$9^{th}$ International Symposium on Electronic Theses and Dissertations
$IX^{e}$ Symposium international sur les thèses et mémoires électroniques
June 7 – 10 Juin, 2006, Quebec City / Québec, Canada

# Added values to E-theses

***A preliminary version***

**Sayeed Choudhury (sayeed@jhu.edu) Johns Hopkins University, Baltimore, USA**

**Eva Müller (eva.muller@ub.uu.se) Uppsala University, Sweden**

**Abstract**

Putting theses and dissertations online offers more than just the possibility to manage, store, organize and disseminate digital materials created by an institution and its community members.
By developing new services which take advantage of the characteristics of electronic materials new values can be added to the original documents at the same time as some shortcomings can be overcome.
The aim of this paper is to raise awareness of the main issues involved when the infrastructures supporting some of these new services are being constructed.

## *E-theses, e-publishing and repositories*

It seems clear that electronic publishing systems and institutional repositories are becoming an established part of a general-purpose infrastructure within our universities and that e-theses are currently an important portion of the content stored within them.
Though, the focus is mostly on the text based content. In the future it is important to support also other content such as multimedia and to support developments related to e-science.
The building of a bridging infrastructure between institutional repositories and infrastructure supporting preservation and curation of research data is one of challenges which we will be dealing with in the coming years.

## *Added values to e-theses*

### *Added values to e-theses by linking to related resources or source data on which the e-theses are based*

One of the most fundamental aspects of scientific scholarly communication is the ability to cite and examine data in a consistent manner. Without this ability, the very essence of the scientific method, with its requirement of validating results, becomes compromised. The JISC Briefing Paper “Data Deluge” (http://www.jisc.ac.uk/index.cfm?name=pub_datadeluge) describes the unparalleled growth in data from e-science projects. For example, large-scale astronomy projects such as the Sloan Digital Sky Survey (http://www.sdss.org) at Johns Hopkins University have gathered data at unprecedented rates, raising new challenges and opportunities. Through the International Virtual Observatory Alliance (IVOA, http://www.ivoa.net), astronomers have developed community-wide metadata standards and web services that facilitate access and querying of these datasets. While science has always been data-driven, e-science, and increasingly computational humanities, has raised the importance of data to new levels.
With the IVOA and other e-science initiatives, it has become possible for graduate students and new researchers to access datasets for their research activities and for their e-publications. Given this rapid rise in data and the accompanying use in e-publications, it has become critical to develop a robust, viable data curation strategy. The JISC Briefing Paper “e-Science Data Curation”

(http://www.jisc.ac.uk/index.cfm?name=pub_escience) describes the fragility of these datasets, aspects of data curation and a set of recommendations to meet this important scholarly need.

*Added values to e-theses by providing long term preservation and curation*

While there are major questions to address and technological and policy issues to explore, there are also notable components of infrastructure that might provide the necessary base for data curation, especially as it relates to e-publications such as e-theses. The digital repository represents a software foundation for digital preservation. It should be noted that storing content in a repository represents a necessary, but insufficient condition for digital preservation. Repositories provide an opportunity to manage digital content, including datasets, in such a way that consideration of digital preservation is a possibility.

While some institutions rely solely on the institutional repositories, in some European countries there are experiences of building of infrastructure systems that support long term preservation in cooperation with national libraries or archives.

It is becoming increasingly possible to build interfaces between repositories and applications or web services. As an example, Uppsala University and Johns Hopkins University are working together to evaluate e-publishing systems, including Uppsala's DiVA, and appropriate connections to repositories. Working with astronomers at Johns Hopkins, the Sheridan Libraries are examining how the disseminators from the Fedora repository software can be interfaced with the web services offered through the Virtual Observatory framework. Such examinations and possible connections offer a technology infrastructure that can accommodate the appropriate policies for data curation as it relates to e-theses.

*Added values to e-theses by resolving legal issues*

Building seamless interfaces between repositories and applications or web services raises questions of importance for solutions for managing digital rights in this environment in a clear and effective way. The implementation of machine readable licenses based on rights management languages are a possible technical solution.

Moreover, in a time when we are experiencing a growing number of students producing e-theses, the question of plagiarism and how to deal with it is of immediate importance.

Some universities in Sweden have implemented systems to analyze papers and theses for potential plagiarism. Such systems are provided by commercial companies like Urkund (www.urkund.com) or GenuineText (www.genuinetext.com). Additionally, cooperation between these companies and institutions running repositories – like DiVA-group members prevents plagiarism through in-depth indexing.

## *Impact on the libraries*

As universities develop institutional repositories, the unit providing the infrastructure is most typically the library. It is important to be aware how these new roles and services affect libraries when it comes both to human resources and cooperation within the organization and outside it. In this particular context, the concept of "data scientist" is most relevant. This term has been used to describe an individual with significant knowledge and expertise with both digital libraries and a specific discipline. That is, these individuals represent the human interface between the repository experts and the researchers. There is a great need to train such individuals and to recognize their potential contributions within both the library and the scholarly realms.

In addition to cooperation with the academics we serve with digital programs, libraries will need to develop or strengthen their relationships with the information technology providers inside and outside their universities. The challenges of data curation as it relates to e-theses are both numerous and complex. In order to operate repositories successfully within university settings, it will be necessary to develop well defined and complementary roles with the central information technology units. For this particular relationship, it is extremely important to clarify that digital preservation entails more than storage of bits;

it includes the overarching policy framework and processes that support long-term access to data. There are also examples – such as DiVA in Sweden - where a number of universities run and develop services together

Moving beyond the organizational confines of the university, it will become essential to develop relationships with both non-profit and for-profit partners. For example in the US, Portico (http://www.portico.org) has recently secured a set of agreements with publishers for archiving e-journals. In Europe, some National Libraries have established cooperation in this field with both publishers and university repositories. Finally, several for-profit companies support existing e-science projects, and have expressed an interest in repository projects. It would be worthwhile for libraries to explore and define conditions under which it is mutually beneficial to work with such companies.

## List of references (confirmed according the guidelines comes in the final version)

## Author's biographies:

**G. Sayeed Choudhury** is the Associate Director for Library Digital Programs and Hodson Director of the Digital Knowledge Center at the Sheridan Libraries of Johns Hopkins University. He serves as principal investigator for projects funded through the National Science Foundation, Institute of Museum and Library Services, and the Mellon Foundation. He has oversight for the digital library activities and services provided by the Sheridan Libraries at Johns Hopkins University.

**Eva Müller** is a librarian at Uppsala University Library (UUL), Sweden and the Director of its Electronic Publishing Centre (EPC). She has been active in planning and developing of library information services at UUL since 1993. As Head of the Information Development Department she was in charge of UUL's digital library program. Since 2000 Eva has run the everyday work of the EPC and leads its Research and Development group. Her current work and research interest is in the field of electronic publishing and repositories and focuses on the development of an integrated infrastructure supporting long-term preservation and access to digital published materials.