" KNOWLEDGE LIBRARIAN"
AN INTERNATIONAL PEER REVIEWED BILINGUAL E-JOURNAL OF LIBRARY AND INFORMATION SCIENCE
Volume: 02, Issue: 05, Sept. – Oct. 2015 *eISSN NO. 2394-2479*
Impact Factor (IIFS) - 0.331

A PREMIER GUIDE OF SobekCM FOR THE LIBRARIANS

Ali Akbar Petiwala

Librarian

RBK International School

Mumbai , Maharashtra, India.

ABSTRACT

The chapter describes basic concept of Institutional Repository (IR) and it's installation on Microsoft Windows® 7 Professional, Enterprise, or Windows® 7 Ultimate. With time computers have progressed and developed and so have the operating systems (OS). Multiprocessing & Multitasking OS such as Microsoft Windows 7, 8(Commercial) and Linux (freely distributed) are widely used. Mostly Open Source Software (OSS) is designed to run on Linux because of the availability of pre-defined software freely available on Linux. But sometimes unfriendly Graphical User Interface (GUI) and concept of terminal based query, user finds it difficult to operate certain software on Linux. So after a pilot study, an attempt was made of introducing an OSS SobekCM based on Windows OS for IR.

Keywords: Digital Library, Institutional Repository, Digitization, Open Source Software (OSS), SobekCM

1. INTRODUCTION

Institutional Repository is a concept for collecting, managing, publishing and preserving scholarly works created in digital form by faculty, students and researcher in individual universities, colleges and even schools. It is totally digital in nature. Apart from intellectual output, administrative documents and teaching materials coming out of the normal academic life may also be a part of IR. A single organization or multiple institutes constituting an organization can setup an IR. Its access can be open to all members or certain limitations can be defined by policy.

IR is a source for managing intellectual resources properly. Research material is hosted and managed on an IR server using special software like DSpace, Greenstone, Fedora and E-Print etc. can be accessed on the LAN/intranet + Internet/private network.

2. BENEFITS OF IR

- Proper dissemination of resources leads for the long term preservation of scholarly work.
- Authors are benefited by acting as contributor.
- Institution of course can act as a sponsoring body.
- User can browse and search the documents in IR. Users can access the full/partial text as defined by policy. In case of partial access, users can avail 'Request a copy' service.

- Grey literature like teaching materials, unpublished documents can be accessed from IR.

3. WHAT TO UPLOAD?

- Published Research Material like Journal articles, Book chapters, Conference papers etc.
- Unpublished Research Material like preprints, working papers, Thesis/dissertations, technical reports, progress/status reports, and committee reports presentations, teaching materials, audio/video clips, audio-books, e-books etc.

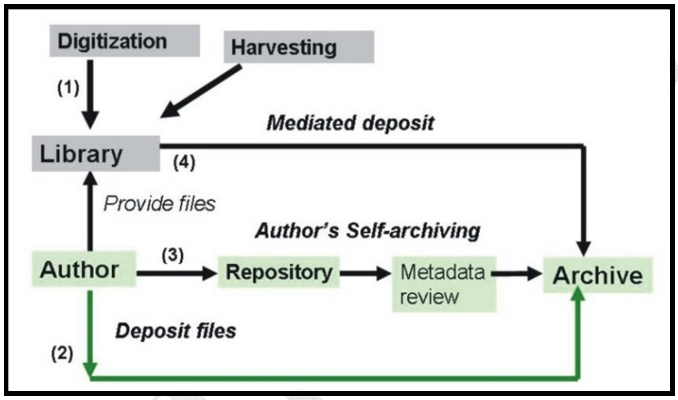

Figure: 1 *–Explains the workflow of IR works and process of archiving*

4. SETTING UP IR USING SOBEKCM

SobekCM is the IR software engine which powers digital repositories by allowing users to discover online resources via semantic and full-text searches. For each digital resource in the repository there are a plethora of display options selected by appropriately authenticated use. It includes online metadata editing and online submission to support institutional repositories. More than just a digital repository, it's also a workflow management tool. It supports nearly all file formats for upload.

SobekCM provides a single solution for universities, colleges, museums, archives, and other heritage organizations to support the discoverability and access to digital resources and facilitate the creation of digital scholarly works. SobekCM comes with unrivalled file and metadata support. It has built-in viewers for all standard resource types from books, slides, and photographs to newspapers and electronic theses and dissertations. Multiple files are accepted without any restrictions. This flexibility, along with an innovative set of ways to hierarchically organize resources, allows complete control and customization of user's experience.

In addition to excellent file type support, SobekCM offers the most comprehensive support for different metadata types of any hosted solution. While the system is natively built upon MARC library records and METS/MODS digital resource files, it can also support EAD's for archival collections, VRACore for visual resources, Darwin Core for taxonomic materials, GML for mapping, IEEE-LOM for learning objects and much more.

SobekCM software was developed at the George A. Smathers Library, University of Florida by a team of Mark Sullivan with on-going effort spanning several departments. Today it is one of the housing content for more than hundred institutions. It is released as OSS under the GNU GPL license and can be downloaded from the SobekCM Software Download Site.

Year	Accomplishment
2006	SobekCM First Released • Display layer over Greenstone Digital Library • Written in C#, served by Windows IIS • Based on MODS/METS
2011	Version 3.0 Released • Second major rewrite • No longer dependent on Greenstone Digital Library • Integrated tracking and workflow • SobekCM Released as Open Source
2013	Version 4.0 Released • HTML5 / CSS3 • Online Quality Control
2014	*Sobek Digital Hosting & Consulting* created to offer hosted solution

***Figure: 2** - Brief History of SobekCM Development*

5. Installing SobekCM on Windows 7(Widely used Microsoft OS)

5.1 Basic Server Specifications

Since the digital repository was written in C# predominantly, it only operates under a Microsoft Windows / IIS environment to date. None of the hardware for this system needs to be particularly robust. It is recommended to have at least 4GB of memory and 80GB of disk space on the web server. Hard drive space needed for this is mostly dependent on how much full text will be uploaded.

5.2 Pre-requisite Software

Listed below is the pre-requisites software which is needed to be installed before running SobekCM setup file.

a. .NET Framework Version 4.5

.Net is a new framework for developing web-based and windows-based applications within the Microsoft environment. It's a programming infrastructure created by Microsoft for

building, deploying, and running applications and services that use .NET technologies, such as desktop applications and Web services.

Link to download: ***https://www.microsoft.com/en-in/download/details.aspx?id=30653***

b. Microsoft SQL Server with Full Text Searching

Microsoft SQL Server is a relational database management system (RDBMS) from Microsoft that's designed for the enterprise environment. SQL Server runs on T-SQL (Transact -SQL), a set of programming extensions from Sybase and Microsoft that add several features to standard SQL, including transaction control, exception and error handling, row processing, and declared variables. The product is said to provide enhanced flexibility, scalability, reliability, and security to database applications, and to make them easier to create and deploy, thus reducing the complexity and tedium involved in database management. SQL Server also includes more administrative support. The full-text searching components of the database must also be installed. The free version of Microsoft SQL (Express) is sufficient for smaller libraries. Ex. Microsoft SQL Server2008 R2 RTM - Express with Advanced Services.

Link to download: ***http://www.microsoft.com/en-us/download/details.aspx?id=25174***

c. Internet Information Services (IIS)

IIS 7, 7.5 provides a redesigned World Wide Web Publishing Service architecture that can help to achieve better performance, reliability, scalability, and security for web sites. IIS is one of the most powerful web servers provided by Microsoft that is able to host and run web applications. It supports the following protocols: FTP, FTPS, SMTP, NNTP, HTTP/HTTPS. Web sites can be hosted on IIS; or can be used as an FTP site also. By default, IIS is not installed on Windows® 7 Professional, Enterprise, or Windows® 7 Ultimate. IIS can be install by clicking on Windows Features in Advanced Options under Programs in Control Panel. The web application requires IIS 7 or 7.5 to be installed on web server. This is included in most Window machines.

d. Java Runtime Environment (JRE)

The Java Runtime Environment (JRE), also known as Java Runtime, a set of programming tools for developing Java applications. The Java Runtime Environment provides the minimum requirements for executing a Java application; it consists of the Java Virtual Machine (JVM), core classes, and supporting files. Latest Oracle's Java Runtime Environment is a must-have installation. It's required to run several Web-based services and such software like Apache-Tomcat.

Link to download: http://www.oracle.com/technetwork/java/javase/downloads/jre7-downloads-1880261.html

e. Apache Tomcat

Apache Tomcat is an open source software implementation of the Java Servlet and Java Server Pages technologies. Tomcat, or officially named Apache Tomcat is a light-weight web container used for deploying and running web application based on Java. Just like apache for PHP as container, tomcat's role is becoming a runtime for java web application. There are so many web containers in the world but I choose tomcat as this tutorial since tomcat is quite easy to be understood and the server management is not that difficult. So for beginner, tomcat is a good kick-start to a real application server. For development, tomcat is a best friend since it has a very quick start-up and it's reliable for small number of applications.

(Note: Apache Tomcat requires the installation of the latest Java JRE.)

Link to download: ***http://www.us.apache.org/dist/tomcat/tomcat-7/v7.0.64/bin/apache-tomcat-7.0.64.exe***

f. SobekCM

SobekCM (alternately Sobek and SobekCM) is an Open Source software engine and suite of associated tools for digital libraries and digital repositories for galleries, libraries, archives, museums, colleges, universities, scholarly research projects as with the Digital Humanities, Research Data Collections, and more. SobekCM is named in part for the Egyptian crocodile god, Sobek. SobekCM's name comes in part of the mythical Sobek's role as a collector of

lost things, or "the idea that as a crocodile, Sobek is the best suited to collecting items upon the Nile."

Link to download: http://sobekrepository.org/software/download

- ✓ *Please note that for security reasons, it's advisable to make at least one service/user account with password. At the time of installation, keep it handy.*

5.3 INSTALLATION STEPS

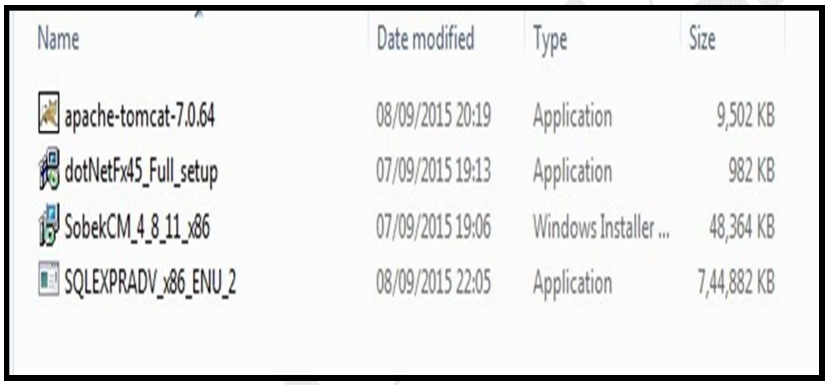

***Figure: 3** - Before attempting installation please check that all files are downloaded and saved in one folder as shown above.*

Step 1 - JRE

- After downloading the required software from the above given link.
- After the download is complete, click Run again.
- When the User Account Control Screen comes up, hit Continue.
- Click Next.
- Accept the license agreement by clicking Install.
- JRE is now installed.

Step 2 - Microsoft .NET 4.5 Framework

- After downloading the required software from the above given link.
- After the download is complete, run the installer setup file.
- It will start downloading in-built required files from internet.
- Follow the on-screen instruction.
- .NET 4.5 is now installed.

Step 3 - Microsoft IIS

- Go to [Start/Control Panel/Programs and Features]
- Click in the left panel on "Turn Windows features on or off"
- Expand "Internet Information Services" and enable these options:
- Web Management Tools, IIS6 Management Compatibility, check "II6 Scripting tools"
- Web Management Tools, IIS Management console
- Web Management Tools, IIS Management Scripts and Tools
- Web Management Tools, IIS Management Service
- At 'World Wide Web Services/ Application Development Features', enable :
- .NET Extensibility
- ASP
- ASP.NET
- At 'World Wide Web Services/Common Http features', enable:

- Static Content
- Click "OK" and let Windows update. This may take some time.

Step 4 - Apache Tomcat

- Once the setup file is downloaded from the above given link, double click on "apache-tomcat-7.0.64.exe" and follow the on-screen instructions.
- While being prompted for choosing destination directory, choose the default path.
- Follow the default on-screen instruction until the installation is finished
- (Note: Apache Tomcat requires the installation of the latest Java JRE.)

Step 5 - Microsoft SQL Server 2008 R2 Express Edition

- Double click on the installer. It will take some time to extract. After that a window will open as follow.
- From this window click on the Installation menu (See Left). Now there will be appear some option under Installation. From right, click on the option New SQL Server stand-alone installation or add features to an existing installation.
- Note: After Clicking on that a window may appear as bellow. Just click on the button Run Program of that.
- Now setup will take a moment to complete some operation. To continue there cannot be any operation with status failed. If any operation fails then it will not be possible to continue installation. Click OK

- Check accept the terms and click NEXT
- Now Click Install button. Some setup support files will be installed.
- Now again it needs some operation to be completed. Every operation must be passed or warning or skipped. If any operation fails then it is not possible to continue. Click Next.
- Now select the features which is needed. Here I am selecting Database Engine Services, Full-Text Search Business Intelligence Development Studio, Management Tools basic & SQL l Client Connectivity SDK. Click Next.
- Click Next.
- Now a window will appear and it will show the space required and available. Click Next.
- Here select the account name from the dropdown list for SQL Server Database Engine-NT AUTHORITY\NETWORK SERVICE and Click Next.
- Select "mixed mode" option.
- For mixed mode I have to provide Strong password(Ex. Passw0rd!) for the root account (username "sa" default) and adding my current user by which i will also be able to access database engine without default administrator account (username : sa).
- Click Next
- Now again every operation must be passed or skipped. If there is any operation with status failed then installation will not be possible. Click Next
- Now finally it is going to start the installation. Click Install

- Now wait till installation is complete. Setup is complete. Click Next
- Now Close this window
- At last Exit this window. Installation is complete

✓ *If the database is on a different server, enable TCP/IP on that server.*

Step 6 - SobekCM

- After Completion of Steps (1-5), Double Click on the SobekCM.exe setup file.
- On first window simply click on next.
- Accept the terms and click on next.
- Choose Full Install option which will enable all the services and install all necessary files.
- As shown in figure 4, enter Computer/Database Server name and Database name.

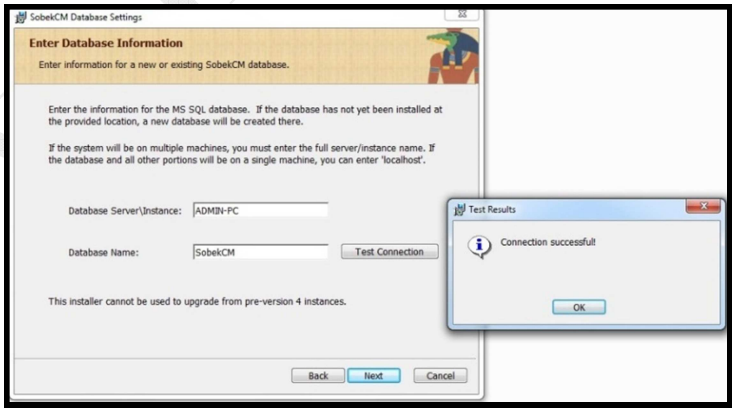

Figure: 4

- Next screen will be as shown in figure 5, it indicates that previous all installation and settings were successful. Click Next.

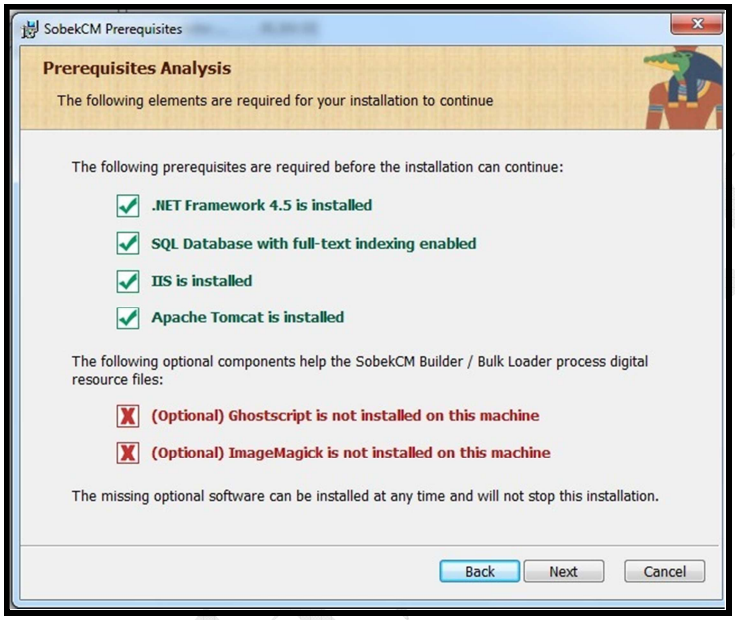

Figure: 5

- On next screen, provide Institution name and its Abbreviation. Click Next.
- As shown in figure 6, it will prompt for system administrator account details and password. After entering. Click Next

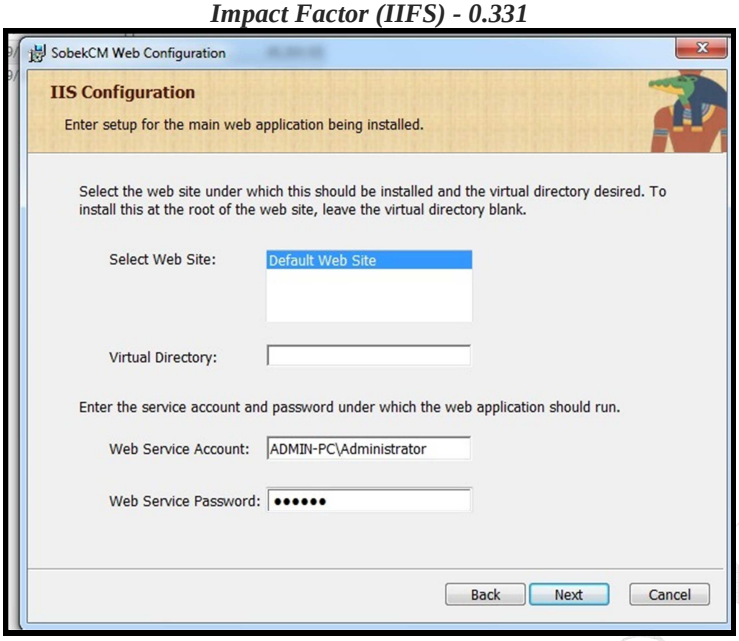

Figure: 6

- On Builder Configuration screen, enter administrator account name and password.
- Click on Next and click on finish
- Launch SobekCM

✓ *Finally SobekCM is installed and along with that it has created one Database which can be viewed from SQL Server Management Studio and it has also created web-based files in C:\inetpub\wwwroot.*

6. Working on SobekCM

- Click on Finish, it will redirect to one of default browser with the localhost link so that it can carry out few remaining configuration like registration shown in figure 7 below.
- Select Register Now and registration screen will be shown. Enter personal information on the screen below and hit SAVE.

Figure: 7

- First launch of SobekCM repository after the registration is completed shown in figure 8

Figure: 8

- Select the link in the upper right corner of the header for 'myHome'. This will bring up the login screen below in figure 9.

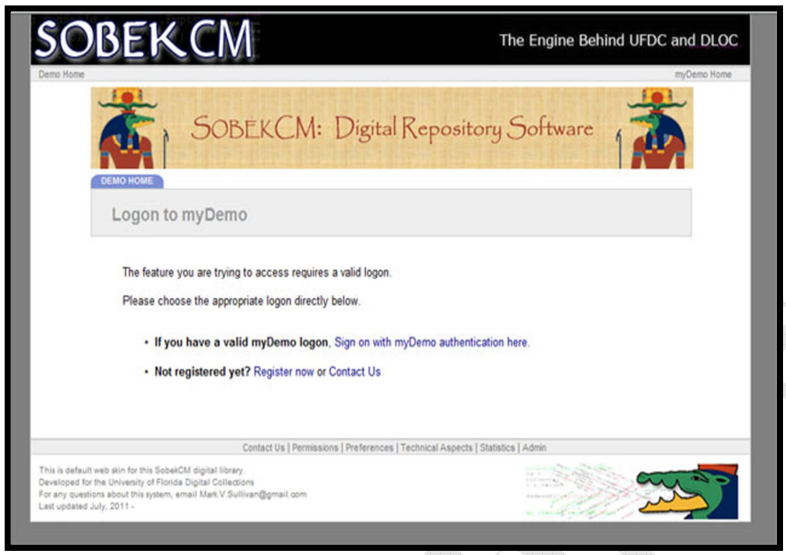

Figure: 9

- After logged-in, click on accept and start uploading.

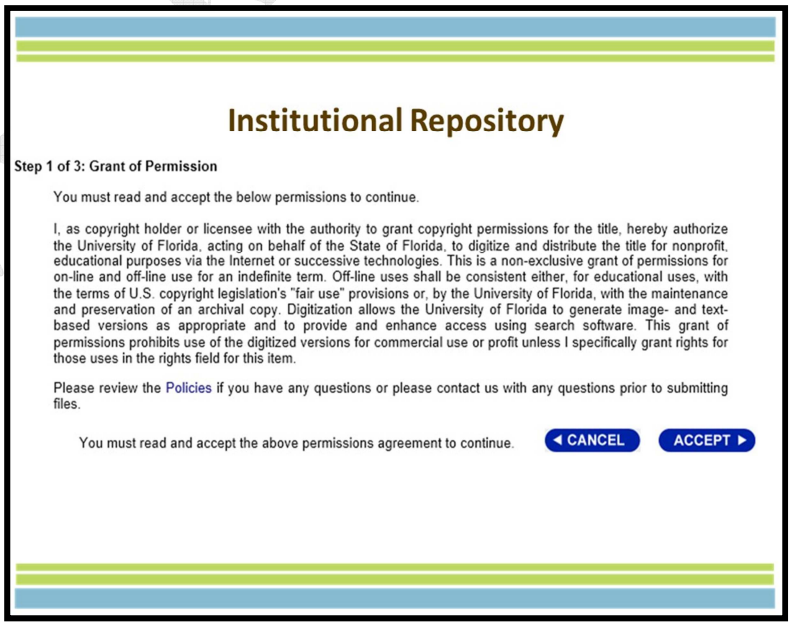

Figure: 10

- After entering required details as shown in the figure 11, Click Next

Figure: 11

- Next screen will allow to browse any format file and once it's done, click on upload tab. Click Finish.
- Next window will confirm that files are uploaded successfully and it can be edit or view the same files as shown in figure 12.

Figure: 12

- Simultaneously subject-wise collection can be created and it will display main webpage as shown in figure 13.

Figure: 13

- Figure 14 shows few examples on how different file formats will come-up as an output.

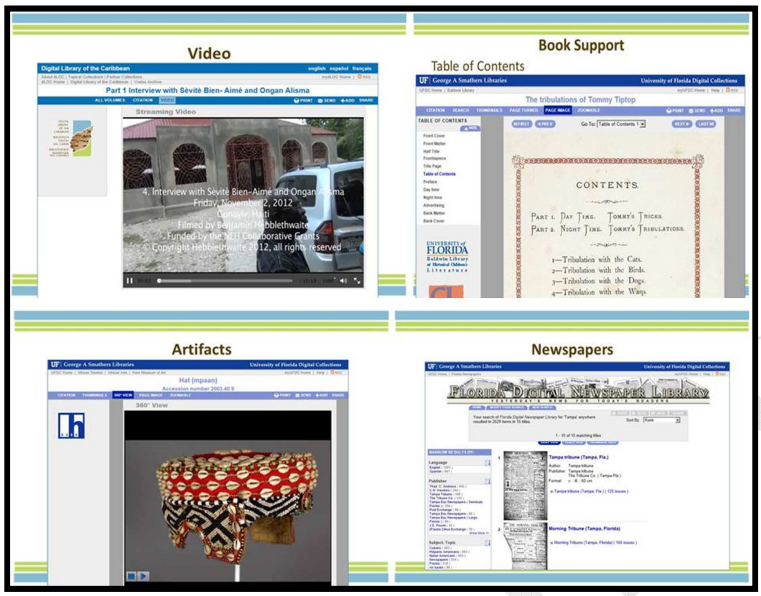

Figure: 14

7. SobekCM BACKUP

7.1 Why backup?

It is almost certain that you will be storing valuable information in your SokemCM server. Even assuming that the SobekCM code is 100% bug free, it is more likely that drives & fans break. Sometimes Computers get stolen. Server rooms get flooded. Without proper backups this could be a disaster.

It is strongly recommended that we take regularly backup of SobekCM files and MSSQL Server database. Always keep multiple sets of backups and try to keep a recent backup physically separate from the archive -either in another room or ideally another site.

7.2 What to backup?

MS SQL Server database i.e. SobekCM (Through MS SQL Management Studio 2008 Utility)

SobekCM web application files (Located in C:\inetpub\wwwroot)

7.3 How to take backup?

7.3.1 MS SQL Server Database Backup

- Go To Start
- Go to All Programs
- Go to Microsoft SQL Server 2008
- Click on SQL Server Management Studio
- Server Name = Computer Name on which the SQL Server lies
- Authentication = SQL Server Authentication
- Login = sa
- Password = (SA Password)
- On Left-Hand Side, select the Database under the Databases option of which Backup has to be taken
- Right-click the database, select Task, select Backup.
- In Source -> Database = Database of which Backup has to be taken.
- In Destination -> Click 'Add' and Select the path on which Backup needs to be stored (Note: Make sure that backup is stored on some safe path which is more safe than other paths regarding virus attack.)
- Click 'OK.'
- On Successful completion it will give a success message otherwise it will give warning.

7.3.2 Web application files

- Take the entire "wwwroot" folder as backup file.

8 CONCLUSION

From the above guide it is concluded that building an Institutional Repository has now became a mandatory task for Libraries as technology continues to evolve exponentially and as the amount of electronic information mushrooms. An Institutional Repository is an online locus for collecting, preserving, and disseminating, in digital form, the intellectual output of an institution. Institutional Repositories represent a certain basic cost to an institution but there are ways of minimising this. A Repository that is filled by self-archiving – that is, by researchers depositing their own articles – is far less costly to an institution than one where the library does all the depositing work. SobekCM, therefore, was designed as an open source application that institutions and organizations could run with relatively few resources. SobekCM helps the Librarians to implement IR in an easier manner because its OSS based on Windows OS. This research also concludes that Library professionals are generally facing problems while dealing with open source software based on Linux OS. Thus the SobekCM followed the librarian's inclination to create a system that would be as easy as possible to implement and use.

9 References

- **SobekCM Website (http://sobekrepository.org)**