

Digital Archiving in the Tuberculosis Research Centre: A Case Study

R Rathinasabapati

Abstract

Increasing amounts of research outputs on tuberculosis exist in print as well as in digital form; collecting and preserving those materials serves multiple purposes; particularly the research papers on specific disease should reach at least the researchers those working on the same area to explore further; developing an Institutional Repository will remove the access barrier in future. This paper reveals the experience faced while developing an Institutional archives at the Tuberculosis Research Centre.

Keywords: Digital Archive, Dspace, Institutional Repository, Clinical Trials

1. Introduction

1.1 About TB

Tuberculosis (TB) is contagious and spreads through the air. If not treated, each person with active TB can infect on average 10 to 15 people a year. Like the common cold, it spreads through the air. Two billion people-one third of the world's population- are infected with TB bacilli, the microbes that cause TB. One in every 10 of those people will become sick with active TB in his or her lifetime. When infectious people cough, sneeze, talk or spit, they propel TB germs, known as bacilli, into the air. A total of 1.6 million people died from TB in 2005, equal to about 4400 deaths a day. TB is a worldwide pandemic. The vast majority of TB deaths are in the developing world, with more than half occurring in Asia. Although the highest rates per capita are in Africa (28% of all TB cases), half of all new cases are in six Asian countries (Bangladesh, China, India, Indonesia, Pakistan and the Philippines) (WHO 2008). India with its population of over 1000 million is estimated to

account for nearly 30 percent of global tuberculosis burden. TB affects the most productive age group of 15-54 years and hence is a major barrier to Indian economic development; costing Rs.12,000 crore a year.

1.2 Importance of TB

The landmark discovery of the causative organism of TB by Sir Robert Koch was announced on 24th March 1882. It had taken almost three thousand years to understand the full nature of tuberculosis. A number of seemingly reliable treatments were tried to master what seemed an unstoppable disease. X-Ray becomes the cornerstone for diagnosing TB cases in 1895. Between 1945 and 1960 a remarkable development was made and more patients were cured using the new antibiotics. Towards the end of 19th century, TB had penetrated the Indian population and had become quite common. Many institutions in India working towards the controlling of the disease, viz., Tuberculosis Research Centre, Chennai, National Tuberculosis Institute in India, Bangalore, Vallabhai Patel Chest Institute, Delhi and Lala Ram Swarup Hospital for TB and Chest Medicine, New



Delhi etc. A series of path breaking findings of the studies carried on different aspects of TB at TRC viz., Epidemiological, Operational, Sociological and Bacteriological.

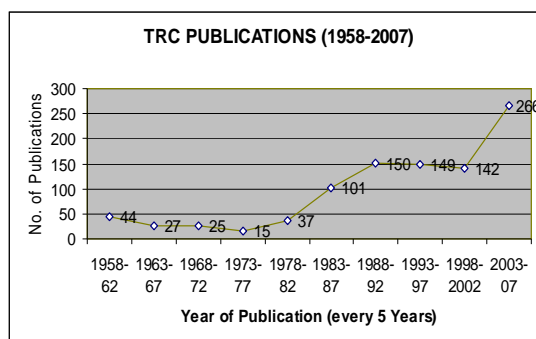
2. About TRC

The Tuberculosis Research Centre was established at Madras in 1956 under the joint auspices of the Indian Council of Medical Research (ICMR), the Madras State Government, the World Health Organization (WHO) and the British Medical Research Council (BMRC) as temporary project to conduct controlled studies designed to provide information on domiciliary chemotherapy in the treatment of pulmonary tuberculosis. In 1964 the Centre was made a permanent establishment under the ICMR. The Tuberculosis Research Centre is a large and complex institution. Six hundred staff members located in Chennai, Tiruvallur and Madurai support the vision and mission of the institution. The Tuberculosis Research Centre is at an important stage in its development. It is poised to become an international centre of excellence in biomedical research after having carved a niche for itself in the field of chemotherapy of tuberculosis. TB control programmes need better diagnostic tools to help detect the disease early and determine its spread in the community. Now the centre has unique role to play in contributing to tuberculosis research. It is in the forefront of training and dissemination of information relevant to control programmes.

3. Collection of Literature At TRC

Tuberculosis Research Centre has unique data sets in publication as well as in Clinical Trials conducted since 1958; those are available in print format. Some of them are unavailable or unnoticed and scattered. Those require a customized approach

for proper preservation and dissemination; collecting and preserving of those materials serves multiple purposes; particularly the research papers on specific disease should reach at least the researchers those working on the same area to explore further; many of the Indian papers are not accessible; developing an Institutional Repository will remove the access barrier in future. It is decided to undertake a study on research output of TRC to build digital archive. The total publications of TRC from 1958 to 2007 cover 952 papers. The below Table shows the publishing patterns of TRC since 1958-2007 (every 5 year-wise)



4. Objectives

This study attempts to

- 4.1 develop an Institutional Repository for Tuberculosis Research Centre
- 4.2 digitize TRC research publications
- 4.3 knowledge management

5. Justification

Increasing amounts of research outputs on tuberculosis exist in print as well as in digital form; collecting and preserving those materials serves multiple purposes; further many of the Indian papers are not accessible; developing an Institutional Repository will remove the access barrier in future. This study suggests developing

an Institutional archives which serves the following main purposes:

- 5.1 it allows researchers to access not so easily available Indian papers
- 5.2 it brings out India Centric Specific Subject Repository
- 5.3 it helps to ensure the long term preservation of scholarly articles
- 5.4 it leads to space saving for better library management
- 5.5 it serves as a single access point for the research output of TRC and helps in dissemination

6. Methodology

This study carried out in four phases:

- Phase I** deals with identification, collection and consolidation of the research publications of the Tuberculosis Research Centre
- Phase II** converts print articles in to an electronic format (PDF)
- Phase III** Assess hardware and software requirements
- Phase IV** develop an Institutional Repository for the Tuberculosis Research Centre

7. Output

The main research work needs a real framework to acquire data. The published papers were started collecting from all scientists of Tuberculosis Research Centre (TRC). A significant research challenge is being faced, particularly given the need to sufficient time for collecting research papers of TRC; often it is found that some scientists who lost their own publication(s) which was in print format. It was made simultaneous attempts with other authors (apart from first author) and publisher(s)

for requesting those articles. Efforts have been put for making consolidation of papers since 1958. So far the consolidation process made till 2007 which covers 726 papers (Phase-I). Since 1958 all the archives which are available in print format have been converted into electronic format (PDF) using Adobe Capture software (Phase-II).

The hardware and software requirements have been assessed. After thorough study was made and experienced, it is decided to develop the Institutional Repository for TRC on **DSpace** software package (DSpace is an open source software, jointly developed by MIT Libraries and HP Labs. It captures data in any format like text, video, audio and data. It indexes the work and provides feasibility to the users to search and retrieve the items. It preserves the digital work for long term). This software allows the communities and users to publish their articles remotely on the archives. It requires minimum configuration for better operation with Windows/Linux, Pentium 4, 40 GB HDD, 510 MB RAM (Phase-III).

It is found that DSpace on Linux needs a programmer's help for installation. Hence I started building an "Institutional Repository for TRC" on **DSpace on Windows** on my Laptop (Phase-IV). The archives have been classified under three heads namely, Clinical Trials; Reports and Scholarship archive.

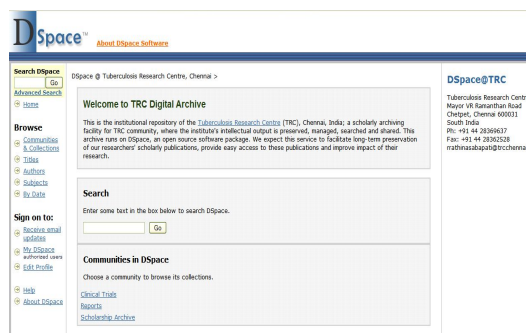


Figure1. Front Page of TRC Digital Archive

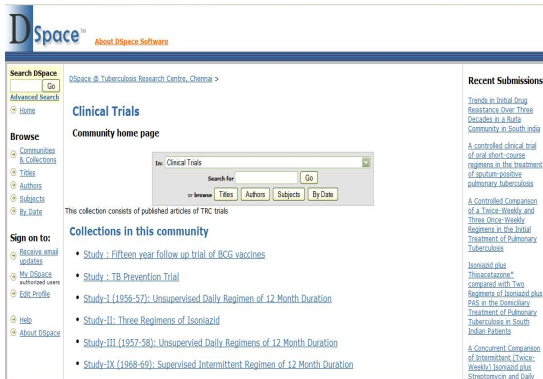


Figure2: Community: Clinical Trial



Figure3: Community: Reports

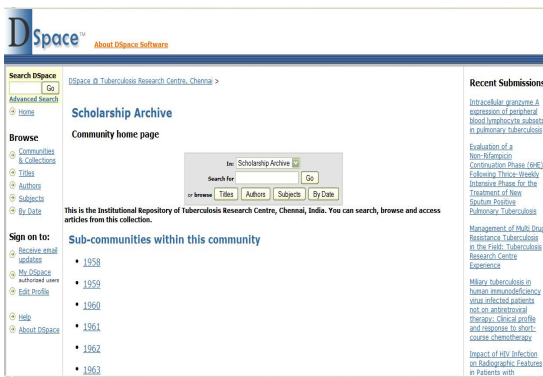


Figure 4: Community: Scholarship Archive

8. Conclusion

Dspace software is jointly developed by Hewlett Packard and MIT Libraries, Cambridge. It is made available since late December 2002. It has been

developed based on the experience gained by Eprints. DSpace has a well defined data model such as, Community + Collection + item + Metadata + Bundle. It supports Dublin Core metadata elements. DSpace is also to serve as a repository back up for future developments to address long term preservation and remote/online access issues. The basic web application and server(s) like Java, Apache, PostgreSQL, Tomcat can be loaded by a sound programmer only. I had difficult experience to find a person for installing DSpace. As suggested by Dr. V Kumaraswami, the Officer-in-Charge, Tuberculosis Research Centre, Chennai, through M/s M.S. Swaminathan Research Foundation I identified a resource centre named National Resource Centre – Free/Open Source Software (NRC-FOSS) at the AU-KBC Research Centre, MIT Campus, Anna University, Chrompet, Chennai 600044. Dr. Srinivasan, the Professor and the Project Scientist. AU-KBC Research Centre helped through his assistant Mr. Mohamed Imran, Programmer, to load “DSpace on Windows” (ready built CD for DSpace supplied by IIMK during its workshop) on Laptop and “DSpace on Linux” on our office PC. While started working on Laptop it gave Internal Server Error. Since Dr. Srinivasan was not familiar with DSpace, I requested help from MIT Libraries and DSpace help mail. Unexpectedly I have received a mail saying that to get in touch with any local persons (Dr. Naina Pandita, NIC, New Delhi and Dr. ARD Prasad, DRTC, ISI, Bangalore). Both of them have been approached. Through Dr. Naina I received help from Mr. Surinder Kumar, NIC to make proper SMTP server configuration. Then it started functioning. Unluckily the DSpace on Linux has not functioning well; while “submitting” it gets stopped. Even made several configuration setup guided by Mr. Surinder Kumar, it was not successful. He added that since

the import and export facility are available on DSpace, anyone can work on any platform. Hence it was decided to start building an “Institutional Repository for TRC” on **DSpace on Windows**. Of course this is a continuous process. After this study completed it is planned to continue the documentation of the TRC paper in future also.

Acknowledgement

I thank to our Officer-in-Charge Dr. V Kumaraswami for his support for developing the Institutional Repository for TRC. I also thank to Dr. A Amudhavalli, Professor and Head, Department of Information Science, University of Madras for her inducement to write this paper and encouragement for enhancing the digital archive at TRC. I extend my profound thanks to Mrs. V Shanthi for her immense cooperation on paper collection.

References

1. TRC: A Review of Research Activities 1956-1966
2. TRC Annual Report January 2001-March 2002
3. TRC Annual Report 2002-2003
4. TRC Annual Reports 1958-2007 for data collection

5. <http://www.sparceurope.org/Repositories/> (13/1/2009)
6. <http://www.micklebring.com/oakwood/ch18.htm> (15/01/2009)
7. <http://www.tbandu.co.uk/tbhistory.html> (15/01/2009)
8. Bates, JH and Stead WW, “ the history of tuberculosis as a global epidemic” - Medical Clinics of North America 1993 Nov;77(6):1205-17
9. Toman’s Tuberculosis: Case detection, treatment, and monitoring 2nd Ed., Ed. By Frieden, T, 2004, WHO
10. <http://www.who.int/mediacentre/factsheets/fs104/en/> © WHO: 2008 (16/01/2009)
11. <http://www.who.int/features/factfiles/tuberculosis/en/index.html> (16/01/2009)

About Author

Mr. R Rathinasabapati, Library & Information Officer, Tuberculosis Research Centre, Mayor VR Ramanathan Road, Chetpet, Chennai 600031.
E-mail: rrathinasabapati@trcchennai.in