JINR Document Server - Open Access Repository of the Joint Institute for Nuclear Research: Current Status and Future Plans

I.A. Filozova[1], S.V. Kuniaev[2], G.G. Musulmanbekov[1], R.N. Semenov[1], G.S. Shestakova[1], P.V. Ustenko[2], T.N. Zaikina[1]

[1] e-mail: fia@jinr.ru, Laboratory of Information Technologies, JINR, Dubna

[2]e-mail: svk@jinr.ru, Automated Management Information Systems Department, JINR, Dubna

Introduction

Nowadays various institutions and universities around the world create their own repositories, depositing there different kinds of scientific and educational documents making them open for the world community. Open Access to Research — a way to make scientific results available to all scientific and educational community by the Internet. The annual growth of the numbers of repositories and deposited in them records are observed according to statistics given by the Registry of Open Access Repositories (ROAR – http://roar.eprints.org). Today there are 2453 Open Access (OA) repositories with scientific research documents created in the frameworks Open Archive Initiatives (OAI) [1]. It contains about 21 millions or records. Besides, such kind of initiative has been put forward in education, as well, Open Educational Resources [2]. Open Access in science is a way to collect, preserve the intellectual output of scientific organization and disseminate it over the world. This is the aim of the Open Access repository of the Joint Institute for Nuclear Research, Jinr Document Server (JDS) which started its functionality in 2009 [3, 4]. In this paper we describe some peculiarities of filling and depositing documents, possibilities of the visualization of search and navigation and ways of the further development of the information services at JDS.

JINR Document Server Collections

Building the institutional repository has as its objects to make accessible the scientific and technical results of JINR researchers for the international scientific community, to increase the level of informational support of JINR employees by granting an access to other scientific OA archives and to estimate the efficiency of scientific activity of JINR employees. JDS has been built as an OAI-compliant repository to realize these goals. JDS functionality, supported by the CDS Invenio software [5], covers all the aspects of modern digital library management. JDS, integrated into the global network of repositories ROAR, makes its content available for everyone anywhere at any time. The content of JDS is composed of the following objects:

1. The research and scientific-related documents of the following types: publications issued in co-authorship with JINR researchers; archive documents that describe all the essential stages of the JINR research activity;

2. Tutorials, various educational video, audio and text materials for the students and young scientists;

3. Documents providing informational support for scientific and technological researches performed in JINR.

The objects stored in the repository are grouped into collections: published articles, preprints, books, theses, conference, proceedings, presentations and reports, reports, dissertation abstracts, clippings from newspapers and magazines, photography, audio and video materials [6]. All these collections are arranged hierarchically into two trees: the basic, or regular, and virtual. The basic tree (left column in Table 1 in the JDS is formed according to a classification feature – the genre of scientific publications, and the virtual tree – by the subject of publications (right column in Table 1). The subject collection allows one to perform selective search. It may be advantageous to present a different, orthogonal point of view on nodes of the regular tree, based on the other attributes.

The developed user interface of JDS provides a wide range of information services: search and navigation, creation of groups by interest saving the searched results, personal and group bookshelves, deposit of manuscripts and arrangement the discussions on them, sending out notices and messages.

The main point for a newly created repository is how to fill its content by the documents of interest issued before. We used various methods of filling and updating the JDS content of the publications of JINR authors: automatic data collection (harvesting) from arXiv.org and CERN Document Server (CDS), other OAI – compliant archives; semi-automatic collection of documents from retrieval databases SPIRES, ADS, MathSciNet.

The second point is how to deposit manuscripts, preprints and new publications of JINR authors. A little earlier the Personal INformation (PIN) data base has been developed in JINR where research workers deposit their personal information in.

Table 1: JDS Collections in basic and subject trees

Narrow by collection:		Focus on:
Articles & Preprints(32,736)	Articles(14,867)	JINR Articles & Preprints(23,652)
Preprints(15,403)	Conference Papers(3,248)	JINR Published Articles(12,744)
Books and Reports(1,653)	Reports(166)	JINR Preprints(13,640) JINR Conferences(177)
Books(1,487)		
Conferences, Presentations and Talks(19,427)		JINR Annual Reports(137) JINR(12) VBLHEP(13)
Conferences Announcements(4,099)		BLTP(19) FLNR(16) FLNP(28) DLNP(10) LIT(20)
Conferences Proceedings(15,326)	Lectures(0)	LRB(12) UC(7)
Notes of Schools and Seminars(0)	Talks(2)	
Notes(0)		
INDICOSEARCH(0)		High Energy Experiments in JINR(179) FASA-3(1)
INDICOSEARCH.events(0)		MARUSYA(3) EDBIZ(0) BECQUEREL(0)
INDICOSEARCH.contribs(0)		NUCLOTRON & NUCLOTRON-M(164) NIS(0)
INDICOPUBLIC(0)		NICA/MPD(19)
Handbooks & Manuals(0)		Heavy Ion Physics Experiments in JINR(9)
		ACCULINNA(1) DRIBS(1) DRIBS-2(0)
		CORSET-DEMON(0) MASHA(0) VASSILISSA(7)
Theses(8) [restricted]		Non-accelerator Neutrino Physics & Astrophysics(162)
Multimedia (3) Press(3) Audio(0) Videos(0)		BAIKAL(106) EDELWEISS & EDELWEISS-II(10)
Pictures(0) Posters(0)		GERDA(9) GEMMA & GEMMA-II(5) LESI(0)
		NBulletins(0) STL Bulletins(0) EMO(32)
Bulletins(0) STL Bulletins(0)		External Experiments(1,055) SPS(219) FAIR(59)
		RHIC(321) LHC(459)

In addition to other personal data (affiliation, collaborations, participation in various projects, experiments, teaching, grants etc.) PIN includes their publications (bibliography and full texts). With the aim not to force the authors to deposit their manuscripts and publications twice (in PIN and JDS) we set the communication channel to import data from PIN to JDS. It delivers PIN data in MARCXML format which are then uploaded into JDS. Furthermore, preprints issued by JINR Publishing department (bibliography and full texts) are uploaded into JDS without author's participation. Nevertheless, authors can (by their desire) deposit their manuscripts and publications in JDS in the 'self-archiving' or by proxy mode. Collection 'INDICOSEARCH' with its subcollections is intended for searching in the administrative data base ADMIN, composing various events like committees, lectures, reports, meetings, conferences etc.

The data base was created and is managed by the elaborated in CERN 'INDICO' software [7]. Inasmuch as Indico provides OAI–compliant output format, it is possible to harvest data from this database.

Visualization of Search and Navigation

The design and usage of visual interfaces to digital libraries is becoming an active and challenging field of information visualization. The aim of visualization is to make the usage of digital library resources more efficient in reduced search time, provide better understanding of a complex data set, reveal relationship among objects (authors, documents), make data to be viewed from different perspectives. Visualization of search and navigation allows analysing the search results more efficiently. Thus a visual interface for search and navigation in a digital library should meet the following requirements: plain representation of search results for better identification; finding of interrelations among documents; improved search engine; graphical vision and navigation in a digital library; mapping of user's operation with available documents with the aim of better functionality of a digital library.

With the aim of usage in JDS the following tools designed for information visualization were analysed: Java Universal Network / Graph Framework (JUNG); JGraphT; JavaScript InfoVis Toolkit (JIT); Graphviz; Prefuse (set of software tools for creating rich interactive data visualizations). The package Prefuse, as meeting all our requirements, has been chosen to visualize the JDS resources [8].

Two visual prototypes were developed by Prefuse on the part of real JDS data. The tree-map method proposed in paper [9] is applied to display the collection of JDS documents by the subject (Fig. 1). Each rectangle (collection) of the tree-map has a certain color; the larger the area of the rectangle, the more articles in the topic. The next level of the hierarchy displays records, and a final — publication in the form of interactive display rectan-
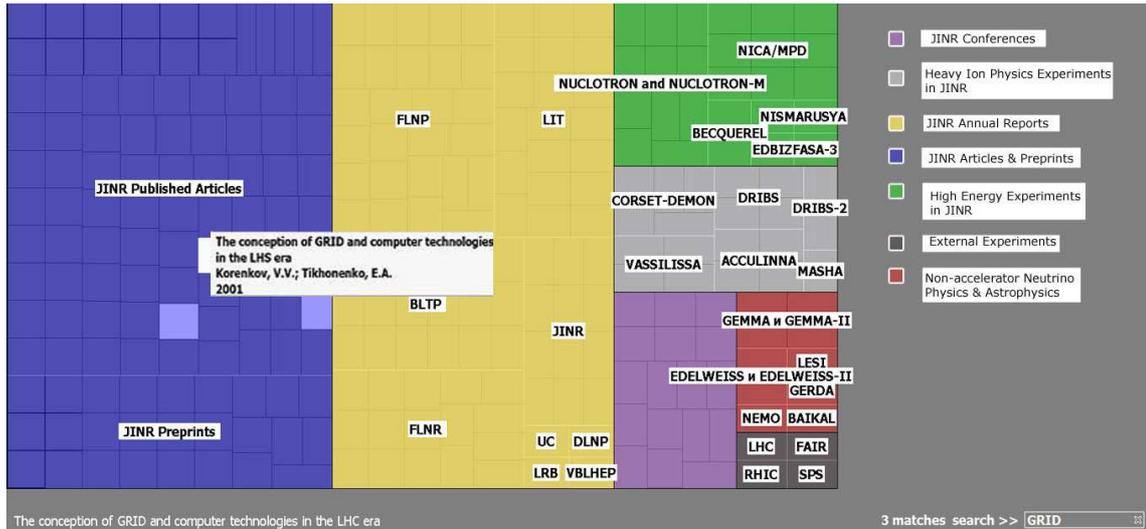
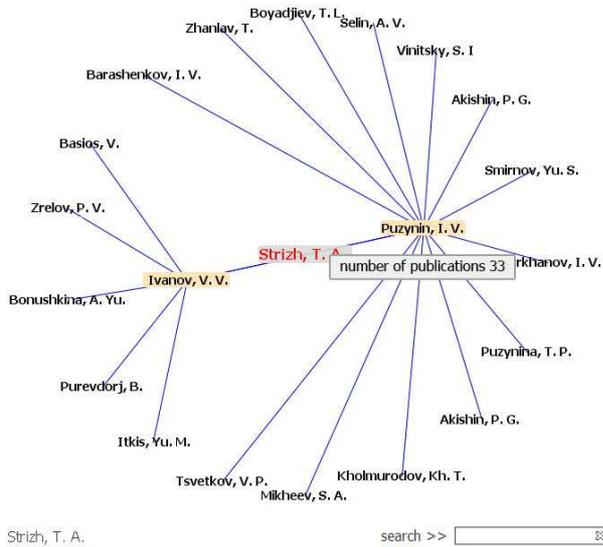Figure 1: Visualization of JDS Resources: Tree Map Method

Figure 2: Visualization of JDS Resources: Radial Graph Method

gles. The visual representation also contains the search path and found document which is highlighted. The publication is a reference to itself. When user points cursor at the publication, the information about the author and publication date is displayed. Visualization created with the usage of the threaded tree allows one to explore the repository content by subjects. The search result sets out in more light tone that allows one to estimate visually the number of publications that the search finds. It supports zooming and panning which allow one to display more detailed information.

Usage of the radial graph method demonstrates a relationship between authors and co-authors of publications (Fig. 2). An author appears at the center, co-authors on the second concentric circle, and co-co-authors on the third. To avoid over-saturation, the only three levels of hierarchy are uses in rendering. An author and his co-authors are highlighted by appropriate color. When cursor selects the author/co-author the information about the number of his publications are displayed.

JDS: What is next?

With the aim to minimize the author's efforts to deposit his publications we are planning to arrange the reverse channel delivering documents (bibliography and full texts) from JDS to PIN. The program of the visualization interface is continuing – visualization of search results, statistics and monitoring. JDS users may join in groups by interests and are able to interact with each other through the service tool WebGroup to discuss on actual publications using the module WebComment. The module WebComment provides a socio-oriented tool for the discussed document ranking by readers. WebMessage facilitates clustering of users in groups via a web-forum. JDS has a custom module WebStat, providing statistics collection by some parameters such as the number of calls, downloads, citations, etc. Thus, all these service tools, WebStat, Web-Basket, WebGroup, WebMessage, WebComment facilitate to form a social network within the scientific community in the framework of the information system. The following elements for visualization of the user's groups as a social network will be developed: browsing the groups and its members, browsing the detailed information about group members, browsing relations between the groups, visual search by the user.

Results of the research, scientific and engineering efforts represented in publications have semantic re-

lations between them via quoting mechanism. Description of these relationships and their properties opens up new possibilities for studying the document corpus of digital libraries. Publications, as well as collections of publications, contain other semantic linkages which reflect the logic of author's thought on certain subjects. These linkages can be studied and described. There are other types of linkages too. For example, the linkages relating to personal profile – the author of publications, his affiliation, participation in the collaborations, experiments, projects, etc. So, the personal profiles of authors are formed in the framework of the digital library. These profiles are a basis of a scientific social network. By analysing these linkages, we can obtain important information about motivations of these interactions. For example, intensity of relations between authors, working in a scientific area (subject project) illustrates the degree of activity in this area, etc. We see the following directions of JDS development: creation of the JINR social science network, elaboration of semantic search and navigation.

References

[1] Open Archives Initiative:
http://www.openarchives.org.

[2] Open Educational Resources:
http://en.wikiversity.org/
wiki/Open_educational_resources

[3] I.A. Filozova, V.V. Korenkov, G. Musulmanbekov. Towards Open Access Publishing at JINR. – Proceedings of XXII International Symposium on Nuclear Electronics and Computing (NEC'2009), Varna, Bulgaria, Sept. 7-14, 2009. – Dubna: JINR, 2010. – p.124-128 - (JINR; E10,11-2010-22).

[4] V.F. Borisovsky, V.V. Korenkov, S.V. Kuniaev, G. Musulmanbekov, E.G. Nikonov, I.A. Filozova. On Open Access Archive for publications of JINR staff members. Proceedings of XI Russian Conference RCDL'2009, Petrozavodsk, Russia, Sept.17-21, 2009. – Petrozavodsk: Karelia Scientific Center of Russian Academy Science, 2009. – p. 451-458, (in russian).

[5] CDS Invenio: http://invenio-software.org/

[6] V.F. Borisovski, I.A. Filozova, S.V. Kuniaev, G. Musulmanbekov, P. Ustenko, T. Zaikina, G. Shestakova. Open Access Archive of scientific publications: JINR Document Server. Proceedings of XII Russian Conference RCDL'2010, Kazan, Russia, Oct.13-17, 2010. – Kazan: Kazan State University, 2010. – p. 162-167, (in russian).

[7] INDICO: http:// http://indico-software.org/

[8] J. Heer, S.K. Card, J.A. Landay. Prefuse: a toolkit for interactive information visualization. Portland Oregon: SIGCHI Conference on Human Factors in Computing Systems, 2005.

[9] Jonson B., Shneiderman B. Treemaps: a space-filling approach to the visualization of hierarchical information structure. Proc. of the Second Internat. IEEE Visualization Conf. – 1991.

[10] Card S. K., Mackinlay J. D., Shneiderman B. Readings in Information Visualization: Using Vision to Think. – San Francisco: Morgan Kaufmann, 1999.