Chapter 2

Literature Review

“If you have an apple and I have an apple and we exchange these apples then you and I will still each have one apple. But if you have an idea and I have an idea and we exchange these ideas, then each of us will have two ideas.”

George Bernard Shaw (Phi Kappa Phi, 1952)
Structure

2.0 Introduction
2.1 Facets of Institutional Digital Repository (IDR) System
   2.1.1 Advocacy and Promotion
   2.1.2 Content Recruitment
   2.1.3 Digital Preservation
   2.1.4 Interoperability and Open Access (OA)
   2.1.5 Legal and Intellectual Property Right (IPR)
   2.1.6 Metadata
   2.1.7 Models
   2.1.8 Multilinguality
   2.1.9 Quality
   2.1.10 Software
   2.1.11 System management and Administrative parameters
   2.1.12 User Interface
   2.1.13 Versions identification and management
   2.1.14 Workflow management
2.2 Indian Scenario
2.3 Findings of Literature Review: Gaps and Overlaps
2.4 Need of Present Research
2.0 Introduction

The literature review aims to place the current investigation within the existing research literature, both in terms of how the literature review has impacted the investigation, and how the research undertaken will contribute to the professional understanding of the subject. Conducting a literature review for relevant research on the open access repositories (institutional repositories in particular), was a challenge in itself. A large amount of the work done in these subjects has theoretically explored the issues but there is little research on the actual instances. This research study attempts to give a clear idea to the topic selected for the study. This research is a relevant topic with current importance. Because there is no much study are conducted earlier in this area. As IRs is still in a growth phase, the building and development of an IR has yet to be addressed empirically in the literature (Swanepoel, 2005) and the literature review may be used as an integral part of the research process in particular when the research topic is relatively young, providing an initial foundation for a new research topic (Levy & Ellis, 2006). As this is a relatively new area of enquiry, one of the main attractions is that there is plenty of scope for research but one of the drawbacks is discovering the relevant literature. It enables one to be aware of the past and current trends in any particular branch of research and devotes to examine the review of works relating to various aspects of repository movement throughout the world. This review is not just an analysis of the relevant literature that was collected, but it is also an exercise that attempts to describe the key elements and characteristics of this mainly uncharted territory. This literature review therefore, not only analyses and synthesizes the available previous work on the subject (Levy & Ellis, 2006), but also gathers and collectively overviews literature on these issues providing an initial foundation for this new research topic.

The purposes of literature review is two fold – (i) to carry out to identify, locate and synthesize research reports, articles, books and Web-resources; and (ii) to put this research oriented book into proper perspectives. Under such circumstances, the objectives of this chapter on literature review are:

- To discuss different facets of institutional digital repositories (IDRs) and shows how they are related to the development of IDRs;
- To review recent research trends on various facets of IDRs;
- To examine Indian scenario in the domain of IDRs;
- To identify gaps and overlaps of existing literature, if any; and
- To explain the need of present research for the University of Burdwan.

Literature search for this study has been done by examining general and specialized bibliographic tools, indexing and abstracting tools, directories, Web portal of journals, electronic databases, and other appropriate references tools.
2.1 Facets of Institutional Digital Repository (IDR) System

This particular section has discussed some of the facets involved in setting up and running an IDR system. This research work divides the whole array of digital repository into two major sections for analyzing the research trend of the field under study. The first section examines the different facets identified throughout the literature and the second section is confined to India and reviews existing literature on Indian context.

2.1.1 Advocacy and Promotion

One aspect of the literature that is directly related to the ‘back-end’ processes of IR management is the discussion of advocacy, marketing for IRs. As IRs, OA, and self-archiving are unfamiliar and relatively new concepts to the academic community, advocacy has become a crucial aspect of any IR project (Chan, Kwok & Yip, 2005). Leary, Lundstrom & Martin (2012) rightly pointed out that continued marketing leads to continued growth of the IR, making it all the more necessary that the IR runs smoothly. There are a number of strategies (e.g. brochures, campus newsletters, campus newsletters/newspapers etc.) that existing institutional repositories have used, or can use to encourage acceptance and uptake of their repository (Mark & Shearer, 2006; Horwood et al., 2004; Ashworth, Mackie & Nixon, 2004; Pinfield, Gardner & MacColl, 2002; Barton & Waters, 2004–2005). Setting up an archive is one thing, getting users to participate in its ongoing development is quite another. Many authors (Jones, Andrew & MacColl, 2006; Mackie, 2004; Markey et al., 2007; Bevan, 2007; Greig, 2004) described several methods to populate IR. Gierveld (2006) considers IR as a product that needs to attract a market. The other experts (Dill & Palmer, 2005) proposed some promotional ideas for the implementation of an institutional repository. Another study (Morgan & Team IDR, 2006) identified a number of things about promoting awareness, marketing, and advocacy on issues surrounding scholarly communication. This study (Weenink, Waaijers & Godtsenhoven, 2007) investigates the challenges in populating repositories particularly in European context based on six good practices that influences the development of IR.

Johnson (2008) is of the view that ‘advocacy is a route to achieve the crucial goal, real cultural change’. Markland & Brophy (2005) are of the view that for an institutional repository to be successful, cultural change must be achieved through advocacy or ‘getting the right message to the right people with the tone and contents varied by audience’ (Johnson, 2008). Several other experts (Aschenbrenner et al., 2008; Lynch, 2003; Westell, 2006) put emphasis on collaboration and participation of stakeholders or community members as it may be a new concept to them. The stakeholders include both academic and non-academic staff; those involved in
teaching or research; and both postgraduate and undergraduate students. Each of these groups contains potential authors and readers of material. The contributions of authors, in particular, are critical to the success of an IDR. Thibodeau (2007) puts emphasis on collaboration among stakeholders and concluded that if an institutional repository works in isolation, it might signify failure because it does not share and transfer knowledge.

A number of different advocacy strategies can be used, including top-down and bottom-up, blanket and targeted approaches (DRIVER, 2008). But there is a serious lack of visibility (McKay, 2007; Davis & Conolly, 2007) and usage (Kim, 2006; Woodland & Ng, 2006) among authors and information seekers. Ignorance among researchers about copyright and plagiarism issues may restrict adoption of open access repositories (Suber, 2008; Abrizah, 2009). Another group of authors (Davis & Connelly, 2007; Watson, 2007) put emphasis on faculty attitudes, motivations, and behaviors for non-participation in institutional repositories and encourage academic communities to deposit their work. Other experts (ALPSP, 2002; Bentum, 2001; Gonzalez & Porcel, 2007) reported that quality of contents is another problem of faculty participation. For faculty and students, JISC stresses the advantages of “free sharing of information, encouraging collaboration and the widespread communication of institutional education and research activity” (JISC, 2005). A few other experts (Shearer, 2003; Kim, 2006) suggest that the success of IRs will be determined eventually by their uptake and use by researchers and argues that one of the measures for IR usefulness is contribution of contents. And the repository will need to be effectively marketed to members of the university community in order to acquire contents (Kiran & Chia, 2009). Other experts (Hubbard, 2003; Ware, 2004) identified cultural rather than technological factors limit the use and development of IRs. Another study (Bailey et al., 2006a) reported that contents recruitment was also one of the biggest challenges in IR development.

Another report (Gonzalez & Porcel, 2007) is in support and Pickton & McKnight (2006) put emphasis not only on academic staff but also on library staff and mentioned possible mechanisms for promotion of IR. Supporting staff in deposition process through the use of a system of ‘mediated deposit’ is one solution. Another study (Jenkins, Breakstone & Hixson, 2005) discussed in detail various attempts that they had made at the University of Oregon to increase user awareness about their IR. They further suggested that reference librarians have a vital role to play in helping to recruit authors to submit their contents to IR, as well as in educating users to search such repositories effectively and retrieve the scholarly contents from them. Horwood et al. (2004) briefly outline how they do this at the University of Melbourne, and refer readers to Nixon’s (Nixon, 2002) paper for further details. Ashworth, Mackie & Nixon (2004) also reported that they had to do this at the University of Glasgow.

Providing incentives such as awarding prizes to top depositors has also been found to be effective for the development of IR (Cullen & Chawner, 2008). Xia et al. (2012)
showed how this strategy for encouraging faculty to participate is working in University of Minho in Portugal. Another useful report (House of Commons Science and Technology Select Committee, 2004) recommended for compulsory deposit policy of public funded research of all levels of education. Another group of authors (Pinfield, 2005; Harnad, 2006a, 2006c; Swan & Brown, 2005; Sale, 2006b; Suber, 2006; Rowland et al., 2004) strongly support this view. Peter Suber (Suber, 2008) in another paper recommends three principles, two of which relate specifically to institutional repositories. The first principle suggests that the university provide open access to all research output. He recommends the use of mandatory language regarding university expectations; faculty and staff education and assistance; and incentives to use the repository.

2.1.2 Content Recruitment

“Recruitment of content, not technology, is the greatest barrier to success” (Gibbons, 2004b) and is the central issue for most IRs (Westrienen & Lynch, 2005). Many libraries are facing the challenges to develop and manage contents of an IR. Studies consistently report that recruiting contents is difficult (Heery & Anderson, 2005; Davis & Connolly, 2007; Salo, 2008). Several studies (Westell, 2006; Sale, 2006a; Jingfeng, 2006; Kingsley, 2008) identified several questions and issues (mandate; integration with planning; funding model; relationship with digitization centers; interoperability; measurement; promotion; and preservation strategy, personal name identification etc.) those need to be considered before recruiting contents for an IDR system. Recruiting contents for IRs has been difficult strategies that IR staff have enlisted (Crow 2002b; Chan, 2004; Shearer, 2004, 2005; Mackie, 2004; Barton & Waters, 2004–2005; Jenkins, Breakstone & Hixson, 2005; Graham, Skaggs & Stevens, 2005; Chan, Kwok & Yip, 2005; Bell, Foster & Gibbons, 2005). Contents recruitment was also identified as one of the biggest challenges in IR development (Bailey et al., 2006a). He also identified several motivating factors in the ARL survey. These motivators include implicit values of librarianship; collecting and preserving an institution’s scholarship and ensuring that it is accessible. Carr & Brody (2007) state that the key to a successful repository is sustained deposits. To achieve sustained deposits, community engagement is necessary (Henty, 2007). A few studies (Davis & Connolly, 2007; Foster & Gibbons, 2005; Rieh et al., 2007b) reported that faculty’s lack of understanding of the IR as one of the problems of recruiting contents. In the same article, Davis & Connolly (2007) concluded that use of IR by faculties depend on disciplinary cultures and reward structures while Foster & Gibbons (2005) mentioned three ways (deposit process, created researcher pages for faculty, promoting IR in the faculty’s language etc.) of promoting contents of IR. The other experts (Xia et al., 2012) identified additional issues that prevented many faculty members from contributing to their institution’s repository included limited technical skill levels and a lack of additional time to deposit research articles.
Aggressive marketing (Troll Covey, 2011) and providing value-added services are necessary to increase faculty participation (Bankier & Perciali, 2008). Giesecke (2011) is not enough; additional incentives must be built in as well. The other studies (ALPSP, 2002; Bentum, 2001; Gonzalez & Porcel, 2007) reported that the process depends on quality of contents.

Another group of authors (Hassen, 2006; Mark & Shearer, 2006; Kim, 2006) identified and discussed several others recruitment strategies (accessibility, publicity, professional recognition, contents harvesting, usage, feedback) for an IR. Jenkins, Breakstone & Hixson (2005) described recruitment approaches through creation of user communities by the librarian. Genoni (2004) suggests that collection development plans should be developed to prioritize contents added to a collection. Local efforts to embed the repository within the research processes of an institution are not always successful. Feijen et al. (2007) suggest that what is needed is the embedding of repository use in research and research publication processes on a large scale. Another study (Pettijohn & Neville, 2003) reported how libraries have responded to this paradigmatic shift by pioneering new collection development strategies, and then examines the changing responsibilities of collection development librarians in an electronic environment.

2.1.3 Digital Preservation

Digital Preservation is another critical issue of digital library environment and has been currently a major concern for the information management, technological and scientific communities in all domains. The issue about addressing long-term access to contents is very strong in the literature and preservation is perceived as one of the key issues that can determinate the success of IDR. The initiative to support digital preservation and asset management in institutions has been discussed in several studies (Pinfield & James, 2003; Smith & Moore, 2007; Weenink, Waaijers & Godtsenhoven, 2007; Wheatley, 2004; Gibbons, 2004). Several key initiatives were addressing digital preservation issue such as SHERPA (http://www.sherpadp.org.uk/index.html) and the Digital Preservation Coalition (DPC) were the most prominently initiatives. The DPC (http://www.dpconline.org) offers a wealth of information and practical guidance on digital preservation. JISC Digital Preservation and Records Management Programme launched the Digital Curation Centre and support the Digital Preservation Coalition, which aims to develop different approaches to achieve long-term preservation (Carpenter, 2005). Another initiatives has been taken by Stanford University developed the LOCKSS (Lots of Copies Keeps Stuff Safe), an OAI compliant software system that exploits peer-to-peer networking technology to maintain electronic publications copies (Eaton, 2005). The CLOCKSS initiative (jointly taken by Library of Congress and Stanford University) is a collaborative, community initiative to build a trusted, large-scale, dark archive. It is intended to
Designing Institutional Digital Repository for The University of Burdwan: A FLOSS Based Prototype

provide a decentralized and secure solution to long-term archiving, based on the LOCKSS technical infrastructure (http://www.lockss.org/clockss/). Another excellent source for learning about digital preservation was the PADI (Preserving Access to Digital Information) site published by the Australian National University. The site offers a concise guide to the most often discussed strategies for digital preservation – migration, adherence to standards, encapsulation and emulation—along with additional resources for in-depth information (http://www.nla.gov.au/padi/about.html). Another approach is taken by Portico (http://www.portico.org/) which aims to preserve e-journals.

Digital preservation is a complex process and there are still many unsolved issues which make it a challenging task for IDR administrators. Several authors discussed the issues and challenges of digital preservation facing institutional repositories (Jones, Andrew & MacColl, 2006; SPARC, 2002a; Bradley, 2005; Yakel, 2007; Cordeiro, 2004; Lin, Ramaiah & Wal, 2003) highlighting current research and standards efforts. A few authors (Hockx-Yu, 2006; Lor, 2005; Patel & Simon, 2007) described various other issues (i.e. technical, organisational, economic, political, legal & ethical) and another report (PALS report) explained different approaches (Ware, 2004) of digital preservation. Another important issue such as fear of technological obsolesce and need of preservation was discussed in few articles (Pinfield & James, 2003; Jones & Beagrie, 2002).

Preservation of digital format or file format is more complex than that of paper-based information, mainly due to the rapid advances in technology. The issues raised by the long-term preservation of digital objects are very far from solved. There are clear differences between file formats because a file format that is good for access today may not be a format that is easy to migrate, but a format that is easy to migrate may not be easy to read. Some IR systems are format-neutral—the system accepts deposits of any digital format whereas other systems are hard-coded— to only accept formats of certain types. This vital issue has been discussed in several studies (Abrams, 2004; Rosenthal, 2010; Thompson, 2010; Hitchcock & Tarrant, 2011; Jones, Andrew & MacColl, 2006; Bailey et al., 2006a). Another group of authors (Curtis, 2006; James et al., 2003; Weenink, Waaijers & Godtsenhoven, 2007; Morgan & Team IDR, 2006) argued different file format for different types of digital objects. Several other experts (Aschenbrenner & Kaiser, 2005; Pinfield, 2002; Cervone, 2004) concluded that PDF as preferred format is likely to guarantee the preservation of the document as is widely used in many repositories. Smith (2002) compared the changeability of the digital format with a volatile and fickle object and emphasizes that society in general does not understand or appreciate the complexity of the problem.

Another issue highlighted in the literature is sustainability and long term preservation of digital object (Bullock, 1999; Harmsen, 2008; Hockx-Yu, 2006; Stanescu, 2005; Wheatley, 2004). Long-term preservation of scholarly contents is an essential role of IRs (Lynch, 2003). Not a one-time event, long-term preservation consists of
specific functions such as ingesting digital objects in which metadata is created, storing such objects and associated metadata, monitoring technology obsolescence, and evaluating the usage of digital objects (Fyffe et al., 2004). Shaon & Woolf (2011) investigates the requirements for ensuring sustained access to environmental data from the perspective of a preservation-aware SDI.

Jones, Andrew & MacColl (2006) stated that ‘one objective of many repositories is to provide items in perpetuity’. The APSR (Australian Partnership for Sustainable Resources) discussion paper on sustainability issues by Bradley (2005) goes into considerable detail on issues related to sustainability and long-term preservation, and steps needed to achieve this. The SPARC (2002b) report is in support and addresses preservation in perpetuity in more detail. Another paper addresses the question of whether or not e-prints should be preserved (Pinfield & James, 2003). Another group of authors (Calanag, Tabata & Sugimoto, 2004; Dondorp & Meer, 2003; Ware, 2004) described different projects and techniques of digital preservation along with relevant preservation standards and best practices for digital objects. Another survey report (Li & Banach, 2011) reveals the challenges and opportunities of implementing digital preservation for IRs in a complex environment with rapidly evolving technology and examines the current practices and standards of digital preservation of IR materials.


Many IRs employed shared standards such as the Open Archival Information System (OAIS) reference model and the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The OAIS model provides a comprehensive framework of all the functions required for digital preservation including ingest, storage, retrieval, and long-term preservation of digital objects. And OAI-PMH is a mechanism for harvesting extensible markup language (XML)-based metadata from repositories, and it therefore makes possible interoperable search and retrieval among repositories (Branin, 2005).

The other experts (Calanag, Tabata & Sugimoto, 2004; Chilvers & Feather, 1998; James et al., 2003; Cantara, 2006) shared their experience of making metadata as another solution of digital preservation. Lavoie & Gartner (2005) provided a definition of preservation metadata, described its role in the digital preservation process, and reviewed a number of existing preservation metadata initiatives, with an emphasis on identifying points of convergence and divergence among them. Another researcher (Day, 1997; Groenewald & Breytenbach, 2011) suggests how the concept of metadata could be extended to provide information in the specific field of digital preservation. Other researchers (Pugin, Hankinson & Fujinaga, 2012) discussed a
new Web-based cataloguing system for the global music heritage materials. Another expert (Goh et al., 2006) mentioned that preservation not only refers to the preservation of metadata but also to the use of quality control measures to ensure integrity, and persistent documentation identification for migration purposes.

Smith (2002) discussed several preservation techniques and mentioned that there are at least four approaches to preservation (migration; technology preservation; emulation; and persistent object preservation) and there is significant research going on under nine headings. Another expert (Kuny, 1997) identified the gap in terms of the three main methods of preservation namely technology preservation, technology emulation and information migration. Other expert (Granger, 2000) describe and focus on "Emulation" as one of the technique of preservation and Wheatley (2001) on "Migration" as another technique. Han (2004) suggested preservation of digital objects through software. Another study (Jantz & Giarlo, 2005; Chen, 2007) described architecture and technological aspects and put emphasis on new methods, policies, standards, and technologies of digital preservation that can be integrated into an operational digital preservation framework. Another group of authors (Hitchcock et al., 2005; Ferreira, Baptista & Ramalho, 2006) described preservation service architecture. Stanescu (2005) also remarks how implementing models for achieving preservation must be a priority: he claims that preservation plans should be based on objective analysis of risk trends (for file types, software and hardware) rather than on individuals’ opinions and experiences. Another project at the University of Michigan has developed its own preservation plan and provides three levels of support for scholarly materials and describes best practices for different file formats (Deep Blue, n. d.). The other authors describe the set of components that are necessary to build a Service-Oriented Architecture (SOA) to enable cultural heritage institutions to carry out digital preservation with minimum human intervention (Ferreira, Baptista & Ramalho, 2006). Another group of authors (Hitchcock et al., 2005, 2007; Dondorp & Meer, 2003; James et al., 2003) described different OAIS models. Another blue book (Consultative Committee for Space Data Systems, 2002) has more detailed recommendations about best practice and appropriate work flows for each stage of the preservation process. Another question is who should be responsible of this duty? Cervone (2004) identifies two stakeholders: publishers and librarians. The SPARC report addresses preservation in perpetuity in more detail and proposes to lay the task in the hands of librarians, those professionally prepared (Crow, 2002a).

2.1.4 Interoperability and Open Access (OA)

Crow (2002a) reported that IRs is by their nature usually designed to be OA and a primary goal is to be able to link up with other similar archives. He further suggested that systems (IR) must be able to support interoperability and able to harvest metadata
through multiple search engines and other standard discovery tools in order to provide broader access to the academic and research community, users outside the institution. It should be interoperable, which refers to the capability of a computer hardware or software system to communicate and work effectively with another system in the exchange of data, usually a system of a different type, designed and produced by a different vendor (Reitz, 2006). Westell (2006) states that interoperability indicates openness on behalf of the institution or the library which is willing to contribute to national and international scholarship.

Several research studies (Hunter & Guy, 2004; Horwood et al., 2004; Ginsparg, Luce & Van de Sompel, 1999; Van de Sompel & Lagoze, 2000; Mazurek et al., 2006; Eaton, 2008; Alipour-Hafezi et al., 2010; Khazraee et al., 2011; Miller, 2004; Westell, 2006) explained the necessity of OAI Protocol for Metadata Harvesting (OAI-PMH) in digital repository environment and described Open Archives (OA) as a tool by which interoperability can be achieved (Alexander & Gautam, 2004). Interoperability between digital repositories can be achieved in various ways and on various levels. “To achieve this, the parties need to agree on the structure, the syntax and the semantics of the data and/or metadata objects, and they need to either comply with common formats or to provide a means for metadata mapping and/or for object conversion” (Aschenbrenner & Kaiser, 2005). Suleman (2001) reported that the success of making interoperability for digital library depends on vigilance in specification of the protocol as well as standardization of implementation. Peset et al. (2007) highlighted the situation of the development of repositories that use the Open Archives Initiative (OAI) protocol for data collection.

The other experts (Jerez et al., 2004) focused on the multi-faceted use of the OAI-PMH to access stored contents in repository architecture. Bell & Lewis (2006) described OAI-PMH as a part of automated export/import process of software components and being a part of the process it encourages the preservation of multiple copies for safety of the contents. Another group of authors explained its metadata formats (Van de Sompel et al., 2004; Van de Sompel, Young & Hickey, 2003) whereas others expert explain its origins in promoting E-Prints, the details of its technical standard for metadata harvesting, the applications of this standard (Lagoze & Van de Sompel, 2001). Warner (2001) showed in practical through Perl code how OAI-PMH can be used to expose and harvest metadata from Archives. Prom (2003) outlined a method for exposing deep, hierarchical metadata from encoded archival description (EAD) files and assesses some theoretical and practical issues that will need to be confronted by institutions choosing to provide or harvest OAI records generated from EAD files. Kaczmarek & Naun (2005) explained the construction of a meta search service model based on the Z39.50/OAI Gateway Profile and concentrated on the technical aspects of building such a service model rather than on the usability of the interface or on questions of interoperability at the metadata level. Van de Sompel & Lagoze (2000) described the Santa Fe Convention of the OAI, a set of relatively simple but potentially quite powerful interoperability agreements that
facilitate the creation of mediator services. Pieper & Summann (2006) reported the activities of Bielefeld University Library in establishing OAI based repository servers and provides an overview of the functionalities of BASE (Bielefeld Academic Search Engine) and gives insight into the challenges that have to be faced when harvesting and integrating resources from multiple OAI servers.

2.1.5 Legal and Intellectual Property Right (IPR)

Intellectual property and copyright has become a critical issue in digital environment. Intellectual property and copyright legislation are major and controversial issues and present a number of challenges for institutions implementing repositories (Jones, Andrew & MacColl, 2006; Jones, 2007). They have discussed this issue and suggested that the institution may be one of three legal positions: Common carrier - absolute immunity; Distributor - not subject to liability unless the have specific knowledge; and Publisher - liable for what they choose to publish. The copyright issues have been extensively reviewed in RoMEO (Rights METadata for Open Archiving) project (Gadd, Oppenheim & Probets, 2003a). This RoMEO Project has compiled a list of many journals' "Copyright Policies" about "self-archiving." The project investigate the rights issues surrounding the 'self-archiving' of research in the United Kingdom academic community under the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). In another paper, they recommend the use of Creative Commons (CC) licenses to express the rights attached to individual research papers (Gadd, Oppenheim & Probets, 2004). Another project "A guide to developing open access through your digital repository" covers important issues (e.g. licensing, technical considerations such as software and metadata) relating to copy right and provides valuable directions to the repository managers ((Pappalardo & Fitzgerald, 2007). This report contains important sections that discuss pre-prints and post-prints and copyright issues. ‘Third party copyright’ was recently raised as an issue by several people representing institutions that are in the process of implementing repositories. This vital issue requires more focus hence the creation of this section and has been discussed very briefly in the literature (Jones & Andrew, 2005; Gadd, Oppenheim & Probets, 2003b).

One of the more obvious issues in digital library environment is copy right issue both for publishers and for authors whether to deposit their research into the repository. Willinsky (2002) reviewed the specifics of publishers' contracts with editors and authors, as well as the larger spirit of copyright law. The Study (Swan & Brown, 2005) noted that authors are not always aware of the full copyright implications of their work and had fewer concerns over the use and potential abuse of their work (i.e. plagiarism, integrity and commercial use) (Gadd, Oppenheim & Probets, 2003c). In another article they concluded that material placed in an IR will be owned by the institution, the author or, in the case of a post print, a publisher (Gadd, Oppenheim &
Another study (Hoorn & van der Graaf, 2006) report that 71% prefer to keep copyright, 2% prefer to transfer copyright to publishers, 23% are neutral about the choice between the two, and 4% do not know.

Another legal issue is plagiarism and most of graduate students and research supervisors expressed concern about the risk of plagiarism (Copeland, Penman & Milne, 2005; Greig, 2005; Friend, 1998; Jewell, Oldfield & Reeves, 2006). They further argued that plagiarism and misuse is facilitated by making theses and dissertations electronically accessible through the web.

Copyright or intellectual property right (IPR) is seen as a significant barrier often confronting institutions, which have adopted ETD (Electronic Theses and Dissertation) programs (Copeland & Penman, 2004; Leung, 2005). The issue of copyright is critical to the development of IRs, especially with regard to journal articles and materials for which others may own the copyright. Greig (2004) provides a useful overview of how to populate an IR and stay within the legal limits (http://www.arl.org/sparc/meetings/ir04/presentations/greig.html). Another Study (Morgan & Team IDR, 2006) recommended that copyrights for things like articles, working papers, technical reports, etc. are owned by the faculty and students who wrote them until they are transferred in writing.

Another report (ALPSP, 2002) suggests that authors and faculties are now aware of the IPR issues and they consider it important to retain copyright with them. They don’t want their materials to be used by others and want to keep self-archiving rights and rights for personal educational use with them (Kling & McKim, 2000; Bennett, 1999). Probets & Jenkins (2006) pointed out that the complexity of IPR is of great concern to all involved with the management and accessibility of intellectual property whereas Denning (1995) reported a few cases of plagiarism of digital material. Another study (Weenink, Waaijers & Godtsenhoven, 2007; Joint, 2006a) outlined some useful insights into how to manage copyright and other intellectual rights aspects of digital library collections relevant for digital repositories development. Lynch (1994) suggested solutions like dedicated server, document digest algorithms, and cryptographic signatures to overcome some of these problems. The working librarians find it a difficult task to protect author's right in proper way. Librarian must be well acquainted with the IPR issues, particularly rights management legalities for developing IDR. Several authors (Chepesiuk, 1997; Collins & Berge, 1994; Crawford, 1998; Jasperse, 1994; Perryman, 1991; Sasse & Winkler, 1993; von Ungern-Stenberg & Lindquist, 1995) have discussed this issue and addressed various problems faced by the librarians in the digital library environment. Hoorn (2005) put emphasis on and considers ways librarians can support scholars in managing the demands of copyright so as to respond to the needs of scholarly communication. The other authors (Besek et al., 2008) reviewed current copy right and related laws and their impact on preservation. The study also recommends for drafting national policies and adapting laws to allow digital preservation. The Audit Checklist (RLG &
NARA, 2005) points out that it is the responsibility of the IR to have a mechanism in place to track and verify the rights and restrictions applicable to a digital item.

2.1.6 Metadata

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource (National Information Standards Organization, 2004). Standardized metadata is important because it enables end users to easily and effectively search, find and retrieve information from the repository (Pinfield, Gardner & MacColl, 2002; Jones, Andrew & MacColl, 2006; Jones, 2007). A number of decisions need to be made by the digital repository regarding metadata of various types. ‘Administrative, descriptive, technical, structural and preservation metadata, using appropriate standards, are used to ensure adequate description and control over the long term’ (DCC, 2008). Several authors (Burk et al., 2007; Dunsire, 2008; Simeoni, 2004) discussed issues, principles and methods of metadata creation and harvesting in digital repository. Burnett, Ng & Park (1999) described and compared the contributions towards the evolution of metadata by both the fields’ library science and computer science. Weibel (1995) provided one of the first major introductions to the basic concepts of metadata usage in the digital library (DL) environment. Other experts (Chilvers & Feather, 1998; Calanag, Tabata & Sugimoto, 2004) discuss role of metadata in preserving digital data. A few other experts (Robertson, 2005; Ochoa & Duval, 2009; Park, 2009) focused on quality of metadata. Currier (2008) gives an overview of current initiatives in standards for educational metadata. Another expert (Zschocke & Beniest, 2011) introduced a process for assuring the creation of quality educational metadata based on the ISO/IEC 19796-1 standard to describe the agricultural learning resources resources in the repository of the Consultative Group on International Agricultural Research (CGIAR).

Gibbons (2004) reported that existing IDR systems differ widely in the handling of metadata. Some repositories implement additional or extended metadata schemas for domain specific datasets and others have used or adapted existing schemas, such as Dublin Core or MARC. In some systems, the metadata schema is hard-coded, and others support some degree of modification. Some IRs can provide for a different metadata schema with each collection, and others require the same metadata schema across the entire repository. He concluded that in the end it is up the institution to select what the best schemas for their needs and to decide which to support in their repository. A UKOLN article outlines a checklist of guidelines to assist organizations in choosing a metadata schema for a repository (http://www.ukoln.ac.uk/qa-focus/documents/briefings/briefing-63/html/). Another expert (Park & Richard, 2011) assessed the metadata element sets of electronic theses and dissertations that are currently used at Canadian academic institutional repositories. Day (1999) returns to
the topic of an earlier metadata column and reviews recent activities related to the development of metadata schemes for digital preservation. Other researchers (Groenewald & Breytenbach, 2011) investigated the awareness about digital preservation and discussed the use of metadata principles and the implementation of tools for the preservation of documents stored on personal computers. Dappert & Enders (2008) showed how it could be used a combination of METS, PREMIS and MODS to represent eJournal Archival Information Packages in a write-once archival system and reported the use of METS structural, PREMIS preservation and MODS descriptive metadata for the British Library's eJournal system.

2.1.7 Models

Libraries have been struggling with how to fund the IDR in a sustainable way. Branin (2003) in his draft paper on Institutional Repositories in ‘Encyclopedia of Library and Information Science’ described conceptual model and standards for a digital repository and basic components of an IR. Recent research (Rieh et al., 2007a) indicates that there are a wide variety of ways, many informal, that this is being accomplished. Prosser (2003a, 2003b) describes IR as a new model that is taking advantage of the new technology and discusses a possible future for scholarly communications. Swan et al. (2005a, 2005b) describe a delivery, management and access model for e-prints and OA journal contents for UK higher education. Rowland et al. (2004) reviewed possible models for implementing OA to research reports in IRs and OA journals. They concluded that ‘harvesting model’ was preferable to either a centralized national service or a completely decentralized service for the UK.

Developing sustainable business model is a critical concern for any publishers and McLennan (2009) focuses on the overview of business models for OA journals as stated in SPARC report entitled ‘Income Models for Open Access: An Overview of Current Practice’. Cockerill (2006) discusses various issues related to economically sustainability of OA publishing. The other researchers (Bird, 2008; Albanese, 2005) discuss the paid OA initiative and started experimenting with an 'author-side payment' OA model. IR as a management model (Kim, 2007) and as a consortial model (Moyle, Stockley & Tonkin, 2007) could be a guiding tool for others to vitalize their existing IRs has been highlighted in the literature. Peters (2002) is in support and provides a good overview of the various types of digital repositories-individual, discipline-based, institutional, consortial and national and discusses the value of repositories maintained by consortia as an alternative model. Ware (2004) in the PALS report and Jones, Andrew & MacColl (2006) note the early successes achieved by repositories in physics (arXiv), computer science (Networked Computer Science Technical Reference Library), economics (EconPapers), cognitive science (CogPrints) and more notably in the PubMed central model. Another study (Crow, 2002a) identifies four components of scholarly publishing: registration, certification,
awareness and archiving. Saracevic (2000) proposed a conceptual evaluation model for digital libraries, and emphasizes both the researcher’s and the librarian’s perspective. He further identified few criteria and grouped them into four categories: usability involving studies of contents, process and format; system features including technology performance and process/algorithm performance; usage consisting of material use and usage patterns; and ethnographic and others. Another study by Itsumura (2000) proposed other evaluation frameworks with the four categories of contents, system, uses/users and management/policy while Larsen (2002) describes the criteria of three categories excluding management and policy.

2.1.8 Multilinguality

Digital Libraries (DLs) with especially in online network environment are expected to serve multicultural and multilingual communities (Chen, 2007). Library community has been dealing with multilingual access to information since 1970’s (Brendler, 1970) and information professionals are concerned about providing multilingual information services to the user communities (Zielinska, 1976). The most important feature of any multilingual digital library system is that it allows browsing and searching across two or more different languages. It brings together collections from various countries, regions, cultures and provides access on a global scale (Yang, Wei & Li, 2008; Maeda et al., 1998) and preserves cultural heritage (Nichols et al., 2005). The question of multilingual access and multilingual information retrieval is becoming increasingly relevant in digital library environment. It is often found that user communities like to have information in their local languages or scripts. So it is important to identify their needs and expectations about multilingual information processing in digital library environment (Wu, He & Luo, 2012).

Language issues in digital libraries are multifarious (Borgman, 1997). Thus managing multilingual information and provision of multilingual information services has become a big challenge for librarian in digital library environment. Several researchers (Dartois et al., 1997) reported that most research and development activities concentrated on different issues of multilingual environments. Crossing the language barrier is one problem and is concerned with the translation of resources. Several experts (Chung et al., 2004; Wang et al., 2006) present a solution to the problem of missing dictionary terms in a query translation. The other challenges are data management and representation of information (Klavans & Schaible, 1998); interoperability (Fox & Marchionini, 1998); development (Hutchinson et al., 2005); management and storage of contents and metadata (Karvounarakis & Kapidakis, 2000). To solve this problem, Maeda et al. (1998) developed a technology to enable viewing of multilingual documents in a Web browser. Another group of experts (Kramer, Nikolai & Habeck, 1997; McCulloch, Shiri & Nicholson, 2005; Yang, Wei
Another study (Budzise-Weaver, Chen & Mitchell, 2012) suggested strategies for building and sustaining multilingual information access for digital libraries. Peters & Picchi (1997) described mainly two issues (e.g. multiple language recognition, manipulation and display; multilingual or cross-language search and retrieval) that must be addressed in order to implement a multilingual interface for a Digital Library system. Borgman et al. (1996) reported problems towards designing digital libraries for distributed environments necessary for interoperability with other systems and services. In another paper he drew attention to the different multi-lingual issues such as medium, culture, and language and concluded with design challenges and technical matters involved in designing digital libraries (Borgman, 1997). Borbinhal & Machado (2005) described ‘DEPTAL’, as a collection-centric multilingual framework not only in the user and administration interfaces, but also in the handling of the contents. Another researcher (Baker, 1997) expressed concern over conceptual and organizational problem of maintaining metadata standard in multiple languages. Anderson (2003) provides an overview of the progress made in bringing historic scripts to the Unicode standard and also warns of the danger of failing to pursue such work actively.

Supporting searching and display in multiple languages is an increasingly important issue for all digital libraries accessible on the Internet. Searching for information in multi-media digital libraries is more complex than text-only searching. As Croft (1995) noted in an earlier D-Lib issue, general solutions to multi-media indexing are very difficult, and those that do exist tend to be of limited utility. Large & Moukdad (2000) gave an overview of multilingual information access issues in relation to the Web. Wellisch (1978) advocated the support of machine-readable catalogue formats for multilingual documents. He pointed out the problem of language experts for cataloguing a record in different language scripts and concluded that universal bibliographic control of documents cannot be achieved excluding multilingual documents.

2.1.9 Quality

Review is an essential part of the existing scientific and scholarly publishing process. The development and usability of any IDR system depends much on quality of resources. But still quality of contents is a low-prioritized aspect of digital libraries and digital repositories. Several experts (ALPSP, 2002; Bentum, 2001; Gonzalez & Porcel, 2007) reported that it (quality) presents another obstacle to faculty participation in IRs, as authors give importance only on peer view quality articles. The contents are the most important factor that has been cited by researchers to show
the success of a repository (Macha & de Jager, 2011). So, guaranteeing the quality of the contents of the repository is important to win the willingness of the authors and to achieve the intended outer institutional visibility (Gonzalez & Porcel, 2007). It is found that authors were overwhelmingly in favour of traditional peer review for guaranteed quality. It is also an important issue for most researchers as most of the readers want high quality, peer reviewed articles in their respective field (Swan & Brown, 2003). There are different grades of perceived quality. Scholars do not treat all peer reviewed reports as equally trustworthy; rather they rely upon a variety of processes and markers, which are dependent upon everything from the structure of the discipline itself to the social networks that the readers are embedded in (Kling & McKim, 1999). Jones (2007) reported that validation and verification of the data are an important factor. There are two levels of this task: the metadata about a particular scholarly work and the organizational and contextual information about the work. Several studies reported that all research-based work is not of high quality especially where pre-prints are concerned. This study recommended that it should be reviewed from the Dean or Head of the Department (http://www.lib.unimelb.edu.au/eprints/collectionpolicy.htm). Another study (Kling, Spector & McKim, 2002) argued that reputation of the department, in case of non-peer view articles, ensures the quality of the documents. Others expert (Wang & Wang, 2012) put emphasis on selection of persons with corresponding background knowledge to review the achievements submitted by researchers, which will identify the effectiveness, appraise the importance, and ensure the quality of IR. This view is supported by other researchers (Lee, Goh & Chua, 2010) and suggested that it can be carried out by the knowledge management expert with in the institution using a standardized checklist tool. Day (2003) suggested that as focus of an IR is on contents (peer-reviewed or not), the choice being left to those who develop their collection policies. Harnad (2003) recommended that work should be submitted to a refereed journal, and then the post print should be self-archived in a digital repository. Another study (Pinfield & James, 2003) proposed a mechanism for clear labeling of pre- and post-prints into two sections or in separate areas of the site. He concluded that departments can employ a kite mark, stamp system or metadata entry to show the level of quality of individual works in the IR. This will help users differentiate certified from non-certified content. SPARC (2002a) recommended for differentiating between preprints and published peer-reviewed research. Other experts (Probets & Jenkins, 2006) support this view and suggested to segregate pre- and post-prints into separate repositories. This study also put emphasis on quality of documentation as one of the evaluative criteria for any repository and this difficult issue should be made by individual institution and not mandated by international standards (Genoni, 2004). A few other experts (Robertson, 2005; Ochoa & Duval, 2009; Park, 2009) focused on quality of metadata because low-quality metadata compromise the effectiveness of services that repositories provide to their users. Other experts (Zschocke & Beniest, 2011) put emphasis on quality of educational metadata based on the ISO/IEC 19796-1 standard to describe the agricultural learning resources.
2.1.10 Software

The selection of software for any IDR system is crucial and this technical issue has been discussed by many researchers (Shoeb, 2010; Payne & Singh, 2010; Fay, 2010; Joint, 2006b; Crow, 2004; Falk, 2003). Several studies (Jones, Andrew & MacColl, 2006; Jones, 2007) discussed IR software systems and system feature and provided valuable advice on the topic of software selection. Pruett & Choi (2013) reported that libraries are now considering migration to an open source integrated library system. Barton & Waters (2004-2005) described in detail various technical aspects such as level of customization, file formats, metadata standards, system administration etc. for evaluating software for an IR system. It depends both on technical and non-technical issues (Gibbons, 2004c; Powell, 2005) and on internal and external issues like usability, interoperability, support costs, and the ability to migrate materials over time, metadata standards (DeRidder, 2007). Lynch (2006) stated that factors such as repository contains, how it will be used, the features that are wanted, and the local technical environment need to be considered because even the ‘best’ software may not have every feature an institution wants. Cervone (2006) expressed concern over the issues regarding security and authentication, long-term cost and maintenance, vendor viability as well as training and documentation of the software. Another study (Singarella, 2005; Rankin, 2005) put emphasis on organizational goals and objectives and on organizational IT strategy (Powell, 2005) whereas Bevan (2007) concluded type of resources as other criteria for software selection. Beier & Velden (2004) pointed out that it should be compatible with existing products. Other researchers (Goh et al., 2006) developed an alternative set of criteria such as contents management, content acquisition, metadata, search, access control and security, report and inquiry, preservation, interoperability, user interface, standards compliance, automatic tools and support for evaluating open source software for repositories. A survey by Primary Research Group Inc. (2007) reported that selection of software varied markedly between different regions in the world such as the United States and Europe, and between developed and developing countries. Fortunately, a few brave institutions are sharing their methods for evaluating software solutions. One such institution is the University of Arizona (Han, 2004) whose evaluation strategy is more useful than the particular findings. JISC also has a very good site describing the evaluations made by its FAIR (Focus on Access to Institutional Resources) participants (http://www.jisc.ac.uk/index.cfm?name= fairsynthesis_repol). Another report (OARINZ Project, 2006) discussed and evaluated six repository systems and recommended that DSpace could be accommodated within the national network of New Zealand because DSpace scored well in the overall evaluation.

Technological issues have been discussed in several publications. There were two articles which compared IR software’s (Budapest OAI Guide, 2004; Canadian Association of Research Libraries, n.d.). Open Society Institute (2004) provided a detailed and very valuable analysis of the existing repository software, including a
comparison between the two top performers, DSpace & EPrints. Its “Feature & Functionality Table” is an excellent guide to evaluating any software for managing digital repositories. Nixon (2003) shared practical experience using both EPrints and DSpace and compared between the two systems. Witten et al. (2005) described particularly two popular systems (Greenstone & DSpace) and compared their similarities and differences and also discussed possible opportunities for combining the advantages of the two. Another study (Prudlo, 2005) discussed three most widely known repository software packages and compared in terms of cost, underlying technology etc. Another experts (Marill & Luczak, 2012) evaluated repository software at the National Library of Medicine. Other experts (Smith et al., 2003; Tansley et al., 2003) restrict themselves to the design and functionality of the open-source DSpace IR whereas Rieh et al. (2007b) advocated that DSpace is the most prevalent system both in terms of pilot-testing and implementation. The other two survey report concluded that most CARL (Shearer, 2004) institutions and ARL-member libraries (Bailey et al., 2006a) are using DSpace. Another study (Deng & Reese, 2009) presented methods for customized mapping and metadata transfer from DSpace to Online Computer Library Center (OCLC). Another expert (Liu & Zhou, 2011) discussed several technical issues relating to implementing and using DigiTool, proprietary software by Ex Libris.

2.1.11 System management and Administrative parameters

Generally the questions may arise who will manage and the IDR? Or where will it be located? It is found that (OpenDOAR, 2012; ROAR, 2012) the very large majority of IDR sites studied in this research were run by the library (or information services) at the host institutions. There are a number of reasons cited as to why the library is the appropriate locus for the leadership of such projects (MIT, 2003; Branin, 2002; Joint, 2006c; Chang, 2003). Libraries are essential component of nation’s information infrastructure (Borgman, 2003a) and their staff plays a significant role in the development of institutional repositories (Bailey, 2005). "Linking people to resources" (O'Brien, 2005) has been the task of information specialists for many years. The ARL SPEC Kit survey results reveal that the library plays a critical role in initiating, planning, and implementing IRs (Bailey et al., 2006a). They further quoted by saying that “All respondents, implementers and planners alike, indicate that the library has been a driving force in the creation of or planning for an IR”. Pelizzari (2005) advocates the library as “the standard bearer for the IR.” Rockman (2005) quoted by saying that reference librarians are “natural partners to be involved with institutional repositories”. Quint (2002) urges librarians to assert themselves as the leaders in IR implementation. Another useful report (Rieh et al., 2007b) show that librarians take the lead in IR pilot testing and implementation, and funding for the IR almost always comes from the library. Another study (Phillips, Carr & Teal, 2005; John, 2005) identified the marketing role of librarian in promoting IR system.
A study by Pelizzari (2003) indicates over 70% of the respondents singled out the library as the structure to be given the mandate of managing an institutional archive. The other studies (De Beer, 2005; Kaur & Ping, 2009), the respondents in such studies were of the opinion that libraries should own and manage the IR. In most cases responsibility will likely fall to an information technology (IT) unit, to a library, or to a combination of these units. Gibbons (2004) advocates a partnership of librarians, instructional technologies training staff, and computer services staff for the IR effort. Ware (2004) asserts that IRs start with a partnership between the library, the institution’s instructional technology unit, and a vice president’s office (e.g., bursar or provost for academic studies). Allard, Mack & Feltner-Reichert (2005) identify librarians in these roles: learning how the IR works; leading the IR implementation effort; developing policy (especially with regard to defining collections); leading anything that pertains to metadata, reviewing submissions to the IR; and training contributors. Additional roles include evaluating IR systems, being an advocate for the IR, recruiting contents, and serving as advisory contributors on intellectual property issues (Chan, Kwok & Yip, 2005).

2.1.12 User Interface

Carefully consider several technical aspects of an IDR’s user interface when selecting an IDR system. Gibbons (2004) identified several features of user interface and one of them is ease of use, no matter how sophisticated and elegant the backend of an IR system may be. The report also identified another feature of the interface that administrator should consider is the level of branding and customization available. It is practically useless without a usable interface on the front end. He concluded that the interface of an IR should reflect the culture and image of the institution. Hunter & Day (2005) discussed several practical issues of an IR’s user interface and concluded that these issues are to be properly examined and reviewed before selecting an IR system. A few authors (Yee et al., 2003; Callahan & Koenemann, 2000) compared and evaluated different type of interfaces for browsing and searching digital image collections in electronic environment. Hornbaek & Frokjaer (2001) made a comparative study of three interfaces (linear interface, fisheye interface and overview+ detail interface) for reading electronic documents and Bekaert & Van de Sompel (2005) described two other interfaces based on two formal specifications (i.e., OAI-PMH and the NISO OpenURL Framework for Context-Sensitive Services (OpenURL Standard). Designing multilingual user interface to meet the requirements of the local users is crucial and has been suggested by many experts (Del Gado & Nielse; 1996; Head, 1999; Rskin, 2000). Another group of authors (Phillips et al., 2005; Hunter & Day, 2005; Rieh et al., 2007b; Feijen et al., 2007) advocated for adopting own interfaces and customizing it rather than using the default user interface that comes with any software. Chen & Dumais (2000) reported the development of a user interface that organized web search results into hierarchical categories. Peters &
Picchi (1997) pointed out mainly two issues (multiple language recognition, manipulation and display; multilingual or cross-language search and retrieval) that need to be addressed in order to implement a multilingual interface for a Digital Library system.

2.1.13 Versions identification and management

The problem is multifaceted and growing and the difficulties associated with version identification and management existed long before the digital age. It is found that multiple versions and versioning issues have existed for centuries. Richardson (2005b) describes how IR aims to include post-prints. Quint (2006) supports this view and concludes that less evaluated contents may reach more users than expertly reviewed contents. Another study (Pinfield & James, 2003) recognises that e-prints regularly contain more information than published versions. Johnson (2001) reported that the official version of articles remains the peer-reviewed version. Academics often want the final version to be ingested whereas repository managers prefer to ingest everything to reflect the research process.

The software development sector is extremely important when considering version control and version management techniques. Another study (Allsop, Somerville & Shipsey, 2007) identified a number of techniques are promoted by the institution to assist in the version control of documents. These include using version numbers, naming conventions, read-only tags and version control tables. The process of software development necessitates the creation and management of many versions of source code, configuration files and documentation. The Version Identification Framework (VIF) provides practical advice and recommendations to authors and contents creators, repository managers and those involved with repository software on how to identify versions better (http://www2.lse.ac.uk/library/vif/). Puplett (2008) outlines the issues associated with versions in IRs, and discusses the solutions being developed by the Version Identification Framework (VIF) Project. He further discusses the number of potential versions created as part of a contemporary research project, and the potential number of relationships between iterations, variations, and manifestations. He suggests that versioning problems arise from reader confusion over the publication status of the contents they find in a repository. The problem of version identification can only be solved if the repository platforms that people use are able to identify version status and version relations of the objects that they store to the outside world (http://www2.lse.ac.uk/library/vif/framework/Software Development/index.html). Another study (Brace, 2008) explains why giving time to versioning within a repository is worthwhile and outlines the best practice to implement. The most significant issue for repository managers then, is how best to organize these multiple versions, and how best to describe them so they can be properly found. Decisions also need to be made over which version should be
Chapter 2: Literature Review

ingested. Nagel (2006) suggests using subversion to help with personal information management of documents stored on different computers and storage media. Pitkow & Jones (1995) present a prototype web publishing environment, emphasising that version control should be supported, permitting multiple revision of documents, and graceful recovery from system and user errors.

2.1.14 Workflow management

Workflow design is core aspects of the operational responsibility of any IR system. Madsen & Oleen (2013) reported that as an IR matures it will face the challenge of how to design workflow. Morrow & Mower (2009) address the need for a workflow manager as a way to coordinate multiple persons in the deposit of numerous articles. Another survey report by IRs managers (Hanlon & Ramirez, 2011) indicates that a majority of IRs follow a mediated deposit process. The workflow problems relating to non-textual documents particularly those related to data curation have also been the focus of recent literature. Recent studies (Yoon & Tibbo, 2011; Witt, 2008) address the workflow challenges of “non-traditional” IR deposits (i.e. non-article/ETD contents). The other researchers (Boock & Kunda, 2009) compare the workflows for depositing electronic theses and dissertations in the Oregon State University.

Jones, Andrew & MacColl (2006) shared their own experience building the Edinburgh repository and discussed workflow diagrams including post-submission workflows, such as assigning items to categories, contents verification procedures, cataloguing steps and the verification and augmentation of metadata, short and long-term storage procedures etc. Aschembrenner & Kaiser (2005) described eleven workflows as identified by the DSEP process model and implemented by each of the reUSE demonstrators separately. Several authors (Green, Dolphin & Sherratt, 2007; Beier & Velden, 2004) explained the need for customization of workflow pattern for making a document publicly available. Jones (n.d.) described different methods used to customize workflow steps and concluded that customization process depends on different types of documents. Another study (Hulse, Cheverie & Dygert, 2007) advocated that workflow design should be based on a collection-by-collection basis and on lifecycles of the full range of institutional information resources (Campbell, Blinco & Mason, 2004). Nolan & Costanza (2006) shared practical experience in developing workflow and concluded that workflow (administrators, editors, reviewers, authors) for processing materials is based on the particular IR software.
2.2 Indian Scenario

This section examined open access movement (OAM) and open access repository (OAR) development from the perspective of India context. Several key issues from general to specific have been consulted on this aspect. Over the past seven to eight years, the implementation of OARs has been growing rapidly and the publications on OARs have flourished accordingly.

The open access initiatives and the development of OAM in India have been described by many researchers (Arunachalam, 2008; Gaur, Munshi & Murthy, 2004; Kataria, 2007b; Sreekumar, 2006; Sreekumar et al., 2007; Arunachalam, 2004a, 2004b, 2006; Hirwade & Rajyalakshmi, 2006; Doctor, 2005; Kirsop, 2007; Sahu & Parmar, 2006; Ramachandran, 2003; Tripathi & Tripathi, 2007; Fernandez, 2006; Swan, 2008; Satyanarayana & Babu, 2007; Urs, 2007; Ghosh, 2011, Sawant, 2009; Lal, 2008; Balaram, 2008; Kirsop & Chan, 2005; Sridhar, 2007; Roy, 2010). A few studies highlighted recent trends in the OAM and discussed the significance of those trends for information access in developing countries (Das, Sen & Dutta, 2005; Gobbur, 2007; Chan & Costa, 2005; Ghosh & Das, 2007; Sharma, 2007). Several other studies (Nazima & Devib, 2008; Mittal & Mahesh, 2008; Venkadesan, 2009; Gul, Shah & Baghwan, 2010) described growth and development of IR in our country. Another study (Roy, Biswas & Mukhopadhyay, 2012a) focuses on the OA initiatives and describes some of the current OA channels and the areas where India has made significant progress. A book published by UNESCO, authored by Anup Kumar Das (Das, 2008) gave a brief overview of IRs in India. Another study (Bhat, 2010) focused mainly in specific discipline like computer Science and IT (Information Technology) and specific area like universities and research institutes (Krishnamurthy & Kemparaju, 2011). Another expert (Sawant, 2011a) focused on planning, pilot testing, system implementation, exploratory activities conducted before implementation. In another paper she investigates the experience, contribution and opinions of users of respective (Sawant, 2012a). Another study finally makes recommendations for their successful implementation in academic institutions (Jain, 2011). A broad outline on the growth of OA archiving in developing countries focusing mainly on India and its benefits had been provided by Chan and his colleagues on Scidev.net (Chan, Kirsop & Arunachalam, 2005). One more article by Chan & Kirsop (2001) and a few articles by Professor Arunachalam were also important in this regard (Arunachalam, 2005, 2006). Specific Indian initiatives especially open archive initiatives were mentioned in this context. One of the important research papers was by Fernandez (2006) who evaluated the growth and development of online research repositories in India within the broader framework of OA. She conducted interviews with information professionals responsible for creation and maintenance of online research repositories in India. Swan (2008) focused on how OA can help to resolve the problems of maximizing the visibility, and thus the uptake and use, of Indian research outputs and how self archiving can provide a boost.
to OA movement. Another study (Sahu, Goswami & Choudhury, 2013) looks at the use of the repository based on the repository log data. Another study (Arunachalam, 2008; Hirwade & Rajyalakshmi, 2006) discussed two ways of achieving OA vehicles viz. open access journals (OAJ) and open access archives (OAA) and focused on various fronts channels or where open access is making good progress and deliberated different issues of OA. In this context two government organizations (National Knowledge Commission, 2007; University Grants Commission, 2005) have been playing vital role in promoting OA contents of public funded research materials. Bangalore Declaration (2006) drafted a model OA policy to support full access of publicly-funded research outputs.

The needs of IR in Indian context have been identified by many professionals (Rajashekar, 2003; Arunachalam, 2004a, 2006, 2008; Prasad & Guha, 2005; Sreekumar et al., 2007; Patel, Vijayakumar & Murthy, 2005; Gaur, Munshi & Murthy, 2004; Sreekumar, 2006; Doctor, 2005; Kataria, 2007b; Swan, 2008; Fernandez, 2006; Deoghuria & Roy, 2007; Mittal & Mahesh, 2008; Suleta & Sangeeta, 2006; Varatharajan & Chandrashekara, 2007; Chakravarty & Mahajan, 2006; Meitei & Devi, 2009). Several other experts (Vinayagamoorthy, Ramesh Babu & Gopalakrishnan, 2006) highlights the present digital library initiatives in engineering educational institutions in Tamil Nadu. Another study (Mukherjee & Nazim, 2011) analyzes the present trend of institutional archives worldwide. Roy & Mukhopadhyay (2010) put emphasis solely on the development of learning object repositories (LORs) in India. Recent research (Babu et al., 2012) finds out the rationale for IR categories and various developmental challenges are underlined. Hashim & Jan (2011) examines five web based OARs for the purpose of identifying their strength and limitations, using pre-defined standard parameters. Another study (Roy, Biswas & Mukhopadhyay, 2011) provide an overview of Indian IDRs and compare in respect of contents, types, objects uploaded, software used etc. In other paper they give an overview of current state of OARs in Asian countries with special reference to SAARC countries (Roy, Biswas & Mukhopadhyay, 2012b). Another study (Roy, Biswas & Mukhopadhyay, 2013) compares OARs against different criteria, shows ranking of Indian IDRs in World’s repositories and also describes several policy issues. A few researchers (Ghosh, 2009; Harnad & Swan, 2008) discussed the Indian contribution to OA and put emphasis on formulating and adopting a national OA self-archiving mandate for all of its research institutions and funders. The other researches (Swan, 2008; Singh & Pandita, 2005) advocated for mandatory policies that encourage authors to make their work OA and other experts (Sahu & Parmar, 2006) advocated the development of IR with government bodies. Many professional body and government organizations like INFLIBNET, UGC advocated for OA and supported mandating OA to all publicly funded research (Chand et al., 2004; Narang et al., 2005). Several other studies (Madhan, Rao & Awasthi, 2006; Vagiswari & Birdie, 2007; Swan, 2008; Srinivasan, Patil & Rajan, 2007) reported that IRs is essential for Indian research and advocated to populate IR in different ways. The other researchers (Khanna et al., 2013) describes a health
Designing Institutional Digital Repository for The University of Burdwan: A FLOSS Based Prototype

A number of problems come while designing OARs for an institution and have been discussed in the literature (Arunchalam, 2004a, 2006; Chan, Kirsop & Arunchalam, 2005; Tripathi & Tripathi, 2007; Kataria, 2007a; Deoghuria & Roy, 2007; Chand et al., 2004; Patel, Vijayakumar & Murthy, 2005; Senthil & Moorthy, 2006; Srinivasan, Patil & Rajan, 2007; Anuradha, 2005; Sreekumar et al., 2007; Nazim, 2009). Kumar (2012) reported that there are many technical, social, and expertise issues required for establishing IR such as selection of hardware, communication bandwidth, suitable software, international standards and documentation, including operational aspects like loading software, uploading data, set-up test server, manage process, maintenance and load material advocacy and preservation. Another study (Venkatesh, Nageswara Rao & Kalpavall, 2012) broadly covers digitization issues related to OSS including their merits and demerits. Doctor (2008a) determines the number of simultaneous users of an institutional knowledge repository in India context. Some other experts (Deoghuria & Roy, 2007; Hirwade & Rajyalakshmi, 2006; Satyanarayana & Babu, 2007; Srinivasan, Patil & Rajan, 2007; Singh & Pandita, 2005) pointed out that the development of IR depends much on active participation of key stakeholders of the institution and concluded that authors, librarians, researchers are not aware of the OA publishing. Sahu & Arya (2013) analyze the awareness of OA publishing among researchers and faculty members of the concerned institutions. Kataria (2007a) reported that organizational commitment is needed for the development of IR. A few researches (Deoghuria & Roy, 2007; Kataria, 2007b; Singh & Pandita, 2005) argued that stakeholders have also a role to populate IR. Fernandez (2006) described both top-down and bottom-up methods to promote repositories.

Several technical problems have been discussed in literature. Rajendran, Babu & Gopalakrishnan (2005) mentioned technical problems like font issues, lack of standard format, and reproduction of graphics. Rajashekar (2003) described several other problems like contents related standard and specifications, metadata standard, workflow pattern, OAI compliance software and preservation of digital objects. Several other authors (Satyanarayana & Babu, 2007; Singh & Pandita, 2005) put emphasis on quality of research materials and on lack of expertise (Kataria, 2007b; Hirwade & Rajyalakshmi, 2006; Satyanarayana & Babu, 2007) is another problem for developing IR. The lack of infrastructural facilities like hardware, obsolete technology, search engine and connectivity of high bandwidth have been identified as
another problems by many other researchers (Singh & Pandita, 2005; Sreekumar, 2006; Urs, 2007; Hirwade & Rajyalakshmi, 2006; Narang et al., 2005; Rajendran, Babu & Gopalakrishnan, 2005). The other experts (Das, Sen & Dutta, 2005) identified that web server does not work all time and the uniform resource locator (URL) has been changed in due course of time. This lack of accessibility has somewhat become a barrier to access, use and cite local research. IR has made the often limited dissemination and access of information extremely possible now in developing countries (Ghosh & Das, 2007).

Formulating policies (Table 4.2 of chapter 4) regarding different issues of IR is essential for smooth functioning of a repository. Several studies (Kataria, 2007b; Sreekumar et al., 2007; Madalli, 2005; Madhan, Rao & Awasthi, 2006; Srinivasan, Patil & Rajan, 2007; Satyanarayana & Babu, 2007; Urs, 2007) proved that formulating policies (contents development, contents recruitment, right management, preservation interoperability standard, multilinguality, workflow pattern etc.) are essential for developing IR. Swan (2008) advocated for mandatory policies that encourages authors to make their work OA. Sawant (2012b) investigates various issues concerning the management of IRs. Another paper (Das, Dutta & Sen, 2007) explores the policy frameworks, strategic dimensions and analyses SWOT (strengths, weaknesses, opportunities and threats) of existing ETD initiatives in India. Fernandez (2006) points out different issues like fund, submission policies. Ghosh (2007a) discussed the subject coverage, number of items, access policy, browse or search option, value added services etc.

Digital environment makes the copyright protection a difficult task. A number of issues and concerns are associated with the usage of digital information. A group of authors considered copy right or right management as most challenging issues of digital library environment and to be aware of this issue while depositing objects to the repository (Urs, 2007; Vagiswari & Birdie, 2007; Madhan, Rao & Awasthi, 2006; Madalli, 2005; Kataria, 2007b; Singh & Pandita, 2005; Doctor, 2008b). A few other studies (Satyanarayana & Babu, 2007; Rajendran, Babu & Gopalakrishnan, 2005; Narang et al., 2005; Srinivasan, Patil & Rajan, 2007; Rajashekar, 2003) reported that copy right and plagiarism issue are also responsible for the development of IR. Hombal & Prasad (2012) suggested that this issue may be controlled by imposing restriction on access using license and encryption methods. Another expert (Shashi Nath et al., 2008) discussed copyright issues in the context of populating IR. Moorthy & Karisiddappa (2005) discussed different issues and concerns of librarians in the face of the intellectual property laws like perpetual dependency; societal rights, pricing, access and ownership and highlights adverse impact of IPRs acts on knowledge society and open source initiatives. Yaranal & Ramesha (2012) explains the role of libraries and library professionals in managing the resources within the limitation of Intellectual Property Rights (IPR).
The organization and management of contents is crucial. Recruitment of contents is another major problem for the IDR development. This issue has been raised by several researchers (Madhan, Rao & Awasthi, 2006; Sreekumar et al., 2007). Other experts (Urs, 2007; Das, Sen & Dutta, 2005) focused on types of contents and on its quality (Satyanarayana & Babu, 2007; Singh & Pandita, 2005). University Grants Commission (2005) reported that most of the IRs achieves this (quality) through software having a submission buffer. Narang et al. (2005) advocated PDF format to build up e-theses collection. Rajashekar (2003) highlighted contents related standard and specifications.

Digital Preservation is not a new concern - it has been with us since the introduction of computers into our lives in the sixties. A few authors (Urs, 2007; Rajashekar, 2003; Katre, 2011; Chakravarty, 2010; Chowdhury, 2010) expressed concern over the preservation of contents in digital format. Another paper (Madalli, Barve & Amin, 2012) presents an analytical study along with observations regarding digital preservation support available in existing open-source digital library software (OSS-DL) based on test beds created for that purpose. Another expert (Barv, 2007) discussed several issues of digital preservation and recommended preferred formats for long term preservation of digital objects. Another report (University Grants Commission, 2005) concluded that the default formats accepted by most IR software are PDF, Postscript, ASCII, and HTML. Rajendra, Babu & Gopalakrishnan (2005) focused mainly on preserving print version of the theses and dissertations into electronic form. Patel, Vijayakumar & Murthy (2005) proposed IR as an economical way to save contents in digital form and format for long term purpose. Gaur & Tripathi (2012) highlights various problems and projects initiated. Another paper (Singh, 2012) discusses the initiatives taken by Indian government for digital preservation of cultural heritage resources and manuscripts. Another expert (Katre, 2012) made a comparison between the American and Indian digital preservation programmes based on the essential building blocks. Chandra & Gokhale (2012) describes OAIS reference model and identifies the tools and develops the processes to implement the preservation services and actions.

A few other authors (Narang et al., 2005; Satyanarayana & Babu, 2007) described IR as an information system model through which information flows everywhere with just the click of button. Another researcher (Rajashekar, 2003, 2004) proposed IRs as OA publishing model that organize OA resources and improve the quality of Indian research whereas Patel, Vijayakumar & Murthy (2005) proposed IR as an economical way to save digital contents. Another study (Vijayakumar, Murthy & Khan, 2006) proposed a prototype model for Indian universities to preserve electronic theses and dissertations.

Designing IR system with OSS has been discussed in several papers. A number of authors (Alexander & Gautam, 2004; Mittal & Mahesh, 2008; Das, Sen & Dutta, 2005; Jayakumar et al., 2007; Soundararajan et al., 2007; Jayakanth et al., 2008; Jain
Chapter 2: Literature Review

& Shrivastava, 2008; Mishra et al., 2007; Anuradha, 2005; Narang et al., 2005; Madhan, Rao & Awasthi, 2006; Singh, Pandita, & Dash, 2007; Fernandez, 2006; Doctor, 2005; Lihitkar & Lihitkar, 2011; Sawant, 2011b; Chindalia, 2008; University Grants Commission, 2005; Barve & Dahibhate, 2012; Deka, 2006) highlighted the state of IR especially the use of OSS. Another study (Jayakanth, Minj & Dastidar, 2012) reported that many academic and R&D establishments have made it mandatory to set up IRs using OSS. The other researchers (Salve, Lihitkar & Lihitkar, 2012; Kamble, Raj & Sangeeta, 2012; Giri & Sengar, 2011) describe features of some of the popular software packages for developing their digital libraries. Kumar (2008) highlights problems of selection, installation and maintenance of OSS. A few studies considered that the selection of software is also vital issue and depends on different technical and non technical issues (Madalli, 2005) and Krishnamurthy (2007) put emphasis on several other issues like lower costs, greater accessibility, long-term preservation etc. Chand et al. (2004) identified the need of standard and protocol of OSS whereas Singh & Pandita (2005) advocated for OAI compliance IR software. Another group of authors (Madalli, 2005; Laxminarsaiah & Rajgoli, 2007b; Jose, 2007; Thakuria, 2008) advocated that DSpace is the most popular IR software in the open source domain and is widely used digital library software in India. Sawant (2012b) is in support by saying that 79 per cent of the institutions in India had used the DSpace. Alam & Pandey (2010) provides a mechanism for the development of a digital library using OSS like Greenstone. Sonkar et al. (2005) discusses in detail all the issues related to the development of digital library of newspaper clippings and implementation of Greenstone software in developing such collection. Biswas & Paul (2010) compared two popular IR software (DSpace & Greenstone). Another study (Sastry & Reddy, 2010) deals with the technical comparison between DSpace and Greenstone. Lihitkar (2011) highlights the comparison of features, function and usability of GSDL, Dspace and Ganesh. Another study (Patil, Kanamadi & Gopale, 2008) tries to make out the comparison, features, function and usability of OSS like DSpace, Greenstone and EPrints. The other experts (Singh, Witt & Salo, 2010) outlines the design of a comparative analysis of the four IR software packages like EPrints, DSpace, Fedora and Zentity in a comparative manner and concluded by saying that the output of this study will be highly useful for repository developers, repository managers, and especially those who are selecting a repository for the first time. Another study (Karmakar, Das & Thakuria, 2010) analyzed and compared two popular software Dspace and Eprints. Another expert (Lihitkar & Lihitkar, 2012) prepared a ranking of the software based on the assigned points for each criteria. The other experts (Mulla, Chandrashekara & Talawar, 2010) put emphasis on usage and performance of various library software modules in engineering colleges of Karnataka.

Several other experts (Doctor, 2007; Jayakanth et al., 2008; Jobish et al., 2005; Krishnamurthy, 2005; Laxminarsaiah & Rajgoli, 2007a; Madalli, 2003; Sutradhar, 2006; Chand et al., 2004; Patil, Vijayakumar & Murthy, 2005; Senthil & Moorthy, 2006; Srinivasan, Patil & Rajan, 2007; Anuradha, 2005; Sreekumar et al., 2007;
Narang et al., 2005; Singh & Pandita, 2005; Doctor & Ramachandran, 2008; Vijayakumar, Murthy & Khan, 2006; Shewale, 2012; Jayakanth, Minj & Dastidar, 2012) have shared their practical experience in developing and maintaining IR at their own institution using OSS. Another study (Cherukodan, Santhosh Kumar & Humayoon Kabir, 2013) describe the design and development of a digital library at Cochin University of Science and Technology (CUSAT) using DSpace.

The other studies (Senthil & Moorthy, 2006; Kumar, 2005; Prasad, 2006) explain the need for customization of IR software and mention areas where customization is needed. Laxminarsaiah & Rajgoli (2007a) explained how to customize the metadata fields according to institutional requirements. Another study (Srinivasan, Patil & Rajan, 2007) focused on designing homepage customization of submission page.

Another technical issue is interoperability which ensures that outputs are discoverable. The need for developing interoperable OAR have been suggested by many researchers (Sreekumar et al., 2007; Narang et al., 2005; Rajashekar, 2003; Singh & Pandita, 2005; Chand et al., 2004; Awasthi & Jaiswal, 2008) in Indian context and creation and use of Open Archives (OA) is one approach to achieve some degree of interoperability (Alexander & Gautam, 2004). Prasad & Guha (2005) discuss the need of interoperability in digital library arena and explore the potential of the OAI-PMH protocol and illustrate its structural and functional models. The study also reported that most of the IDRs have adopted the existing metadata protocol OAI-PMH. Another researcher (Amin, 2003) proposed for OAI compliant repositories and described OAI-PMH framework and its technical architecture. Other experts (Hirwade & Hirwade, 2006; Hirwade & Bherwani, 2011) describe metadata, OAI-PMH and major metadata harvesting services in India. Another study (Singh, Pandita & Dash, 2008) explained how metadata is harvested (extracted) from Data Providers (Repositories) by Service Providers (Search Engines) using this model. Another research paper (Sarkar & Mukhopadhyay, 2010) presents a method of metadata harvesting from different OAI-PMH compliant repositories containing electronic theses and dissertations. This study has developed and described in details harvesting framework using PKP harvester. Another paper (Deshmukh, Bhavsar & Bhavsar, 2012) covers OSS available for federated search and gives some federated search applications of public domain.

It is essential to have a metadata schema for different types of object for easy access of information (Sreekumar et al., 2007; Prasad, 2006; Laxminarsaiah & Rajgoli, 2007a; Madalli, 2005; Rajashekar, 2003). Another expert (Hirwade, 2011) investigated the metadata standards available worldwide and analyzed of these standards. Narang et al. (2005) mentioned three metadata schemas MARC, Dunlin Core and ETD-MS. University Grants Commission (2005) reported that most of the digital repositories follow the qualified Dublin Core standards and unqualified Dublin Core standards and also suggested using other metadata schema to meet the needs of users to provide better access and management of different type of digital objects. Pal
(2010) discusses some emerging issues on metadata as a mechanism of resource discovery and its impact on precision of search results in a distributed network environment. Prasad (2006) made an attempt to accommodate additional metadata formats like IMS-LOM. Another researcher (Patra, 2008) described metadata requirements for setting up a digital repository in ceramics resources.

With the advent of Unicode and the development of information technology, the question of multilingual access and multilingual information retrieval is becoming increasingly relevant in digital library environment. Managing digital objects in Indic script is discussed in the literature (Maitra, 2002; Prasad, 2003; Mukhopadhyay & Azim, 2006). India is considered as a multi-religious, multi-cultural, multilingual country, which has massive amount of precious literature in many languages. India is a multilingual country with twenty-two constitutionally recognized Indian Languages and with many variations of dialects. There are 428 languages listed of which, 415 are living languages and 13 are extinct (Gordon, 2005). It is essential for any IR system to have Unicode standard multilingual environment for searching theses and dissertations (Karande, 2007; Patel & Madalli, 2007; Narang et al., 2005). Biswas (2005) highlights the development of Unicode standard, discusses some of the issues that must be addressed in order to implement a multilingual interface for a digital library system and the problems associated with character encoding of multilingual text in particular. The other paper (Madalli & Patel, 2009) aims to discuss the various issues involved in Indian languages computing, particularly Telugu, like creating, displaying, searching and retrieving digital contents. Another study (Mukhopadhyay & Azim, 2006; Mukhopadhyay, 2007; Das et al., 2005) developed Bengali script based digital library through the application of open standards and OSS. In another papers, Mukhopadhyay (2006a, 2006b) attempts to design Floss based software framework for public library based and Web-enabled multilingual community information service. Chandrakar (2002) described ‘Unicode’ standard UTF-32 that uses four bytes (32 bits) to encode all possible (millions) characters. In another paper he explains Unicode as a multilingual standard and the related technology available for localizing the Indian language materials is discussed (Chandrakar, 2004). Another study (Tripathi, 2004) focused on problems related to cross-lingual information retrieval due to the use of non-standard metadata schema by many of the libraries. He concluded that there is no significant development in Indian language machine translation. In another paper he described Saraswati, a cross-lingual Sanskrit Digital Library hosted at Banaras Hindu University (Tripathi, 2009). Karande (2007) proposed an implementation of multilingual search engine using a Universal Networking Language (UNL). Tripathi & Tripathi (2007) explained difficulties towards creating digital libraries in Indian languages. Another study (Vaidya & Jayakanth, 2006) explained the configuration changes that are needed for Eprints.org software to store multilingual contents and to create a multilingual user interface.

Designing user interface for any IR system is also a crucial issue (Singh & Pandita, 2005; Vaidya & Jayakanth, 2006). Jayakumar et al. (2007) explains a prototype IR
User Interface which was developed in-house at Indira Gandhi Centre for Atomic Research. Rajashekar (2003) described various workflow pattern of the IR system. Another expert (Sreekumar et al., 2007) is in support. University Grants Commission (2005) advocated for Web interface having provision for customization.

2.3 Findings of Literature Review: Gaps and Overlaps

Up to date most of the IDR studies trend to only focusing on the beautiful side of IDR, such as how many items are collected or what IDR can do in theory, and people do not want to touch the hard part – everyday work, the services, the usefulness of the service in practice and the quality of the IDR. Some think IDR has achieved more than they need and some believed it has not met the service level. This review looked at what studies have been done, what can be learned from the studies and what the gaps are.

A. Existing policy

A successful IDR can not be developed without giving serious consideration to its overall structure and design. This overall structure and governance can be initiated by developing IDR policies (as suggested in Table 4.2 of section 4.2.1 of chapter 4). Repository policies need to have clear explanations and examples but several policy issues are missing and have not been discussed in the literature which need to be developed in the line of global recommendations.

B. Advocacy and Promotion

The success of IDR system much depends on user’s participation and voluntary involvement with the system. There should have publicity and branding policy of IDR resources. But studies lack focus on how IDR will increase in quantity, in order to become competitive with other providers. Another issue training and documentation (for end users, authors and administrators) has not been discussed in literature.

C. Contents

There are many decisions associated with the contents policy. There will be descriptive requirements, preservation issues, workflow pattern associated with different types of documents. There is no study as to how it will be structured. Consideration should be made of the types of checks and who is likely to be performing it. All these issues are not properly discussed in literature. No discussion about retrospective information or contents has been reported.
D. **Legal Framework and Licensing Model**

IDR don’t have huge legal problems in their setting up but can face heavy issues in the development of a critical mass, consisting of journal papers written by researchers, faculty members which are commercially published. There are three parties involved in this system namely author, institution and publisher. It is not possible to have a single straightforward licensing policy as different publishers have different licenses. It would therefore, be desirable to have a customizable licensing system for items submitted to the repository. Studies do not have licensing policy for the legal issues in regard to uploading and accessing the documents. Even studies do not propose any licensing models for rights management of IDR System.

E. **Access Control and Rights Management**

IDR system must have mechanisms to restrict access to the information when OA is premature or otherwise not desirable. But a few studies have been conducted and further studies are required in this area.

F. **Resource organization and management**

For any IDR system, data standardization tools like standard lists, code lists or vocabulary control devices need to be followed but studies do not propose common standard, methods, tools and techniques in the organization of IDR resources. As a result services are not based on internationally agreed upon data standard.

G. **Contents’s Quality**

Especially for the quality of the IDR contents, only a few studies have been done in this area. Validation and verification of the data are an important factor and there are two levels of this task: the metadata about a particular scholarly work and the organizational and contextual information about the work. These issues have been neglected in the literature. Studies do not clearly explain how quality is assured or who will measure the contents’s quality. So, further studies are urgently needed in this area.

H. **Metadata Schema**

Studies do not recommend any domain specific metadata schema suitable for different types of objects. It would therefore, be desirable to have a flexible metadata schema that will be to able to combine the different requirements for each of the possible contents domains to produce a metadata set to be collected which is exactly appropriate. Another issue, authority control is a large issue in metadata submission. But this issue has not been discussed in literature.
I. Indexing services and standards

Another key issue is to develop common platforms to make IDR interoperable with other system in order to import and export of resources. Users aren’t interested in browsing each IDR, there is a need to integrate resource discovery/search tool that can be built using OAI-PMH metadata structure. This issue has not been properly discussed in literature. Studies are urgently needed in this area.

J. Resource identification for perpetual access

Identifying and locating online contents is a key issue in repository systems to grant long-term access. IDRs need to assure the permanence of object names in the repository. It provides access to the contents on the same timescale and shows the relationship with other items is another. Each object should have a unique and persistent identifier such as +CNRI, handle, DOI, etc. But this problem still exists and it is barely treated in the literature, as a mere technical aspect. So studies are urgently needed in this area.

K. User Interface

There are hardly any studies have been done in this area. Any IDR system demands customizable, accessible, web service interfaces so that repository can participate in distributed application systems. Development of multilingual interface and mechanism to access multimedia learning objects has been ignored in literature. Studies are urgently needed in this area.

L. Multilinguality

This issue has been badly neglected in the literature. Crossing the language barrier is not the only challenge facing the multilingual digital libraries. Even studies do not recommend managing multilingual database. There is limited scope for multilingual search provision and access to multi format learning objects. Studies lack focus on multilingual data format that will support Universal Character Set (UCS). Additional problems and challenges are related to: data management and representation of information, interoperability (linking between systems) etc. Multilingual document indexing is also challenging because each language has different characteristics and rules. All these issues have not been discussed in the literature.

M. Model

There remains a lot of work to be done to determine best funding and business models however a few studies have been done in this area. Studies are urgently needed in this area.
N. **Standard compliance**

As the repository field is a new and developing one, some parts of the system have more agreed standards than others. Some standards, such as OAI-PMH, are part of the software and are well developed. The standards on metadata are still under development. All system is able to provide Dublin Core output, but this is not descriptive enough to be able to ensure complete interoperability and transfer of data between different systems. The system should be based on common standards in respect of software, data formats, import/export, access and network etc. but all these issues have not been properly discussed in the literature.

O. **Preservation and Curation approaches**

Studies do not discuss any specific plan for preservation along with backup strategy. Only a few studies discuss about file preservation. There is no study about in-house preservation system and these services are still under development. There are two technical issues: the first being to ensure that the physical item remains intact at the bitstream level and secondly that the digital object remains understandable has not been properly discussed. This challenge remains a long way from being solved. A coordinated strategy like OAIS (Open Archival Information Systems) reference model, the de facto standard for digital archive architecture is required to ensure long-term preservation of the IDR contents.

P. **Software**

Studies recommended using OSS for IDR system but there is no uniformity in using software for IDR system. Even software is not up to the global standard and demands customization to meet the local requirements. There should be options to allow the system to plug directly into local system. But studies do not recommend pluggable/modular software architecture that could be customized as per requirements.

Q. **Workflow standard**

This issue has not been examined broadly in literature. Studies do not recommend any standardized workflow pattern for different contents type. Generally, studies do not discuss workflow stages so that all verification and refinements can be done in a logical way.

R. **Role of Librarian**

It is not clear from the study as to who will manage IDR's system and what role and responsibilities librarian will perform in this distributed system.
S. Metadata Harvesting

Another technical issue like metadata harvesting from different OAI-PMH compliant repositories has been neglected in literature and no framework has so far been developed. There is no such mechanism of extracting metadata from multiple repositories.

In addition, not a single study recommends Indic script based subject access system (vide section 5.4 of chapter 5) in software framework for searching and indexing specific subject categories. Even, the existing study does not recommend use of social networking software (e.g. Blog, RSS etc.) in repository system for scholars.

In India, these problems are very much clear as there is a lack of published documents and almost all the documents are theoretical in nature and do not cover any policy in details. Only a few studies are institute specific but do not recommend any guidelines and best practices.

2.4 Need of Present Research

There are many studies, journal articles, papers and reports, survey on IDRs. Generally the works discuss mainly the pros and cons of IDRs. Only a few studies have been conducted so far that produces standards and strategies and suggests best practices. But all studies are theoretical in nature. There has been little research that discusses different issues and facets of IDRs, its technical architecture along with practical implementation for a specific organization. So there is an extensive gap in the literatures and several technical as well as non-technical issues have not been covered. Only a few technical papers have been published in India and almost all the studies are descriptive in nature. These documents do not provide any direction towards the building up IDR for a specific organization. There is no existing roadmap for how best to make progress. As it (IDR) is a flourishing area so indepth study is required for designing a policy document for a specific organization. This research work will help administrators in developing university specific repository policy based on global standards and procedures for managing disciplinary and institutional collections and will give a better sense of how to handle some operational issues when implementing a repository. In addition, there are many initiatives and recommendations at national and international levels by various organizations, funding agencies and learned societies that support observations made in previous section. The study will become a link between the research proposed and the studies already done in this area. Lastly, it will provide a handy guide to the topic and will give comprehensive knowledge to the professionals keeping them up to date with what is current in the field.