

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal)

Libraries at University of Nebraska-Lincoln

Summer 4-30-2022

Creating Digital Archive using Dspace: A Case Study of National Agricultural Research Centre (NARC), Pakistan

faheem Akbar
narclib@yahoo.com

Follow this and additional works at: <https://digitalcommons.unl.edu/libphilprac>

Akbar, faheem, "Creating Digital Archive using Dspace: A Case Study of National Agricultural Research Centre (NARC), Pakistan" (2022). *Library Philosophy and Practice (e-journal)*. 7121.
<https://digitalcommons.unl.edu/libphilprac/7121>

Creating Digital Archive using Dspace: A Case Study of National Agricultural Research Centre (NARC), Pakistan

Faheem Akbar*

Abstract:

One of the objectives of PARC is to generate, acquire and disseminate agricultural information for expeditious utilization of research results. The basic aim of project is to develop digital repository (full-text database) of PARC and PARC scientist's publications and making this archive public domain by giving its access on internet encouraging its use by the scientific community and farmers. The repository would be developed using common standards and best practices to ensure its wide accessibility and retrieval/navigation in simple form to facilitate the users. The repository is not only including current research publications of PARC and its scientists but all publications both conventional and non conventional are included in the archive. International metadata standards, data exchange protocol and agricultural vocabularies and thesauri especially AGROVOC and AGRIS standards are used for cataloguing, classification and indexing, etc. The Repository will not only help in improving the accessibility, availability and use of PARC research outputs by researchers, research managers and farmers but also facilitate its preservation for future generation. As no significant work has been done in Pakistan on digital archive and only few institutions recently have started building digital repository so it would be a model for other institutions in Pakistan. The improved access to PARC research outputs will lead to better research which will help in the development of better production technology packages for the benefit of farming community as well as the overall prosperity of the nation.

Dspace an internationally community based developed software is used for management of publications in digital format. The software was selected after testing different software's used for archiving (i.e Dspace, Greenstone, Eprints). Java applet was developed for conversion of Data from DB-Text format to XML format. Data is initially entered in DB-Text and then this data is converted to XML and imported in the main archive. Two servers are configured for Dspace the Main server is Linux based and the backup up server is Windows based as Dspace support multi flat forms. XAMP is used as used web-server on both servers. For securing the data of Digital Archive Backup of data in different formats is obtained and stored on external hard drives and computer system in library. For collection of research publications from PARC scientists and centres/institutions librarians of the respective centres are nominated as focal person they collect publications from the scientists of their centre and send it to DSI for inclusion in Digital Archive. Scientists are also contacted directly through emails for sending their publication for Digital Archive. To-date the Archive contains about 7850 different types of research publications which will be updated regularly. The archive is linked with PARC website and accessible online at <http://175.107.15.226:8080/jspui/>

Keywords: Keywords: Digital Archives, Agriculture Archive, Pakistan Agricultural Research Council

Introduction and Problem Statement: Pakistan Agricultural Research Council (PARC) is the premier institution for agricultural research and development in Pakistan and has the mandate to undertake aid, promote and coordinate agricultural research in the country. Research is a continuous process which generates information therefore hundreds of research articles, documents, reports, feasibility studies; thesis, proceedings, etc. are produced by PARC scientists every year on various aspects of agriculture. This published material can help scientists, planners and policy makers to set priorities in the agricultural research and development of the country. Similarly, it can help farmers to increase production. But unfortunately due to the lack of well-organized system, very few of such articles/documents are available for the use of researchers, planners, policy makers, university students and farmers and most of these are lying with the concerned Institutes, Program and scientists. Some time scientists of the same institutes do not have information of publication published by his/her fellow scientists. So there is a need to keep record of all PARC and its scientist's publications by developing a well organized procedure for collection, processing, storage and retrieval of full-text information to support agricultural research and development in Pakistan. The Directorate of Scientific Information (DSI), NARC has made some efforts in this regard by developing "Pakistan Agriculture Database" comprising bibliographic information with abstracts of literature published in Pakistan or elsewhere in the world about Pakistan's agriculture. Database presently contains about 76,000 records out of which about 7200 records are on PARC and PARC scientist's publications. But the availability of full-text material is limited and most of these are not easily and widely available and every time requester needs to request concerned scientists or NARC Library for sending reprints. So there is a need of developing mechanism for their preservation and quick access when required by scientists and farmers.

Literature Reviewed:

During last few years many academic and research institutions have built digital repositories of the institution's books, papers, theses, documents, etc. The University of Texas Digital Repository (UTDR) <http://repositories.lib.utexas.edu/> provides open, online access to the products of the University's research and scholarship work. Pakistan Research Repository <http://eprints.hec.gov.pk> is a project of HEC which aim to maintain a digital archive (full-text) of all Ph.D. theses produced indigenously to promote the intellectual output of Pakistani institutions. The Borlaug Digital Archive <http://borlaugarchives.tamu.edu/> provides scholars and the general public with access to the papers of Dr. Norman E. Borlaug is an effort to build on his lifetime research from the year 1940 until his death in 2009. In the year 2010 CABI has developed "Global Agricultural Research Repository" www.cabi.org/GARA with the aim to archive agricultural knowledge on behalf of developing countries around the world. Repository presently contains 2800 full –text records.

Senior Librarian (Directorate of Scientific Information & Publications), NARC, Islamabad

Methodology, including Experimental Design and Data Analysis:

Computer hardware were installed and configured. Software available for creation of full text databases were tested and Dspace was selected based on its compatibility with existing PARC fiber optic network Pilot database was created for presentations to information professionals and users for finalization of software, metadata standards and procedure for entry of full-text books/articles. All PARC Centres, Institutes, Directorates and Programs were contacted for getting copies of their scientist's publications. Individual scientists' especially retired persons were also visited for collection of publications. All the collected publications were properly classified, indexed and entered their information in the database after scanning their contents, converting the data in PDF and XML formats. A Script in JAVA was created for converting bibliographic information according to metadata format. For the classification and indexing, AGRIS and CABI standards were followed using AGROVOC and CABI thesaurus. All type of publications, i.e. articles, monographs, reports, proceedings of seminars, thesis of PIASA students, contents of Pakistan Journal of Agricultural Research, chapters of books, etc. were entered. Initially, access of the repository was provided on PARC Local Area Network after configuring and connecting the Digital Archive servers with PARC network. Finally it was linked with PARC website <http://www.parc.gov.pk> for giving access on internet after web programming. Repository contains approximately 8000 full-text publications.

Selection of Software for Digital Archive:

Software selection is an important aspect in Management of Digital Contents. A lot of software's are available for purpose some of them commercial and some are free of cost. Software's available for creation of full text databases would be tested and suitable software would be selected based on its compatibility with existing PARC fiber optic network. Preference would be given to open source software like Greenstone, EPrints or DSpace.

Creation of Pilot Database:

Pilot database would be created for presentations to information professionals and users for finalization of software, metadata standards and procedure for entry of full-text books/articles.

Collection/ Scanning and Formatting of Publications:

All PARC Centres, Institutes, Directorates and Programs would be contacted for getting copies of their scientist's publications. Individual scientists' especially retired persons would also be visited for collection of publications. All the collected publications would be properly classified, indexed and entered their information in the database after scanning their contents, converting the data in PDF and XML formats. JAVA script would be created for converting bibliographic information according to metadata format. For the classification and indexing, AGRIS and CABI standards would be followed using AGROVOC and CABI thesaurus. All type of publications, i.e. articles, monographs, reports, proceedings of seminars, thesis of PIASA students, contents of Pakistan Journal of Agricultural Research, chapters of books, etc. would be entered.

Migration of Bibliographic Data:

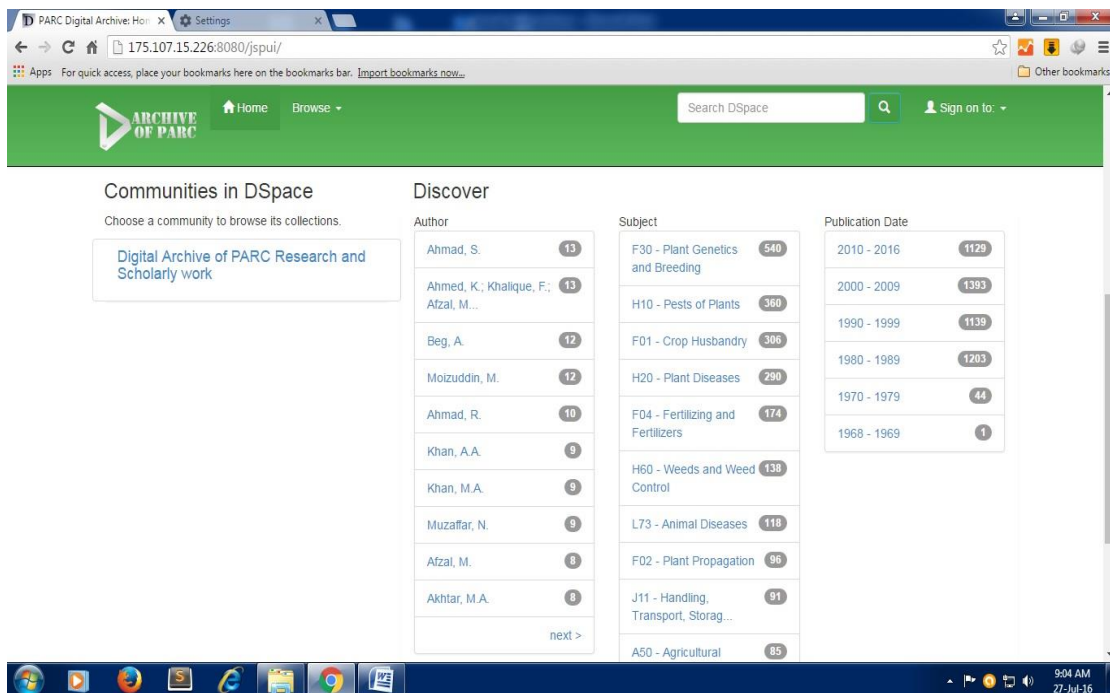
DSI, NARC has been working as liaison office for AGRIS Database of FAO is preparing bibliographic on the standards followed for Digital Archive of PARC. The same data was utilized for Digital Archive with minor changes. The data was migrated to Dspace environment after making necessary changes.

Access of Digital Archive on PARC LAN and Internet:

In the 1st phase access of Digital Archive was provided on PARC Local Area Network (LAN). For wider accessibility of archive contents, access was provided on Internet.

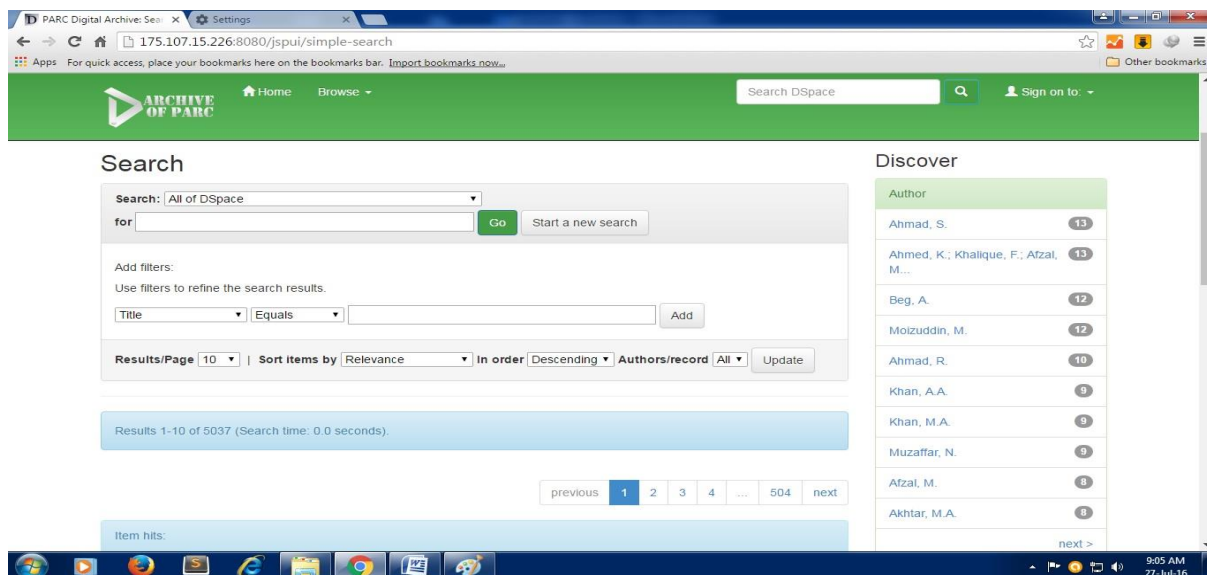
Customization of Dspace User interface:

The user interface of Dspace was customized for better look and compatibility with PARC website. The default color scheme was change from blue to green.



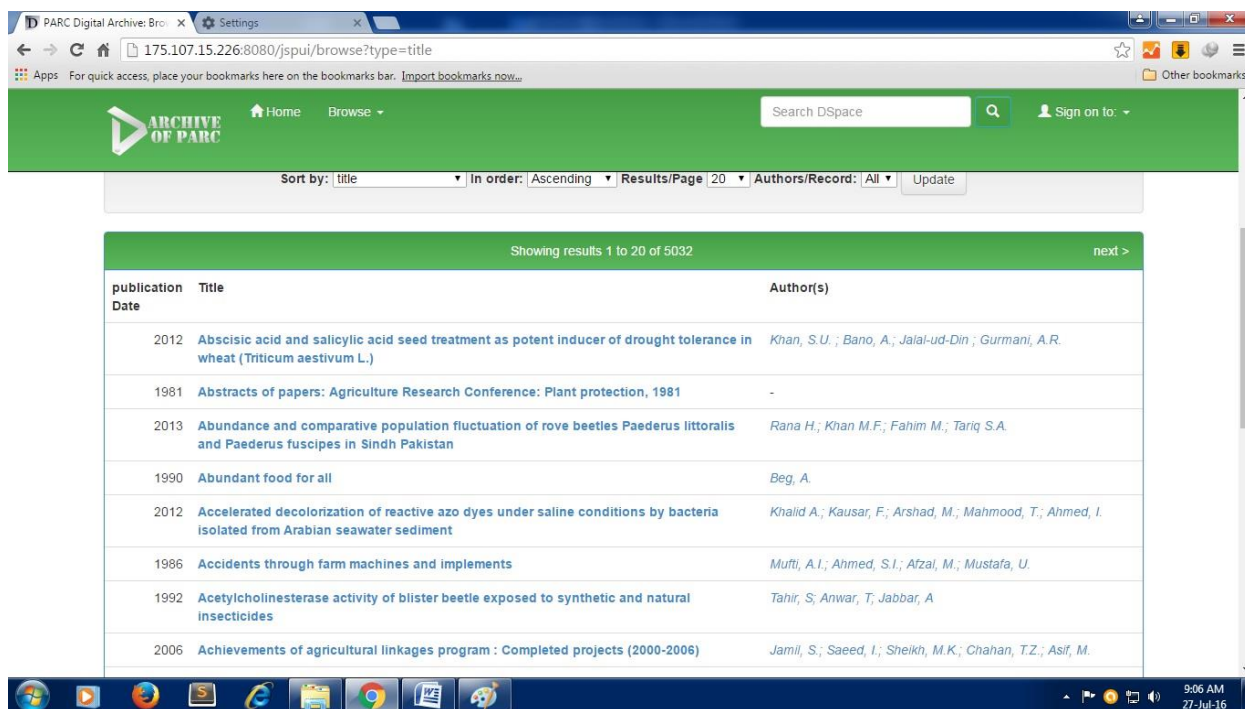
PARC Digital Archive Home Page

Advanced Search filter for searching data from different field was customizing.



PARC Digital Archive Search Screen

Display format was designed for full record display by developing xmlui module interface. Main menu width position of “carousel” slider display was adjusted.



PARC Digital Archive browsed by title

Senior Librarian (Directorate of Scientific Information & Publications), NARC, Islamabad

Collection of Publications:

For collection of publications PARC scientists were contacted through emails, circulars and personal visits. Research reports, research articles, project reports thesis and proceedings were collected from scientists and they were requested to send the remaining and new publications either in soft or hard form to Directorate of Scientific Information, NARC for inclusion in PARC Digital Archive. Baluchistan Agricultural Research and Development Centre, Quetta and Southernzone Agricultural Research Centre (SARC), Karachi were visited for collection of publications for inclusion in PARC Digital Archive. Research Journals available at NARC library were also search for research publications of PARC scientists. During the project duration 7850 (+) research publications were collected for Digital Archive of PARC Research and Scholarly Work. Out these publications about 35% publications were in soft format while the remaining 65% publication were in hard format which were scanned, formatted in Adobe Photoshop, and converted into pdf format.



Figure: Document Scanning using Bookeye 4



Document Scanning and Formatting using (HP Scanner 8500)

Bibliographic data of all collected publications was entered in Inmagic based database after classification, allotment of subject heading and keywords using AGROVOC and FAO AGRIS database standards. The data was then converted into XML format using Java program. The converted data along with full text article in pdf format was then imported in Dspace based Digital Archive.

Conversion and Correction of Data:

Initially data was entered in the INMAGIC software by defining the same database structure used in Dspace for facilitation in spell check and avoiding duplication. Being working as liaison office for AGRIS of FAO the already compiled bibliographic data was used along with the full text article. For conversion of data from INMAGIC format to xml a small program in JAVA language was created. In the first step data was converted to Excel format and then to required XML by creating Dublin_core.xml. The whole data of the project was passed through this process. For amendment and necessary correction in Dspace the Batch import and export facility of Dspace was used.

Conclusion:

The creation of repository will help in preservation and dissemination of valuable literature generated by PARC for the benefits of researchers, research managers and farmers as well as for Senior Librarian (Directorate of Scientific Information & Publications), NARC, Islamabad

future generation. The repository has been developed; old publications have been entered in the full-text database and contain approximately 8000 full-text publications. The repository is linked with PARC website <http://www.parc.gov.pk> for and accessible on internet.

References:

- References Buchanan, S. and Cousins, F. (2012). Evaluating the Strategic Plans of Public Libraries: An Inspection-based Approach. *Library & Information Science Research*, 34, 125-130.
- Chaudhry, Abdus Sattar and Jiun, T. P. (2005) Enhancing access to digital information resources on heritage: A case of development of a taxonomy at the Integrated Museum and Archives System in Singapore. *Journal of Documentation*, 61(6), 751-776.
- PREMIS Editorial Committee, "PREMIS Data Dictionary for Preservation Metadata," v.2, March 2008, <http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>
- Duranti, L. (2001). Concepts, principles, and methods for the management of electronic records. *Information Society*, 17(4), 271-293.
- National Diet Library, "Long-term Preservation and Use of Digital Information," <http://www.ndl.go.jp/jp/aboutus/preservation.html>.
- Ferro, Nicola and Silvello, Gianmaria. (2013). NESTOR: A formal model for digital archives. *Information Processing & Management*. 49(6), 1206-1240.