

Demo Portal of Information and Documentation in Science and Technology

Doina Banciu, Dora Coardos

National Institute for Research and Development in Informatics - ICI Bucharest

doina.banciu@ici.ro, coardos@ici.ro

Abstract: The paper presents part of the achievements of scientific research on information-documentation and communication systems. In fact, it is about designing and creating an electronic data warehouse, part of a demo pilot system for information-documentation and communication. This data warehouse contains information about research project achievements that have been obtained through national research programmes. The portal allows computer users to online access to projects in the fields of science and technology, and their results. Also, the paper gives a broad description of an on-line training system that aims at teaching the necessary skills and abilities required in the information distribution and documentation process.

Keywords: information and documentation, communication, data warehouse, CERIF standards, training system, Web portal.

1. Introduction

It is widely agreed that, in this century, digital information [5] and computer networks are the main engines of economic growth and social development. Innovation is the guarantor of development.

The environment in which the research is conducted and the research results is disseminated known profound transformations, related to new information and communication technologies that enable new opportunities and changes in research practices.

To have a real impact in the real life, research results must be disseminated, known and assimilated by the potential users. This implies the need to use modern methods to access to scientific information.

Access to global scientific and technological information sources is a necessity and managing those sources has become an economic and political challenge in order to ensure unrestricted access for every citizen.

The field of documentation and information software in science and technology is very complex in the world, with important features in various developed countries.

We cannot speak of the existence of national systems to focus their databases for research and development results of projects financed from the national budget or international programs. We can speak rather of distributed information, located to developers, or in the databases of electronic publications, or to organizers of events that intends to disseminate research results.

The creation of information systems for Research and Development is signaled as a strong tendency both in USA and in EU member states. The majority of these systems are based on a close cooperation among institutions that coordinate [2], at a national level [10], the R&D activities, and organisations directly involved in those activities, such as institutes or major research laboratories along with universities.

The information and documentation systems have behind well-structured relational databases, with interfaces equipped with search engines adequate to the data bases. Usually the access (even to summary information) can not be done without filling out an online registration form in the system.

Only a few countries have developed national research portals with similar characteristics. In most cases these portals are managed by a subordinate structure for technology transfer and innovation. Extremely well-classified specific information is made available on websites in the UK, France, Switzerland, Canada, Japan etc.

A significant example of a dedicated computer information system is the CORDIS portal [9] that contains information pertaining to the results of R&D projects conducted under concluded European framework programmes.

Complying with the worldwide trends and the EU recommendations, in Romania it is very important to create and implement information systems for R&D in various fields, accessible through Internet, integrated into an informational network for the

Romanian R&D, integrated, in turn, to the European informational network for R&D.

Achievements in our country in this area, by promoting local research sites or Internet portals, are comparable to those of the European and international level.

At a national level, the information related to results of state-supported R&D projects is stored in databases managed by the authorities that conducted the programmes. This information is not yet accessible to the public nor to the Romanian scientific community.

Because of that we have taken into consideration the implementation of a "*Demo Portal of Information and Documentation in Science and Technology*" that would take this information, store it into a database within the system and make it available to those who are interested, through a Web portal-type platform.

2. System Presentation

Systems used in information and documentation process are backed by well-structured, relational databases, have interfaces equipped with search engines for these databases, and usually, the access (even to summary information) cannot be provided without on-line registration.

The system for information-documentation in science and technology is designed and created on two main components:

- SNIUREST - National educational system for training in usage of electronic resources on accessing the international fund of literature on science and technology;
- SIDCDEST - The data warehouse-type system for information-documentation, for access and usage of the national fund of literature and achievements in science and technology.

Both components will be accessed through a Web portal-type platform. The two main components have been designed in a unitary manner, they are compatible and coherent.

2.1 Short presentation of the SNIUREST component

National educational system for usage of electronic resources in accessing the international fund of literature on science and

technology is associated with a shared computer system, designed to serve a nationwide network of support entities.

Creating a network of support entities in information-documentation and communication activities requires the creation of an information and communication environment, an electronic network that would include the support entities. In this respect the support entities will be connected through an informational network accessible through the Internet. A central server provides services for this network. The network has functions that assist the training of staff to access external information resources in the field of science and technology.

The system offers search functions oriented on multiple criteria, based mainly on categories of information that will be defined (patents, articles, books, scientific reports, etc.) on type of information depository, searchable by authors, country, availability, etc. Data related to available information resources will be stored in a unique database, on the server.

The system has also a component of on-going training for the network's members, aiming to the acquisition of the necessary skills and abilities required in the information and documentation field. For this purpose, the system also includes a platform for distance training, platform that is to be created within this project.

At this level, within the system, the Repertoire of electronic pedagogical resources in information-documentation and communication training will be created and implemented. This will be stored in the database, on the main server. The contact data and activities for the support entities will also be stored there. The system will be provided with a separate module for web-based maintenance.

The aim of the *Repertoire of electronic pedagogical resources* is on one hand improvement of quality of continuous training, through top technological specialisation, through offering insider news and through improving the assets needed for training in usage of electronic resources for information-documentation and communication [3], and, on the other hand, the portal will be

accessible to anyone interested in the training process in this field.

As presentation of electronic resources for storing the system inside the database, there were proposed: *The title, the link itself and a short description* of the content (where this applies)

On the curriculum aspect, this has to contain newcomers' courses and advanced level courses.

These courses refer to:

- *Information and Communication Technologies,*
- *New technologies in Information and Communication – basic notions about Web pages;*
- *Information Literacy.*

2.2 Short presentation of the SIDCDEST component

The data warehouse-type system for information-documentation, for access and usage of the national fund of literature and achievements in science and technology is based on two main components:

- *Digital pilot platform* for information-documentation and communication on national research programmes and projects, and their results;
- *Pilot database* containing the national fund of literature and achievements in science and technology, with reference to the international indexing systems.

The digital pilot platform is employed both in extracting and displaying information from the database to the user, and in up-dating on-line the database by the personnel authorised in this respect.

The platform will contain automated updating modules with information from other databases with which it would collaborate on-line. In this respect web services will be used.

The platform offers users a unique, ergonomic, intuitive, interactive interface and having a solid search-engine for the database.

The access in the system will be governed by defining and implementing a system of access rights, oriented on several users' categories, namely:

- Technical administrator;

- Contents administrator;
- Personnel that has the right to update data stored on line;
- Common user.

Pilot database containing the national fund of literature and achievements in science and technology, with reference to the international indexing systems is situated on the first level of system's architecture (the data level). On the same level the database management server is located.

The database is a relational-type one, and it contains restrictions that would ensure the integrity and the coherence of the data. Data updating and adding new information will be achieved through the digital platform described above. For uploading information into the database the following methods will be used:

- Automated feeding, using *web services* over the Internet;
- Feeding by introducing data in on-line forms provided by the information system that is to be created;
- Feeding by uploading files on-line through the interface provided by the system; the uploaded files will be processed automatically on the application level, useful data being automatically extracted and introduced in the database;
- Manual feed, through the web-based maintenance module that will be created within the project;
- Manual feed, directly on the server (*back-end*).

3. Solutions Used in Designing and Creating the System

In designing the system the characteristics that define a modern and efficient system were taken into account, including:

- A modular structure, opened, flexible and scalable;
- Usage of Web technologies that allow access through Internet using merely a Web browser.

Due to the dynamic character of Web applications [11] and the complexity of applications required by the system we have chosen a *three level client/server architecture*.

The *three-tier* architecture model is characterised by:

- *The data level*, where we have the database and the database maintenance server. At this level, information is stored and extracted. Separating the data on a dedicated layer improves scalability and global performances;
- *Application level*, or application server (connecting level) in charge with application functions by undertaking complex processing of the information;
- *Data presentation level* includes the Web server and the client (a standard Web browser) is the highest level of the application and deals with displaying the information, exchange of data / results with the inferior level and is represented by the interface visualised by the client through an Internet browser.

Graphically, such architecture can be represented as it can be seen in Figure 1.

fund of literature and achievements in science and technology.

At database level, the database will be maintained by a fast and reliable dedicated server. We are considering using one of the *MySQL* servers, *MySQL Server*. All these servers provide backup and data restore mechanisms, so that the developers may define a coherent and reliable strategy of periodically saving the data.

At the application level *PHP* software developing environment will be used, given the great degree of interoperability [7] with other environments and the interfaces that it offers to all kinds of database servers [1]. On a limited scale we look into using *JavaScript* modules.

The data presentation level will be realised using *HTML* (directly or as a result of *PHP* programming). Control of display styling will be ensured by *CSS (Cascading Style Sheets)*.

Hardware-wise, the system for information-documentation in science and technology implies a platform composed by two servers (solid and reliable, one as application and Web server, the other as database server), with a limited number of local station for

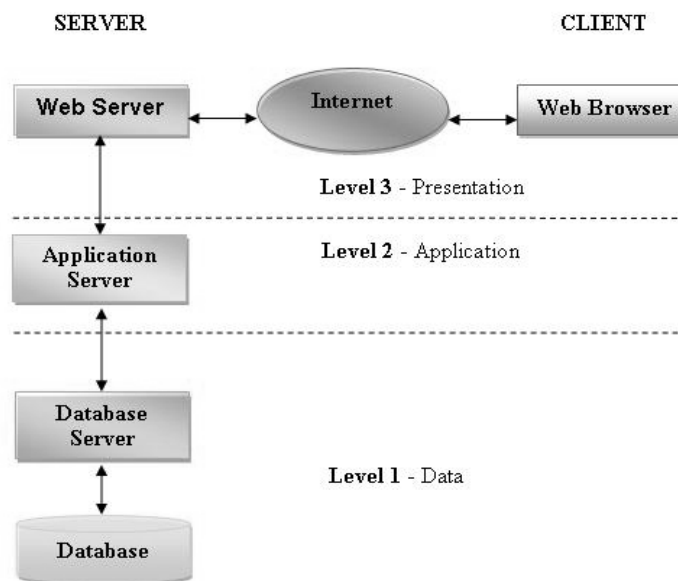


Figure 1. *Three-tier* architecture model

On levels 2 and 3 the pilot digital platform for information-documentation and communication concerning national research programmes and projects and their results is realised and implemented, while on level 1 there will be the pilot database containing the national

administration / off-line data processing, workstations for the support entities registered to the network created (one workstation for each support entity).

The support entities designated for the demo period are: the libraries of Bucharest, Cluj, Iași and Timișoara Universities, and as a research library: the Romanian Academy Library.

4. System Database

The system's database is a relational-type, accessible through the Web.

In designing the system database the EU recommendations contained in the CERIF standard (Common European Research Information Format) have been taken into consideration. The format offered by the CERIF standard [4] is a modern and efficient solution in achieving interconnection of various databases that contain information related to European Research.

CERIF is a model created as a support for the management of information contained by research projects and their outcomes, providing a set of rules that facilitate interoperability amongst various systems of

information-documentation and communication.

From the point of view of the concept, the CERIF model is composed by entities and relationships. Among the CERIF entities one can point out the following types: (a) basic entities that form the core of the model, (b) result entities, and (c) second level entities. Structure of this model with entity types is illustrated in Figure 2.

Based on this conceptual model, the database for the digital pilot demo platform for information-documentation and communication of the system proposed for accomplishment is going to contain, two main types of entities, namely, the *core* (projects, people, organizations) and the *results* (publications, products, patents), the level (c) being represented by the *Programme* entity (funding sources).

The database core consists of *People*, *Organisations* and *Projects*. Figure 3 presents the basic entities, relationships between them and recursive relations.

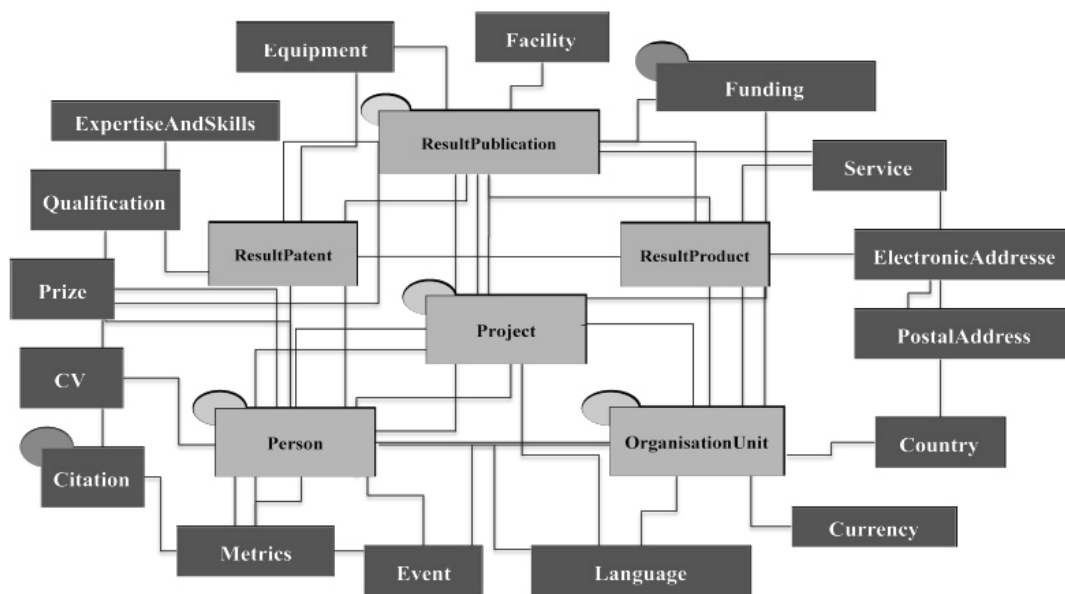


Figure 2. The conceptual CERIF model

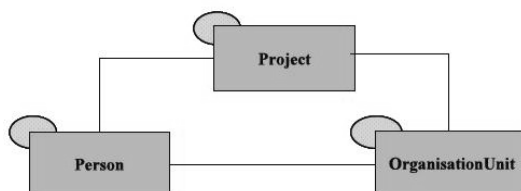


Figure 3. Database core structure

Each entity has recursive links to it and maintains relationships with the other entities of the core. The detailed graph of the database is represented in the Figure 4.

The entities of level (b) - *Results* in the CERIF model are *ResultsPublications*, *ResultsPatents* and *ResultsProducts*. Figure 5 illustrates entities and types of relations existing at this level, entities representing the result of research undertaken during the projects.

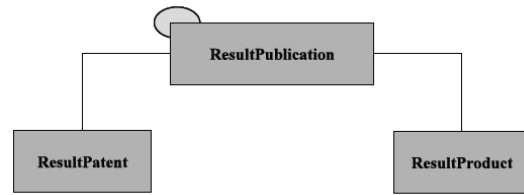


Figure 5. Entities of the **Results** level

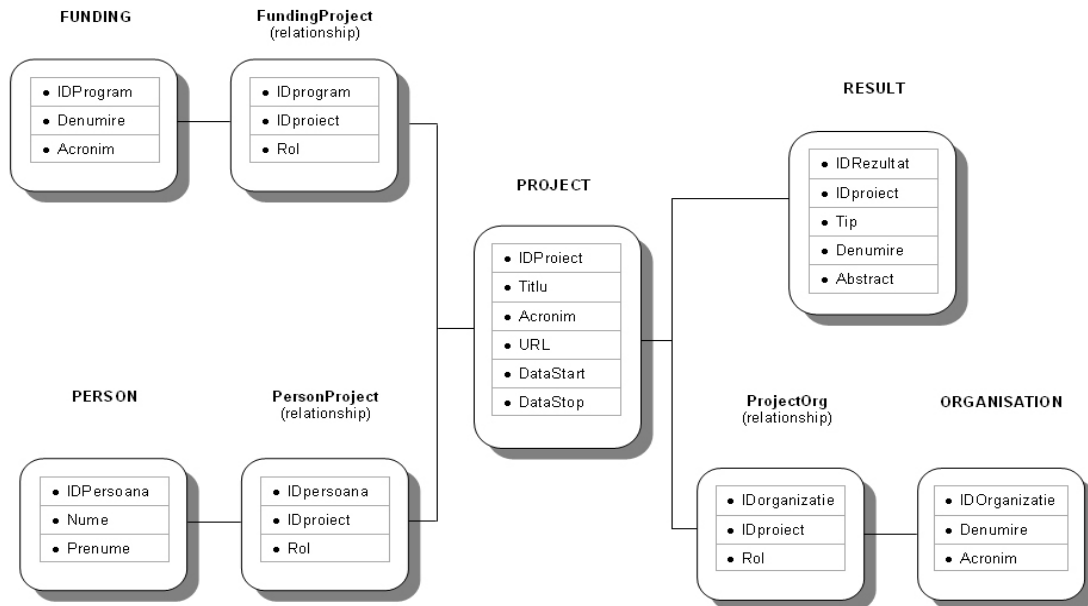


Figure 4. Detailed graph of the database

Results - type entities allow storage of information related to identification and description of types of research results undertaken during the projects.

In the database structure, *Results*-type entities are correlated with the *Projects*, *Organisations* and *People* entities, as it can be seen in Figure 6.

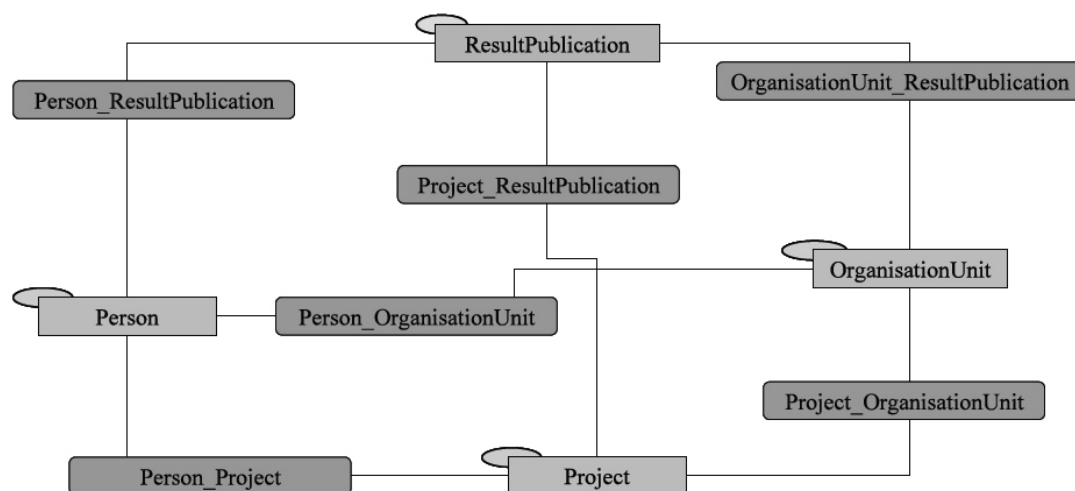


Figure 6. The linking entities within the database

Results - type entities allow storage of information that enable the identification and description of types of research results undertaken during the projects.

4. Portal Presentation

The main goal of the Web portal is to ensure a simple, logical and intuitive interaction, between the system and the user, centred as much as possible on the user.

The portal offers registration services, authentication and authorisation, grants access to resources, and enables data security and protection according to the user's access rights.

Supplementary for administrators and developers there are other services, such as, personalisation, security and protection against Internet threats, as well as, special formats for data exchange and advanced management functions.

- Communication management (editing and distributing newsletters, creating mailing lists, discussion forums, FAQ);
- System management (the administrator would have access to all resources, both through the Web interface where it should be identified through a username and password and directly into the system (back-end)).

The start page of the Web portal is represented in Figure 7.

Regardless of the selected component, all pages are designed in the same manner.

For each of the system's component, on the top page, the user finds, in the central area of the page, a general presentation of the selected system, than, on the top and on the left side, there are horizontal and vertical menus for system navigation. To the left there are the main sections of the application,

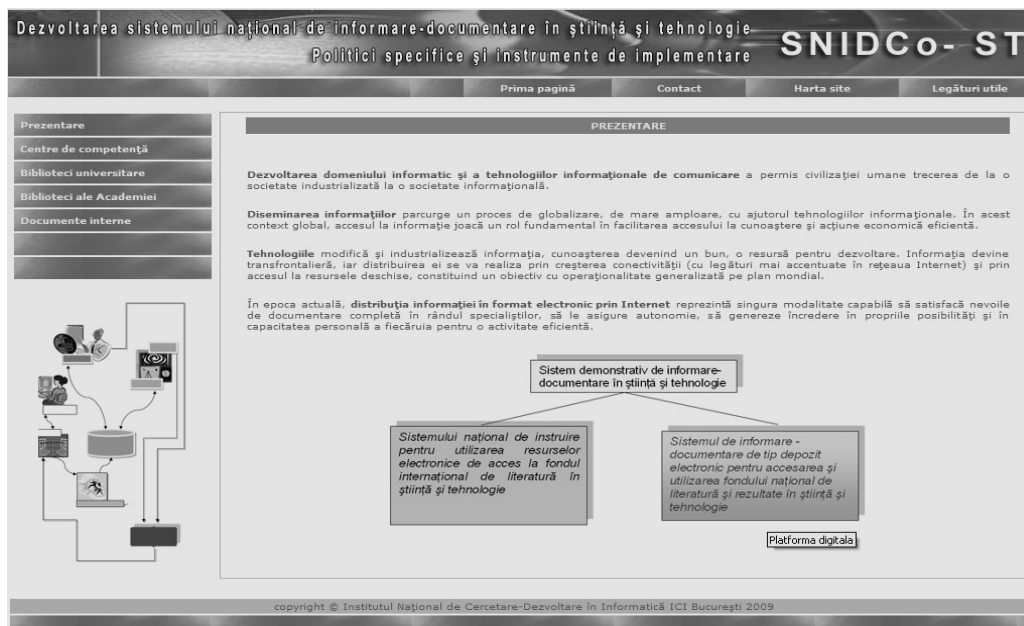


Figure 7. The start page of the Web portal

The main functions of the *information-documentation system's interface* are the following:

- Information management (collecting, processing, storage, delivery to the users, in a suggestive and useful manner);
- User management (registration, granting access rights, personalisation, interactivity, securing private data);

allowing different access of users to the information available through the training system.

The central area is a dynamic zone, changing upon the option selected in the user's menu and containing the specified information.

In the lower part there is information regarding copywriting. The upper and the lower part are static and don't change from one page to another.

Regardless of the selected component, according to the top page of each component, a visitor can *register*, if it is his/her first visit, choosing a username and a password, or can *authenticate*, if he/she has registered previously. Registration is not compulsory for common visitors that only want mere information.

Exploiting the system implies usage of *dialog interfaces* with users, through two key elements: forms (screen layouts) and results presentation reports.

One of the key components of the *data warehouse-type system for information-documentation, for access and usage of the national fund of literature and achievements, in science and technology* is offered by the *on-line data acquisition module*.

- Organisations participating in R&D&I programmes;
- People participating in R&D&I programmes.

Each Web form can be accessed individually. Each blank field of the acquisition form contains a description of the field itself, avoiding confusion.

Validation of data collected through Web forms depends on the type of the blank field, as follows:

- For the *hidden list* type, the user is lead into choosing one of the valid variants offered by the Web form, eliminating the risk of introducing void data;
- For the *text* type, there have been created

Figure 8. Data acquisition form

The on-line data acquisition represents a modern method of data harvesting, available – on the Web – through *on-line forms*. A Web form contains blanks on a Web page and the user fills information in. The communication manner ensures the interface among applications, components and application modules. By gathering necessary information for uploading on-line the system database, the Web interface contains the following types of forms, used in collecting information [8] in regard with:

- Research-Development and Innovation programmes;
- R&D&I programmes' modules;
- R&D&I projects;

two functions: one that tests the compulsory fields for null value and the other that tests the value of numerical fields in order to avoid non-numerical values.

Fields marked with [*] are compulsory. Figure 8 presents such a data acquisition form.

Another essential component of the interface deals with *information search and display*, regarding research projects, upon different criteria, such as: type of program, organization, type of project, the person involved, the project results, etc.

Types of allowed searches: *simple search* - the user can use a single searching criterion, *advanced search* – the user uses simultaneously one, two or three searching

criteria, linked by logical operators and *browsing search*.

In the Figure 9 we present an example of advanced search.

purely informative each visitor being able to opt to access one of the two main components of the system.

Figure 9. Advanced search

The interface can be accessed from any point of the Internet, without any restrictions related to IP addresses or networks the users connect from.

5. Conclusions

The intended portal has a high degree of complexity, aiming to create a national network of support entities that would conduct teaching and training activities in the field of information-documentation and communication.

The modern solution adopted is to introduce Web services as a mechanism, safe and efficient, in data exchange over the Internet, between the information system proposed and those that exist in the entities depository of the scientific and technical information.

The Web interface ensures the interactive side of the application, being the method of access from user to the entire information-documentation system.

The system offers a security mechanism, based on definition by the system administrator of information access rights for users in relation with the role these have within the system.

The first page of the portal for information-documentation in science and technology is

The system addresses professionals in information and communication science, as well as, users, especially, within the scientific community.

Acknowledgement

This work was supported by the research project funded by ANCS under 15/2009.

REFERENCES

1. ANGHEL, T., **Dezvoltarea aplicațiilor web folosind PHP și AJAX**, Editura Edusoft, 2008.
2. **ARIPIST - Aquitaine-Limousin Réseau des Professionnels de l'Information Scientifique et Technique**, <http://www.arpist.cnrs.fr>
3. BANCIU, D., **e-Romania - A Citizens' Gateway towards Public Information**, Studies in Informatics and Control, 2009, Vol.18, Issue 3, pp. 205-211.
4. CERIF 2008 - **1.0 Full Data Model (FDM)**. http://www.eurocris.org/fileadmin/cerif-2008/CERIF2008_1.0_FDM.pdf
5. Comisia Europeană, **Europe's Digital Competitiveness Report – Main**

- Achievements of the i2010 Strategy 2005-2009**, 4 August 2009, [http://www.epractice.eu/files/Europe's Digital Competitiveness Report - Main achievements of the i2010 strategy 2005-2009.pdf](http://www.epractice.eu/files/Europe's%20Digital%20Competitiveness%20Report%20-%20Main%20achievements%20of%20the%20i2010%20strategy%202005-2009.pdf)
6. CALVARY, G., C. PRIBEANU, G. SANTUCCI, J. VANDERDONCKT (Eds.): **Computer-Aided Design Of User Interfaces V**, Proceedings of the Sixth International Conference on Computer-Aided Design of User Interfaces, CADUI 2006 6-8 June 2006, Bucharest, Romania. Springer 2008, ISBN 978-1-4020-5819-6
 7. IVAN, O., **Interoperabilitatea bazelor de date de cercetare**, Al V-lea Forum Deschis pentru Inovare și Transfer Tehnologic, 14 - 15 martie 2006, București, http://www.technologyforum.ro/Docs/C2006_Oana_Ivan.pdf
 8. **Planul Național de Cercetare, Dezvoltare și Inovare pentru perioada 2007-2013**, http://www.mct.ro/img/files_up/1188313421PN2%20ro.pdf
 9. Portal CORDIS, http://cordis.europa.eu/home_en.html
 10. RENATIS - **Réseau National des Professionnels de l'Information Scientifique**, <http://renatis.cnrs.fr/>
 11. ULLMAN, L., **PHP si MySQL pentru site-uri WEB dinamice**, Editura Teora, 2007.