Chapter 4

4.1 Introduction

National Digital Repository System (NDRS) is basically an aggregation of resources available in various digital repositories. It is to take care of the maintenance of direct services to end users, or directs users straight to the required resource available in various institutional repositories in the country, offers an enhanced metadata to the end users, facilitates preservation through appropriate metadata provision and/or content package and maintains the control of over the content while releasing metadata. It also provides control and personalization of content access through RSS feed, provides the exposure for content providers, preservation and metadata enhancement capabilities to support the long-term storage and access to the content.

It offers various value-added services to enhance aggregated metadata, amalgamate collections, brokering role to facilitate access, to multiple source of research and other materials to aid discovery and possible exploitation through the building of value added services on top. It provides a single point of information for statistics about access and downloads of data (1).

Based on the current developments and challenges in the repositories of digital content, the main research questions in this study are to evolve *national digital repository so as to follow best model and practices derived from the experience of the other system in the world and making it suitable to Indian context for accessing information in large repositories and to provide assistance to users without impairing usability.* This is a quite broad research question. In order to narrow the scope, focus of the research on following important challenges related to the present study.

4.2 Challenges in NDRS

The keen challenges of NDRS are:

4.2.1 Large Information Repositories

As repositories are growing larger, it is not expected that users to search each and every repository individually on Internet or Internet search engines; consequently it becomes necessary to consolidate all the institutional repositories as a National Digital repository to make it useful and for better visibility for the nation.

4.2.2 Complex Data Models

In order to make use of data stored in a various repositories, it is important to understand how the data can be accessed at a single search (data model). As model complexity grows, this task becomes more difficult – especially for users with limited training and background knowledge. Writing an exact query is only one way in which a repository can be accessed. Does it make sense to support different strategies and can this be done without going to each and every individual IRs. All Internet search engines are not covering all the repositories (2); Whether NDRS may able to achieve this task of covering a large number of repositories.

4.2.3 Unpublished and Institutionally Created Digital Information or Publications to be Covered

Though the challenges mentioned are more focused than the overall research study, they are still too broad and encompassing for the in-depth study. For this reason, study used some of the fundamental ideas to address the key challenges to make it simple and usable for large information repositories. By developing a model, implementing and evaluating several national repositories, based on several ideas, it is expected to gain insight into the research question and the key challenges. The

study covers all related ideas for NDRS as a vital part of successful step in developing comprehensive repository evaluation frameworks. One such measurement is to compare universe of possible IRs and NDRS for total research output of the country. It also helps for detailed tracking of contribution institutional scholar's output, ultimately of the country and additionally, it is helpful for comparing actual participation pattern of the different IRs elsewhere in the world for their possible participation.

Another measurable facet of participation is the distribution of a repository's total content, per contributor or institutions. By applying simple analysis on NDRS – a frequencies and distributions of authors and papers contributed in the country. Would such a perspective on a digital repository be useful? The measurement involves, if only in IRs, what is already in a single repository, would the data for analysis be simple to obtain from many repositories through NDRS? There are many questions as to (3):

- Whether this consolidation would become important for all the individual contributors or institutions?
- Whether contributor distributions a useful analytical metric for repositories?
- Do these distribution patterns differ for institutional and disciplinary repositories?
- Do contributor distribution patterns differ for mandatory-deposit and voluntary-deposit institutions?

4.2.4 All the Questions are Considered in Developing NDRS

Open Access (OA) repository, most of these have very little content in them, such as published research articles, working papers, theses and preprints. In fact institutional repositories are becoming backbone as institutions are collecting and exposing the research outputs of every research-active institution. At the same

time, awareness about the Institution Repositories or NDRS to the researcher is poor, or are to those aware of IR not informed of its benefits, or are aware of its benefits do not provide it for their work. These changes calls for corresponding changes in academic policies and procedures of institutions, to make repositories embedded in the culture of research-based organizations (1).

Research funders with public money, are funding as moral case. Even the funders are not yet developed policies. However the promising development on Open Access is that some of the individual institutions and universities are pressing ahead with repositories and very few are developing policies (4). Also there are software for interoperability, repository establishment, data repositories, legal aspects of OA, preservation, e-learning and machine services continue to push forward our understanding and knowledge of how Open Access can develop and establish good practice in active developmental programmes. There are strategies and support access to and use of repositories, with a view to the establishment of a national repository services infrastructure or framework. In addition there are sustainable technical and organisational models to support user-oriented services across digital repositories. This study is to find the merits and demerits of various implemented strategies to support access and use of repositories in the establishment of a national repository services infrastructure or framework. This study encompasses analysis of clusters that cover data, e-learning, preservation, legal and policy issues, machine services and integrating infrastructure (1).

The recommendations of the study are based on the followings (but not limited these factors only):

- **User requirements study** – this includes both distilling existing systems/ practice and detailed primary study
- **Analysis and review of the Institutional repository and its service** - already in operation with respect to organisational requirements for building viable and sustainable repository services on a national scale

- **Technical architecture and infrastructure for national repository**: examines the limitations of current technical standards with respect to the demands of interoperability of a range of repositories and repository types, and the possible technical solutions
- **Business modeling for NDRS**: development is visualized with an attention to scalability, viability and sustainability

The NDRS model is developed and predicated upon free-to-use information and to provide access to distributed IRs to be linked with interoperable data providers so that maximal amounts of content can be searched whilst the locations of individual IRs.

NDRS goal is to link repositories in such a way that services can be built upon them that provide value and utility to the user. Some of the repositories are based at departmental level, or institution-wide but serving several separate purposes - individual repositories for thesis, eprints, and research data and so on. The NDRS requires to be networked effectively and building upon them a set of services:

- Collect and present information
- Facilitate searching and access
- Maintain compliance with legal, intellectual property and copyright bounds
- Enable resources to be shared,
- Provide and use common standards
- Facilitate the deposition of information
- Encourage and motivate authors to participate by depositing their work
- Have overt and identifiable benefits to authors and users
- Be viable and sustainable

In some of the cases Open Access content may themselves display an array of business models, some being free-to-use and others paid-for. The goal is to develop a linking model that is workable, connects Open Access repositories together effectively, and permits service providers to develop their offerings over the whole corpus of Open Access material.

4.3 NDRS Technical Model

Many of the national repositories have developed along different lines according to national requirements and attributes with a broad-scope of national database of Open Access content that can be added to, searched, mined, re-used, exploited for specific interest groups and built upon overtime.

Some of the national repositories start with a few organizations and later extend to the remaining research institutions/ universities. Each institution/ university need to have repository-exposing content to OAI harvesters provides a search interface for users. This need to be made as mandate in the country, and require researchers to deposit the details of published articles in their own institutions. NDRS links all the repositories thereby researchers deposit only once and metadata is migrated between systems.

Some repositories link the institutional repositories, some harvests OAI-compliant content from the repositories and provides the search interface and also have built subject-specific services on the system. Institutional repositories in collaboration with the NDRS of the country need to preserve text-based, video, audio, moving images and like. It can be modeled in any of one three model defined.

4.3.1 Centralized Model

In the centralized model, authors would deposit their e-prints on to central archive. It would provide the interface through which readers would search, browse and retrieve articles. The metadata from these articles would made available to other service providers through various protocols such as OAI-PMH as data providers, RSS, SRW/SRU, and Opensearch for use by other service providers (5). It has been depicted in the figure 4.1.

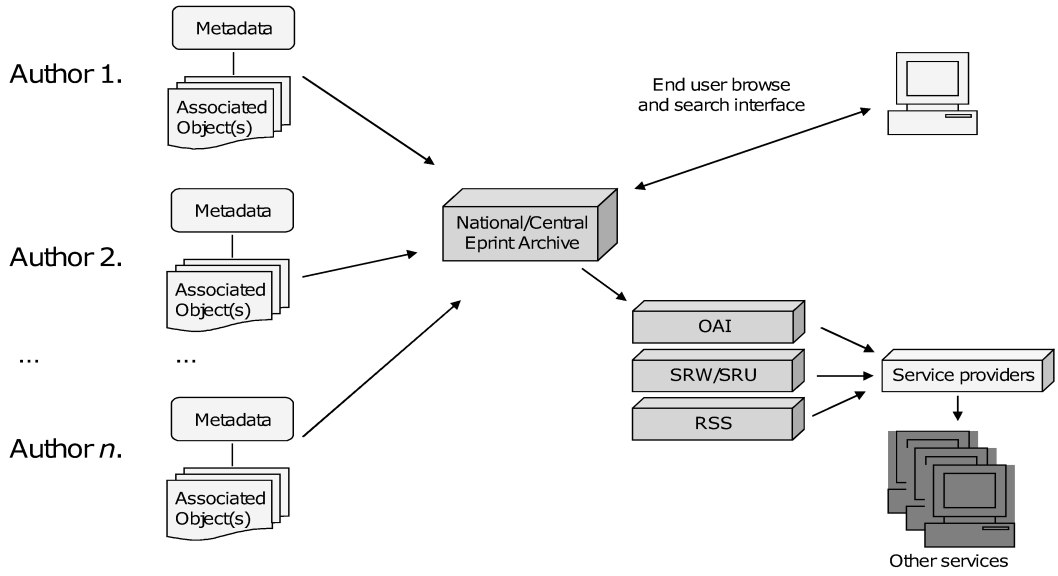

Figure 4.1: Centralized Model

Reproduced from Swan, A. (2005). Delivery Management and Access Model for E-prints and Open Access Journals in UK further and Higher Education. Retrieved from website: http://eprints.ecs.soton.ac.uk/11000/1/Eprints_LP_paper.pdf

The Pros and Cons of the centralized model are given below:

4.3.1.1 Advantages of Centralized Model

The Centralized model offers the following advantages. These are:

- Here, the hub will interact with the individual server, thus reduces the amount of connections with each other as central hub has to maintain a single connection.
- The centralization of harvesting and searching tasks enables the hub to encapsulate the business logic such as changes in protocols, extension and continual refactoring of the core logic without having to replicate by each individual server.
- The centralized hub can be made most up-to-date source of data for other institutions to build on it.

- The centralized hub would harvest data from the individual institution, thus reduces the amount of traffic to the individual server by removing the need for individual institutions to harvest data from other nodes.

4.3.1.2 Disadvantages of Centralized Model

The centralized model has the following disadvantages:

- It creates a single point of failure; if the hub goes down then service is lost. This risk could be mitigated by creating multiple mirror sites
- Nodes have to wait for the hub to respond. This risk could be mitigated in the selection or development of the hub by concentrating efforts on Quality of Service as a measure of success.
- Information latency would be another factor which depends on mechanism of gathering content from the respective institutions.
- To provide long term sustainability, the centralized hub has to make a provision for strong business model as the cost of maintaining central infrastructure would be quite heavy.

4.3.2 Distributed Model

In this model, metadata and contents would be residing in their respective servers. Metadata is cross-searched ‘on the fly’ (5).

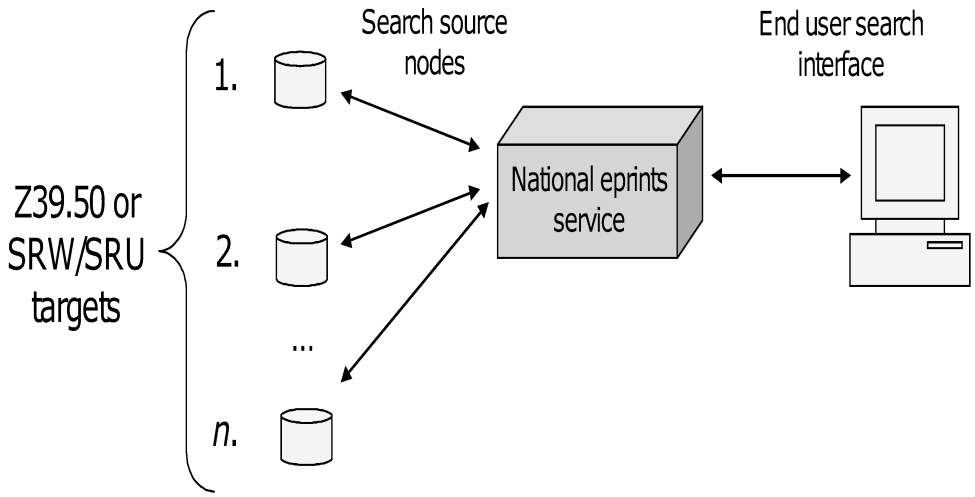

Figure 4.2 Distributed Model

Reproduced from Swan, A. (2005). Delivery Management and Access Model for E-prints and Open Access Journals in UK further and Higher Education. Retrieved from website: http://eprints.ecs.soton.ac.uk/11000/1/Eprints_LP_paper.pdf

It offers certain advantages and disadvantages as given below:

4.3.2.1 Advantages of Distributed Model

- It would provide all the time metadata and contents up-to-date and access to contents is immediate.
- There is no need of replication of metadata
- This is true autonomy in that it is up to an individual institution as to whether they want to connect to another institution
- There is no one point of failure
- Searching scripts can be customized giving the institution 100% visibility
- Searching can be conducted locally without worrying about time lag involved with communication between a central system and end nodes.

4.3.2.2 Disadvantages of Distributed Model

- There is no scope of refinement of metadata as metadata is cross-searched on the fly. The presentation of metadata would entirely depend on what the local repository would make available to end users[swan]
- The more connections you have, the more maintenance is required. If each node on this type of network was to reference every other node the number of connections is exponential to the number of nodes
- If one server was to change their URL then all servers connected to it would have to modify their reference to point to the new location.
- Various developments such as change of protocol, software etc taken at institutional level results in the functionality of the distributed model.

4.3.3 Harvesting Model

In this model, the service provider would harvest and store metadata from available Institutional Repository/e-print archives and open access journals (who become data providers), using the Open Archives Initiative Protocol for metadata Harvesting (OAI-PMH). Harvesting model would have the interface through which user can browse and retrieve results. The metadata from the IR server would be exposed via various protocols such as OAI-PMH, SRW/SRU and RSS - as data provider to other service providers. It is a hybrid model wherein metadata is harvested into a central searchable database and the content is distributed in Institutional repositories (5).

It offers the following advantages and disadvantages as given below:

4.3.3.1 Advantages of Harvesting Model

The advantages of this model are:

- The OAI-PMH is a standard protocol which is easy to implement

- OAI-PMH is a flexible in nature. Though it employs, at present, unqualified Dublin Core metadata schema, however, other richer, more complex, metadata schemes can also be employed
- The OAI-PMH is primarily designed to allow metadata exchange and the sharing of scholarly knowledge
- A large number of institutional and subject-based archives employ software that supports the OAI-PMH.
- Harvesting can be carried out by automatic scheduled tasks, thus minimum efforts are required to harvest metadata from other institutions.
- Search providers can enhance various value added services such as statistics, subject classification, browsing and searching capabilities etc.
- It can form the basis for an overall programme of continuing development and improvement

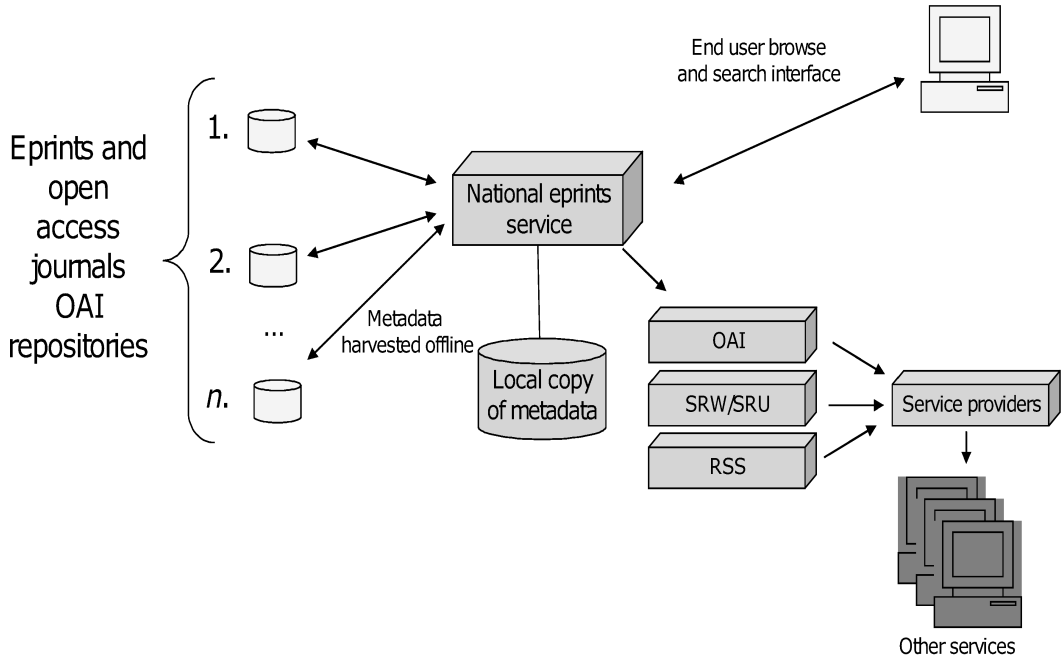

Figure 4.3 Harvesting Model

Reproduced from Swan, A. (2005). Delivery Management and Access Model for E-prints and Open Access Journals in UK further and Higher Education. Retrieved from: http://eprints.ecs.soton.ac.uk/11000/1/Eprints_LP_paper.pdf

4.3.3.2 Disadvantages of Harvesting Model

- At present, OAI-PMH employs only Unqualified DC, which is mandated as the minimum metadata standard for use by the OAI, is the only metadata scheme. It is a lowest common denominator which lacks semantic richness and limits the possibilities of providing enhancements
- The metadata exposed by the service may not always be the very latest version of that metadata.

Now question arises which model should be adopted as national level system, here, and the comparison has been made between harvesting model with distributed system to understand the pros and cons of both the model.

4.3.4 Comparison between OAI and Z39.50

Z39.50 is the American National Standard Information Retrieval Application Service Definition and Protocol Specification for Open Systems Interconnection. It defines a standard way for two computers to communicate for the purpose of information retrieval. It supports information retrieval in a distributed, client and server environment where a computer operating as a client submits a search request (i.e., a query) to another computer acting as an information server. It can be implemented on any platform. This means that Z39.50 enables with different operating systems, hardware, and search engines, database management systems to interoperate and work together seamlessly. There was a debate as to why not use the existing Z39.50 protocol, which is used for the search and transfer of metadata (http://www.cni.org/pub/NISO/docs/Z39.50-brochure/50.brochure.part01.html).

The harvesting approach based on OAI-PMH protocol might look operationally much different to the Z39.50, but both achieve the functionality of federated search. The federated searches allow users to gather information from multiple related resources through a single interface. The basic difference between the two protocols is in the search approach. The Z39.50 allows clients to search multiple information servers in a single search interface in real time, gathers the results, eliminates or

cluster duplicates, sorts the resulting records and presents the results to users whereas the OAI-PMH allows bulk transfer of metadata from the repositories to the Service Providers' database. Hence the clients do not need search multiple data providers in real time rather they search the metadata database of the Service Provider who collect and aggregate the metadata from different data providers.

There were number of reasons to adopt a new protocol rather than implementing archaic protocol as of today. Some of the reasons for using OAI-PMH protocol in establishing NDRS are given as follows (6):

- It has been proven that it is very difficult to create high-quality federated search services across large numbers of autonomous information servers through Z39.50 for several reasons such as some servers in the whole system would remain down due to network bandwidth and performance tends to be constrained by the performance of the slowest individual server participating in the federation of servers.
- Retrieval accuracy is a problem as different servers interpret Z39.50 queries differently due to lack of specificity in the standard, leading to semantic inconsistencies as a search is processed at different servers.
- There are scaling problems in the management of searches that are run at large numbers of servers; and performance would be affected making based federated search performance sensitive to participating server response time, result size, and network bandwidth.

Moreover, implementation PMH is very simple since one does not need a different port like Z39.50 (which uses port 210). It works over the HTTP, which any web server listens. It means one can use common Linux programs such as wget or curl to harvest the metadata from repositories. One does not need a special toolkit like Yaz for Z39.50 to work with Z39.50 standard (6). According to Lynch (7) "These two protocols are really meant for different purposes, with very different design parameters, although they can both be used as building blocks in the construction of similar services, such as federated searching. Neither is a substitute for the other […] and we should not think about the world becoming partitioned between Z39.50-

based resources and Metadata Harvesting protocol resources, but rather about bridges and gateways."

As distributed model offers the advantage of providing up-to-date metadata as it focused on immediate access to source location of metadata, including content and the major disadvantage is the scalability of this model when accessing large number of repositories simultaneously. However, in case of harvesting model, up-to-date access can be mitigated by more frequent harvesting of repositories and also provide more refined metadata to users. Harvesting model offers much more advantages than distributed model. Thus, it is proposed that harvesting model to be preferred over distributed model.

Further, an attempt has been made to compare Harvesting Model vs. Centralized Model to understand nuances of both the systems.

4.3.5 Comparison between Centralized and Harvesting Model

Central Subject-based e-print archives have been set up largely as a result of the efforts of eminent scientists such as Paul Ginsparg in setting up ArXiv, CogPrints set up by Stevan Harnad, PubMed Central under the influence of its former director, Harold Varmus. These have been hosted at institutions where these scientists worked and are populated with articles as a result of advocacy within the appropriate subject communities. It is anticipated that this sort of activity will continue in establishing more subject-based e-print archives to their subject community. Certainly the National level repository would harvest from existing subject-based archives and also those set up in the future (8).

Institutional or departmental e-print archives, however, will be far more important to the success of the national level services. Subject-based archives have been successful in their particular fields, but still cover only a fraction of the total research

output and this is likely to remain the case even if the numbers of such archives increase. Moreover, this approach is the 'wrong way round' with respect to e-print provision since for cultural reasons distributed, institution-based archives are much more likely to fill quickly, particularly if institutions adopt mandatory policies across all disciplines, something they are likely to do when the advantages of such archives to institutions become clear (8).

The potential for centralized archives as a basis for a national open access service is suboptimal because:

- The number of centralized (subject-based) archives is tiny and it is speculative that subject-based archives covering the whole spectrum of scholarly research will be set up within a reasonable time (those that operate at the moment are all in the sciences) (8).
- The number of articles deposited in them has grown only slowly over the last few years. It requires a lot of advocacy in filling up central archives within a subject community, something which may only ever work on the basis of persuasion and appeal, since there is no discipline-based power to mandate content provision. In practice, this is a tried-and-tested approach that has met with only limited success (8).

Keeping in view of aforementioned points, advantages of having centralized model is not very high as compared to disadvantages. The OAI employs a philosophy whose time has come, and it has gained worldwide acceptance. It makes it helps in sharing information about scholarly resources and also offers enhanced resource discovery tools. In view of this, it is recommended that the harvesting model should be adopted to serve as the basis of National Digital Repository System.

4.3.6 Aggregation Model (NDRS) Requirement and Services

The developments of any system or model, digital era, can be a user-centric (around users) or a technology-centric (around technology) or a content-centric (around content). It is important to take into account of potential users of the system. The followings stakeholders identified in the development of NDRS should have some practicable responsibilities:

- Repository Manager
- End users as searchers / readers
- End users as content providers

4.3.6.1 Repository Manager (RM)

Repository Manager is to manage repositories of an institution, where repositories can be subject-based, object-type or special collection repositories (such as museum collections) (9). The responsibility of developing repository involves the following tasks of:

- Installation of software
- Customization (look & feel) of the software
- Adding digital content
- Creation of Metadata
  - Use of controlled vocabulary system
  - Name Authority Systems
- Selection of File Formats
- Digital Preservation
- Data Exposure
- Manage Access and Authentication

Repository Manager needs to understand the user's problem and their requirements such as Access and Authentication (keeping in view of copyright of the publisher), providing access to the owner of the document including end user services, Access control, advocacy and usage. The Repository manager has to find out usage of

resources so as to encourage users to deposit the content to repository and at the same time to build confidence among the senior management of their institute.

In context of digital preservation, repository manager requires devising policy for a long-term preservation; attend the change of file format, required technical standards and exposing metadata in appropriate forms to achieve the proper visibility for the repository's content. The knowledge of Intellectual Property Rights (IPR) particularly copyright issues (as users are always wary of these issues) is important to get content from researcher/content creator and attract them. The sustainability of repository is an important issue, where repository manager has to formulate a plan for business model, in want of additional fund, after initial funding from public funded bodies or other funders.

4.3.6.2 End Users as Searchers / Readers

Users search information via discovery services such as Google, Scirus etc or via subject or object type-specific portals or discovery services, which are different among different scholarly disciplines and even the way information is accessed, used and deposited in repositories, and nature of the information itself (10). Majority of users are accessing the primary data through the prefer route of Google as demonstrated in JISC study on time-based media collections examination of the log files for the Southampton ECS repository which showed only 11% of searchers entered via repository itself, rest coming in via Google and other aggregator services (11) where librarians observation is that majority of their students and researchers feels that 'if Google can't find it, it isn't there'.

4.3.6.2.1 Access

For some users, subject-specific or object type-specific specialized discovery services will be the preferred route. However, the literatures in institutional repositories are now also gaining importance. Web portal of National repository system should cater

to user needs in searching wherein unsophisticated users - search by item type, keyword, field; experienced users - use Proximity operators, Boolean operators, nesting and so forth. It has also recommended that there should be authority control over the keywords that human cataloguers may assign to describe the digital object. These same keywords may be used by the visitors of the repository for searching and browsing of the repositories (12). However, those users who search using simple strategies require a simple, uncluttered interface to maximize ease of retrieval. Alerts are important to users through e-mail, though, RSS may be more useful.

Users need to extract information from repositories in various forms and styles to suit particular purposes such as simple sort by date, to include various permutations with document type and date, item type and so forth depending on the kind of requirements by the users. Though, these issues are relatively simple software developments yet they make the difference between adequate and extremely useful for the end user.

4.3.6.2.2 Searching

Users may need simple search without much of noise or irrelevant material for which technology needs to attain better level of performance and pragmatism. There is a feeling among end users that repository should be bigger in scope and size to be searched also at the same time to provide consistent high quality metadata formats to support cross-resource searching (13). Users may not always know the types of material relevant to their needs and at that time, resource discovery services become very important while searching across repositories to return all item types in response to a search on a specific topic.

It is mentioned that end users gather information from various resources to repurpose the content like teaching, which can be delivered and populated through Course Management Systems, Virtual Research Environments, Virtual Learning

Environments and over time it is expected that repositories and search services will become part of such systems with all the interoperable components linked (14).

4.3.6.2.3 Access and Authentication

In context of personalization of search and information, users tend to look out locally at institutional level, rather than at a centralized system. Also users want to access materials available over Internet outside campus; it is possible to access through authentication tools.

As this study is confined to Open Access content, the question of payment does not arise, nonetheless, there is a need to discuss about the payment to access the materials, wherein metadata are open access but the full object is not, as is the case with material where royalty fees might be applicable. From the user's point of view, where payment procedures are necessary and it must be made very simple and trusted. Wherever payment is required, repositories must authenticate users, before allowing access to their content and services must provide the means to implement these systems. Revenue collection, accounting, and reporting functions must be provided by services either specific to individual repositories or acting as cross-repository services (15).

4.3.6.2.4 Value-Addition

In work-related environment, users find it increasingly difficult to manage information and they value the processed information. Repository Managers (RM) who can add this sort of value to content residing in repositories will gain immediate recognition. There is plenty of scope for RM where Open Access research repositories are concerned, since content is available for aggregating in new ways, re-publishing and make it a useful information product to user communities (1).

4.3.6.3 End Users as Content Providers

Content providers are, in most cases, the same people as the searchers/readers, but their role as providers of content to Open Access Repositories, need encouragement on several scores. There are large proportion of researchers and teachers, who have content they may wish to share, do not have a proper web space to deposit that material. There is a prospect for national-level repository that would accept articles from researchers whose institution does not yet have a repository of its own. It is suggested that repository being envisaged as primarily for research output and its associated objects, hence need to think of open repositories. In fact researchers got dissuaded from depositing their work due to lack of knowledge about copyright and other associated issues (1). End users are interested to deposit metadata along with the digital only once to either institutional repository or funder's repository and they want that process to be as simple as possible. To further ease of entry, simple metadata creation can be done by author, if rich metadata are required then skilled mediators are necessary. JISC has funded Simple Web Service Offering Repository Deposit (SWORD) project to make the process of author deposit easy by entering only once their metadata and it is also replicated in funder's repository (16).

Many a times researchers face difficulty even to upload document in just 'PDF' format, it is necessary to train the users or to allow depositors to submit in any format and a mediator reformats the document. The deposit process should be embedded in workflow and to make this as simple as possible it should take place at the authors' own institutions in order to maximize efficient and compliance by authors. The deposit must also make sense with respect to each subject discipline; in other words, if specific requirements are made regarding metadata, these must fit into the concepts of a discipline as closely as possible. There should be advice on deposit process, educate on metadata creation, help with format issues to researchers (17).

4.3.6.3.1 Intellectual Property Rights (IPR)

The knowledge of copyright with authors is not sufficient to understand the intricacy of publishers' copyright agreement as they are frequently confusing (or even opaque) and can vary from journal to journal even within the same publishing house. Researchers are always having anxieties about infringing copyright agreement. It is one of the barriers to spontaneous self-archiving. Service provider should prepare guidelines on copyright with respect to repositories which may play an important role (1).

4.3.6.3.2 Ownership

The sense of ownership of their material by authors can be very pronounced. Authors regard their output as their intellectual capital and are anxious that their work should remain associated with them in case it enters the Open Access corpus. Repositories themselves may also wish to 'brand' their content; in some cases this is for marketing reasons alone, but in others there may be regulatory reasons for having this in place. The Identity of the user and purposes of using the data need to take this requirement into account (1).

4.3.6.3.3 Usage

There are some of the advantages in providing usage statistics to authors, where authors feels informative, and encouraged to deposit further articles because of the increased visibility.

4.3.6.3.4 Impact

The usage data just inform authors and encourage submission, whereas citation analysis of the services can provide impact very effectively. For example, BASE is one of the biggest OAI-Search-Engines having interface with Google Scholars to provide statistics about each articles in its repository (http://www.base-search.net).

4.3.6.3.5 Peer Review

Peer review for the scholarly literature has traditionally been carried out by scholars, within a process of manuscript and publishing management by publishers. Now, there is a new approach and form for peer review i.e., by posting publication, commentary, pre-publication on repository for open peer review and there are variations in some of those themes. Some learned society publishers are involved in this, and some are already using repositories as a submission tool, encouraging authors to alert the publisher when an article has been deposited as a preprint so that the publisher can take on and manage the peer review process and formally publish the article subsequently. This constitutes an early-indicator of how publishers may themselves provide services across repositories in the future (http://www.earlham.edu/~peters/fos/2008/07/nature-will-deposit-into-disciplinary.html).

4.3.7 Type of Contents Covered in Various IR and NDRS

Research Communities along with the developer of IRs have to decide the kind of contents to be deposited in the IRs. Some of the types of contents found in the IRs are:

- EPrints – Preprints/Postprints
- Working Papers and Reports
- Conference papers & Proceedings
- Electronic Thesis and Dissertations
- Data Sets
- Supplementary Materials
- Online and Overlay Journals
- Books
- Learning Objects
- Multimedia Collections

- Electronic Portfolios

Some of the important NDRS existing in various countries are given below:

4.3.8 Various Existing National Level System

At national level, a network of institutional repositories increases the exposure of national research output, and allows services, such as enhanced searching, and statistics generation to be developed using economies of scale. Some of the initiatives in various countries based on harvesting model are described below:

4.3.8.1 France-HAL (Hyper Articles On Line)

(http://hal.archives-ouvertes.fr/index.php?langue=en)

It is an Open Archives repositories of France launched in 2001 and operated by CNRS (Centre National pour la Recherche Scientifique) with individual portals, either thematic or institutional. It is a multi-disciplinary open access archive for the deposit and dissemination of scientific research papers both published and unpublished and PhD dissertation. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers (18).

4.3.8.2 ePrints UK (http://eprints-uk.rdn.ac.uk/)

ePrints UK is a two-year JISC-funded project under the Focus on Access to Institutional Resources (FAIR) Programme which began in July 2002. The aim of the project is to develop a national service provider repository of e-print records based at the University of Bath derived by harvesting metadata from institutional and subject-based e-prints archives using the Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH). ARC harvesting software hosted on the Resource Discovery Network (RDN) servers at UKOLN has been used to harvest metadata from interoperable IRs. They are developing suitable SOAP (Simple Object Access Protocol) interfaces to pass the metadata (and full text) to external Web services for

enhancement, augmentation, or validation of the metadata. Access to ePrints UK service is available through central website for the project, providing a search interface to all the enhanced, harvested metadata (19).

4.3.8.3 Ireland's National Portal for Open Access to Research (http://www.rian.ie)

Ireland's National Portal for Open Access to Irish published research from the seven universities under the Irish Universities Association (IUA) launched in 2010. The IUA Libraries National Research Portal will make Irish research output freely available to the global research community by creating repositories of research papers within each University. The project has established a national portal on the Web which provides a single view of the country's research publications. The aggregated content will make further value-added features such as statistical analysis and will allow other agencies to harvest normalized metadata for better search results.

4.3.8.4 DAREnet (http://www.darenet.org/)

The Digital Academy Repository (DARE) initiative led by SURF in the Netherlands has implemented institutional repositories at all Dutch Universities. DAREnet is the OAI service provider that provides a search service across these repositories. A subset of DAREnet is the Cream of Science service (http://www.creamofscience.org/) which focuses on all the publications produced by the top 207 academics in the Netherlands (of which ~60% can be accessed freely on open access). A third service being developed, 'the Promise of Science' offers access to doctoral theses. National services are being complemented by smaller, subject-oriented service providers including DARC (Distributed Africana Repositories Community). Holland is currently the only country in the world so far with this level of countrywide open access end-user services.

4.3.8.5 DRIVER (http://www.driver-repository.eu/)

Digital Repository Infrastructure Vision for European Research (DRIVER) is launched in 2006 and funded under the e-Infrastructures call of the European Commission's 7th framework programme. It is considered the largest initiative of its kind in helping to enhance repository development worldwide. It offers sophisticated services and functionalities for researchers, administrators and the general public.

One can access the network of freely accessible digital repositories with content across academic disciplines with over 2,500,000 scientific publications, found in journal articles, dissertations, books, lectures, reports, etc., harvested regularly from more than 249 repositories, from 33 countries**.** D-NET, an Open Source Software for repository networks is used for harvesting metadata from the various IRs around the globe (20).

4.3.8.6 ARROW (http://www.monash.edu.au/eresearch/capabilities/arrow.html)

ARROW (Australian Research Repositories Online to the World) is launched in 2003 and project ended in 2008. It is a digital archive for Monash University's research output and contains open access published articles, working papers, conference proceedings, historic photographs and PhD theses. In particular, the National Library of Australia is responsible for the ARROW Discovery Service, which provides OAI service provider functionality across a number of repositories. In order to ensure consistency within this service the NLA has laid down policies on metadata preparation and record submission to researchers. Images are harvested separately from within ARROW for use by Picture Australia and Music Australia.

4.3.8.7 CARL Institutional Repositories

The Canadian Association of Research Libraries supports the systematic archiving of, and access to digital research output of Canadian academic organizations into institutional repositories.

The Canada Association of Research libraries (CARL) began the Institutional Repository Project in 2003 to provide support for Canadian implementers of institutional repositories (IRs). As a result, over 80% of CARL members have implemented an IR at their institution. CARL harvester harvest data from 23 archives across New Zealand and has 140905 items in their database (21).

4.3.9 International Digital Repository System

Apart from existing national level initiatives, there are certain organizations which archive data from various interoperable open archives from the world. Some of the well noted initiatives are given below:

4.3.9.1 OAIster (http://oaister.umdl.umich.edu/)

OAIster was initially funded by the Mellon Foundation in 2001 to investigate the development of OAI service providers. OAIster harvests around 900 repositories all over the globe. The inconsistency of metadata that occurs across repositories has led to OAIster putting in place a number of normalization procedures to facilitate access. OAIster is exposing its harvested metadata to Yahoo! and providing an SRU search interface onto the aggregation (22).

4.3.9.2 ARC (A Cross-Archive Search Service) http://arc.cs.odu.edu)

It was the first end-user federated search service based on OAI-PMH which was established in 2000 at Old Dominion University. It has been developed as a proof-of-concept other service provider to develop similar kind of services, as a result, several service providers taken ARC software to develop similar kind of services. It supports simple search, advanced search, interactive search, an annotation service, and browse/navigation over the search results (23).

4.3.9.3 Scientific Commons (http://en.scientificcommons.org)

Scientific Commons.org is a search engine for scientific publications made available via open access repositories. It is launched in 2007 and uses the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) for isolation of scientific content. It provides access to 25 million documents from around 100 different repositories around the world (24).

4.3.9.4 BASE (http://base.ub.uni-bielefeld.de/en)

The BASE (Bielefeld Academic Search Engine) service at the University of Bielefeld in Germany combines OAI data providers alongside other sources of information and provides a FAST-based search interface across these resources. The service is launched in 2004. This is an example of OAI-PMH being used in tandem with web search engine technology. It has been linked to Google Scholar from search results, furthering this connection and allowing the end-user to pursue their investigations (25).

4.3.9.5 Philpapers (http://philpapers.org/)

PhilPapers is basically a directory of online philosophy articles and books by academic philosophers. Members monitor online journals and publications, and the site also accepts direct paper uploads. There are now about 374,972 items in the database. It was launched in 2009 and has more than 3,000 registered users. It does selective harvesting from the archives listed in OpenDOAR. The service is being funded by JISC.

4.3.9.6 Registry of Open Access Repositories (ROAR) (http://www.roar.org)

ROAR provides information about 2283 repositories registered with it. It promotes the development of open access by providing timely information about the growth and status of repositories throughout the world. One can search the content of all the repositories registered in ROAR by using the Google Custom Search Engine

4.3.9.7 Open Directory of Open Access Repositories (OpenDOAR) (http://www.opendoar.org)

It is an authoritative directory of academic open access repositories which is duly checked the information provided by repository manager. It has over 1800 repositories listed in its database. One can search the content of all the repositories registered in ROAR by using the Google Custom Search Engine.

4.3.9.8 Avano (http://www.ifremer.fr/avano/)

It offers an access to 279550 electronic resources about the marine and aquatic sciences. At present, it harvests using OIA-PMH from 297 Open Archives around the world.

4.3.10 Publisher Repository

It is a publisher's repository, with the published version available but subject immediate access upon publication or to an embargo ranging from six months to one year (26). Some of the best known publisher's repositories are given below:

4.3.10.1 Directory of Open Access Journal (DOAJ) (http://www.doaj.org)

The Directory aims to be comprehensive and cover all open access scientific and scholarly journals that use a quality control system to guarantee the content. It provides access to journal literature without any delay to users. This service covers free, full text, quality controlled scientific and scholarly journals. There are now 5585 journals in the directory and around 2400 journals are searchable at article level.

4.3.10.2 Highwire Press (http://highwire.stanford.edu/lists/freeart.dtl)

HighWire provides online hosting and technology solutions to publishers, management tools for librarians and unique search functionality for researchers. It one of the largest archives providing more than two millions of free full-text science articles to end users.

4.3.10.3 Public Library of Science (PLoS) (http://www.plus.org)

It is a non-profit open-access scientific publishing project aimed at creating a library of open access journals and other scientific literature under an open content license. It launched its first journal PLoS Biology in October 2003 and totally publishing seven peer reviewed journals.

4.3.10.4 BioMed Central (http://www.biomedcentral.org)

Biomed Central is an open access publisher committed to providing immediate open access to peer-reviewed biomedical research. It publishes over 200 online peer reviewed journals covering the whole of biology and medicine. These are open access and available online without charge or any other barriers to access.

4.3.10.5 Access to Global Online Research in Agriculture (AGORA) (http://www.aginternetwork.org/en/)

The AGORA program, set up by the Food and Agriculture Organization of the UN (FAO) together with major publishers, enables developing countries to gain access to an outstanding digital library collection in the fields of food, agriculture, environmental science and related social sciences. AGORA provides a collection of 1278 journals to institutions in 107 countries. It is designed to enhance the scholarship of the many thousands of students, faculty and researchers in agriculture and life sciences in the developing world.

4.3.10.6 HINARI (http://www.who.int/hinari/about/en/)

The Programme for Access to Health Research (HINARI) provides free or very low cost online access to the major journals in biomedical and related social sciences to local, not-for-profit institutions in developing countries. It was launched in January 2002, with some 1500 journals from 6 major publishers. Since that time, the numbers of participating publishers and of journals and other full-text resources has grown continuously. As of now more than 150 publishers are offering more than 7,000 journals in HINARI.

4.3.10.7 MedKnow (http://www.medknow.com/aboutus.asp)

Medknow Publications provides access to around 150 peer-reviewed, online/print online journals in the area of STM. It is the largest open access publisher publishing

on behalf of learned societies and associations. It follows 'fee-less-free' model of open access publishing and provides immediate free access to the electronic editions of the journals as well majority of which do not charge the author or author's institution for submission, processing or publication of the articles.

4.3.10.8 MedInd (http://medind.nic.in)

National Informatics Centre provides online access to 40 full-texts of Indian biomedical periodicals to the users. Participating publishers have signed memorandum of understanding (MOU) with NIC to provide online availability of their journals on line for access to users.

4.3.10.9 Open J-Gate (http://www.openj-gate.com/Footer/About.aspx)

Open J-Gate is an electronic gateway to global journal literature in open access domain. Launched in 2006, Open J-Gate is the contribution of Informatics (India) Ltd to promote open access. It provides seamless access to millions of journal articles available online. It is also a database of journal literature, indexed from 7623 open access journals, with links to full text at Publisher sites.

4.3.11 Search Engines

4.3.11.1 Google Scholar (http://scholar.google.com/)

It is a freely accessible web search engine that indexes the full text of scholarly literature across an array of publishing formats and disciplines. It was released in beta in November 2004. It indexes peer-reviewed online journals of Europe and

America's largest scholarly publishers as well scholarly content from interoperable open archives across the world.

4.3.11.2 Scirus (http:// www.scirus.com/)

The Scirus web search engine from Elsevier, using the FAST search engine technology also used by BASE, offers similar cross-resource search capability through web crawling, including access to open access repositories where these are known.

4.3.11.3 Yahoo (http://www.yahoo.com)

It is a general based search engine which crawled content from the world. It is harvesting scholarly content from the various open archives across the world.

4.3.12 Centralized Subject Based Repository

These repositories, such as PubMed Central or RePEc, restrict themselves to certain fields, where they are usually very prominent. Many are also long established – arXiv, for example, predates the World Wide Web. Some of the very important subject based repositories are given below:

4.3.12.1 ArXiv (http:// http://arxiv.org/)

ArXiv is an e-print service in the fields of physics, mathematics, non-linear science, computer science, quantitative biology, quantitative finance and statistics and is funded by The National Science Foundation. It is operated at Cornell University.

4.3.12.2 Social Science Research Network (SSRN) (http://www.ssrn.com/)

Social Science Research Network (SSRN) is devoted to the rapid worldwide dissemination of social science research and is composed of a number of specialized research networks in each of the social sciences. The SSRN eLibrary consists of two parts: an Abstract Database containing abstracts on over 308,400 scholarly working papers and forthcoming papers and an Electronic Paper Collection currently containing over 248,000 downloadable full text documents in Adobe Acrobat 'PDF' format. It also includes the research papers of a number of Fee Based Partner Publications.

4.3.12.3 ELIS (http://eprints.rclis.org/)

E-LIS was formed in 2003 for the deposit of documents in the Library and Information Science (LIS) domain. It relies on the voluntary work of individuals from a wide range of backgrounds and is non-commercial.

4.3.12.4 CogPrint (http://cogprints.org/)

CogPrint**,** an electronic archive for self-archive papers in any area of Psychology, Neuroscience, and Linguistics, and many areas of Computer Science as well as any other portions of the physical, social and mathematical sciences that are pertinent to the study of cognition. This site is powered by EPrints 3, free software developed by the University of Southampton.

4.3.12.5 PubMed Central (http://www.ncbi.nlm.nih.gov/pmc/)

PubMed Central is a free digital archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health (NIH) developed and managed by NIH's National Center for Biotechnology Information (NCBI) in the National Library of

Medicine (NLM). With PubMed Central, NLM has taken the lead in preserving and maintaining access to the electronic literature.

4.3.14 Setting up National Digital Repository System (NRRS)

NDRS or Aggregation system is based on harvesting system which comprises of the following components (27, 28). These are:

- Storage System
- Selection of Harvesting System
- End User Service

4.3.14.1 Storage System

In order to have a functional National Aggregation System, content/metadata has to be stored on a high end server from which queries can be performed. The major key requirements for the storage mechanism are scalability, storage and speed. Further, it depends upon the number and size of the repositories to be harvested which may consists of many thousands to millions of objects to index. The system should be selected in such a way that it should able to scale to store data at least for a period of five years.

As NDRS will expose data for reuse by other service providers, therefore, slow or resource-intensive response by the storage layer will result in a slow response by the exposure and subsequently a poor experience for users. It is recommended that high end servers (minimum 32 GB RAM) may be deployed for harvesting metadata from data providers.

4.3.14.2 Selection of Harvesting System

In order to avoid vendor-lock-in and to achieve the neutrality design principle, the harvest layer must be able to harvest data from any institutional repository following OAI-PMH interoperability framework. Currently there are a number of Open Source Projects that offer this functionality. These projects have been evaluated based on certain criteria such as design, technology complexity, community and documentation. Some of the harvester systems evaluated based on these criteria are given below:

4.3.14.3 Open Source Harvester Evaluation

4.3.14.3.1 PKP Harvester (http://pkp.sfu.ca/?q=harvester)

It has been developed as part of the Public Knowledge Project (Canadian government funded research initiative) forms a part the PKP "Open Systems", it harvest and indexes metadata and providing a searchable web based interface to end users. It releases its projects under the open source GNU Public Licence. It is based on PHP implementation of an OAI-PMH Harvester. This application installs very easily with a standard LAMP (Linux, Apache, MySQL and PHP) stack. It works with common PHP (>4.2 including 5), MySQL (>3.23.23) and Apache (1.3.2x and 2.0.4) versions. There are no external dependencies or unique PHP modules required making it a very clean installation. All configuration information is managed easily via a web front end provided by the system.

This project has a solid Object Oriented Architecture based on the MVC Pattern (Model View Controller). With plug-in classes this project is made for extensions. These extensions can be developed parallel to the core so any modification on either side is protected from conflict. Plug-ins, once developed, are copied and pasted into the plug-in directory and then can be switched on and off via the web front end. PKP project also uses a pre-packaged smarty instance for the separation of form from

function. This project has by far the most active community of all the projects. This project has by far the most developed documentation of all the harvester projects

4.3.14.3.2 ARC (A Cross Archive Search Service) (http://sourceforge.net/projects/oaiarc)

It is designed for harvesting OAI compliant repositories and making them available through a unified search engine and is one of the earliest OAI harvesting implementations. It is maintained by the Old Dominion University Library Research Group and released first stable release in 2006. It is a Java implementation of an OAI-PMH Harvester and installs as a servlet on any server running at least JDK1.4 Tomcat 4.1 and MySQL4.1. It is a bit complex to configure ARC software as compared to PKP Harvester and is managed via configuration files for JDBC and application connections. In the future the harvester could be packaged as a module for IR software and deployed as part of an Enterprise Java Project.

ARC is a Java application the system follows the object-oriented paradigm. Java Server Pages are used to separate form from function allowing stylization without affecting the core code. Additional functionality can be added to the application by anybody having knowledge of Java programming. Following release of version 1.0 activity on the site has increased but still falls well short of the PKP community. ARC system documentation is limited to an installation document.

4.3.14.3.3 Celestial (http://eprints.celestial.org)

It is a Perl based implementation of an OAI harvester and installs as a CGI on any CGI-capable web server. It requires MySQL along with common Perl packages and the OAI Perl libraries. It is as complex to configure as ARC software with a number of configuration files as well as some Apache specific configurations. Unlike ARC and PKP Harvester, Celestial doesn't include any search functionality and as such it is merely a data aggregation tool. The core of Celestial consists of a flat structure of

packaged Perl subroutines. Presentation is highly coupled to business logic making stylization difficult without going through the code. As sustainability is surmountable in developing national level repository system, in general, it is felt the architecture of Celestial is not suited to national digital repository system as it requires strong links with the Celestial team. It has no visible community and has limited documentation.

To conclude, these projects have been evaluated based on certain criteria, where the results of the evaluation determine that both PKP and ARC are the most suitable harvester. However, PKP's Public Knowledge Project Hervester2 would be the best for the national digital repository system. This was mostly because the project was capable of accepting extensions to the harvester without modification to the core code.

4.3.15 End User Service

The National Digital Repository System (NDRS) or a linked-repository network is very good opportunity to open services in various sectors such as digitization, IPR, Technical level etc at various levels. Some of them can be outsourced or others can build upon forming various work group.

4.3.15.1 Digitization Services

It provides simple digitization or XML conversion for repositories or content creators who do not have the resources to do this kind of work.

4.3.15.2 Rights and IPR Advisory Services

It has been recommended that work group should be formed at National level so as to fulfill the requirement of the users. .

4.3.15.3 Technical Advisory Services

Institutions are going to require expert advice on the creation and management of their digital assets. NDRS should conduct various workshops on latest technological software such as DSpace, Greenstone etc to make aware technical know-how about these software. DRTC provides technical inputs on IR software through their DLRG List server (dlrg@t drtc.isibang.ac.in). As e-science and e-research grow, this sort of professional help will be critically important to institutions trying to run repositories.

4.3.15.4 Open Access Advisory Services

Institutions are still unclear or uninformed in many cases about Open Access, the concept, the advantages, the ways to provide it, and how it works. It is suggested that a trusted national body should be formed to provide such a service but in the absence of this, individual organizations will do what they can.

4.3.15.5 Repository Construction Services

Many institutions will build their own repositories as it has happened in most cases, but others are now turning to third parties to do this for them as they either do not have the resources or do not wish to employ them in this way. National Informatics Centre (NIC) is helping many institutions to build their repositories such as the official debates of Rajya Sabha (http://rsdebate.nic.in), Supreme Court of India Judgment (http://judis.nic.in) etc. NDRS should provide in building the institutional repositories for the organizations.

4.3.15.6 Hosting Services

Some institutions may not wish to host their own repository and would prefer to outsource this activity. This is already happening and both public sector and commercial organizations have begun to provide the service. NDRS should have mandate to provide domain name services to various organizations.

4.3.15.7 Pre-Aggregator-Level Services

4.3.15.7.1 Metadata Enhancement Services

This kind of service will increase its importance as more repositories start functioning. There is need of automated metadata enhancement processes to be developed which will hugely enrich the metadata and make searching semantically a possibility.

4.3.15.8 Output-Level Services

This is the level that offers the greatest scope and it is at this level that most new services will arise. The list of services which can be provided through national level repository system is given below:

4.3.15.8.1 Resource Discovery Services (including Subject Portals)

Resource discovery services may return results that include material from across the whole corpus of India such as Open Access material or they may focus on providing subject-specific or object-type specific selections.

4.3.15.8.2 Name Authority /Authentication Services

It is required for administrative reasons, services that can provide identification and authentication procedures will be necessary.

4.3.15.8.3 Preservation Services

Institutions may not be in a position to manage the preservation challenges of some of the content that is produced, nor may they wish to accept responsibility for long-term storage anyway. Third party specialists may provide trusted solutions to these issues. They should give certificate to institutional repository as trusted repositories.

4.3.15.8.4 Overlay Journals (a Subset of Publishing)

Some publishers have already developed overlay journals using repository content. Institutions may also develop overlay journals on their repositories. For example, Lund University hosts the Lund Virtual Medical Journal (http://www.lub.lu.se/epubl_2005_Lund/) provides collection of medicine-related articles by Lund authors taken from institutional repository maintained by Lund Library (29).

4.3.15.8.5 Bridging Services

Services that provide information about repositories and their content to other service using the NDRS networked content will be increasingly important. At International level, there do already exist some important bridging services such as ROAR, OpenDOAR, etc. There is scope for more providing overview services, pointing services, current awareness (about repositories) and mapping services related to repository content.

4.3.15.8.6 Citation Analysis/Research Assessment Services (Meta-Analysis Services)

These activities have enormous scope for growth. It is expected that researchers as well funders would keen to know for ways of monitoring the outcomes of their work. It should further augmented by text mining and data-mining.

4.3.15.8.7 Usage Statistics and Feedback Services

Repository managers, institutions and authors themselves are eager for this sort of feedback information which informs their operations and enables them to advocate and educate within their own community.

4.3.16 Interface to Harvest System Protocols with Data Provider

Interoperability is the ability of a system or a product to work with other systems or product without special effort of customer. In case of repository**,** the following protocols can be adopted to harvest metadata from interoperable data repositories which supports interoperability. These are:

4.3.16.1 OAI-PMH

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) version 2.0 was released in 2002. This version is now established as the standard used by the vast majority of open access repositories exposing their metadata for harvesting to service providers. The OAI model contains the concept of data providers (expose metadata) and service providers to harvest from data provider. The Protocol mandates the minimum use of Dublin Core metadata for harvesting, although there are the possibilities of using alternative metadata formats as well mentioned in the protocol document and also in the associated implementation guidelines. All metadata to be harvested must be in XML format though additional metadata formats do not appear to be used so far, except where there is a specific requirement. There is a standard set of 6 OAI-PMH verbs (request syntax) used to send request to data providers. The request is transmitted according to the rules of HTTP over the Web. Data providers receive these requests and reply with appropriate OAI-PMH responses in valid XML format specified by the OAI-PMH protocol, thus it is acting as pull mechanism. These are:

- **Identify** – used to request information from the repository on whether and how it is configured for harvesting using OAI-PMH
- **ListMetadataFormats** – used to request information about the available metadata formats available for harvesting
- **ListSets** – used to request information about the specific sets of records available
- **ListRecords** – used to harvest metadata records
- **ListIdentifiers** – used to harvest just the headers of records rather than the metadata itself
- **GetRecord** – used to retrieve single records using the record's identifier

Through appropriate OAI-PMH verbs, the harvester is thus able to discover what is available for harvesting and subsequently harvest the required records for its own needs. This creates a copy of the metadata in the XML format at the OAI service provider, the harvester acts as aggregator in this instance. The process can be repeated as necessary in order to aggregators are able to access the most up-to-date content. The harvesting process can capture metadata from across a range of repositories and aggregate the results of this into a single collection (30). The various features of OAI-PMH are given as:

4.3.16.1.1 Flow Control (Resumption Token)

OAI-PMH requests may return list of records, headers (unique identifiers) and sets. In some cases data providers send incomplete lists with resumption Tokens so that the service provider can issue further requests to complete the list. Following optional attributes can be added as resumption Token elements – expirationDate, completeListSize, and cursor.

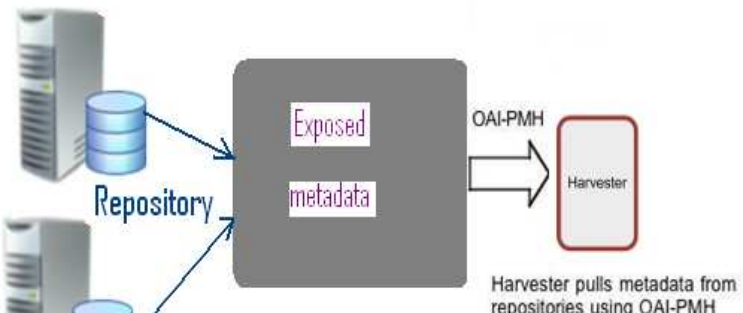

Figure 4.4: Harvesting using OAI-PMH

4.3.16.1.1.2 Errors and Exception Conditions

In case of error or exception conditions, repositories indicates OAI-PMH errors using following Error Codes – badArgument, badResumptionToken, badVerb, cannotDisseminateFormat, idDoesNotExist, noRecordsMatch, noMetadataFormats, noSetHierarchy.

4.3.16.1.1.3 OAI containers

When an Identify request is made to a repository, it can provide additional pieces of information to assist the harvester. These are:

- **rightsManifest** – information about the rights statements that are attached to metadata records within the repository
- **eprints** – a means of providing collection description information about e-print repositories

- **friends** – a means by which a repository can alert a harvester to other repositories that could be harvested
- **branding** – a means by which a repository can convey branding information related to the metadata being harvested
- **gateway** – a means for listing associated gateways through which records can be made available for harvesting.

Containers at the metadata record level can also contain information about the metadata format being used by the repository. The provenance container at record level can contain information about the history of when a record was harvested if implemented.

4.3.16.1.1.4 OAI Selective Harvesting

The Protocol offers three mechanisms by which selective harvesting can take place, rather than a capture of all the metadata records being exposed. These are:

- **By Datestamp** – When the harvester goes back to a repository it has already harvested it needs to know the changes that have taken place rather than reharvest everything again (which can take time depending on the size of the repository. This process is common in established harvesting situations, allowing aggregators to maintain a reasonably up-to-date copy of the relevant information.
- **By Set** – Repositories can allocate metadata records to sets, which may provide additional information about the context of the metadata. Sets can provide organisational granularity or subject classification information. This additional information can be used by end-user services. The level of application of sets in the UK varies considerably at this time. An additional container at the set level can contain a description of the set for further information.

- **By Metadata Dissemination Type** – The metadata format to be harvested has to be specified when carrying out a ListRecords request. This mechanism determines which metadata format the harvester will aggregate and can affect the level of end user service built on top of the harvested metadata.

4.3.16.1.1.5 OAI Static Repositories

There are still number of repositories not adhering to the full OAI-PMH protocol as they do not have the wherewithal to migrate to OAI-PMH compliant IR software. This can be achieved in two ways, first, one may write some middleware programme to convert non-OAI-PMH data into XML format which is in OAI-PMH Complaint. This has been implemented in case of CDS/ISIS based repository; many authors have contributed the source code for OAI-PMH compliant. Greenstone Digital Library (http://www.greenstone.org) is also providing its data in OIA-PMH compliant data which can later be merged with the main repository (31).

Secondly, this is a bit sophisticated option where Open Archive Initiative has made available OAI Static Repositories specification (http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm) open for developers to build gateway for communication between static repository and national data repository system. This is an alternative for small metadata collections (1-5000 records) where it is not possible to configure a repository for full use of OAI-PMH.

4.3.16.2 Harvesting Metadata and Content

4.3.16.2.1 Object Transfer: OA-X

In order to access the content itself from an institutional repository, an extension to the OAI-PMH protocol has been made as OA-X. The OA-X project has extended the OAI-PMH framework to include verbs that allow the collection of digital objects: a

GetObject verb (equivalent to the GetRecord) and a PutObject verb to add an object to the repository. A checksum allows the integrity of the object transfer to be controlled. OA-X is a user innovation that allows OAI-PMH to not only transfer metadata but also the digital objects (32)

4.3.16.2.2 Object Reuse and Exchange (ORE)

Open Archives Initiative Object Reuse and Exchange (OAI-ORE) defines standards for the description and exchange of aggregations of Web resources. These aggregations, sometimes called compound digital objects consisting of text, images, data, and video. The goal of these standards is to expose the rich content in these aggregations to applications that support authoring, deposit, exchange, visualization, reuse, and preservation. Although a motivating use case for the work is the changing nature of scholarship and scholarly communication, and the need for cyber infrastructure to support that scholarship, the intent of the effort is to develop standards that generalize across all web-based information including the increasing popular social networks of "web 2.0 (http://www.openarchives.org/ore/). It is supported by the Mellon Foundation and Microsoft. The project involves major actors in the field of digital repositories development and interoperability, and also from different communities involved with research data (33).

The OAI-ORE standard deals with Dublin Core metadata and associated digital objects. The objective is only to standardize the representation of relations between resources and their components. IR software is providing plug-ins or integrating OAI-ORE along with the repository software.

4.3.16.3 Exposure of Package through OAI-PMH

The purpose of packaging up a compound object is varied. It is primarily a means of transferring materials from one place to another. This could be for transit between two repositories, or for use via an end-user service.

Among packaging standards available, METS has become widely used within the library community as a way of bringing together different metadata records about the same object for transfer and possible delivery as part of resource discovery. The existence of multiple packaging standards inevitably means that for full interoperability between repositories and end-user services the ability to move between and manage multiple standards is valuable (34).

4.3.16.4 RSS/ATOM

RSS (http://en.wikipedia.org/wiki/RSS_file_format) and ATOM (http://www.atomenabled.org/) are widely known as syndication formats to provide news alerts and updates from the repository to the user. A repository offers defined feeds of metadata that individuals subscribe to and access through a browser or desktop tool. Although often commonly perceived and presented as information being pushed to the end-user, they represent an alternative means through which repositories can present their metadata for exposure: the RSS reader then aggregates the metadata by pulling it from the repositories. RSS and ATOM require the repository, or content owner, to take a more active role and clearly lay down what can and cannot be exposed through the respective feeds made available. RSS/ATOM readers select what feeds they wish to receive and aggregate what they are given. Exposing metadata through RSS and ATOM can be considered a more controlled way of exposing metadata for aggregation elsewhere.

4.3.16.5 RSS/ATOM and Content

RSS feeds are intended to be brief, and to connect end-users with greater detail through links back to the main 'repository', however, in certain cases, it is a reference to a web page. However, the level of detail within RSS feeds can be made extensive. As RSS is a common syndication format used by blogs, and it is feasible to include the whole blog post within the RSS feed rather than simply a headline or summary. The ATOM standard provides metadata and also contains content using the same Base64 encoding as MPEG-21 DIDL. As such, it can carry out both syndication and packaging functions. It does offer an alternative means of transferring content for aggregation downstream from the repository, and the potential of using this within end-user services.

4.3.16.6 Web Crawlers

In considering the interfaces that need to be considered by a repository when exposing its metadata and content on open access, it is impossible to ignore the role of web crawlers. Search Engine is another route for a repository's metadata and possibly content to be aggregated and easily exposed to end-users (38).

Exposure of repository metadata and content through web crawlers offers a valuable means of bringing end-users to a repository through commonly used web search engines such as Google and Yahoo!. However, such crawling may not always lead to the level or type of aggregation and exposure that the repository is seeking. It is recommended that the exposure of repository contents within web search engines be examined in closer detail to assess the paths of exposure that exist and the implications for repositories of exposure via this route.

4.3.17 Interface of Aggregation Model (NDRS) to End Users

Aggregation provides a body of metadata - and possibly content - that can be used through end-user services. The main point of access for the end-user to NDRS is likely through a web page. In addition, it may offers number of interfaces to end

users to access the resources of aggregation as listed in table 4.1.

The flexibility of the OAI-PMH aggregation is very apparent, with the ability to feed into many different end-user service scenarios. There are interfaces which have not widely used, for example, the use of OpenURL is largely untapped as an interface, and an assessment of how this might be best used with both repositories and aggregators will be of value. RSS/ATOM aggregations too offer a wide range of options for inclusion within end-user services, though there is scope for tools that allow these aggregations to be re-used beyond simple presentation through a reader.

Web crawler aggregations, predominantly rely on the user going to the web search engine and searching the aggregation at that point. Both OAI-PMH and RSS/ATOM can target the user offers a relative simplicity of access, but fails to take best advantage of the aggregation and the flexibility this offers.

Table 4.1: Options for exposing OAI-PMH aggregations to end-user services

Access point	Remarks
Web interface	Direct access for search and browse enabled through web access onto indexed aggregation. This may involve direct web access or embedding of such access in distributed services elsewhere on the web.
SRW/U	Structured search of an aggregation using distributed search protocols.
RSS/ATOM	The OAI-PMH aggregation can itself by the origin of RSS or ATOM feeds for delivery through the variety of readers available for these standards
OAI-PMH	The OAI-PMH aggregation can itself be harvested for additional aggregation elsewhere

OpenURL	The aggregation can be used as an OpenURL target to facilitate location of individual items
SOAP	A Web services interface that allows the aggregation to be embedded as part of a wider Web services environment
Semantic web	Interfaces that present semantic information about content that can be used to build services upon. Often based on RDF.

4.3.18 Conclusion

Open Archives Initiative, through its basic philosophy of the free-access and a set of technical protocols for the harvesting and aggregation of well-formed descriptive information (commonly called "metadata") about resources. To aggregate the resources, at national level, from Institutional Repository, it requires to harvest metadata (from many "data providers") and develop services (national repository as a "service providers") - an online service that is of value to both the suppliers and users of resources, by connecting one to the other, where access to the resources is not necessarily itself "open". The development of national repository is relatively simple and inexpensive, where discovery services are operating over a pre-selected set of resources – perhaps for a specific discipline. Technical compliance with the protocols for metadata provision and harvesting is not onerous for either data or service providers.

In national repository 'data provider' makes metadata available and 'service provider' collects it from a specific class of data providers. The collection and aggregation of metadata facilitates the development of new publishing initiatives and provide potential gateways to information resources that might otherwise be largely invisible in the over-crowded environment of an ever-growing World Wide Web. The Open Archives Initiative is a second aspect, as a facilitator of institutionally-based resource repositories, particularly encouraging individual Institutions to become the publishers of the work of their own academic staff (and perhaps also their students). Also Open Archives Initiative provides a mechanism for co-operative marketing of the content of these repositories. The thrust of the Open Archives Initiative is to support the development of open access e-print archives,

specifically aimed at the primary research literature published in peer-reviewed journals, on the network in parallel with conventional publication, and thus made available for free access by scholars on a global basis.

Institutional Repositories can be aggregated as a national repository of resources by discipline through metadata aggregation (links the distributed resources (IRs)). The Open Archives Initiative may form an integral part of the “open access” movement for “Free Online Scholarship” (FOS), not only facilitating discovery of resources but also as providing infrastructure for the free and open access to those resources.

The formation of National Digital Repository System (NDRS) in a country is an important step in preservation of data scattered in number of repositories. It is noted that harvesting model is the best option as it offers many more advantages as compared to other models. With advent of interoperability standards such as OAI-PMH inbuilt repository software, the exposure of metadata by data providers is a smooth affair. NDRS can offer many services at various levels to end users.

4.4 National Digital Repository System – Intellectual Property Rights

4.4.1 Introduction

National Digital Repository System provides access to metadata by harvesting and aggregating metadata from data providers, thus helping in discovery of resources. In this model, the resources themselves may be digital or physical; and access to the resources is not necessarily itself “open”. Therefore, there is a different legal relationship between an author in the scholarly environment and their scientific

output worldwide. Majority of the open access movement, in its commercial and economic factors are linked with Intellectual Property, to support their business model. As per Berlin Declaration, open access is the worldwide electronic distribution of peer-reviewed literature and completely free of charge, free of most copyright and licensing restrictions to all scientists, scholars, teachers, students and other curious minds (http://www.zim.mpg.de/openaccess-berlin/berlin_declaration.pdf). It promotes Internet as a functional instrument for a global scientific knowledge base and for human reflection. It specifies measures for research policy makers, research institutions, funding agencies, libraries, archives and museums could consider when disseminating knowledge widely and readily available to society.

4.4.2 Open Access Movement

There are two more initiatives namely Firstly, Budhapest Open Access Initiatives (BOAI) which focus on principle, strategy and commitment, reflecting on how the separate initiatives in the Open Access movement could work together to achieve broader, deeper and faster success (http://www.soros.org/openaccess/read.shtml). Secondly The Bethesda Statement on Open Access Publishing (June 2003) stimulates discussion on how to proceed with goal of providing open access to the primary scientific literature (http://www.earlham.edu/~peters/fos/bethesda.htm).

According to the Berlin Declaration open access contributions must satisfy two conditions:

- The author(s) and right holder(s) of such contributions grant(s) to all users a free, irrevocable, worldwide right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship (community standards will continue to provide the mechanism for enforcement of proper attribution and responsible

use of the published work, as they do now), as well as the right to make small numbers of printed copies for their personal use.

- A complete version of the work and all supplementary materials, including a copy of the permission as stated above, in an appropriate standard electronic format is deposited (and thus published) in at least one online repository using suitable technical standards that is supported and maintained by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving.

Most of the institutional repositories do not fulfill the two above conditions of the Berlin Declaration, that is, they are only complying with the Open Archives Initiative (OAI) protocol for metadata harvesting, which makes them interoperable. It is expected that institutional repositories also act in accordance with the conditions regarding long-term archiving or the requested permission statement to use the work freely and widely. Therefore, missing this permission statement could have raised question that though users can find a work in a repository, yet they don't know whether they can use the work freely or that they are still bound by the restrictions of copyright. The open access movement employs the followings for delivering open access

4.4.2.1 Self-Archiving

It can be described as making e-prints freely available in digital form on the Internet, referred to as the 'green road', publishing an article in a toll access journal and concurrently archiving it in an institutional open access repository. The most common strategies for self-archiving are depositing an article on the author's personal website, in disciplinary archives, in institutional unit archives or in institutional repositories. This strategy must meet the conditions of open access - the material must be searchable and must fulfill the necessary requirements of a publication, such as the determination of ownership, passing peer review, accessibility and preservation (35).

4.4.2.2 Open Access Publishing

Open access publishing is the so-called ‘golden road’, where author in golden road publishes his article in an open access journal that makes the articles freely accessible online immediately upon publication. Open access refers to free and unrestricted availability without any further implications. In scientific publishing it is usual to keep an article's content static and to associate it with a fixed author (35).

4.4.2.3 Open Content

Open content describes thus any kind of creative work, or content, published under an open content license that explicitly allows copying and modifying of its information by anyone, not exclusively by a single organization, firm or individual. This is different from the idea of open access publishing which usually is defined to include the general permission to modify a given work. Even after open access has been achieved, access barriers of censorship, language, handicap and connectivity could hinder accessibility. Suber states nonetheless that there is no reason to hold off using the term open access; ‘Removing price and permission barriers is a significant plateau worth recognizing with a special name’ (36).

As more and more authors deposit research articles in the repository. Although open access should be immediate rather than delayed and should apply to full text, not just to abstracts or summaries (37). Due to restrictions in earlier agreements with publishers it is possible that the full text is only shown campus wide or that only metadata are visible to end users. It could also be possible that provisions of data protection regulations might apply.

Whether or not an author can make his journal article available via the repository depends on the terms and conditions of the agreement that he signed for publishing his article. If they are too strict or when the author has transferred his copyright,

distributing the article is liable to the permission of the publisher. For that reason an author must consider carefully whether to sign the publishing agreement unaltered or to amend the publishing agreement to the extend his right to deposit the article in the repository (38). Author may deposit a preprint for which he doesn't need to consult his publisher. Depending on the copyright policy of his institution an author can decide himself to archive his preprint or postprint through the use of SHERPA/ROMEO project web site (http://www.sherpa.ac.uk/romeo.php). It gives an overview of publishers' policies about copyright and archiving. Instead of using the permission statement as indicated in the Berlin Declaration there are several other means to express the consent to open access. The use of a Creative Commons licence is an easy, effective and increasingly common way to achieve the task (http://creativecommons.org/).

There are around 286 international academic and other bodies as signatories to the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. By doing so an academic institution declares it is in favor of the principle of open access but committing the institution to actually providing open access needs more concrete steps. They should 'Implement a policy that request their researchers to deposit a copy of all their published articles in an open access repository, encourage their researchers to publish their research articles in open access journals where a suitable journal exists' (39).

Individual universities begin to adapt policies requiring that their researcher employees provide open access. The list of policies of institutional archives in the Registry of OA Repository Material Archiving Policies (ROARMAP) (http:// roarmap.eprints.org/) is growing in number.

4.4.3 Attitude of Stake Holders to Intellectual Property Rights for Open Access

4.4.3.1 Role of Funding Organizations

Many funders have now either made commitments to open access, some are in the process of reviewing their policies and procedures, with a view to opening up access to results of the research to enhance access to research publications, especially when such works are financed by public resources and it is often their priority ensuring that the availability and accessibility of the output of research funded by them is not adversely affected by copyright strategies of publishers. Some of the initiatives of funding agencies are given below:

4.4.3.2 European Research Advisory Board

The European Research Advisory Board (EURAB) has recommended to European Commission that "mandating all researchers funded under FP7 to lodge their publications from EC-funded work in an open access repository as soon as possible after publication, to be made openly accessible within six months at the latest" (http://ec.europa.eu/research/eurab/pdf/eurab_scipub_report_recomm_dec06_en.pdf).

4.4.3.3 Wellcome Trust

It requires from its grantees that they submit an electronic copy of the final manuscript of their research papers into PubMed central and expects authors of research papers to maximize the opportunities to make their results available for free and, where possible, to retain their copyright. It made agreements with several big publishers that will allow authors to comply with the requirements of the Trust (http:// www.wellcome.ac.uk/).

4.4.3.4 Research Councils UK (http://www.rcuk.ac.uk/aboutrcuk/default.htm)

Research Councils UK (RCUK) (partnership with eight UK Research Councils) in his position paper insisted in self archiving and author's pay model. It is author's choice as to where to place his research for publication. Further, it is for the author's

institution to decide whether it is prepared to use funds for any pay charges or publishing fees. With regard to self-archiving, RCUK agrees that their funded researchers should deposit the outputs from research councils funded research in an acceptable repository as designated by the individual research council.

4.4.3.5 Medical Research Council (MRC) (http://mrc.ac.uk/)

MRC stated in 2006 that "electronic copies of any research papers accepted for publication in a peer-reviewed journal, which are supported in whole or in part by MRC funding, are deposited at the earliest opportunity – and certainly within six months – in UK PubMed Central (UKPMC)". It encourages authors to deposit their articles in their own institutional repositories also. Further, it also strongly encourages authors to publish in journals that allow them (or their institutions) to retain ownership of the copyright. It will pay article processing charges where these have been included in applications for MRC grant funding.

4.4.3.6 Deutsche Forschungs Gemeinschaft (DFG) (http://www.dfg.de/en/)

The Deutsche Forschungs Gemeinschaft (DFG) tied open access into its funding policy and permanently reserve a non-exclusive right of exploitation for electronic publication of their research results for the purpose of open access. There is a discipline-specific delay of periods, generally six months to twelve months, before which publication of previously published research results in discipline-specific or institutional electronic archives may be prohibited'.

4.4.4 Position of Author and his Relationships

4.4.4.1 Ownership of Scholarly Works

The copyright owner of a work is the maker of the work. However in case of scholarly works there is always ambiguity about the work of the owner. Now it is more or less generally accepted that the copyright of scholarly works is vested in the

author. Also regarding the question in which the moral rights of a scholarly work are vested there is a difference of opinion; some scholars state that the moral rights are vested in the employer, other scholars adhere to the viewpoint that the maker of the work owns the moral rights.

In the United Kingdom Section, 11(2) of the Copyright, Designs and Patents Act 1988 (CDPA) states that the Author of a work is the first owner of any copyright in it (http://www.opsi.gov.uk/acts/acts1988/Ukpga_19880048_en_2.htm#mdiv11). Employee in the course of his employment makes a literary, dramatic, musical or artistic work; his employer is the first owner of any copyright in the work subject to any agreement to the contrary.

SURF developed a policy for the Dutch universities where Copyright in academic publications remains vested in the author (http://www.surffoundation.nl/en/Pages/default.aspx). Author grants the university a license to use the publication for educational or research purposes without claiming any royalties accruing to him. SURF and JISC (2006) investigated as to how universities in the Netherlands and the United Kingdom deal with copyright in terms of their policies and practices especially with respect to the ownership of scholarly works (40).

4.4.4.2 Author-Publisher Relationship

The author-publisher relationship merely determines which rights an author can exercise himself or which he can exercise towards his university. When publishing an article the author agrees with his publisher on the terms and conditions under which his article is going to be published. Therefore an author should identify the rights he may wish to retain. The basic aim here is to create a balance of rights for the stakeholders involved. A publishing agreement can be an important step in achieving this balance of rights and responsibilities in the process of scholarly communication.

4.4.4.2.1 Licence to Publish

A vital component of the copyright toolbox is the 'Licence to publish' (http://www.surf.nl/copyrighttoolbox). In this licence SURF identifies the issues to be addressed when submitting an article to a journal whilst at the same time it will be deposited in an institutional repository. There is likely the interests of authors and publishers often converge, but sometimes they do not. Author grants the publisher a sole licence, allowing for certain copyright related acts which have an economic or commercial objective with respect to the article where author retains certain rights for various scholarly purposes, such as depositing the article in a repository. Licence to publish' makes no distinction between preprints, post-prints or author's version but stipulates that the published version of the author's article can be disseminated via an institutional or centralized repository immediately after publication in a journal or after an embargo period of a maximum of six months.

4.4.4.2.2 Digital Peer Publishing Licence

(http://www.dipp.nrw.de/lizenzen/dppl/index_html/dppl/DPPL_v2_en_06-2004.pdf)

The 'Licence to publish' is a licence concerning publishing in a traditional journal whereas Digital Peer Publishing License is designed for scholarly content which covers the following attributes to the authors and publishers:

- It covers all aspects of authenticity, citation, bibliographic data and metadata, permanent access and open formats.
- The licence can be used either by publishers of e-journals or by the authors themselves.
- DPPL is modular built, where licence is customized for national law, it is internationally applicable, and it covers three modules:
    - Reading, distributing or accessing verbatim copies, sharing and re-using the work and properly citing if changes are made.

All documents covers basic module subjects to being read, accessed for downloading and distributed unchanged. They do not distinct between scientific or commercial use as licence only concerns delivery of the document in electronic format and the rights concerning a printed version or a version on storage media are not covered.

4.4.4.3 Author Addenda

Transfer or assignment of all rights to a publisher could lead to loss of control by the author over his scholarly output. The author may not able to re-use his work now and in the future, he has to seek the publisher permission for publishing his article in a repository. Therefore, institutions and organizations encourage authors to retain their rights. This can also be achieved by adding an author's addendum to a publishing contract. It is a standardized legal instrument that modifies the publishing agreement and allows the author to keep his rights (41). An addendum specifies what rights an author does or does not have in several key areas.

4.4.4.3.1 Author's Addendum from SPARC

Scholarly Publishing and Academic Resources Coalition (SPARC) is an alliance of academic and research libraries and organizations, working to correct market dysfunctions in the scholarly publishing system (http://www.sparceurope.org/). The Author's Addendum made by SPARC is a form which an author can use to amend the publishing agreement supplied by a publisher (http://www.arl.org/sparc/author/docs/AuthorsAddendum2_1.pdf). It has the following points:

- Author in addition to any right under the publishing agreement retains the right to reproduce, distribute, publicly perform, and publicly display the article in any medium for non-commercial purposes, as well as the right to prepare derivative works and the right to authorize others to make any non-commercial use of the article.
- It requires of the publisher to demonstrate consent by signing the copy and send it to the author.

4.4.4.4 Author-Society Relationship

In the author-society relation the author establishes his relationship with the users of his works. This happens by attaching a permission statement to the work. Author can adopt any one of the following options:

- Copyright holders can either compose their own licence or permission statement or use one of the many open content licences.
- Use of one of the Creative Commons licences is an easy way to make clear to society how a work can be used.
- This use is far more limited than the free, irrevocable worldwide right of access granted under the permission statement of the Berlin Declaration.

The author-society relation also covers the relation between author and repository.

- Author and his institution need to make specific arrangements about the works that an author is going to deposit in the institutional repository.
- Establish obligations and rights of both parties in a formal way.
- For each deposit into the repository the author has to give permission to the repository,
    - Firstly to store and preserve the work,
    - Secondly to make it available under set conditions, the latter depending on the existence of a publishing agreement and its terms and conditions.
- Just as a publishing agreement sets the rights and obligations for publishing an article, so does a deposit licence set the conditions for preserving and making available scholarly works?

SHERPA report on deposit licences for e-prints indicates that the majority of deposit licences cover four topics such as ability of the depositor to legally deposit the eprints, the rights the depositor maintains over the deposited work, the permissions

the repository gains to maintain the deposited work and the conditions under which the repository can remove the e-print (42).

4.4.4.5 Creative Commons

The Creative Commons license was developed at Stanford University in 2001. In the core licensing suite is a total of six licenses to choose from, each of which each permits different uses of the work. They are expressed in three different ways as a plain explanation of the licence together with the relevant icons that indicate the scope of the permitted use, the legal document and the machine readable code. There are six major licenses of the Creative Commons (43) as given below:

4.4.4.5.1 Core Suite Creative Commons Licence

- Users are permitted to copy, distribute, display, and build upon the author's work as long as they name the original maker of the work.
- User can make use of the work commercially.

4.4.4.5.2 Attribution Share Alike licence (by-sa)

- Work can be copied, distributed, displayed, and performed as long as the newly created work is licensed under identical terms.
- User must attribute the original author and can use the work commercially.

4.4.4.5.3 Attribution No Derivatives (by-nd)

- Redistribution of verbatim copy of the work commercially and non-commercially under acknowledgment of the creator.
- A user can copy, distribute, display and perform a work non-commercially under an acknowledgement of the creator.

4.4.4.5.4 Attribution Non-Commercial (by-nc) Licence

- It authorizes others to copy, distribute, display and perform the work, and derivative works based upon it- but for noncommercial purposes only.
- The new work must bear the name of the author but it does not have to be distributed under the same terms and conditions.

4.4.4.5.5 Attribution Non-Commercial Share Alike (by-nc-sa) Licence

- Work may remix, tweak, distribute and build upon a work non-commercially with acknowledgement and further licensing under the same terms.
- All new works will carry the same licence and derivatives will be non-commercial.

4.4.4.5.6 Attribution Non-commercial Non-Derivatives (by-nc-nd) Licence

- Users can redistribute the work under attribution of the original author
- It does not allow for the work to be changed in any way

An institution needs to be aware that it cannot attach Creative Commons licence, therefore, institution repository cannot upload the works without the consent of the copyright owner. It is the decision of the copyright owner to decide under which conditions his work can be re-used. In this context even the repository is also a user and therefore it is authorized to exercise the permitted rights. This would mean that no e-deposit licence is needed for the distribution of works, because that is clearly written out in a Creative Commons licence. However, from a managerial and risk-avoiding point of view, an academic institution should perhaps like to specify some rights not covered in the Creative Commons licence. For instance, the Creative Commons licence does not oblige that the repository guarantees the availability of the work in the future or preserves the work digitally.

4.4.5 Attitudes of Publishers of the Primary Literature

Many publishers were opposing open access movement as it is effected their journal subscription and they were looking for viable alternative economic model to fund publication of their journals. These publishers primarily from the not-for-profit sector believe that open access could undermine the economic base of the journals publishing industry, particularly in the quality control of journals.

Later even the publishers started actively participating in the Open Archives Initiatives. Institute of Physics Publishing (IoPP) was making metadata available for harvesting in accordance with the OAI protocol. Indeed, most publishers now appear to believe that distributing their metadata as widely as possible, used for search purposes and believe useful if it brings those seeking for content back to the publisher's own online resources. Publishers understand the significance of compliance with standards in this respect.

In the long run, if the Open Archives Initiative is successful, and open access repositories of resources become an embedded part of the scholarly communication landscape, then it is expected that all publishers will sit up and take proper notice. This could imply the complete change in business model - a switch from payment for demand to payment for supply. The payment to sustain scholarly communication would not make by readers (or their proxies) but by authors (or their proxies (http://www.biomedcentral.com). Author-subsidized publication is always likely to raise questions about the objective quality of the resources in the minds of readers, even though these are entirely unjustified (44, 45).

4.4.6 Secondary Publishers

The publishers of secondary services related to the literature, abstracting and indexing services and similar aggregators of disparate information, seems was commonly overlooked. It was also felt that business model depends simply on gathering and aggregating information, like for example a metadata intrinsic to resources. Secondary publishers not only aggregate information, they add value to

it in quite specific ways. It was felt that value is sufficient to put the secondary publishers at the heart of the scholarly communication process. Secondary publishers may find the OAI metadata harvesting protocol of considerable assistance to them, if metadata becomes available to facilitate access to resources that are otherwise hard to find. They do not see non-commercial OAI service providers as likely to provide serious competition in the short term as creating added value in discovery services is not trivial (46).

4.4.7 Academic Institutions

Academic institutions have a complex relationship with Open Archives. This emanated an idea as to whether Universities should become the publishers of the research output of their own academics (or others)? Many Universities are well established as publishers. There is no reason why they should not continue to be active publishers of academic content in the network era; indeed the relatively low cost of entry into network publishing makes that ambition a relatively easy one to achieve. However, they will be constrained by the law of copyright in exactly the same way as any other publisher is constrained, and will need to be careful with compliance if they are not to be embarrassed by infringement (46).

4.4.8 Conclusion

Intellectual Property is simply overwhelmed by the ease with which content can be copied and redistributed on the network. Those who depend financially on the protection of copyright in creating and disseminating content have to find some other way of earning their living. The future will be a compromise, somewhere in the middle. As open access repositories of scholarly resources are available to all free at the point of use, therefore, it depends on the attitude of authors rather than users. Thus, it is those who create rather than reader control the scholarly communication process.

Readers find the services provided through Open Archives is useful and worthwhile. The only significant model arXiv and, so far at least, the message is that (for this community) Open Archives have proved a useful adjunct to the published literature rather than a replacement for it. ArXiv is heavily used but so is the equivalent paid for research literature. Open Archives (metadata) services or aggregation model

(NDRS) depends on the data providers, and they may have to follow Intellectual Property issues carefully so as to provide the service for a longer period of time.

4.5 National Digital Repository System (NDRS) - Long Term Preservation

4.5.1 Introduction

The growth of institutional repositories and their valuable digital content raises questions as to how to preserve the content for a long term. Digital preservation is a complex process and there are many unsolved organizational, managerial and technical issues that make digital preservation a challenging task for those managing repositories. It has not been embedded as an integral part of the repositories workflow and also there is neither much experience nor commonly agreed practice, as to how content would be preserved in a long term basis. It has to deal with software/hardware migration, physical deterioration of digital media, metadata, user needs and preferences.

Digital preservation is a process which ensures the usability, durability and intellectual integrity of the information contained in the institutional repositories, which in precise is the storage, maintenance, and accessibility of a digital object over the long term. It is viewed as 'series of managed activities necessary to ensure continued access to digital materials' over the years (47).

As per UNESCO (2003) Charter on the Preservation of the Digital Heritage, institutional repositories are also part of the cultural heritage of a country. Therefore, the scientific output of a country should also need to be preserved. Majority of institutional repositories focus more on collecting material, storing and making it accessible to end users. They are yet not clear about how to perform digital preservation (48).

4.5.2 Process of Creation of Digital Objects

Digital material is created with special software, running on a dedicated technical environment, resulting in files with specific characteristics and behavior. A combination of software and hardware is needed to create these digital objects.

The scalability of the hardware, software and whole life cycle of digital object is important, which means it include technology preservation, technology emulation or data migration (49).

4.5.3 Issues Involved in Digital Preservation

Learning materials are becoming part of IR, ensure its preservation and provide access with the changing hardware and software/file format etc. Guarantee of long term preservation helps authors more confidence in the future accessibility and more incentives to deposit the content. There is general lack of practical implementation of preservation standards, technical knowledge, therefore, the attention required on the following issues:

4.5.3.1 File Format

There are hundreds of file formats (either born digital or digitized object is created) and not all formats are suitable for long-term preservation. Software programs, offer the user a range of options that might influence the preservation of the object. The creator or author might not be aware of these risks. The digital repository manager should seek a balance between easy depositing and costly preservation operations. The file format choice can be influenced to a certain level, by limiting the file formats allowed (to open standard-based file formats, whenever possible) or by performing normalization on the digital objects. Guidelines will add to better understanding and raising awareness of the authors (50).

4.5.3.2 File format Registries

Some information is closely related to the digital object and should be stored together with the object that will be preserved, like information about the use of character sets, supplier of the software, maintenance of the software, e environment needed to run the software, necessary tools to render the software, newer versions or perhaps even the successor. Several initiatives have started to store this information in so called file format registries, to be consulted by

everyone, as they are freely available on the internet. One of the examples is PRONOM, an initiative of the National Archives of the UK, describing the file format information of over 130 file formats (51).

4.5.3.3 Preservation Levels

Not every institution has the means and (technical) opportunities to guarantee long-term preservation of its digital objects. It might be that an institution starts with a digital collection, preserves this at a minimal level, and after a certain period hands the content of the repository over to an organization that is well equipped to perform digital preservation. A 'preservation level' enables the institution to show the user of the object to which level the repository was able to take its responsibility; it shows to which level the institution has preserved the objects in the repository. In this way, an institution can await the moment technology offers new opportunities to treat the object correctly according to a higher preservation level. Although there is not a fixed list of preservation levels, there are some widely accepted levels. The basic preservation level is 'bit stream preservation'; raw data are being stored and kept exactly as they have been delivered. Although this requires a well-qualified IT environment, the future user will not have the guarantee that the object is rendered as it was originally, because the object lacks information about the interpretation of the bit stream (52).

4.5.3.4 Persistent Identifiers

While storing a digital document in the repository, it is important to identify this object uniquely, for now and for the long run. Even if the repository moves to another archive, the objects should keep a unique identifier, the 'persistent identifier'. This identifier enables the researchers and repository managers to identify the object, and use it in the scholarly process. Several systems have been developed such as Uniform Resource Name (URN), the Handle system (http://www.handle.net), the Digital Object Identifier (DOI) and the Persistent Uniform Resource Locator (PURL) (51).

4.5.3.5 Migration

Migration is the preservation strategy whereby the digital object is changed to make it accessible in a new environment. Although migration is a widely accepted preservation strategy, there are some drawbacks. Firstly, once started with migration, the repository will need to perform this action again and again over the years. But an error, once introduced with a migration, might enlarge with new migrations, perhaps leading to a damaged or inaccessible object. Another point is that testing of the results is difficult, time consuming and not yet automated. Migration requires thinking about the characteristics of the objects and it is highly unlikely that after migration all characteristics are still available in the new object (53).

The process of migration from one particular file format to a newer version of the same format needs to be repeated whenever a new version is brought out. Research on this type of migration however, showed that more errors appeared when migrating from one version to the next version of software, then if the migration procedure skipped some versions (54).

4.5.3.6 Emulation

Emulation is the process of bringing digital objects back to life in their original environment on top of a different computer environment. This process is carried out by an emulator, which is, as phrased by the Digital Preservation Test-bed ‘a program that runs on one computer and thereby virtually recreates a different computer’. In this definition the word ‘virtual’ denotes that the emulator functions like the original computer, but differs physically. The original computer is called the ‘target platform’; the computer that executes the emulator is called the ‘host platform’.

Emulation can be done at three different levels: the application software level, the system software (operating system) level and the hardware level. Emulating both application and system software requires knowledge of their design and implementation. These products are complex and very often proprietary, which

makes it difficult to emulate. Another issue with application level emulation is that each application requires a specific emulator (55).

Since emulation is proposed as preservation action in the field of digital preservation, however, there is a debate whether it is only solution for a large class of digital objects or it is too complex and costly affairs to follow. It has still a lot of advantages as well disadvantages in following the emulation (55).

4.5.4 Current Scenario in Digital Preservation

4.5.4.1 OAIS Model

Digital preservation has to deal with software/hardware migration, physical deterioration of digital media, metadata, user needs and preferences. The Consultative Committee for Space Data Systems published a Reference Model for an Open Archival Information System (OAIS) (57) that defined the functions needing to be supported by an archive responsible for preserving content on behalf of a designated community. OAIS introduced the concepts of a Submission Information Package (SIP) and a Dissemination Information Package (DIP) as ways of bringing content into or delivering it from an archive, and outlined the kind of information needing to be included: not just files or pointers to them, but also the reference, provenance, context and fixity information needed for preservation management purposes. This has been described in the figure 4.8.

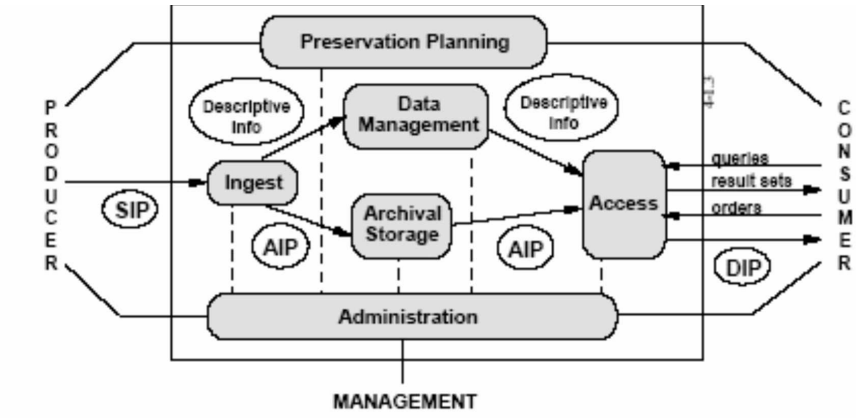

Figure 4.5 OAIS Model.

Reproduced from CCSDC. (2002). Reference model for an Open Archival Information System (OAIS): Recommendation for Space Data System, Blue Book, Washington: CCSDC. Retrieved from CCSDC

website: http://public.ccsds.org/publications/archive/650x0b1.pdf (61)

4.5.4.2 Functions of OAIS Model

OAIS model has the following six functions such as Ingest, Access, Data Management, Archival Storage, Administration and Preservation Planning. A seventh block, that is always present but not part of the figure, is Common Services.

4.5.4.2.1 Ingest

The Ingest function provides the services and functions to cover several activities around the Submission Information Package (the digital object), ultimately resulting in an Archival Information Package (AIP) that is fully prepared to be stored. This means, among other things, that quality assurance checks have been performed and that the necessary descriptive information has been generated to make it possible for users to find the digital object in the archive.

4.5.4.2.2 Archival Storage

This entity provides the services and functions for the storage, maintenance and retrieval of AIPs. Functions like receiving Archival Information Packages from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media on which the holdings are stored, perform routine and special error checking provide disaster recovery capabilities and providing AIPs to Access to fulfill orders.

4.5.4.2.3 Data Management

It provides the services and functions for populating, maintaining and accessing both descriptive Information like descriptive metadata (which facilitates the access to the object) and administrative metadata used to manage the archive. This includes performing queries on the data, providing reports of these queries, performing database updates and administering the archive database functions

4.5.4.2.4 Administration

It provides the services and functions for the overall operation of the archive system. These are:

4.5.4.2.4.1 Access

It provides the services and functions that support Consumers in determining the existence, description, location and availability of information stored in the OAIS, and allowing Consumers to request and receive information products.

4.5.4.2.4.2 Preservation Planning

It provides the services and functions for monitoring the environment of the OAIS and providing recommendations to ensure that the information stored in the OAIS

remains accessible to the Designated User Community over the long term, even if the original computing environment becomes obsolete.

4.5.4.2.4.3 Common Services

It needs no extra description in the OAIS model, as this function is so pervasive, but refers to issues like security, network and operating system services.

4.5.4.2.4.4 The Information Package

As the functions and actors of the OAIS have passed in review, the attention should now go to the subject of the archive: the digital object, or in the OAIS terminology, the 'Information Package'. The "Information Package is a conceptual container of two types of information, called Content Information", which is the actual digital object that the repository wants to preserve, and the "Preservation Description Information (PDI). The Content Information and PDI are viewed as being encapsulated and identifiable by the Packaging Information. The resulting package is viewed as being discoverable by virtue of the Descriptive Information.

This described information package is the subject of the OAIS archive. But in the model of OAIS, the information package gets different names, depending on its role in the functional entities of the archive. As shown in figure 1, the Information Package is called SIP (Submission Information Package) when the Information Package, coming from the producer, is sent to the OAIS archive. Once submitted and archived, it is called an Archival Information Package (AIP). When a member of the Designated Community via descriptive information retrieves one AIP or a set of related AIPs, he will see the Dissemination Information package (DIP). In all cases the information package consists of the digital object but the information of the PDI is also connected.

4.5.5 OAIS Model for Institutional Repositories

An attempt has been made to present the scenario of digital preservation and role of metadata standards such as Metadata Encoding and Transfer System (METS) in the context of institutional repositories. DSpace, institutional repository software has implemented METS Submission Information Package (SIP) and Dissemination Information Package (DIP) as a wrapper for containing descriptive, technical, structural metadata and its associated files.

A number of IRs software are using METS for exchanging metadata as well for preserving metadata to understand the changes happened in the life cycle of document. DSpace is open source IR software which is taking special care in preservation of bit stream of the digital object as well metadata in exporting in METS format.

4.5.5.1 DSpace - METS: Case Study

DSpace (http://www.dspace.org) is an open source software platform that enables organizations to capture and describe digital material using a submission workflow module, or a variety of programmatic ingest options. It distributes an organization's digital assets over the web through a search and retrieval system and preserves digital assets over long term.

Each DSpace site is divided into communities, which can be further divided into sub-communities. Communities contain collections, which are groupings of related content. A collection may appear in more than one community. Each collection is composed of items, which are the basic archival elements of the archive. Each item is owned by one collection. Additionally, an item may appear in additional collections; however every item has one and only one owning collection. Items are further subdivided into named bundles of bit streams. Bit streams are, as the name suggests, streams of bits, usually ordinary computer files. Bit streams that are somehow closely related, for example HTML files and images that compose single HTML document are organized into bundles (58).

Each bitstream is associated with one bitstream format. As preservation service is an important aspect of the DSpace service, it is important to capture the specific formats of files that users submit. In DSpace, a bitstream format is a unique and consistent way to refer to a particular file format. An integral part of a bitstream format is an either implicit or explicit notion of how material in that format can be interpreted. For example, the interpretation for bitstreams encoded in the JPEG standard for still image compression is defined explicitly in the Standard ISO/IEC 10918-1.

Each bitstream format additionally has a support level, indicating how well the hosting institution is likely to be able to preserve content in the format in the future. There are three possible support levels that is, 'Supported', 'Known' and 'Unsupported' that bitstream formats may be assigned by the hosting institution. The host institution should determine the exact meaning of each support level, after careful consideration of costs and requirements.

4.5.5.3 METS Profile

Metadata Encoding and Transmission Standard (METS), a Digital Library Federation initiative, attempts to build upon the work of The Making of America II (MOA2), provides an XML document format for encoding metadata necessary for both management of digital library objects within a repository and exchange of such objects between repositories (or between repositories and their users). Depending on its use, a METS document could be used in the role of Submission Information Package (SIP), Archival Information Package (AIP), or Dissemination Information Package (DIP) within the Open Archival Information System (OAIS) Reference Model (59, 60, 61). A METS document consists of seven major sections:

4.5.5.3.1 METS Header

The METS Header contains metadata describing the METS document itself, including such information as creator, editor, etc.

4.5.5.3.2 Descriptive Metadata

The descriptive metadata section may point to descriptive metadata external to the METS document (e.g., a MARC record in an OPAC or an EAD finding aid maintained on a WWW server), or contain internally embedded descriptive metadata, or both. Multiple instances of both external and internal descriptive metadata may be included in the descriptive metadata section.

4.5.5.3.3 Administrative Metadata

The administrative metadata section provides information regarding how the files were created and stored, intellectual property rights, metadata regarding the original source object from which the digital library object derives, and information regarding the provenance of the files comprising the digital library object (i.e., master/derivative file relationships, and migration/transformation information). As with descriptive metadata, administrative metadata may be either external to the METS document or encoded internally.

4.5.5.3.4 File Section

The file section lists all files containing content which comprise the electronic versions of the digital object. <file> elements may be grouped within <fileGrp> elements, to provide for subdividing the files by object version.

4.5.5.3.5 Structural Map

The structural map is the heart of a METS document. It outlines a hierarchical structure for the digital library object, and links the elements of that structure to content files and metadata that pertain to each element.

4.5.5.3.6 Structural Links

The Structural Links section of METS allows METS creators to record the existence of hyperlinks between nodes in the hierarchy outlined in the Structural Map. This is of particular value in using METS to archive Websites.

4.5.5.3.7 Behavior

A behavior section can be used to associate executable behaviors with content in the METS object. Each behavior within a behavior section has an interface definition element that represents an abstract definition of the set of behaviors represented by a particular behavior section. Each behavior also has a mechanism element which identifies a module of executable code that implements and runs the behaviors defined abstractly by the interface definition.

4.5.5.4 METS Profile in DSpace

METS document represent one DSpace item only. DSpace has defined a set of technical metadata elements for preservation and administration. This metadata includes a unique identifier, check sum, checksum type, mime type, file size, creation date and file path originally assigned to the file. In case, this data exist in the system, it should be included within the METS document. Inclusion of technical metadata should occur in the techMD element and should conform to PREMIS data dictionary. PREMIS is a Preservation Metadata – Implementation Strategies has been developed by working groups formed by OCLC and RLG (62). The working group has produced data dictionary for preservation metadata. It provides a means of encoding the data elements needed for preservation management purposes using the OAIS model as a framework. The data elements are expressed in the form of

four XML schemas – one for objects, one for rights, one for events and one for agents. These can be combined in a PREMIS container as a stand-alone representation of an object. However, PREMIS is mainly concerned with preservation metadata. It does not deal with the description of the object and its structure for discovery purposes or the recording of technical metadata specific to the file format.

DSpace includes a package disseminator and matching ingester for the DSpace METS SIP (Submission Information Package) format. Packagers are software modules that translate between DSpace Item objects and a self-contained external representation, or "package". A package ingester interprets, or ingests the package and creates an Item. A package disseminator writes out the contents of an Item in the package format. A package is typically an archive file such as a Zip or "tar" file, including a manifest document which contains metadata and a description of the package contents. The METS is a typical packaging standard. A package might also be a single document or media file that contains its own metadata, such as a PDF document with embedded descriptive metadata.

The plugin name is METS by default, and it uses MODS for descriptive metadata and PREMIS for preservation metadata. In DSpace, one can invoke command-line tool which gives access to the Packager plugins. It can INGEST a package to create a new DSpace Item, or DISSEMINATE an Item as a package. To see all its options, one may give the command as follows:

[DSPACE]/bin/packager –help //[DSPACE] indicates folder where DSpace has been installed. This mode also displays a list of the names of package ingesters and disseminators that are available.

4.5.5.4.1 Ingesting

To ingest a package from a file, give the following command:

[DSPACE]/bin/packager -e USER -c HANDLE -t PACKAGER PATH

Where USER is the e-mail address of the E-Person under whose authority this runs; HANDLE is the Handle of the collection into which the Item is added, PACKAGER is the plugin name of the package ingester to use, and PATH is the path to the file to ingest (or "-" to read from the standard input).

Here is an example that loads a PDF file with internal metadata as a package:

/dspace/bin/packager -e florey@mit.edu -c 1721.2/13 -t pdf thesis.pdf

4.5.5.4.2 Disseminating

To disseminate an Item as a package, give the command:

[DSPACE]/bin/packager -e USER -d -i HANDLE -t PACKAGER PATH

Where USER is the e-mail address of the E-Person under whose authority this runs; HANDLE is the Handle of the Item to disseminate; PACKAGER is the plugin name of the package disseminator to use; and PATH is the path to the file to create (or "-" to write to the standard output). This example writes an Item out as a METS package in the file "454.zip":

/dspace/bin/packager -e florey@mit.edu -d -i 1721.2/454 -t METS 454.zip

Here is the output of the mets.xml after unzipped 454 files and only header and preservation metadata portion has been taken as an example to display the METS record.

<?xml version="1.0" encoding="utf-8" standalone="no"?>

<mets ID="mets_1" OBJID="hdl:123456789/10" LABEL="DSpace Item" PROFILE="DSpace METS SIP Profile 1.0" xmlns="http://www.loc.gov/METS/"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"

xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd">

<metsHdr CREATEDATE="2008-06-12T17:31:48"> **//Header Section of METS**

<agent ROLE="CUSTODIAN" TYPE="ORGANIZATION">

<name>DSpace at My University</name>

</agent>

</metsHdr>

<amdSec ID="license_5"> **//Administrative Section of METS**

<rightsMD ID="rights_4"> **//Rights Metadata of METS document**

<mdRef LOCTYPE="URL" xlink:type="simple" xlink:href="depositlicense_59.txt"
MDTYPE="OTHER" OTHERMDTYPE="DSpace Deposit License" MIMETYPE="text/plain"/>

</rightsMD>

</amdSec>

<amdSec ID="techMd_for_bitstream_27"> **//Preservation Metadata of METS**

<techMD ID="tech_7">

<mdWrap MDTYPE="PREMIS">

<xmlData xmlns:premis="http://www.loc.gov/standards/premis"
xsi:schemaLocation="http://www.loc.gov/standards/premis
http://www.loc.gov/standards/premis/PREMIS-v1-0.xsd"><premis:premis
xmlns:premis="http://www.loc.gov/standards/premis">

<premis:object>

<premis:objectIdentifier>

<premis:objectIdentifierType>URL</premis:objectIdentifierType>

<premis:objectIdentifierValue>http://localhost:8080/jspui/bitstream/123456789%2F10/27/0
9-02-1997_ISC_B.pdf</premis:objectIdentifierValue>

</premis:objectIdentifier>

<premis:objectCategory>File</premis:objectCategory>

<premis:objectCharacteristics>

<premis:fixity>

<premis:messageDigestAlgorithm>MD5</premis:messageDigestAlgorithm>

<premis:messageDigest>a14a60c98309c8df7bce9116ef31bab8</premis:messageDigest>

</premis:fixity>

<premis:size>2790875</premis:size>

<premis:format>

<premis:formatDesignation>

<premis:formatName>application/pdf</premis:formatName>

</premis:formatDesignation>

</premis:format>

</premis:objectCharacteristics>

<premis:originalName>09-02-1997_ISC_B.pdf</premis:originalName>

</premis:object>

</premis:premis></xmlData>

</mdWrap>

</techMD>

</amdSec>

Figure 4.6 DIPS as taken as an example from DSpace Institutional Repository

4.5.6 **Discussion**

Institution repositories must have policy in terms of content management, rights management, long term preservation policy etc. Peter (63) discovered that about two thirds of Open Access repositories did not have publicly stated policies for the permitted re-use of deposited items and OpenDOAR has provided a tool

(http://opendoar.org/tools/en/policies.php) to IRS manager to provide their IRS policy for such things as submission of items, long term preservation, etc. Various issues relating to digital preservation can be resolved by adopting metadata standard approach. METS, as it has been widely used in institutional repository software can be used as a means of transmitting a representation of an object (physical or digital or partially digital) from one system to another. It can:

- Fully describe the object and its components.
- Encode the metadata needed to aid its preservation and future access.
- Represent the physical and/or logical structure of highly complex objects.
- Represent collections of objects, even where these objects are not stored in the same repository.
- Support a range of submission and dissemination scenarios.

4.5.7 Conclusion

Digital preservation requires new workflow, new skills and close cooperation across different professions ranging from traditional preservation management to computing science. Joint Information System Committee (JISC) (64), RLG-OCLC (53) and OAIS model are providing framework to implement preservation strategy and institutional repositories software a mean to implement these preservation strategy. Since Australian National Library has developed METS profile as submission and dissemination package for integration of various services in Australia (65), it is expected that similar effort at National level in the country may be initiated so that single METS profile keeping in view of local need may be developed or adopting the existing profile as developed by international organizations for the purpose of preserving metadata of digital object and may also be used for the purpose of interoperability across digital repositories.

4.6 National Digital Repository System – Business Model

4.6.1 Introduction

There is a momentum in establishing institutional repositories in the country and it is expected to gather pace further. As the repository network matures, services will develop that both provide leverage for the investment that has been made and offer institutions and end users a growing range of options suitable for their particular needs. These services may have their roots in publicly-funded projects or be newly developed offerings from the commercial sector. The outcome of the repository depends upon the amount of digital contents it attracts or it remain nearly empty; it may raise the profile of an institution becoming one of the best institutional repositories or it may contribute to its obscurity. However, the main purpose of the institutional repositories is not to gain revenues out of it, it is meant for visibility and rapid access to its resources to end users within institutions as well to the worldwide. Therefore, the issue is to what kind of business models one should adopt so that visibility and sustainability of these repositories should be maintained and also at the same time, the national level system which is responsible to harvest data from these repositories

4.6.2 Typology of Business Models for NDRS

There are numbers of attempt made by various authors to develop a typology of business models for web-based businesses (66). In a study on repository services we reduced the extensive lists produced by these authors to a simpler list of five operational models that seemed applicable to repository-related developments (1). These are:

- Institutional Model: institutions own and run the business to further their own goals and strategies;
- Public funded – public bodies sponsor the business for the public good;
- Community model - Members of community invest their own resources to sustain the service.

- Subscription model - It exploits opportunities to sell access to a range of content services. In repository services, subscription models are likely to be targeted at national or regional or individual level repositories in which the services are provided free of cost at the point of use governed by authentication processes. An example of this type of service is Thomson's web citation index which indexes institutional repositories (http://science.thomsonreuters.com/press/2005/8298416/). Institution pays an annual subscription for access to this purpose.
- Commercial Model – Here, the business runs on a commercial basis (other than subscription based): a number of sub-types are covered by this term, for example an advertising model.

All these models seem to be appropriate in the context of digital repositories and all are in use. The institutional model is the one most commonly used for institutional repositories. The community model also applies in some cases where a number of institutions collaborate on a repository. An example of such collaboration is the White Rose consortium comprising the universities of Sheffield, Leeds and York in the UK. The public sponsor model is the one adopted in France, where the HAL (Hyper-Article en Ligne) (http://hal.archives-ouvertes.fr/) repository platform is funded by the Centre for Direct Scientific Communication (Centre pour la communication scientifique directe, CCSD) (http://ccsd.cnrs.fr/) of CNRS, the national science funder. The subscription model is represented, for example, by repositories that lease space or hosting facilities to other institutions that pay annually for the service. Southampton University's School of Electronics & Computer Science (http://www.ecs.soton.ac.uk/) provides this kind of facility. The commercial model is exemplified by repositories that offer additional, one-off, paid-for services such as digitization or the sale of electronic theses. The University of Utrecht (http://www.uu.nl/university/library/EN/igitur/mission/Pages/default.aspx) is providing sale of electronic thesis.

As it is evident that the majority of digital repositories are operating on a non-commercial basis so far, institutional repository manager can think of earning extra revenue by providing expertise and services that others are willing to pay for it.

There are numbers of organizations not willing to take the responsibilities of all tasks associated with digital repositories and will outsource all or part of the work to third party.

4.6.3 Cost of Establishing Network of e-Prints Archive

The annual costs of repository services will vary hugely from service to service. It is difficult to arrive at reliable cost figures at this stage though a few examples may serve to give an in-principle idea of the running costs for a service. It involves setup and running costs in providing repository services to the end users. Swan has given the set up costs and running up costs of the four typical ranging from mid range to a large repositories is given in table 4.2 and a certain level of cost as shown in table 4.3 is also determined by the researcher based on the experience of handling Rajya Sabha official debates (http://rsdebate.nic.in) in Indian context which is one of the largest repositories containing around five lakhs records. It involves hardware, customization software, staff required in enhancing metadata quality and maintenance of server etc. Apart from setting up and present cost, Repository managers need to plan for the possibility of increasing costs in the future involved in software development, , migration to newer systems of system, file formats, increase in content and thereafter enhanced metadata, Development of services for the repository, continued advocacy and marketing for repositories. A repository will need to adapt as technologies, user behaviors and external influences change, and all are likely to change considerably over the medium term.

Table 4.2 Comparative set-up and running costs of sample repositories
reproduced from Swan, A., Needham, P., Probets, S., Muir, A., O'Brien, A., Oppenheim, C., Rachel Hardy, R., Rowland, F. (2004)
http://www.jisc.ac.uk/uploaded_documents/ACF1E88.pdf (34)

Institution Running	Set up costs	Present Costs
MIT (DSpace)	$1.8m grant	Staff $225,000
	3 FTE staff	Operating Costs $25,000
	$400,000 system equipment	Systems equipment $35,000
	Total = $2.4-2.5m	**Annual running costs $285,000**
National University Of Ireland, Maynooth	Grant to hire Computer Science student for set up and customization 6 months	1 FTE staff member for upkeep and maintenance
	Grant for €5,000 for server	
	Total €20,000	**Total €30,000**
Queens Qspace CARL	Software free	
	Server space at Institution	Library staff: $25,000
	Programmer for 12 months: $50,000	ITS Staff: $25,000
	Staff costs for advocacy work with faculty	
	Hardware: $2,065	
	Total Can$52,065	**Total Can$50,000**
SHERPA: Nottingham	Software: Free	Maintenance absorbed within HEI costs: 5 FTE days per annum
	Standard Server: £1,500	Coordination and collection of material £30,000

	Installation 2-5 FTE days £600	3 year update of hardware and software: 2-5 FTE days and £3,900
	Initial customization 15 FTE days £1,800	
	Total £3,900	**Total £33,900**

Table 4.3 Estimation cost of NDRS

Institution	Setup costs	Running cost
National level Institute based in India	Software free (DSpace) Customization software for a period of six months Rs. 10 lakh (outsourced to some agency)	NDRS Manager Rs. 12 lakh
	High end Server Rs. 10 lakh Hosting & Storage 1TB Rs. 7 lakh	Metadata enhancement two library staff: Rs. 5 lakh
	Programmer for 12 months: Rs. 6 lakh	IT staff: Rs. 6 lakh
		Hosting charges Maintenance of software Rs. 12 lakh
		Attending Training/ Conference/Organize Advocacy programmes and Providing technical services to the members. Rs 10 lakh
	Total Rs. 38 lakh	Total Rs. 45 lakh

4.6.4 Adoption of Business Model for NDRS

NDRS has adopted the business model of two national-level repository organizations such as DAREnet (http://www.darenet.org/) and SHERPA (http://www.sherpa.ac.uk/index.html). At the data level, institutions collect, store and retain control over their own intellectual property in digital form. This is important for the institutions and ensures that the provision research content remains the responsibility of the data provider sector. It is in the interest of institutions to provide such access to its own outputs and the national network simply organizes and enables institutions to do this. At the service level, services may be developed at an institution for end users of that institution, or may serve a national audience or even a global one. Some services may aid the ingest process: for example, there may be services that advise on intellectual property issues, on metadata creation and enhancement, on technology, on preservation, or offer repository hosting facilities. SHERPA DP (digital preservation) is an example of such a service; SHERPA's RoMEO (http://www.sherpa.ac.uk/romeo/) provides information on publisher permissions and research funder open access policies respectively, helping repository managers and authors understand their rights and obligations with respect to making their work open access. Other services operate above the data layer and offer things as subject portals, theses collections, or the collected outputs from a particular set of institutions. SHERPA has progressed on successive rounds of public funding through JISC in the UK. Similarly, DAREnet has also been funded over some years by the SURF organization in the Netherlands.

Publishers are using subscription model to sell their journals and online resources for many years. Leona Carpenter expressed that community based model is getting support from the community in terms of ideas, effort or money. Open source software has been able to survive based on strong community support. Open archives could be seen as a similar type of virtual community. White Rose consortium comprising the universities of Sheffield, Leeds and York in the UK is working on community model (66).

4.6.5 Conclusion

Business models are a method for reflecting real world processes and are very much needed for strategic decision maker to plan in considering the future and functioning of their digital repositories in terms of the practicalities of supply and demand, of extending markets or perhaps their strategic goals or mission statement. They will clearly be aware that in order to convert their logistics stream into a revenue stream they will need to put forward a value proposition that will attract customers, funding or sponsorship. As the prime function of establishing national level system is to share knowledge among the end users. It has been suggested that certain service should be supported by public funded money. To conclude, it is suggested that National level system should be publically funded and institutional repository may be supported by respective organization with a provision for providing technical and other related services to member institutions.

4.6.6 References

1. Swan, A., Awre, C. (2006). *Linking UK repositories: Technical and organizational models to support user-oriented services across institutional and other digital repositories: Scoping study report*. Retrieved from JISC website: http://www.jisc.ac.uk/uploaded_documents/Linking_UK_repositories_report.pdf [Accessed on 15th August, 2011]

2. Hauglid, JO. (2004). *User Interfaces for accessing information in digital repositories. Norwegian University of Science and Technology*, Trondheim: Rådsalen, Hovedbygget. Retrieved from NTNU website: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.155.4094 [Accessed on 15th August, 2011]

3. Thomas, C., McDonald RH. (2007). Measuring and Comparing Participation Patterns In Digital Repositories: Repositories by the Numbers, Part 1. *D-Lib Magazine*, *13*(9/10). Retrieved from D-lib website: http://www.dlib.org/dlib/september07/mcdonald/09mcdonald.html [Accessed on 15th August, 2011]

4. *ROARMAP*. (2011). http://roarmap.eprints.org/

5. Swan, A., Needham, P., Probets, S., Muir, A., O'Brien, A., Oppenheim, C., Hardy, R., Rowland, F. (2005). *Delivery, Management and Access Model for E-prints and Open Access Journals within Further and Higher Education.* Retrieved from University of Southampton website: http://eprints.ecs.soton.ac.uk/11001/1/E-prints_delivery_model.pdf

6. Purushothama, GM., Bhandi, MK. (2006). *Metadata Harvesting and the Open Archives Initiative*, In: 4th International Convention CALIBER-2006, Gulbarga, 2-4 February, 2006. Retrieved from Shodhganga website: http://shodhganga.inflibnet.ac.in/dxml/bitstream/handle/1944/535/7(cal%2006).pdf?sequence=1 [Accessed on 15th August, 2011]

7. Lynch, A. (2001). Metadata Harvesting and the Open Archives Initiative, *ARL Monthly Report*, 217, Retrieved from ARL website: http://www.arl.org/resources/pubs/br/br217/br217mhp.shtml [Accessed on 15th August, 2011]

8. Swan, A., Needham, P., Probets, S., Muir, A., O'Brien, A., Oppenheim, C., Rachel Hardy, R., Rowland, F. (2004). Delivery, Management and Access Model for E-prints and Open Access Journals within Further and Higher Education,
Cranfield: EPIC. Retrieved from JISC website: http://www.jisc.ac.uk/uploaded_documents/ACF1E88.pdf [Accessed on 15th August, 2011]

9. Robinson, M. (2007). *Institutional Repositories: Staff and Skills requirements*. Retrieved from SHERPA web site: http://www.sherpa.ac.uk/documents/sherpaplusdocs/notts-Repository%20Staff%20and%20Skills.pdf [Accessed on 15th August, 2011]

10. Day, M. (2003). *Prospects for institutional e-print repositories in the United Kingdom: ePrints UK supporting study, no. 1.* Retrieved from RDN website: http://www.rdn.ac.uk/projects/eprints-uk/docs/studies/impact [Accessed on 15th August, 2011]

11. Carr, L. (2006). *Use of navigational tools in a repository.* Retrieved from University of Southampton website: http://www.ecs.soton.ac.uk/~harnad/Hypermail/Amsci/5170.html [Accessed on 15th August, 2011]

12. Ferreira, M., Baptista, AA. (2005). *Use of Taxonomies as a way to achieve Interoperability and improved Resource Discovery in DSpace-based Repositories*. Retrieved from Citeseer website: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.121.3788 [Accessed on 15th August, 2011]

13. Halbert, M., Butler, J., Farley, L., Furlough, M., Sandore, B., Walsh, J., Milewicz, L. (2006). *DLF-Aquifer Services Institutional Survey report: Aquifer Services Working Group report on the Institutional User-Services Survey results.* Washington DC: Digital Library Federation. Retrieved from DLF website: http://www.diglib.org/aquifer/SWGisrfinal.pdf [Accessed on 15th August, 2011]

14. Swan, A., Brown, S. (2005). Open Access self-archiving: pp1-104. An author study. Published by the JISC. Retrieved from website: http://www.keyperspectives.co.uk/openaccessarchive/reports/Open%20Access%20II%20(author%20survey%20on%20self%20archiving)%202005.pdf [Accessed on 15th August, 2011]

15. Dahl, M., Banerjee, K., Spalti M. (2006). *Digital Libraries: Integrating Content and Systems.* Oxford: Chandos Publishing.

16. *SWORD.* (2011). Retrieved from JISC Info Net Website: http://www.jiscinfonet.ac.uk/infokits/repositories/technical-framework/sword

17. Sparks, S. (2005). *JISC Disciplinary Differences Report.* Retrieved from JISC website: http://www.jisc.ac.uk/Uploaded_Documents/Disciplinary%20Differences%20and%20Needs.doc [Accessed on 15th August, 2011]

18. Baruch, P. (2007). Open Access Developments in France: the HAL Open Archives System Learned *Publishing*, *20*(4), 267-282. Retrieved from HAL website: http://hal.archives-ouvertes.fr/docs/00/17/64/28/PDF/OA_Developments_in_France_final.pdf [Accessed on 15th August, 2011]

19. Martin, R. (2003). ePrints UK: Developing a national e-prints archive. *Ariadne*, 35. Retrieved from Ariadne website: http://www.ariadne.ac.uk/issue35/martin/ [Accessed on 15th August, 2011]

20. Peters, D., Lossau N. (2011). DRIVER: building a sustainable infrastructure for global repositories, *Electronic Library*, *29(2)*, 249–260.

21. *The CARL Institutional Repositories Program: Vision, Priorities, and Projects* (2011). Retrieved from CARL website: http://www.carl-

abrc.ca/projects/institutional_repositories/member_resources-e.html [Accessed on 15th August, 2011]

22. *OAIster.* (2011). Retrieved from OCLC website: http://www.oclc.org/support/documentation/firstsearch/databases/dbdetails/details/OAIster.htm [Accessed on 15th August, 2011]

23. McKiernan, G. (2003). Open archives initiative service providers. Part I: science and technology, *Library Hi Tech News*, 20(9).

24. *Scientific Commons: A Community for Scientific Information.* Retrieved from Open Education Resources web site: http://openeducation.zunia.org/post/scientific-commons-a-community-for-scientific-information/ [Accessed on 15th August, 2011]

25. Pieper, D., Summann, F. (2006). Bielefeld Academic Search Engine (BASE): An end-user oriented institutional repository search service", *Library Hi Tech*, *24*(46), 14 – 6.

26. Armbruster, C., Romary, L. (2009). *Comparing Repository Types: Challenges and Barriers for Subject-Based Repositories, Research Repositories, National Repository Systems and Institutional Repositories in Serving Scholarly Communication.* Retrieved from SSRN website: http://ssrn.com/abstract=1506905 [Accessed on 15th August, 2011]

27. Thornley, K. (2006). *OARINZ Project: Technical Definition of the Harvester/Hub. September* New Zealand: CPIT.

28. Thornley, K. (2006). *Institutional Repository Network Blueprint.* New Zealand: CPIT.

29. Özek, YH. (2005). Lund Virtual Medical Journal Makes Self-Archiving Attractive and Easy for Authors. *D-Lib Magazine*, *11*(10). Retrieved from D-Lib website: http://www.dlib.org/dlib/october05/ozek/10ozek.html [Accessed on 15th August, 2011]

30. Prasad, ARD. , Guha, N. (2005). Interoperability and the OAI-PMH In: *DRTC-HP International Workshop on Building Digital Libraries using DSpace*, DRTC, Bangalore from 7-11 March, 2005. Retrieved from LDL, DRTC website: http://drtc.isibang.ac.in:8080/xmlui/handle/1849/245 [Accessed on 15th August, 2011]

31. Stefka, K., Betti, GL, Castellani, F., Keizer, J. (2008). Achieving OAI PMH compliancy for CDS/ISIS databases, Electronic *Library*, 26.

32. Foulonneau, M., André, F. (2008). Investigative Study of Standards for Digital Repositories and Related Services, Amsterdam: Amsterdam University Press. Retrieved from website: http://dare.uva.nl/document/93727 [Accessed on 15th August, 2011]

33. Tarrant, D., O'Steen, B., Brody, T., Hitchcock, S., Jefferies N., Carr L. (2009). Using OAI-ORI to Transform Digital Libraries into Interoperable Storage and Service Applications. *Code{4}lib Journal*, 6 Retrieved from Code{4}lib Journal website: http://journal.code4lib.org/articles/1062 [Accessed on 15th August, 2011]

34. Swan, A., Needham, P., Probets, S., Muir, A., Oppenheim, C., O'Brien, A., Hardy, R. and Rowland, F. (2005). *Delivery, Management and Access Model for E-prints and Open Access Journals within Further and Higher Education: Technical Report*. Retrieved from University of Southampton website: http://eprints.ecs.soton.ac.uk/11001/ [Accessed on 15th August, 2011]

35. Harnad, S., Tim B, François V., Carr, L., Hitchcock, S., Gingras, Y., …, Hiff, E. (2004). The Access/Impact Problem and the Green and Gold Roads to Open Access, *Serials review 30*/4. Retrieved from University of Southampton website: http://eprints.ecs.soton.ac.uk/10209/ [Accessed on 15th August, 2011]

36. Musakali, JJ., Rotich, RC. (2009). Open Access in African Publishing Industry: Opportunities and Challenges In *KMAfrica 2009 Dakar Conference Paper* Retrieved from http://www.kmafrica.com/?q=node/153 [Accessed on 15th August, 2011]

37. Suber, P. (2002). *Open Access Overview*. Retrieved from SPARC Open Access Newsletter website: http://www.earlham.edu/~peters/fos/overview.htm. [Accessed on 15th August, 2011]

38. Pappalardo, KM., Fitzgerald, AM., Fitzgerald, BF., Kiel-Chisholm, SD., O'Brien, D., Auston, A. (2007). *A Guide to Developing Open Access through Your Digital Repository.* Retrieved from http://eprints.qut.edu.au/9671/1/9671.pdf [Accessed on 15th August, 2011]

39. Weenink, K., Waaijers, L., Godtsenhoven KV. [Eds] (2007). *A DRIVER'S Guide to European Repositories*, Amsterdam: Amsterdam University Press. Retrieved from http://dare.uva.nl/document/93898 [Accessed on 15th August, 2011]

40. Mossink, W., Ralph W. (2006). *Report on institutional copyright policies in the Netherlands & the UK.* Retrieved from SURF web site: http://www.surffoundation.nl/Auteursrechten/nl/landschap/relaties/auteurinstelling/Documents/Final%20OZ%2006%206795%20Report%20on%20Institutional%20Copyrights%20Policies%20in%20the%20Netherlands%20%20the%20UK.pdf [Accessed on 15th August, 2011]

41. Peter H. (2006). Author Addenda: An examination of five alternatives, D-*Lib Magazine 12* /11. Retrieved from D-Lib website: http://www.dlib.org/dlib/november06/hirtle/11hirtle.html [Accessed on 15th August, 2011]

42. Knight, G. (2004). *Report on a deposit licence for E-prints.* Retrieved from SHERPA website: http://www.sherpa.ac.uk/documents/D4-2_Report_on_a_deposit_licence_for_E-prints.pdf [Accessed on 15th August, 2011]

43. Creative Commons (2011). Retrieved from Wikipedia web site: http://en.wikipedia.org/wiki/cite_note-CCLicenses1-14 [Accessed on 15th August, 2011]

44. *Harnad, S (2002). Association of Learned and Professional Publishers (ALPSP) Statement on BOAI,* Retrieved from website: http://users.ecs.soton.ac.uk/harnad/Hypermail/Amsci/1861.html

45. *American Scientist Forum email list* (September *1998)* http://listserver.sigmaxi.org/sc/wa.exe?A0=september98-forum&D=0&F=l [Accessed on 15th August, 2011]

46. *Bide, M. (2002). Open Archives and Intellectual Property: incompatible world views?* Retrieved from OAF website: http://www.oaforum.org/otherfiles/oaf_d42_cser1_bide.pdf [Accessed on 15th August, 2011]

47. Jones, M., Neil B. (2008). The Preservation Management of Digital Material Handbook [online text], UK: Digital Preservation Coalition. Retrieved from the website: http://www.dpconline.org/graphics/handbook/ [Accessed on 15th August, 2011]

48. *UNESCO Portal.* (2011). Retrieved from UNESCO website: http://portal.unesco.org/ci/en/ev.php-RRL_ID=13366&URL_DO=DO_TOPIC&URL_SECTION=201.html. [Accessed on 15th August, 2011]

49. Saha, B. (2006). *Digital Archiving and Preservation: Present Scenario*. In *4th Convention PLANNER -2006*, Mizoram Univ., Aizawl, 09-10 November, 2006. Retrieved from Shodhganga website: http://shodhganga.inflibnet.ac.in/dxml/bitstream/handle/1944/1258/265-274.pdf?sequence=1 [Accessed on 15th August, 2011]

50. Hockx-Yu, H. (2006). *Digital preservation in the context of institutional repositories*. Retrieved from E-LIS website: http://eprints.rclis.org/archive/00007351/ [Accessed on 15th August, 2011]

51. Weenink K., Waaijers, L., Godtsenhoven KV. [Eds.] (2008). *A DRIVER's Guide to European Repositories*, Amsterdam: Amsterdam University Press. Retrieved from DRIVER website: http://www.driver-repository.eu/PublicDocs/D7.2_1.1.pdf [Accessed on 15th August, 2011]

52. Bradley, K. (2006). *Digital Sustainability and Digital Repositories.* Retrieved from website: http://unpan1.un.org/intradoc/groups/public/documents/APCITY/UNPAN023378.pdf [Accessed on 15th August, 2011]

53. *Trusted Digital Repositories: Attributes and Responsibilities* (2002), California: Research Library Group (RLG).

54. Mellor, P., Paul, W., Derek, S. (2002).Migration on Request: A practical technique for digital preservation. Retrieved from Camileon website: http://www.si.umich.edu/CAMILEON/reports/reports.html [Accessed on 15th August, 2011]

55. Dutch Archives Testbed Project. (2003). *Digital Preservation Testbed White Paper Emulation: context and current status*, Retrieved from the website: http://www.digitaleduurzaamheid.nl/bibliotheek/docs/white_paper_emulatie_EN.pdf [Accessed on 15th August, 2011]

56. *Digital preservation through emulation* (2006). Retrieved from Digitization 101 website: http://hurstassociates.blogspot.com/2006/11/digital-preservation-through-emulation.html [Accessed on 15th August, 2011]

57. CCSDC (2002). *Reference model for an Open Archival Information System (OAIS): Recommendation for Space Data System, Blue Book*, Washington: CCSDC. Retrieved from CCSDC website: http://public.ccsds.org/publications/archive/650x0b1.pdf

58. Tansley, R., Bass, M., Stuve, D., Branschofsky, M., Chudnov, D., McClellan, G., Smith M. (2003). *The DSpace Institutional Digital Repository System: Current Functionality*.

In: Proceedings of IEEE Joint Conference on Digital Libraries,87-97.

59. Metadata encoding and transmission standard: primer and reference manual. (2011). Retrieved from Library of Congress website: http://www.loc.gov/standards/mets/METS%20Documentation%20final%20070930%20msw.pdf. [Accessed on 15th August, 2011]

60. METS: an introduction and tutorial. (2011). Retrieved from Library of Congress website: http://www.loc.gov/standards/mets/METSoverview.v2.html [Accessed on 15th August, 2011].

61. Bell, J., Lewis S. (2006). Using OAI-PMH and METS for exporting metadata and digital objects between repositories. *Program*, *40*(3), 268-276.

62. *PREMIS data dictionary for preservation metadata*. (2011). Retrieved from Library of Congress website: http://www.loc.gov/standards/premis. [Accessed on 15th August, 2011].

63. Peter, M. (2006). *Moving forward with the OpenDOAR Directory* In 8th International Conference on Current Research Information Systems, Bergen, and 11th - 13th May 2006. Retrieved from SHERPA website: http://opendoar.nottingham.ac.uk/documents/BergenPresentation20060512Handouts.ppt [Accessed on 15th August, 2011]

64. Beagrie N. (2004). The continuing access and digital preservation for the UK Joint Information System Committee (JISC). *D-Lib Magazine, 10*(7/8). http://www.dlib.org/dlib/july04/beagrie/07beagrie.html [Accessed on 15th August, 2011].

65. *Australian METS profile.* (2011). Retrieved from National Library of Australia website: http://www.nla.gov.au/australianmetsprofile. [Accessed on 15th August, 2011]

66. Leona Carpenter, L., Castelli, D., Day, M., Heery, R., Hunter, P., Nicholson, D., Nixon, W., Child, P., Gadd, E. (2003). *Final review of organisational issues*, Bath: University of UKOLON. Retrieved from OAF website: http://www.oaforum.org/otherfiles/oaf_d32_organisational2.pdf [Accessed on 15th August, 2011]