Chapter 1

Introduction

“And I can’t tell you the rest until the journal comes out.”

(Bachrach et al., 1998)
Structure

1.0 Introduction
1.1 Background of the Study
1.2 Objectives of the Study
1.3 Statement of the Research Problems
1.4 Hypotheses
1.5 Methodology
1.6 Significance of the Study
1.7 Definitional Analysis
1.8 Conspectus
1.0 Introduction

Global trends in scholarly communication increasingly gravitate towards a digital environment, and the development of the open access movement (OAM) and electronic publishing have been outgrowths of this process. Scholarly communication, and more specifically scholarly publication, is an important manifestation of knowledge generation and knowledge diffusion that captures and preserves knowledge assets more effectively and transparently than in the past and provides immediate online access of local research outputs globally. The commonly held reasoning behind open access scholarly communication is that greater access to the public-funded research literature, in effect greater knowledge diffusion, will lead to the advancement of science and technology, and will be especially advantageous for developing countries like India.

Scholarly publication process has its roots in the need for communication between academicians, researchers, and scientists. Electronic publishing and electronic access have brought tremendous changes in scholarly communication. Prior to the invention of the printing press, scientists communicated via the exchange of letters and via public debates akin to latter-day conferences. But electronic publishing and new types of digital resources have brought howling changes in traditional scholarly communication system and are changing our notions of what scholarly publishing means by modifying the way researchers produce, communicate and access information. Scholarly communication has changed dramatically after the introduction of the Internet in public domain during early 1990s. It is undeniable that the Internet and World Wide Web (WWW), as distributed information system, have brought many changes, for life beyond scholarly communication, as well as for the process of scholarly communication. With the development of electronic communication, academic research has introduced new technologies and tools for scholarly communication. It facilitates geographically dispersed information access and retrieval through online mechanisms. It has enabled information to be accessed and disseminated far more easily than ever before. Our ways of work and communicating have been and are changing. Said evolution of the scholarly communication system has garnered much attention in the past decade, it’s having been driven by three factors, namely -

1. the ‘serials crisis’ or serial pricing crisis (Cummings et al., 1992; King & Tenopir, 1999; Houghton, Steele & Henty, 2004);
2. the growth in scholarly research output (Crane, 1972; Pouris, 2003); and
3. the advent and ubiquity of the Internet, in the context of the Information Society.

Scholarly communication has been affected by the so-called ‘serials crisis’ in the last 30 years and scholarly journal prices have increased more rapidly than research institutions’ budgets (Greco et al., 2006). Another expert (Carlson, 2002) reported
that ARL (Association of Research Libraries) statistics estimate that journal subscription rates have gone up an average of 8.5 percent per year since 1986, while library budgets have increased 5.6 percent per year. This situation is well described by Peter Suber, who talks about ‘permission crisis’ (Suber, 2003), relating to financial and technical barriers to access (such as subscription fees and login/password requirements) risen by commercial publishers. Economic and technological changes are cited as the major reasons contributing to the notable evolution of the current scholarly communication system (Thorin, 2003; Moller, 2006; Swan, 2007). Pickton (2005) identified several other technological, financial, ethical issues responsible for open access (OA) publishing. In 1998 the ARL founded SPARC (Scholarly Publishing and Academic Resources Coalition) to address the serial crisis and correct the imbalances in the scholarly publishing system (Joseph, 2006).

It is also important to recognize the link between institutional repositories (IRs) and the open access movement (OAM) and how these influencing potential repository contributors. To describe and explain the context in which IRs sit, it is necessary to:

- explain the OAM including the reasons for its emergence; and
- describe the methods by which OA can be approached.

A. Open Access Movement (OAM)

“Knowledge wants to be free” (Arunachalam, 2008).

The open source software, open access and open standards movements are gaining tremendous momentum. OAM also known as ‘Open-Access Publishing’ or Free Online Scholarship is a social movement that includes open source software and the types of social structures that produce it; open standards; open contents including courseware, digital repositories, and the creative commons; academic alliances for creating academic middleware; open intellectual property policies more generally; keeping the Internet architecture open; and a growing sense among academic leaders that the university should serve as a counterbalance to the overly restrictive objectives of the digital rights movement. In the last few years, the movement has gained a lot of strength among the academic and scientific institutions.

Prior to the advent of the Internet, publishers and academic societies dominated scholarly communication, and researchers channeled their research output solely through authoritative publishers and academic societies. The first scientific journal known as ‘Philosophical Transactions’ launched by ‘Royal Society of London’ began in 1665 to enable researchers to share their work quickly and widely and to
establish the priority of researchers investigating the same problem (Thorin, 2003; Swan, 2007).

In these circumstances, very recently an academic movement called the ‘OAM’ was organized as an alternative to the conventional scholarly communication through journals and books to promote free access to research products and to encourage university libraries to be responsive to their members' requests (Yeates, 2003). The OAM arose at the beginning of 1990 as a reaction against the traditional scientific publishing model. The advent of the Internet and digital technologies made it possible for research to be shared in entirely new ways. Gradually, traditional publishing models are shifting as electronic publishing models. Traditional publishing models are shifting as electronic publishing changes the ways that libraries purchase and access scholarly materials (Association of Research Libraries, 2006).

Global trends in scholarly communication increasingly gravitate towards a digital environment, and the development of the OAM and electronic publishing have been outgrowths of this process. Two contemporary developments in particular have helped shape the nature of today’s institutional repositories (IRs): the emerging knowledge management movement; and the maturing, but still rapidly advancing, technology of content or asset management in the digital information system. The knowledge management movement of the 1990s influenced the development of IRs. The major driving force behind the development of IRs has been the dramatic shift in scholarly communication especially within the past five to ten years. Repositories and their relationship with the OAM constitute a new trend in scholarly communication worldwide. This is typically reflected in the success story of open access digital repositories worldwide.

**B. Open Access Roads**

New models of scholarly communication are evolving and different models are being developed to provide access, manage costs, and manage an organization’s scholarly output, especially at colleges and universities. The currently evolving model is referred to as ‘Open Access’. OA is elaborated upon as one current model of these aforesaid changes in scholarly communication. There seem to be no other competing new models at present which either oppose or complement OA, there seems to be at present the traditional (publication) model and OA. The two schools of thought within OA are the journal reform school, and the self-archiving school. These two parallel but complementary paths (Fig. 1.1) for achieving OA are sometimes referred to as the “Gold” (publish in an open access journal (OAJ) and “Green Roads” to open access (publish in a non–OAJ but archive in an open access archive (OAA) (Antelman, 2004; Chan & Costa, 2005; Bailey, 2006b; Harnad, 2005).
Developing repositories in order to provide access to the world’s research output is one solution and has emerged as a new publishing tool. In these contexts - restricted access to knowledge, changing scholarly practices, and support for open access, rapidly evolving information technology - IRs have emerged as a potential solution. Thus, this new Web-based distributed system seems to have spawned two contradicting developments in scholarly publishing: one towards enabling and promoting OA and the other towards restricting it and preserving the barriers to scholarly works.

The rapid growth of digital assets creates challenges for the academic institutions like universities, research organizations in the use, management, archiving and application of digital information and datasets. It is found that about 80-85% of digital intellectual output of universities is never made accessible to the public (The Open Citation Project, 2004). Academic institutions have been grappling with how to manage the digital intellectual output they produce including journal articles, conference papers, reports, theses and dissertations, teaching materials, artwork, research notes, and research data. Universities are finding new ways to capture, manage and disseminate these scholarly electronic resources, and institutional repositories have been proposed as a tool to aid academics to manage and distribute their digital materials. Many universities and research institutions throughout the world are investigating, piloting, and developing systems for building collections of digital resources and learning materials in the shape of open access repositories (Rothery & Bell, 2006) that will enable faculty and researchers to upload and download scholarly literature and use them to share resources with each other either within the institution or across the region, or more widely still.
Clearly, technology has made it easy to create, store and access digital materials. Several broad technological developments in the digital information system were also pushing individuals and their organizations towards the creation of institutional repositories (Lynch, 2003). In this venture open source software (OSS) is one step ahead. The OSS movement has its root in the 1970s and continuing to grow its popularity to date. The movement gained momentum with the release of two open source systems: Eprints, developed by Stephen Harnad and the team at the University of Southampton in late 2000 (Tansley & Harnad, 2000), and MIT's DSpace in 2002 with support from Hewlett Packard (Barton & Walker, 2003). The development of OSS to manage electronic documents created open source alternatives for IDRs development.

This Web-enabled distributed IDR system is a socio-technical concept. It is a multi-faceted domain. It is not all about information storage and retrieval. The basic logic behind this research is organizing and management of resources and bibliographic items to support open access to knowledge. In this sense IDR is also a psycho-social concept. Another facet of this research is developing Unicode-compliant IDR system based on FLOSS (Free/Libre Open Source Software) software framework. IDRs all over the world now see opportunities to create databases that can be accessed via Web in order to provide global access to local research produced by academicians, researcher, scientists working in universities and research organizations. Technology tools in launching digital repositories are not so easy and can’t be mastered given a little time and exposure. No single software solution is presently available to meet the requirements of such complex information systems. However, adding contents to digital repositories is protracted, and scientific and academic communities are not fully aware of potential of open access to information. Library and informational professionals have a daunting task of convincing authors of the importance of OAM. This is the new role of library professionals. Their role should not be confined to traditional information acquisition, organization, dissemination and preservation. They have an active role to play in the publishing process.

Under such circumstances, this study is an attempt to build up a conceptual model that deals with:

- Mechanism of integrated processing of IDR resources of different types along with provision for support services that are interactive, collaborative and participative;
- Designing a set of policies for different IDR components on the basis of global recommendations and ‘best practice’ guidelines.
- Design of Web-enabled software framework for the University of Burdwan on the basis of policies and standards as framed by using open standards and utilizing Free/Libre Open Source Software (FLOSS); and
- Development of Unicode-compliant Bengali script based interface along with mechanisms for multilingual searching, browsing and subject based hierarchical browsing by employing standard classification scheme.
The FLOSS based software framework is designed keeping in view the requirements of the University of Burdwan. This Web-enabled software solution centers on the following attributes:

- Use of internationally agreed upon data standards/formats for managing IR resources supported by data standardization tools including domain-specific vocabulary control device;
- Integrated processing of IDR resources and bibliographic items to ensure integrated search and browse facility from user interface;
- Development of Unicode-compliant mechanism for Bengali script based user interface, processing of Bengali script based IR resource, multilingual searching, and subject access support system;
- Supporting interactive communication processes through Blog, Wiki and Discussion forum; and
- Developing federated search option for multiple IDRs from a single-window search interface.

1.1 Background of the Study

The Open Access Repositories (OARs) movement in general is neglected in India. Only eighty (80) repositories have been developed in the country as per OpenDOAR (Directory of Open Access Repositories), and ROAR (Registry of Open Access Repositories) databases. The concept is not new but it became popular relatively late compared to other developed countries.

Current adoption levels of OARs are pleasingly high. The two major registries of OARs (viz. OpenDOAR & ROAR) show that adoption in developed countries is already very high, with some developing countries also starting to take the advantages that open access confers. The area is in rapid development, and ROAR statistics show a steep incline from around 2001 to the present in the availability of both open archives and open records in those archives available via OAI (Open Archives Initiative).

As per OpenDOAR (http://www.opendoar.org/find.php) database, there are more than 2381 (up to June, 2013) repositories throughout the world (Fig. 1.2). Another database, ROAR (http://roar.eprints.org/view/geoname/) reported that more than 2904 OARs exit. But there is a vast difference between OpenDOAR and ROAR databases in terms of number of repositories registered. This is because many of them do not have their own domain or sub-domain, and for this reason it is not possible to add them in this study. As per the OpenDOAR and ROAR databases, every day one repository is being added to the databases. The following table (1) shows the growth and development of OARs throughout the world.
So research in this area and developing a model IDR for the University of Burdwan through OSS is prime objective of this work. But technically it is not so easy in view of the present scenario. There are several problems (vide section 3.12 of chapter 3) of Indian OARs other than technical and non-technical problems.

But situation is fast changing and a close look reveals a shining picture. Our country has been the most prominent partner in the field of open knowledge movement. Our professionals and working librarians are trying to build up necessary information infrastructure, essential for open access development and our experts are trying to
establish digital library and institutional repository with free open source software (FOSS). India in its own way made its contribution by developing and launching more than 80 repositories (ROAR, 2012). Several agencies, learned societies, professional bodies have come forward and have established IDR for their own. A nation-wide movement has started and government organizations like the UGC (University Grants Commission), NKC (National Knowledge Commission) recommended OA to public funded research. In one word, OA to information is the realization of Ranganathan’s Five Laws of Library Science in the internet world.

In April 2004, there were only four (4) institutional repositories in India (OpenDOAR, 2012; ROAR, 2012) and this number had climbed to more than eighty (80) in 2013 (June) with an average increase of about none (9) new repositories per year. Based on worldwide numbers in both OpenDOAR and ROAR databases, this growth has placed India as the seventh leading nation in IDRs development. Much of this success is undoubtedly connected to government support as well as sponsorship from MHRD (Ministry of Human and Resource Development), UGC, NKC and other professional (e.g. Information and Library Network (INFLIBNET) and research institutes. As a result, a silent revolution is going rapidly in India in the area of repository development. Under such circumstances, it is felt that a research study in this area can be conducted to develop an IDR model based on OSS having Web-enabled and Unicode-compliant software framework that will manage IDR resources available in the organization in different forms and formats. It is expected that this software framework may act as an off-the-shelf solution in managing resources of the universities.

1.2 Objectives of the Study

In view of the background of this research, the objectives may be framed into two groups as follows:

1.2.1 General Objectives

The main objective of the study is to develop a standard IDR model and formulate a policy for the University of Burdwan that will support information needs of the university stakeholders. The general objectives of this study, as a whole, is to develop a Web-enabled distributed open access IDR model for the University of Burdwan, which can manage and organize its resources including Bengali language based interface and provision of multilingual data processing with additional utilities like federated search, interactive communication tools and standard subject access system.
1.2.2 Specific Objectives

The general objective of this study is multi-faceted one. It includes a number of interlinked and interrelated specific objectives. The specific objectives are enumerated below:

A. To design and develop a FLOSS (Free/Libre Open Source Software) based prototype IDR system based on framed architecture by applying or using Open Standards, Universal Character Set and other related Open Source Software on the basis of global and national recommendations;

B. To make IDR system accommodative and flexible in terms of specific requirements of the University of Burdwan and to make system compliant with global standards and policies related to domain;

C. To develop mechanisms for archiving, processing and retrieval (including sophisticated search mechanisms) of a wide variety of information bearing objects available in different forms and formats on the basis of internationally agreed upon data format, data standardization tools and their incorporation into the software framework;

D. To apply open standards in the design and development of the IDR system as far as information organization, retrieval and harvesting are concerned including development of federated searching and interactive communication support as additional utilities for the system; and

E. To develop Unicode-compliant Bengali script based User Interface to support integrated searching and browsing of regional languages based resources and incorporating standard subject category system in Bengali script.

1.3 Statement of the Research Problems

The objectives of this project, as framed in previous section, directed this research study towards a set of definite research problems. The research problems of this study are represented by a descriptive statement as below:

“DESIGNING INSTITUTIONAL DIGITAL REPOSITORY FOR THE UNIVERSITY OF BURDWAN: A FLOSS BASED PROTOTYPE”.
A close examination of this descriptive statement reveals that it is amenable to be converted into a set of five interlinked questions and these are as follows.

A. How to design a theoretical model for University-specific IDR? What should be the standard components and parameters for it? How to link the parameters with the requirements of users and administrators of IDR? How can global recommendations and best practice guidelines in the domain be framed to create a model theoretical prototype?

B. How to convert the theoretical model into a software framework? How to apply open source software and open standards in designing the IDR software framework? What should be the selection criteria for IDR software and related standards for different clusters of the software framework?

C. What are the procedures for making the IDR software framework Unicode-compliant entity? How to develop mechanisms for processing, retrieval and display of Indic script based information objects? What are the ways to develop and incorporate Indic script based subject access system?

D. How to incorporate into the IDR software framework (for the University of Burdwan) utilities like federated searching and interactive communication tools for scholar?

E. How to implement and integrate the prototype IDR in the existing information access system of the University of Burdwan? How to design and apply multilingual user interfaces at different levels and different points of utilization? How to make the prototype IDR interoperable and open for harvesting? And how to make distribution copy of the software framework in the form of ready off-the-shelf product for the benefit of research and academic bodies?

1.4 Hypotheses

This research study seeks to understand what university members believe about IDRs including their perceived benefits, ease of use, and role within greater system of scholarly publication. It aims to support successful implementation and maintenance of the university’s IDR. However, in response to the research problems, as formulated in the foregoing section, this research study centres on the following hypotheses:

A. It is possible to design a theoretical model with carefully crafted techno-organizational components on the basis of global recommendations, best
practice guidelines, and taking into account specific requirements for the University of Burdwan;

B. The theoretical model is amenable to be converted into an accommodative software framework through the application of related open standards and open source IDR software for different clusters of the framework;

C. It is quite feasible to make the software framework Unicode-compliant through the application of an array of FLOSS based multilingual tools for storing, processing and retrieving of digital knowledge objects available in Indic scripts in general and Bengali script in particular;

D. SKOS (Simple Knowledge Organization System), a W3C standard can be utilized for converting any standard classification into Web-enabled access format for its seamless integration with IDR framework;

E. The Unicode-compliant and FLOSS based software framework can act as IDR system for the University of Burdwan in Web-enabled environment and can be integrated seamlessly with the additional utilities like federated searching and interactive communication tools; and

F. The entire model framework with all of its complexities can be converted into ISO image for utilization by other institutes as an off-the-shelf product.

1.5 Methodology

As stated earlier, the general objectives of this study is to develop a multilingual Web-enabled distributed IDR model for the University of Burdwan and specific objectives are directed towards the accomplishments of the general objectives through development of technical components of the model. The methodology to be followed for the proposed research work may be divided into three conceptual areas.

A. Part I - Development of the theoretical model;

B. Part II – Selection of standards and software; and

C. Part III - Implementation on the basis of Part I and Part II of the work.

The proposed study has been divided into three main parts.
Part-I is concerned with development of the theoretical model. Part-II deals with the designing of prototype IDR for the University of Burdwan and Part-III is implementation of the proposed model. The steps under the above mentioned three parts are given below:

**Part I - Towards the Model**

It includes following steps and sub-steps:

Step 1: Understanding IDR, its application and implementation;

Step 2: Study of existing IDR initiatives;

2.1 Global initiatives;

2.2 Indian initiatives; and

2.3 Identification of key features of national and international IDRs and policy making agencies.

Step 3: Identification of requirements for IDR of the University of Burdwan;

3.1 Study of the University Grants Commission guidelines;

3.2 Study of the requirements of the University of Burdwan;

Step 4: Identification of Parameters for developing IDR;

4.1 Content related parameters;

4.2 Standards related parameters;

4.3 Archiving parameters;

4.4 Collection development parameters;

4.5 Collection management parameters;

4.6 Collection organization parameters including subject access system;

4.7 Multilingual parameters;

4.8 Technical parameters;

4.9 Retrieval parameters;
4.10 User interface parameters;

4.11 Workflow parameters;

4.12 Legal parameters; and

4.13 System management and Administration parameters.

Step 5: Development of the model IDR for the University of Burdwan (to be named as Burdwan University Research Archive - BURA) on the basis of identified parameters;

Part II - Towards the Prototype

Step 6: Development of criteria for different facets of BURA;

Step 7: Identification of open standard metadata schemas for each document types;

Step 8: Selection of metadata standard for each document type against the selected criteria;

Step 9: Determination of technical specifications of BURA;

Step 10: Identification of FLOSS based IDR software as per the technical specifications of BURA;

Step 11: Development of software selection criteria for BURA;

Step 12: Selection of IDR software on the basis of selected criteria;

Step 13: Installation of the selected IDR software along with necessary FLOSS based companion and /or dependent software;

Step 14: Incorporation of selected metadata schemas into the prototype BURA; and

Step 15: Development of Unicode-compliant environment for processing Indic script based documents;

Step 16: Development of Indic script based subject access system;

Step 17: Inclusion of federated search mechanism; and

Step 18: Design and inclusion of communication support system for scholars.
**Part III - Towards Implementation**

The final elements of the methodology involve the following steps:

**Step 19:** Implementation of crosswalks and interoperability standards and development of OAI-PMH (Open Archives Initiative-Protocol for Metadata Harvesting) compliant IDR;

**Step 20:** Designing of user interface in two levels;

20.1 End-User access interface with support for subject based browsing; and

20.2 Submission interface (including support for subject authority subsystem).

**Step 21:** Incorporation of vocabulary control devices (existing open standards tools) in two levels;

21.1 End-User searching with support for Indic script based subject-driven browsing; and

21.2 Indexers/Submitters.

**Step 22:** Uploading of records through distributed processing i.e. remote submission process;

**Step 23:** Design of single window embedded search interface for multiple IDRs;

**Step 24:** Development of auto alerting services and platform for scholarly interaction through blogs and discussion forum;

**Step 25:** Beta testing of BURA and final release through Web-hosting; and

**Step 26:** Generating ISO image from BURA software framework for its ready implementation in other institutes.

**1.6 Significance of the Study**

It is important for any research study to justify the need of the research under study i.e. to clarify the following question:
Why the research problem was considered significant for investigation? Or why this area was chosen for research?

Section 1.1 of this chapter explains the situation behind the formulation of research problems of this study. The statement of research problem (section 1.2) and interlinked questions associated with it is also helpful in establishing the significance of this study. The following are the main reasons behind conducting the investigation of this study.

1. It (IDR) is a multi-faceted domain and requires application of different domain specific software for developing a fully functional digital solution. On the other hand, open source software and open standards in different application domains are emerging rapidly as viable alternatives that allow library professionals to work at system level. But no research till date has suggested application of open source software (OSS) for designing university-specific IDR system.

2. Academic institutes like universities, research centres are generating valuable knowledge resources in the form of theses, dissertations, project reports, courseware, pre-prints etc., but till today no model exists in India that deals with organization of intellectual resources created by its members and dissemination of such resources over the web to overcome money, space, and time limitations.

3. Existing IDR systems (mainly research institutes and R & D sector) are not based on global standards and lack standardization in organization of resources. This study, to overcome this limitation, recommends application of internationally agreed upon data formats and deals exclusively with the integration of data standardization tools like vocabulary control devices, code list etc for organizing IDR resources.

4. Multilingual digital libraries are essential in India in order to provide wider access of IDR resources to a wider section of community. In support of this view, local language based interfaces for searching, browsing, registration and multilingual data processing environment is essential for any IDR system. Unicode, as an open standard, provides a solution in this direction. This research deals with the development of Unicode compliant Bengali language based IDR system that provides not only interface in Bengali language but also Bengali script based searching, browsing of IDR resources and services.

5. Existing IDR systems in national and international level lack policy document and do not suggest any guidelines and best practices based on global recommendations. This study, in order to solve this problem recommends a
Designing Institution Digital Repository for The University of Burdwan: A FLOSS based Prototype

university-specific model policy based on international standards and best practices that will help other institutes in planning and implementing the IDR system.

6. Most of the studies are theoretical in nature and do not provide any technical solution and direction towards the building up university-specific IDR system (keeping in view the specific requirements) using open source software (OSS) and based on best practice guidelines, global recommendations etc. But this research work proposes a theatrical framework and provides valuable practical direction towards this technical and other non-technical matters by developing a software framework on the basis of framed architecture using open standards and OSS.

1.7 Definitional Analysis

Before going to the study, it is better to provide a brief overview of key concepts used in this study.

A. Open Access

There is no agreed definition of open access. It (OA) is not a straight-forward concept. It is not a technology; rather it is a philosophy which is based on the freedom of information exchange and access; and it is also an action and a development and it is more a reality. It does not mean “anything goes”. It has many different definitions and approaches, numerous supporters and abundant detractors. It raises a variety of questions. Open Access has variously been referred to as ‘open access’ (Budapest Open Access Initiative, 2002), ‘self-archiving’ (Harnad, 1999), and ‘free online scholarship’ (Suber, 2002a). Generally it refers to the free availability of peer-reviewed literature on the public internet, permitting any user to read, download, copy, distribute, print, search, or link to the full texts of the articles. It means freely available online to anyone anywhere, with no charges imposed for access. Open access is defined as the mode of scholarly communication aiming at wide distribution of scholarly content with neither price nor any other copyright restriction (Chan & Costa, 2005; Yiotis, 2005). Suber (2002b) was of the view that open access to scientific articles means online access without charge to readers or libraries. The concept has existed since the mid-1990s, with champions such as Stevan Harnad (Harnad, 1994) expressing his ‘subversive proposal’. More recently, OA has found much wider support through out the world mainly after the publication of few important public statements and declarations commonly known as the three Bs, the Budapest, Berlin, and Bethesda. The following two definitions provide a
comprehensive explanation of open access. The Berlin Declaration of Open Access (2003) defines open access as a new mode of scholarly communication through which “the author(s) and right holder(s) of such contribution grant(s) to all users a free, irrevocable, worldwide right of access to, and a license to copy, use, distribute, transmit, and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship”. The Budapest Open Access Initiative (2002), launched in February 2002 and spearheaded by the Soros Foundation’s Open Society Institute states that by Open Access they mean:

“free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself”.

The declaration acknowledges that the literature referred to in their definition is not free to produce, but they add that “… experiments show that the overall costs of providing open access to this literature are far lower than the costs of traditional forms of dissemination.” The Bethesda Statement (2003) had the goal of providing open access to primary scientific literature as quickly as possible.

B. Open Standards

A standard is an agreement between a number of (but not necessarily all) players within a certain area of technology. Recently one question has been attracting great attention: whether a specific standard is of open or a proprietary in nature. “Open standards” has become a very popular term in recent policy debates, but controversy abounds over what it actually means. Open Standard (OS) means different things to different people. In practice, it is difficult to reach an agreed definition. It is merely a feel-good term with no actual technical meaning. Open standards are international accepted protocols (technical specifications) which guarantee interchangeability between and with other ICT-systems. Pountain (2003) defines an open standard as "a standard that is independent of any single institution or manufacturer, and to which users may propose amendments." Another expert (Coyle, 2002) identified three key characteristics: anyone can use the standards to develop software; anyone can acquire the standards for free or without a significant cost; and the standard has been developed in a way in which anyone can participate. Bruce Perens (2004) defines open standard in the following words

“Open standards are transparent, open ended and freely implementable, and accessible over open forum”.
A proprietary standard is characterized by the fact that it is owned by some one who puts restrictions on – or can put restrictions on – users’ access and use. On the other hand, a completely open standard has the following properties:

- It is accessible and free of charge to all (i.e. there is no discrimination between users, and no payment or other considerations are required as a condition of use of the standard);
- It remains accessible and free of charge (i.e. owners will not access limit to the standard at a later date); and
- It is accessible free of charge and documented in all its details (i.e. all aspects of the standard are transparent and documented, and both access to and use of the documentation is free).

The purpose of open standards is to ensure interoperability so that different systems can interact with each other without problems. It facilitates use of the products in new areas and in new ways for the benefit of society.

C. Open Source and Open Source Software

Open Source and Open Source Software (OSS) are both a philosophy and a process. It is not a new idea. It is a philosophy describing the intended use of software and methods of distribution. Open source software is software that includes source code along with the binary version and is usually available at no charge. It is governed by a license under which anyone can access, modify, and further distribute the source code. Open source is a software development model as well as a software distribution model.

The open source movement (OSM) has been in conscious development for nearly two decades but the term “open source” itself has been a relative latecomer. Christine Peterson of the Foresight Institute proposed the term open source in late 1997 during a meeting of small group of open source movement key persons (Raymond, 2001). This group registered the domain name opensource.org, defined “open source”, developed Open Source Initiative (OSI) group, designed OSI certification, and created a list of licenses that meet the standards for open source certification. According to OSI (Open Source Initiative, 2003)

“Open source promotes software reliability and quality by supporting independent peer review and rapid evaluation of source code. To be certified as open source, the license of a program must guarantee the right to read, redistribute, modify, and use it freely”.

On the basis of the discussion and definitions given by several experts, following attributes of OSS may be listed –
C

OSS is typically created and maintained by developers crossing institutional and national boundaries, collaborating by using Internet based communications and development tools;

• OSS development process follows the famous Linu’s law – “Release early, release often and listen to users”;  
• Quality, not profit, drives open source developers who take personal pride in seeing their working solutions adopted; and
• Intellectual property rights to open source software belong to anyone who helps to build it or simply use it and is not locked to any single vendor or institution.

D. Open Access Scholarly Communication

Numerous definitions of Open Access (OA) scholarly communication may be found; the central tenet being that the reader does not pay for access to the research, and that the author may or may not pay for the ‘making available’ of such research. The key concept here is access; that research output is not expensive (i.e. there is a minimal fiscal barrier) and that research output is accessible (i.e. access to the material is not restricted in terms of physical infrastructure nor restricted by authentication systems). Scholarly communication is a broader term reflecting various processes through which scholars exchange information with each other in the course of knowledge creation. The system of scholarly communication has evolved over time since it came into existence. Open Access scholarly communication, and more specifically scholarly publication, is an important manifestation of knowledge generation and diffusion. In the context of this study, open access scholarly communication means the process through which scholars access and/or disseminate scholarly information through Open Access Journals (OAJs) and/or Open Access Archives (OAAAs) as defined above.

E. SKOS (Simple Knowledge Organization System) based Subject Access System

SKOS is a W3C recommendation designed for representation of thesauri, taxonomies, classification schemes, subject-heading systems, or any other type of structured controlled vocabulary. It is part of the Semantic Web family of standards built upon resource description framework (RDF), and its main objective is to enable easy publication and use of such vocabularies as linked data (http://en.wikipedia.org/wiki/Simple_Knowledge_Organization_System). It was developed initially by July 2003 (as open-source project) within the project SWAD-E (Semantic Web Advanced Development for Europe) aimed to define a model for thesaurus compliant with most important ISOs (viz. ISO 2788, ISO 5964). Since
September 2004, the responsibility for the subsequent development was passed to the W3C and on 10 May 2005 was published the first Working Draft.

In library science, it is a common data model or a tool for sharing, organizing, presenting and linking knowledge organization systems via the Web using the resource description framework (RDF). KOS has classically been used in the LIS field as a way of indexing large volumes of documents in order to facilitate document retrieval and navigation. The advent of SKOS now enables the construction of KOS that can be used to index documents and share terminologies on the web. Most KOSs typically contain hierarchical categorizations of concepts, these categories are generally designed with a particular application in mind. One common type of application is a document navigation system where categories are organized in such a way to improve how users browse a document collection. The following are the advantages of SKOS based subject access system:

- Provide a low-cost migration path for porting existing organization systems to the Semantic Web/Data Web;
- Publication of knowledge organization systems;
- Enable easy publication of controlled structured vocabularies for the Semantic Web;
- Provide a lightweight, intuitive conceptual modeling language for developing and sharing new KOSs;
- Mapping support;
- Significant take up and usage;
- Web-oriented representation;
- Re-use & sharing of concepts and their descriptions;
- Linking between concepts from different contexts;
- Extensibility; and
- It can be used on its own, or in combination with more-formal languages such as Bengali

All digital libraries use one or more KOSs and ranges from authority files to classification schemes, thesauri, and ontologies. In IDRs environment, resources are organized in many ways (as stated in section 4.2.1.2 of chapter 4) and no unique system has been developed for organizing and presenting IDR resources in proper perspectives. So, here is the need to use ontology driven SKOS based subject access system for presenting IDR resources in a structured way. Though representing KOS (classification systems) on the Web for publication, presentation and exchange continues to be a challenge within the SKOS framework. The incorporation of Indic script based KOS in any repository system and transformation or implementation of KOS into Web-enabled format is necessary to meet the subject approach of the users. This research work has selected Dewey Decimal Classification (DDC) (DDC 22nd edition – up to 3rd summary) as KOS and modified it in Bengali (as demonstrated in section 5.4 of chapter 5). This KOS has been incorporated in BURA software framework and integrated with administrative interface as well as user interface so that it can be used at the time of indexing and searching.
1.8 Conspectus

The structure of the research study falls into three core parts and includes a total of seven chapters as outlined below. The three parts are as follows:

A. A theoretical background that establishes need of conceptual model and based software framework which supports distributed, integrated and Bengali language based IDR system;

B. An empirical investigation that establishes features, components, architecture of an IDR system or model and develops a software framework on the basis of that model through the application of a group of FLOSS based software, open standards and Universal character set; and

C. Analytical discussion on perceptiveness of features and facilities of the Unicode compliant and FLOSS based software framework in managing various facets of IDR system, and accessing IDR resources and services related with academic institution.

The theoretical background comprises three chapters. Chapter 1 primarily concerns with background of the research and statement of research problems taken up for this study along with its associated elements. It introduces the concept of scholarly communication and OA before stating the problem of the study. The chapter also presents the aim and objectives of the study, research questions, significance of the study, definition of key concepts, a brief outline of the research design, methodology used to carry out this study and the thesis structure.

Chapter 2 furnishes a review of the research results with reference to recent research trends on various facets of IDR system and Indian scenario in the domain of repository movement. It emphasizes the need of present research and identifies limitation of the present system in the literature. The purpose of this review is to position the study within similar works as well as explore the available knowledge in the study discipline.

Chapter 3 presents basic information regarding IDR such as history, features, major initiatives and projects in national and international level. This chapter also explains existing IDR system in India and highlights on problems of the system along with tentative solutions.

Chapter 4 describes policy decisions regarding several key issues such as contents, copyright and licensing model, quality of contents, access to contents, metadata, preservation, multilingual resource management and back up etc. It describes
internationally accepted metadata standards for different types of objects such as learning objects, dissertations and theses etc, and describes different domain specific open source software to facilitate a model where all software can interact with each other seamlessly. The chapter also describes three clusters and software associated with it. It then concentrates on the selection of IDR software on the basis of global parameters to develop a software framework as proposed in the model. The chapter also describes different harvesting software and develops harvesting software framework for BURA that support federated searching in a single-window search interface.

Chapter 5 deals with the practical implementation of the software framework of the model BURA. It deals with the development and customization of the software framework to support Web-enhanced and Unicode-compliant distributed IDR system for the University of Burdwan. This integrated Web-enabled model (BURA - Burdwan University Research Archive) is based on open standards following open source software. This chapter concerns with the customization of user interface in Bengali as well as integration of social software like Blog, RSS in BURA software framework. It describes in details designing and development process of metadata harvesting framework for metadata extraction from different OAI-PMH compliant repositories. The chapter demonstrates the development of ontology driven Indic script based subject access system in the proposed software framework and generates an ISO image for utilization by other institutes or universities as an off-the-shelf product.

Chapter 6 describes different services offered by the system. It explains with the help of various snapshots, the steps involved in administering DSpace, like creating and maintaining Communities, Sub-communities, Collections, E-People; authorizing E-People to submit to the Collections; creating and maintaining E-Groups, service model, searching and browsing of multilingual objects, resource harvesting mechanism etc. The chapter highlights on simple and advanced search techniques including browsing and searching of different Communities and Collections of the software model. It shows the process of browsing, searching and retrieving of resources in Bengali language. This chapter also shows the process of browsing and searching specific subject using subject access system to enhance subject categories. This chapter also describes mechanisms of searching and harvesting metadata from multiple OAI-PMH compliant repositories and deals with integration of social software in BURA software framework.

The chapter 7 is the final chapter of the thesis. In this chapter, the findings arising from the previous six chapters are discussed, common threads gathered and matters of distinction highlighted. Also discusses the limitations inherent in this body of research, identifies possible areas of further study and presents the concluding remarks. Key conclusions are then drawn as to the implications and presents the overall summary and recommendations of the study, including areas for further study.