

#### 3.1 DIGITAL LIBRARIES:

- itc.scix.net. (topic archive) Includes knowledge management, discussions, ratings, collections. Semi open access model - user registration needed for most functions, allows end user input.
- iaps.scix.net. (association archive). Rather minimal implementation of the digital library functionality.
- elpub.scix.net (association archive). Digital archive of the Electronic Publishing conference series.
- europa.scix.net (association archive). Digital archive of the EuropIA association that aims to group together a coherent range of discussions to explore advances in design sciences and technology.



Figure 2: Metdata display of the ITC.SciX.net. Note the built in clustering features.



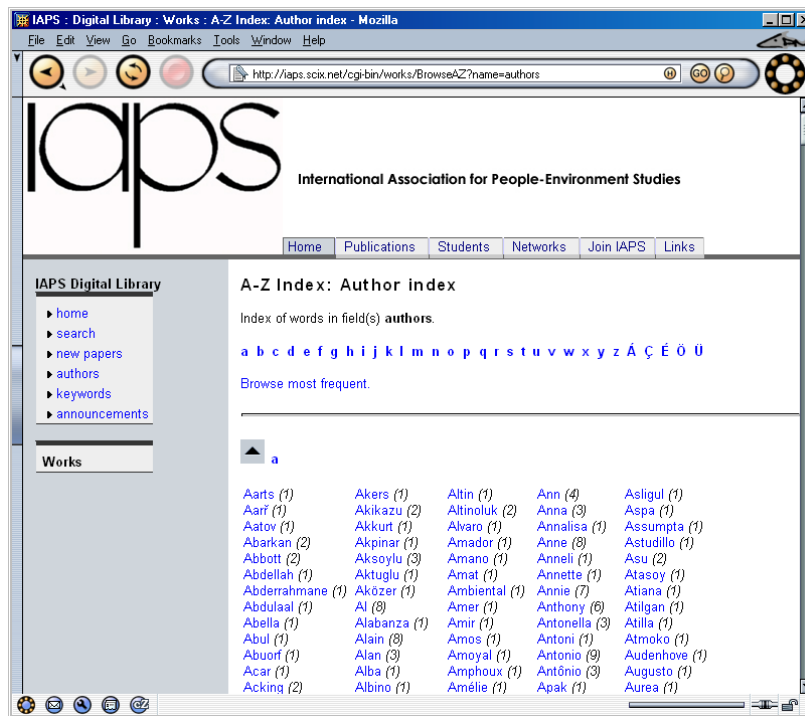


Figure 3: Author index of the IAPS.SciX.net

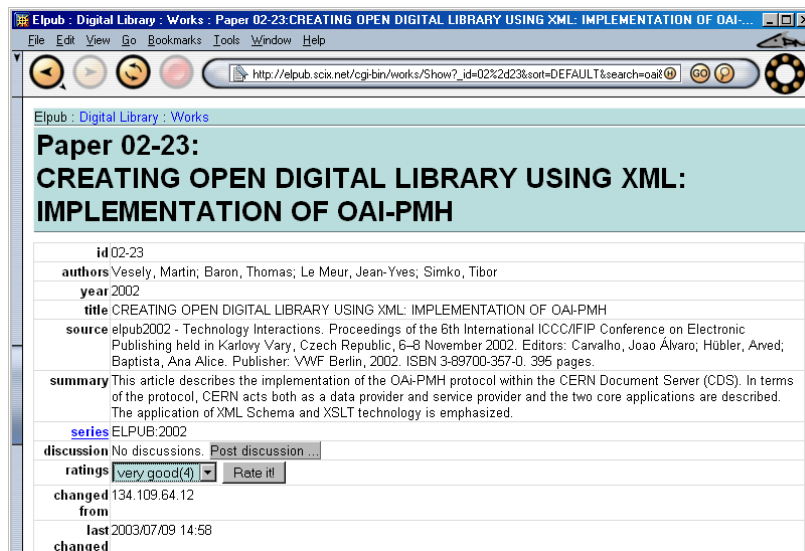


Figure 4: ElPub.SciX.net showing links to ratings and structured discussions in otherwise default style.

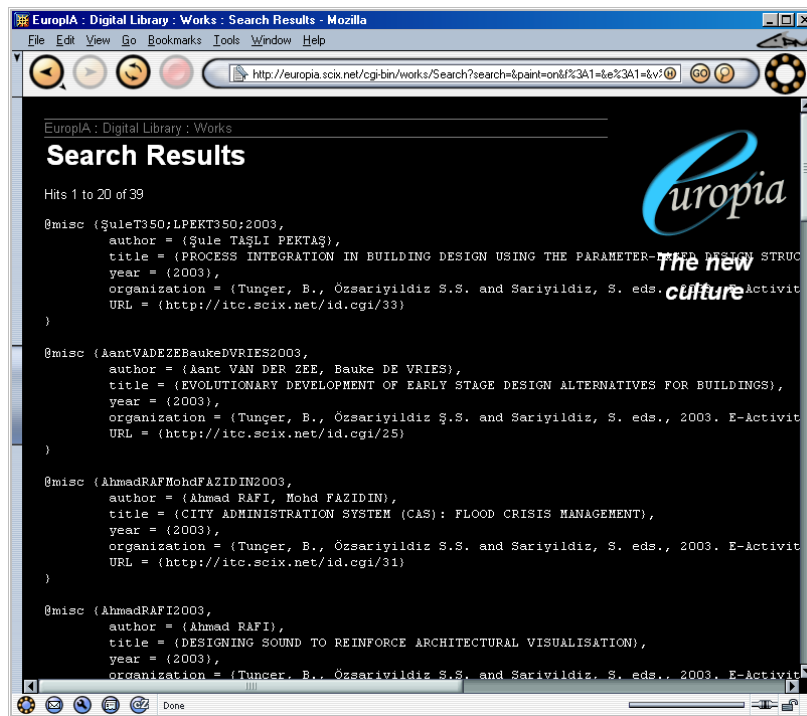


Figure 5: Search results in BibTex format, shown in the EuropIA digital library.

### 3.2 INTERNATIONAL DIGITAL LIBRARIES

These demonstrate that the SciX platform can be efficiently translated to other languages and set up:

- filozofija.scix.net - (institutional archive) Rather bare bones institutional library with philosophical texts in Slovenian languages.
- arhitektur-informatik.scix.net - (topic archive) Support of a German speaking association on the use of IT in Architecture.
- raumplanung.scix.net - (institutional archive) Institutional archive of the TU Vienna.

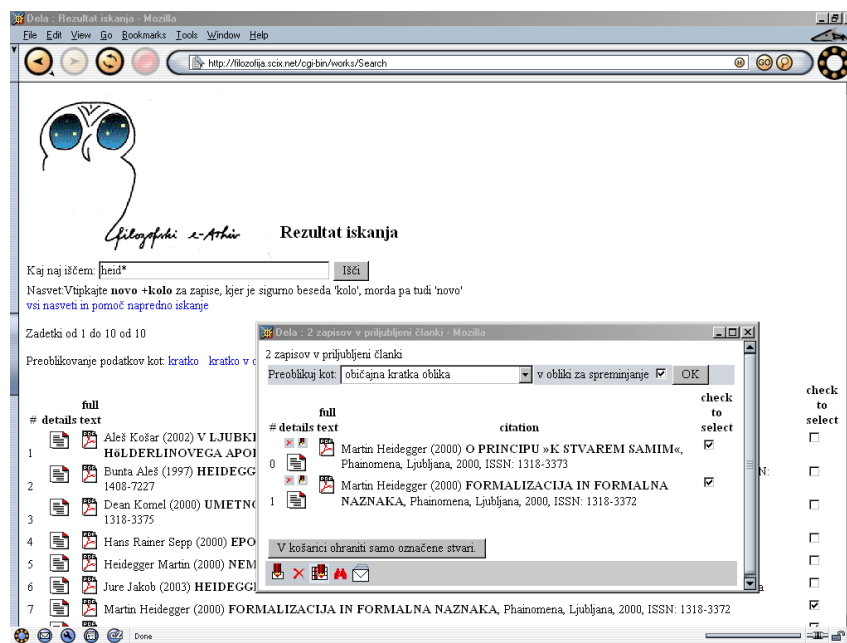


Figure 6: Slovenian philosophical archive - showing the main window as well as a selection of favourite papers.

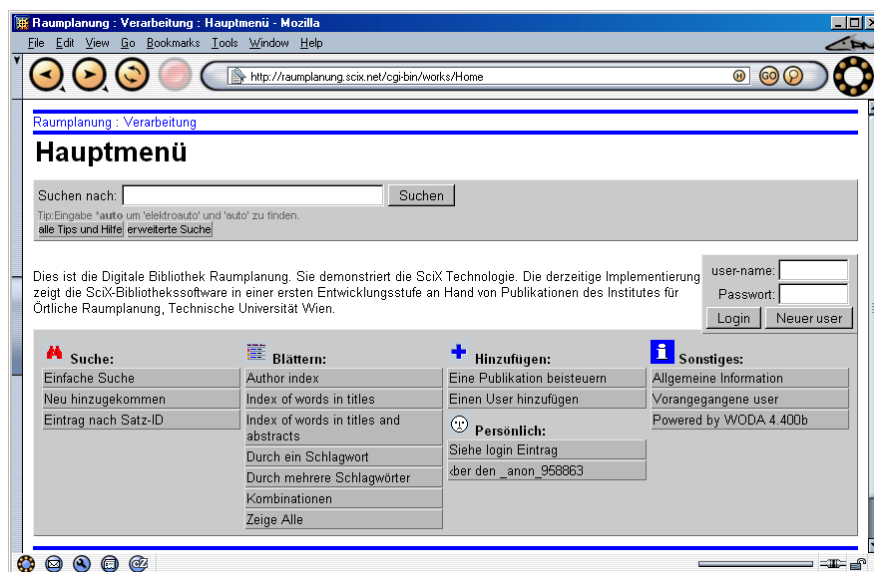


Figure 7: Raumplanung.scix.net - the main menu of a German speaking digital library. Some translations may be less than perfect because they were done by translators unaware of the full context of the translation.

### 3.3 ELECTRONIC JOURNALS

- www.itcon.org. The journal that has been publishing electronically for the last 8 years was transferred to the SciX platform. Implements a very liberal access model - even anonymous users have full access to full content. Does not (yet) implement reviewing. This is otherwise part of the D12 "e-journal infrastructure".

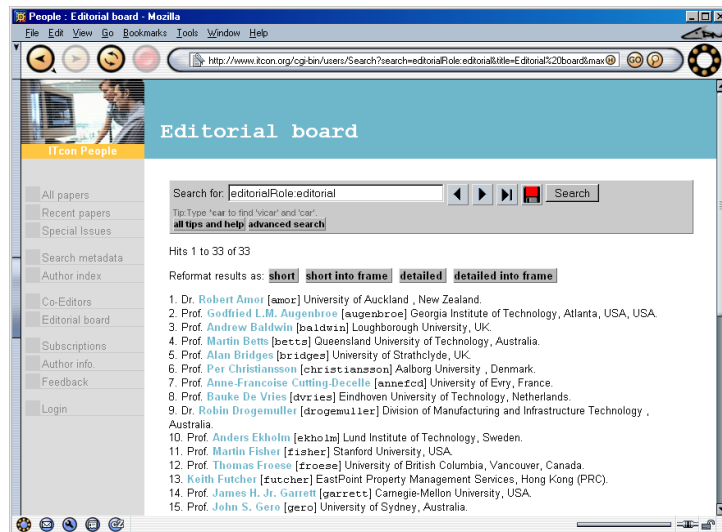


Figure 8: Display of the editorial board that uses the user management service to manage all people related to the journal.

### 3.4 CONFERENCES

- iaps2004.scix.net is a conference where the full workflow of the submitted papers is managed by the SciX platform.

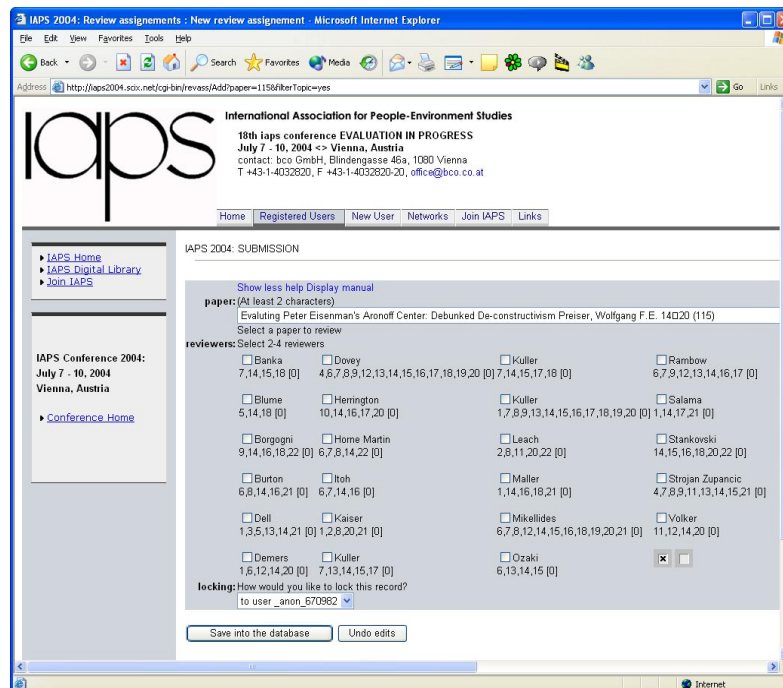


Figure 9: Assignment of a reviewer based on keywords in the paper and expertise of the reviewers. In the iaps2004.scix.net.