Digital library of scientific forums organized by the Faculty of Mathematics and Informatics of the University "St. Cyril and St. Methodius "- Veliko Tarnovo

Tsvetanka Georgieva Trifonova, Kaloyan Zdravkov

Abstract. The increasing needs to expand access to research results and insufficient opportunities for publication in printed issues put a serious challenge to the scientific community. The aim of the paper is to mark the first step towards solving this problem in the Faculty of Mathematics and Informatics of the Veliko Tarnovo University by building a repository of digital documents using open source software.

Keywords: Greenstone, Google Scholar, repository, digital library, open access

1   Introduction.

In recent years with more and more weight is the issue with the possibilities of publishing and providing broad access to results of the researchers. They are not rare cases in which their work is hidden not only for the general public, and even more specialized audience, in whose field work.

The rapid development of the information technology and the global connectivity through the internet expanding opportunities for the scientific community, but does not imply instant solution to the accessibility of publications. Building a personal or institutional website is just a step in the right direction. It is necessary to use tools for structuring and appropriate presentation of the created resources while access to them is provided for a sufficiently long period of time.

The trending initiative "Open Access" (Open Access Initiative) provides scientists the opportunity to avoid dependency of the traditional distribution channels of scientific knowledge. The concept applies to electronic publications, available on the web for free reading, use and distribution. This approach provides a favorable environment for sharing and generating new ideas, provoked widespread discussions and enhance cooperation opportunities. [1]

This paper explored the possibilities of open source systems to create digital libraries for the purpose of applying them to build digital library - repository of scientific forums, organized by the Faculty of Mathematics and Informatics of the University of Veliko Tarnovo.

2   Selecting the software to create a digital library

The characteristics of some of the most popular open source systems to build digital libraries are examined: CDS Invenio, EPrints, DSpace, Fedora and Greenstone Digital Library.

CDS Invenio (formerly CDSWare) was developed by CERN, Switzerland to build and manage their own library of works that are currently over 1.2 million units. A complex content management system with sophisticated modular structure.

Invenio work with the MARC 21 as the base scheme for metadata, allowing define and own conventions for naming fields. Supports standard OAI-PHM. Running the operating system UNIX / Linux and Python / Apache as a Web server. Requires databases MySQL, PostgreSQL or SQLite. The programming language is Python.

Flexible and multifunctional system, complex expansion and customization. Part of the functions required installing software from third party manufacturers.

Released as free software, licensed under the GNU General Public Licence (GPL). [2, 3]

Eprints, developed by the University of Southampton, UK, is also a system for building digital repositories, provided mainly for scientific publications. It works with its own structure of metadata, supporting standard OAI-PHM and allows export formats XML, RSS, DublinCore and METS. Runs under management OS UNIX / Linux and Apache as the web server. Requires database MySQL. The programming language is Perl.

A distinctive feature of this product is the ability to create multiple independent digital libraries within a single installation of the software. As a disadvantage may indicate the static interface, which causes a delay between the publication of electronic documents and the possibility of their use - while regenerate the necessary pages.

Released as free software, licensed under the GNU General Public Licence (GPL). [2, 3]

Fedora is a system for building digital repositories, occurred as a joint project between Cornell University and the University of Virginia, USA. It works with its own internal structure of metadata (FOXML), supports the standard OAI-PHM and enables export formats DublinCore and METS. Requires OS UNIX / Linux and Apache as the web server. Manage MySQL database or Oracle. The programming language is Java.

Only one of the presented systems, which was established on the principle "Service Oriented Architecture". Digital objects are seen not simply as static records, and as a set of objects, properties and behavior.

Fedora is a functional core that does not have a user interface. It is necessary for each implementation to develop a presentation layer to suit specific needs. Another important aspect is the increased requirements of the platform to hardware resources.

Released as free software, licensed under the GNU General Public Licence (GPL). [2, 3]

DSpace is an open source software to build digital libraries, designed primarily for academic and research institutions. The system was developed by HP (Hewlett Packard) in collaboration with MIT (Massachusetts Institute of Technology). One of the

youngest systems, developed entirely with a focus on building institutional repositories. Complex solution including (and working only by) web interface for different user groups.

Works with format DublinCore for internal structure of metadata, and CNRI for permanent identifiers of stored objects. Supports standards OAI-PHM and OpenURL, allows export formats XML and METS.

Playing under the management of operating system UNIX / Linux and Apache as the web server. Manage PostgreSQL database or Oracle. The programming language is Java.

Allows search both in the metadata, so and the full text of the files themselves. Supported formats TXT, DOC, PDF and HTML, with additional installed filters the scope can be expanded.

Supports tools for assembly of the stored objects in the "collections", and each object can participate in one or more of them. The collections can be grouped into "communities" and thus achieve a logical structuring of the digital array in different domains.

The disadvantage is a low flexibility of the software. The system is solid and the change in her behavior is complex to implement.

Released as free software, licensed under the GNU General Public Licence (GPL). [2, 3, 4, 5]

Greenstone was developed within the project of the Waikato University, New Zealand, in collaboration with UNESCO and the NGO Human Info. It is bundled software for building digital libraries. Designed and implemented as the most flexible and convenient system, offers a multilingual interface available in 60 languages. It works with all versions of the operating systems Windows, UNIX / Linux and Mac OS-X, including currently developing a version for Android.

The system provides maximum freedom in terms of interoperability. It is available to download in a version ready bundled with Tomcat server and MySQL database. The programming language is Java.

Provides full support for standards OAI-PHM, METS and Z39.50, it can bidirectional exchange data with other repositories, working with them. It realized an opportunity for direct export to and import from repositories built with DSpace. Works with DublinCore and own formats for internal representation of metadata. It has editor for creating new formats.

There are developed multiple filters for different types of documents: PDF, PostScript, Word, RTF, HTML, TXT, ZIP archives, Excel, PPT, etc., which provides indexed search in the full text of these files. Supports a wide range graphics and audio-video formats.

The user interface is solved with Web access via browser to customers (readers) and local java application management system (incl. Remote Access) and creates collections.

Released as free software, licensed under the GNU General Public Licence (GPL). [2, 3, 4, 5]

Based on the characteristics of the reviewed systems to build digital repositories, presented above, and comparing the possibilities and prospects they offer, for building

a digital library of scientific forums, organized by the Faculty of Mathematics and Informatics (FMI), was selected Greenstone. This solution provides the necessary freedom and flexibility in the implementation of the realization, allows the introduction of objects and metadata for them to be carried out by people without special training and provides the best possible cover of needs the FMI.

3   Indexing in Google Scholar

Google Scholar is a web service that provides an easy and convenient way to search for relevant scientific literature. Important for researchers is the possibility their publications to be more easily found on the Internet. This most directly increases the citing - documents, freely available via electronic channels, have a chance to be read by a much larger number of users than traditional. This is due to the shortening and simplifying the cycle "publishing - reading - citing". Easier access to shared knowledge can be a source of new ideas, confirmation of hypotheses, to cause interest of business and as a result to get financial support for a specific project.

In terms of institutions publication in their own digital repository is incomparably cheaper than traditional print media. This gives a chance for maximum promotion of their activities, the information reaches unimpeded to its audience, in perspective increases the rating of the organized scientific meetings.

The freedom of publication and use of the information do not impose to change the existing rules relating to copyright. Rather includes raising the culture of correctly writing, reference and citation.

Like the standard search engine Google, this service indexes opened documents by strict algorithm. In contrast, however, these rules are fully defined and available in the public domain [6]. Some of the most important requirements are:

- repository have to be with a clean interface, without ads and pop-ups, including login-form;
- the full text of articles or at least abstract must be freely available (lack of both may be grounds article to be rejected);
- the full content should be in HTML or PDF;
- recommended "shallow" structure of the site, each article is no more than 10 clicks from the home page

    It's recommended to use software repositories, which can maintain uniform format metadata.

4 Related works

There have been some test searches1 in larger repositories from which can be drawn appropriate conclusions and to avoid typical mistakes when building a digital library.

In Repository (own work) supported by the "National Academic Library and Information System" (NALIS) are presented resources of three universities and two large libraries. All objects are opened with a minimum set of bibliographic data without an abstracts, which severely limits the opportunities for demand. In the collections of the New Bulgarian University and the American University in Bulgaria lacks full text resources. This makes this dataset a catalog, but not a full resource. There uncovered categories, which generally lacks content.

In the Digital Library of Sofia University (DSpace) Google Scholar has no problem with opening a sought after materials, but unpleasant impression makes the removal of full text resources in some of the digital objects. The indexes in the search engine still available, but the usefulness of the storage decreases, and the next crawl robots can remove invalid links.

The repository of the Union of Mathematicians in Bulgaria is realized in the form of static HTML site, which published volumes of the Conferences of UBM. A major disadvantage to him is the lack of local demand. The references, leading to PDF files with individual papers are placed on the page number in the content, eliminating the possibility of being properly indexed by search engines. In the test queries in Google Scholar proved that the result does not correspond to the specified search to yield gap.

5 Implementation of digital library of scientific forums, organized by the FMI of the University of Veliko Tarnovo.

To build the digital library has been used system Greenstone, version 3. Software installation took place seamlessly, with minimal intervention from the user. The base configuration is sufficient to start work without additional settings. By default there are installed two sample collections for quick orientation in the possibilities of the software.

Greenstone works in addition to pre-defined sets of metadata - Dublin Core, Extracted Greenstone, CDWALite, Australian Government Locator Service, New Zealand Government Locator Service, DSpace (partially - fields not covered by the Dublin Core), etc., and with user-defined private spaces, which is provided the instrument for creating and editing.
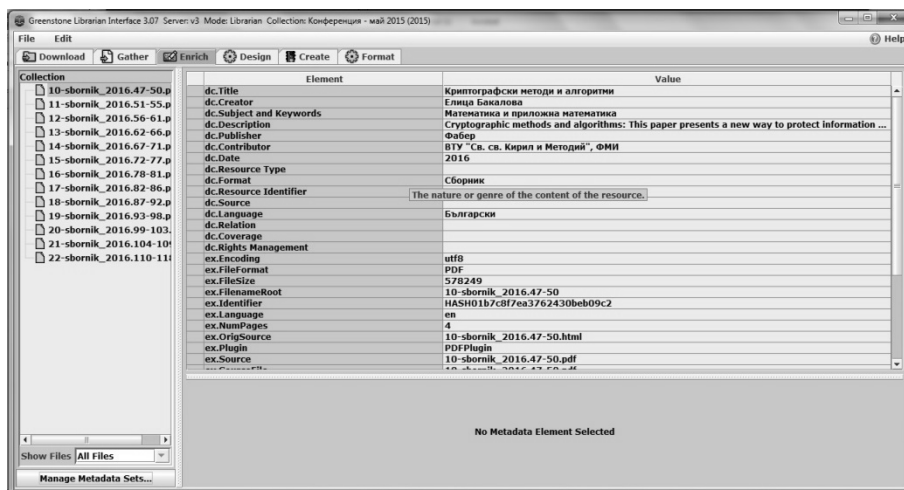

Fig. 1. Introduction of metadata to digital object

Once in defined collection are introduced provided digital objects and for each of them have completed the necessary metadata start a process of "building", where each file is treated with the appropriate filter to extract available meaningful information in text format. Thereafter, it is added as another layer data in the description of the object and allows direct search in full text. In the system is preserved and the original file and the user is able to review the alternative accessible formats.
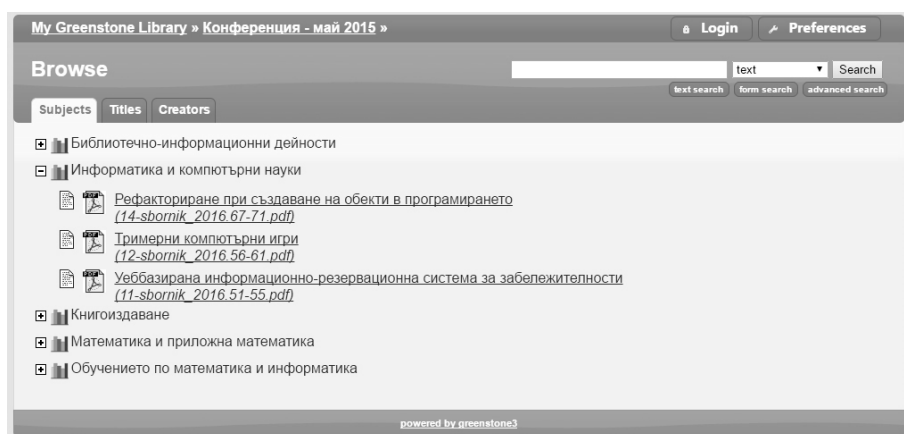

Fig. 2. Review the contents of the collection, arranged by subject

When creating a new collection it is necessary to determine in advance the appropriate set of fields that will be filled with metadata. Depending on the specifics of disclose digital objects is possible to use both fields from different sets, provided that this does not lead to duplication or distorted logic of the information presented. For example, in Dublin Core tags missing the ISSN and ISBN on the item and tag

DC.Publisher proved grossly inadequate when needs to indicate both the issuing institution in which the author, the institution in which the reviewer and the name of science forum. This problem is easily overcome, as to avoid ambiguity add appropriate tags sets of Highwire Press.
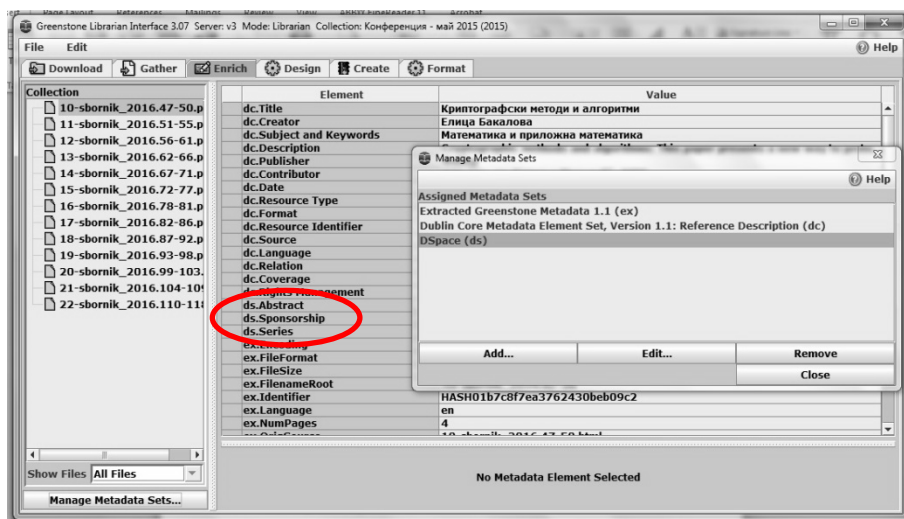

Fig. 3. Added the fields of set DSpace

The flexibility and functionality of Greenstone Digital Library of more basic level provide excellent compatibility with the requirements of indexing service Google Scholar. A good approach is to be constructed custom mixes, containing exhaustive necessary fields to describe each of the types of digital objects, generated by participants in the Repository. This will facilitate the introduction of documents and minimize the possibility of errors, which in turn ensures high quality and availability of information in the repository.

6   Conclusion.

The building, maintaining and updating the repository of digital documents is a necessary step in progression of any scientific institution. With the creation of the described digital library of FMI of St. Cyril and St. Methodius University of Veliko Tarnovo will contribute to qualitative growth and improving creative environment for its faculty and students. As a next stage of development forthcoming to be explored possibilities for application of semantic technologies in search of sources of information in the scientific records.

Bibliography

1. The initiative "Open Access", http://www.iae-bg.com/current/open-access/
2. Kökörčený, M., Bodnárová, HA Comparison of digital libraries systems. Advances in data networks, communications, computers.University of Algarve, Faro, Portugal, 2010, pp. 97-100.
3. Andro, M., Asselin, E., Maisonneuve, M. Digital libraries: Comparison of 10 software. Library Collections, Acquisitions, and Technical Services, Taylor & Francis (Routledge), 2012, 36 (3-4), pp.79-83.
4. Ravikumar, MN, Ramanan, T. Comparison of Greenstone Digital Library and DSpace: Experiences from Digital Library Initiatives at Eastern University, Sri Lanka. Journal of the University Librarians Association of Sri Lanka, 2014, Vol.18, Issue 2, pp. 76-90.
5. Witten, IH, Bainbridge, D., Tansley, R., Huang, Ch., Don, KJ StoneD. A Bridge between Greenstone and Dspace. D-Lib Magazine, 2005, Vol. 11, N 9.
6. Scholar google, by Inclusion Guidelines for Webmasters, https://scholar.google.com/intl/en/scholar/inclusion.html