Research Documentation Portal:
A walkthrough from concept to Reality

Jitendra. R[1]

Information and Communication Technology has improved by leaps and bounds in the recent past. Emerging technologies have also impacted the medical research community by addressing long standing challenges or by refining the already existing processes .There is an old adage: "If it isn't written (published) it doesn't exist". This saying points to the importance and need for clear and consistent documentation which in turn envisages the need for an efficient system for data retrieval as well.

Digitization has become a practical necessity with interventions to provide improved access to information resources, preservation and dissemination of information as required, at any time and any place. Also, digitization narrows down the "Digital Divide" between the haves and have-nots, between individuals and also between developing and developed countries. In this context the objective of digitization is principally to create databases and facilitate their access and also to preserve and disseminate information resources. The tremendous impact of digitization on scholarly communications across the globe not only supports research activities but also to a large extent prevents duplication of research. This not only results in better utilization of funds but also improves the quality of research.

What is digitization

Digitization is primarily the process of converting written and printed records into electronic form. The content may be text, image, audio or a combination of these (multimedia). The output of digitization process is an electronic document that can also be hosted on Internet/Intranet. The electronic document may be in Portable Document Format (PDF) or Tagged Image File Format (TIFF). PDF format is usually used to store information on the Web, as the file size is relatively small and easily downloadable and transferable. PDF also has a unique print/display format that is the same on any platform and can be read using any PDF Reader (downloadable freely from the Internet).

---

1. National Tuberculosis Institute, Bangalore

How it started at NTI

One of the important mandates of National Tuberculosis Institute (NTI), Bangalore is to conduct research studies to provide meaningful inputs to improve the roll out of the Revised National Tuberculosis Control Programme. Several noteworthy studies, a few of which have even been internationally acclaimed, have been undertaken by NTI in vast spectra of fields ranging from epidemiology and operation research to IEC and sociology. NTI, Bangalore, archived its own publications on electronic media for three reasons, namely, (i) archival for the future, (ii) to build a rich repository of research conducted in the institute for reference and guiding research and (iii) facilitate global access to the research conducted by the institute. However the major hurdles faced were: (1) finance, (2) technical know-how, and (3) management of digitization.

Genesis of research compendium at NTI:

The objective of the research compendium was to provide access to researchers, TB workers, administrators, and students to the research conducted by NTI. The research papers especially those published during the early sixties and seventies may not be accessible to many. The summaries of all the studies carried out at NTI since its inception was compiled and published in 1976 and 1977. In 1996, it was decided to update these volumes. Studies carried from 1960 till 1994 were presented in Summaries of NTI Studies Volume-1 while Volume -2 detailed studies carried from 1995 till 2005.

The summaries were arranged topic wise and chronologically, accompanied by key words to make it user friendly. They were categorized under the headings- Epidemiology, Bacteriology, Sociology, BCG, X-ray, Operations Research and Miscellany. Miscellany included papers on Health economics, Health Education, Dissemination of Information, and other projects. Every effort was made to enable the user gain adequate information about the scientific publication without referring to the original paper.

However there was a felt need from researchers, TB Control Programme managers and trainees visiting NTI, to have full text access to the scientific papers for reference from their workstations which is panned across India. To cater to this demand the institute had to adopt a suitable methodology capitalizing on the advances made in the internet technology.

Initiation of the Health Inter Network Project (HIN):

Dr. Kofi Annan, UNO Secretary General in September 2000 envisioned the need to improve information environment of health personnel, professionals, researchers and scientists, and policy makers by bridging the digital divide in the developing countries. This thought translated into launching of 'Health Inter Network Project'. The core elements of the project were content development, connectivity and capacity building. The concept of digitization at NTI got a push when Dr. Fabio Luelmo, WHO and Dr. Shyama Kuruvilla, from Health Information Management and Dissemination, WHO, Geneva, visited NTI in May 2001 to explain the importance of Health Information Network (HIN).

This project assisted in overcoming some of the challenges faced in building of an online research compendium in terms of finance, technical know-how, and management of digitization.

Why India as the Country, TB as the Disease, and NTI as the Institution were chosen for the Project?

A component of HIN is to provide a one-point source for authentic and relevant health information on all health topics- TB being one of them. The reasons for Tuberculosis as a disease were- (a) it is one of the most prevalent public health problems, (b) the disease poses a major barrier to social and economic growth, and (c) TB is a global public health concern. Another reason for selecting TB stemmed from the fact that the world's research output on TB in international databases, such as, PubMed, Citation Index, Biochemistry and Biophysics, was only 5-6 % even though the disease was a global public health concern. India was chosen from among many developing countries for the HIN pilot project because of the priority assigned to the public health programme and the availability of skills and resources requisite for the development of the HIN project.NTI was chosen for the project because of the significant contribution of the institute in conducting operational research on various facets of tuberculosis. A key cornerstone of the initiative is to provide authentic and locally relevant information content.

Thus, an introduction for the HIN project was seeded at the NTI, and the following five institutions were identified for collating data on different aspects, namely, research outputs on TB under Content Development, Digitization, and Journal Custom Content for Consortia.

The five centres were:

1. National Tuberculosis Institute, Bangalore.
2. Tuberculosis Research Centre, (TRC), Chennai.
3. Vallabbhai Patel Chest Institute (VPCI), Delhi.
4. LRS Institute of TB and Respiratory Diseases (LRSITBRD), New Delhi.
5. Central TB Division (CTD), New Delhi.

NTI was chosen as the nodal centre in India for the project. As the digitization work was not feasible with the existing staff at NTI, a Technical Consultant and a Medical Consultant along with a Project Assistant were appointed for advising and assisting on the work. The Project Assistant created the database using UNESCO's Common Communication Format (CCF which is almost similar to Dublin core) as Metadata of the WINISIS database for NTI publications and other scientific papers. The database has MeSH and non-MeSH key words suitable for the Indian context derived on the tree structure of the MeSH. The NTI panel appointed a Medical Consultant with postgraduate qualification and work experience in Revised National TB Control Programme. The following criteria, in consultation with the Technical Consultant, were prepared for selection of full text of articles and other publications for digitization:

• Public Health importance.
• Design, utility and uniqueness.
• Good sample selection and subjected to good statistical analysis.
• Lucidity and specificity in presentation.

The Content of the Summaries of NTI studies Volume 1 &Volume 2 was disseminated by conversion of the same into interactive CDs and was also hosted on the NTI website as a part of the HIN project. Also the process of full text digitization and classification of scientific papers was initiated under the said project.

TB-Net Project, A DBT supported initiative –

Another major boost to the effort was received through the TB-Net project funded by the Department of Bio-technology. The Department of Biotechnology (DBT) has taken up the infrastructure development project to create a network between research institutes to enable sharing and further development of domain expertise at a very low cost. The principal aim of bioinformatics programme is to ensure that India emerges as a key international player in the field of bioinformatics; enabling a greater access to wealth of information created during the

post-genomic era and catalyses the country's attainment of lead position in medical and agricultural biotechnology.

In this direction, DBT has established an extensive bioinformatics network to promote activities in this area across the country and developed new algorithms, software, tools for data mining and data warehousing applications especially related to human, plant and microbial genomes. To initiate the same, it has identified major areas in rice and TB.

The member organizations in the TB-Net project are:

1. National Tuberculosis Institute, Bangalore
2. National Institute for Research in Tuberculosis, Chennai
3. Indian Institute of Science, Bangalore
4. Institute of Bioinformatics, Bangalore
5. JALMA Institute, Agra
6. MANIT, Bhopal.

Under such circumstances, a pilot study was initiated through various organizations and premiere research institutes associated in the field of Tuberculosis. The pilot study was undertaken with the Objective of creation of National level 'TB-Net database'. The institutions identified are to contribute in their relevant area of work expertise in domain of Tuberculosis. Leveraging on the basic ground work already undertaken at NTI, the institute was required to continue the task of creating a research portal on the epidemiological, operational, sociological and monitoring aspects of TB disease control, which is affecting the community in India. The mandate was also to establish a national portal including publications from other TB-Net member institutions as well. This would however be an ongoing activity.

Further National and Zonal level OR committees were formulated to identify and facilitate the research priorities under the programme. Therefore, it was all the more imperative to have a national level documentation for published TB research in India with full text access for members of the consortium for supporting the operations research activities in the country.

Hence an Open Access Repository of publications of researchers of the National Tuberculosis Institute (NTI) - Bangalore, National Institute of Research in Tuberculosis - Chennai, LRS Institute of TB and Respiratory Diseases - Delhi, New Delhi TB Centre - Delhi, State TB Demonstration and Training Centre and research conducted under the National / Zonal / State Task force mechanism was hosted at *http://tbresearch.ntiindia.org.in* . The portal currently

comprises over 1200 published TB research articles classified by 19 subjects from the leading National Institutes, National and Zonal OR Committees, State and district TB cells, Tuberculosis Association of India etc from 1958-2013. The metadata and abstracts for all the articles are available to end users, whereas access to the full text article is available to the registered members. It attempts to collect, preserve and disseminate the intellectual output of these Institutes available in peer-reviewed journals. This is an ongoing activity which looks forward to enrich the database collection by including publications from not only other TB net member institutions but players from the Indian continent at large.

A local flavor of the database at http://ntiresearch.ntiindia.org.in is also hosted which captures other important documents like the research protocols explaining the statistical methodology adopted, work instructions followed etc., for the research conducted at NTI for ready reference.

The emphasis on the universal access of TB care, the challenges of HIV TB co-infection and Multidrug Resistance Tuberculosis etc, necessitates the need for a much more cohesive, operationally feasible methodology to handle this major public health menace. Also noteworthy is the point that India which is categorized as a high TB prevalent country portrays an array of diverse socio-economic, geographic and demographic fronts which needs to be addressed in its own unique ways. The saying -"Think Global Act local" would best capture the true need of the hour in conducting operations research which would enhance the delivery of the Revised National TB Control programme. Hence the creation of the national research portal is one major step and an earnest endeavor to facilitate better utilization of available resources by gaining from the research experiences conducted in the Indian continent and also facilitates better knowledge management in the realms of India's TB research community.

References -

1. Digitization: A Practical Experience at the National Tuberculosis Institute, Bangalore Murthy Sudha S, Sr. Librarian, National Tuberculosis Institute, No.8, Bellary Road, Bangalore 560 003 – www.Indianjournals.com , Year : 2005, Volume : 11, Issue : 2