

Open Access Repository at JINR

I.A. Filozova¹, V.V. Korenkov¹, G.G. Musulmanbekov¹, R.N. Semenov¹, V.F. Borisovski²,
S.V. Kouniaev²

¹e-mail: fia@jinr.ru, Laboratory of Information Technologies, JINR, Dubna; ²e-mail:
vborisov@jinr.ru, Science Organization Department, JINR, Dubna

Introduction

Traditional channels of distribution of scientific results by means of the publications in scientific journals undergo radical changes with the advent of the Internet and a broader access to digital resources. These changes are connected with the transition from a paradigm of the traditional publication to the creation of open archives (repositories) of the scientific literature. This paradigm, which has been put forward in the Declarations of the Budapest and Berlin Initiatives, has received the name “Initiative of Open Access” (OAI – Open Access Initiative) [1]. Availability of any publication for reader means “open access” via the Internet to this publication which can be read, loaded, copied, extended or used for other lawful purposes in the absence of financial, legal and technical barriers. Presently a growing number of the academic institutes all over the world create their own repositories, accumulating and arranging them in the form of open access for the world community. This tendency is a consequence of growing necessity of transition to open access to the scientific literature, defined by the OAI principles: free electronic access for any user, decrease in restrictions on license access and copyrights. In the long term it is supposed that this transition will develop in two directions: i) repositories with Open Access (OA), created by libraries of universities and scientific research institutes and ii) free, reviewed journals with OA.

In the field of high-energy physics (HEP) and related areas the popular open access archive of electronic preprints **arXiv.org** is an efficient place for the free distribution of electronic preprints of scientific papers [2]. However, **arXiv.org** gives limited access to publications, namely, only to pre-prints and it is not in full sense a repository of open access. “Orthogonal” to **arXiv.org** OA repositories, like CERN Document Server (CDS) [3], which additionally make available refereed post-print publications in agreement with the policy of scientific journals. OA repositories, joined in network, offer a range of library management services such as a user-friendly interface, powerful search functions and collaborative features.

In JINR the necessity of creating its own archive (repository) as well as its inclusion in the international system of archives of open access within the OAI has ripened. Our aim is to build, similar to

CDS, an Open Access repository on the basis of the JINR Document Server (JDS).

Open Access to Research

Open Access to Research — a way to make scientific results available to all scientific community by the Internet, so that any person can get access to product from any place and at any time at the own choice. OA does not cancel the copyright and does not contradict it. The personal non-property author rights are not alienated and remain with him irrespective of the publication way. The decision about the OA publication representation is accepted by the author voluntarily. OA is realized in two ways: publication in open access journals; depositing documents in public scientific archives and repositories (institutional and subject). The basic principles of OA are declared in the documents of Budapest Declaration Open Access Initiative and Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities.

Open Access is beneficial for researchers, institutions and society as a whole. For scientists and researchers it results in expansion readership and increasing readability, increasing publication citation, scientific impact, growth of the author popularity and fastening of a scientific priority.

For organization it leads to management of their digital resources, increasing the scientific prestige of the organizations. For society – to return on investment in research, removing barriers to information sharing, creation of additional information services for different categories of users. Further trends in developing the Open Access Initiative are the conversion of the commercial journals into open access ones and changing the subscription policy, as proposed by the SCOAP³ consortium.

Goals and Requirements

JINR repository is intended to be a place on network for collection, preservation and dissemination of digital scientific and research documents and located on the platform of JINR Document Server, JDS. It is implemented as an institutional repository in the OAI framework, the content of which will be composed of the following products: publications issued by JINR researchers and in co-authorship with scientists from other institutions; of archive documents

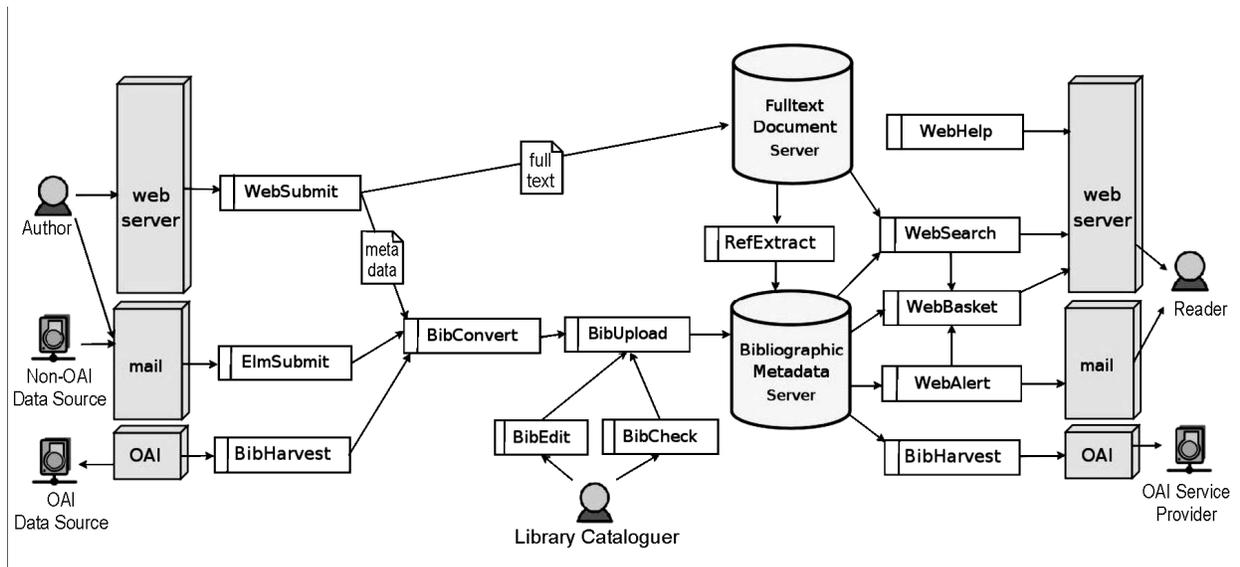


Figure 1: Invenio Architecture

describing all essential stages the JINR research activity; documents providing informational support for scientific and technological research performed in JINR. Building the institutional repository has as its objects to make accessible the scientific and research results of JINR employees for the international scientific community; to increase the efficiency of using information resources of JINR Publishing Department and Scientific and Technical Library; to increase the level of informational support of JINR employees by granting an access to other scientific OA archives; to estimate the efficiency of scientific activity of JINR employees. To achieve these goals we should determine the requirements to the software for creation and management of the OA repository as follows: possibility to harvest and upload documents from other archives; subject classification; wide set of management services for authors and users; web-interface; access rights differentiation; integration with the international catalogues, registers; integration with JINR internal information resources; multilingual interface support; metadata support; support for OAI-PMH [4]; possibility of any format files loading; possibility of the open reviewing and discussions for all interested users even before article acceptance in reviewed journal.

Software

There are a lot of digital library software packages on the market designed to build and manage OA repositories. Most popular among them are EPrints [5] and DSpace [6]. Several institutions collaborating with CERN implemented the package CDS Invenio [7]. The CDS Invenio is the choice for building and managing the JINR repository.

This choice is motivated by the close partnership between JINR and CERN in theoretical and experimental researches in high energy physics and the existing tendency of consecutive unification of information resources of both institutes. CDS Invenio is an integrated digital library system, applications suit that provides the framework and tools for building and managing an autonomous digital library server. The software is free, licensed under the GNU General Public License (GPL). For the time being, CDS Invenio is installed on JINR Document Server, JDS. Installation packages include OS Linux SL5, Web-server Apache 2.2.3, DBMS Mysql 5.0.45, Python 2.4.3, CDS Invenio 0.99.1. CDS Invenio covers all aspects of digital library management. It complies with the Open Archives Initiative metadata harvesting protocol (OAI-PMH) and uses MARC 21 as bibliographic standard. Its flexibility and performance make it a comprehensive solution for the management of document repositories of moderate to large size. Data acquisition is performed from three different sources: direct author submission (using email or the web interface), OAI and non-OAI harvesting (Fig. 1). The metadata is immediately converted into a standard internal metadata representation (MARCXML) whereas full texts are converted into PDF and directly submitted into the document server. Upon upload into the bibliographic server, metadata can be the subject of quality assessment procedures by library cataloguers. Metadata is additionally enriched with citation extraction from the relevant full-texts. The bibliographic server can then be queried to generate indexes, ranks, clusters and formats

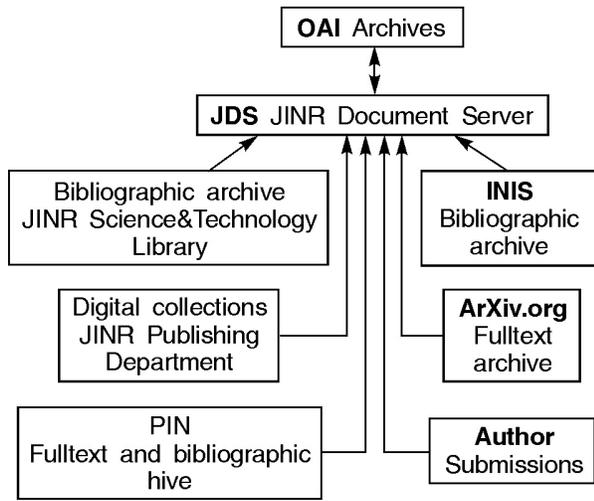


Figure 2: OA and non-OA document sources for JDS of bibliography, suitable for fast retrieval. The information is finally delivered, to users and OAI service providers, through OAI-PMH requests, email alerts and the web search engine. In addition, the web interface offers access to personalized collections of documents (reader baskets or bookshelf), documentation and statistics.

Structure of the JINR Document Server

As a document server JDS should represent well described and interoperable document collections. All harvestable and deposited documents will be structured in several ad-hoc collections covering articles and preprints, books and conference proceedings, dissertations and reports, presentations and talks, handbooks and manuals and various kinds of multimedia. These collections are arranged in a basic (regular) collection tree by classification principle. The regular collection tree is enlarged by the subject collection tree that 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 subject tree involves the following sections: Theoretical Physics, Experimental Physics, Accelerators, Condensed Matter, Informational Technologies, Nanotechnologies, Computational Physics, JINR Laboratories and Divisions, Experiments with JINR participation, Experiments in JINR, Basic Facilities in JINR. As a OAI document provider, JDS can provide for collections arranged according to arXiv.org specifications.

The potential sources of the bibliographic and factual information for uploading at JDS (Fig. 2) are others AI-archives; bibliographic archive of JINR Science and Technology Library; fulltext archive of JINR Publishing Department; INIS bibliographic archive; fulltext archive of arXiv.org;

fulltext and bibliographic archives of the informational system PIN; submission of documents by authors. Delivery of documents into JDS from OAI-repositories is performed by metadata harvesting. The BibHarvest module handles metadata gathering and delivery between OAI-PMH compliant repositories. It enables to configure OAI metadata harvester for eventual periodical batch upload of data – data sources definition to harvest, periodicity, data transformation type before uploading them into CDS Invenio. Receiving documents from non-OAI data sources is performed by consecutive usage of the modules BibConvert and Bibupload. BibConvert utility enables one to convert metadata records from various metadata formats into another metadata format supported by the CDS Invenio. It is designed to process XML harvested metadata records, converting them into MARC21 before they are uploaded into the database. However, BibConvert is flexible enough to deal also with other structured metadata according to your needs, and offers a way to actually insert what you want into the database.

After uploading documents a library cataloguer has the possibility to manipulate bibliographic data directly, edit a single record, do global replacements, and other cataloguing tasks (modules BibEdit, BibCheck). Uploaded publications can be subjected to active discussion and reviewing that can be conventionally done online. The CDS Invenio provides a comprehensive tool for commenting, reviewing and messaging that will allow users and groups to discuss content and share knowledge privately and publicly. Usage and citation statistics of any publication can be derived from the repository.

Conclusion

The currently created JINR’s institutional repository, JDS, based on the OAI paradigm, makes available on-the-fly the results of scientific research performed at JINR for physical community over the world. JINR scientists get not only open access to a vast amount of documents in the field of particle physics and related areas and also a set of advanced user and library-oriented services.

References

- [1] Open Archives Initiative. <http://www.openarchives.org>
- [2] arXiv.org, <http://www.arXiv.org>
- [3] CERN Document Server, CDS. <http://cdsweb.cern.ch/>
- [4] OAI Release of the Protocol for Metadata Harvesting. <http://www.openarchives.org/news/oaiv2press020614.html>
- [5] Official Web-site EPrints: <http://www.eprints.org/>
- [6] Official Web-site DSpace: <http://www.dspace.org/>
- [7] CDSware Overview. <http://cdsware.cern.ch/invenio/index.html>