Towards the Introduction of an Institutional Repository: Basic Principles and Concepts

Joachim Pfister* and Hans-Dieter Zimmermann**

Joachim.Pfister@fh-htwchur.ch, **Hans-Dieter.Zimmermann@fh-htwchur.ch*
University of Applied Sciences HTW Chur, Swiss Institute for Information Research SII,
Switzerland, http://www.informationswissenschaft.ch

Abstract

In this article the principles and benefits of open access are described which can be achieved with the help of institutional repositories. The advantages of a repository are depicted as well as the spread throughout the European Union. The process of installing an institutional repository organizationally is shown as short report on the ongoing efforts at the Swiss Institute for Information Research in Chur. The approach to install a widely accepted repository for the University of Applied Sciences Chur and its difficulties are portrayed as well as the process and methods chosen for the evaluation of the candidate software packages.

Motivation

The motivation for establishing an institutional repository at the University of Applied Sciences Chur is based on the needs of several stakeholders. First of all, the library of our university was looking for a tool helping to archive the digital versions of theses. On the one hand, access to electronic documents should be restricted to users within the university's network, e.g. for faculty members, researchers, as well as for the students. In some cases, e.g. due to confidentiality of the content, it may be necessary to limit access even further. On the other hand selected theses should be made publicly available online. Therefore, the library's intention is directed towards improving the access to information, and archiving these digital works. Secondly, the Swiss Institute for Information Research (SII) is looking for an efficient and convenient way to publish research work like conference or journal papers, research reports or other studies, as well as the issues of a paper series published by the institute electronically, and make them easily searchable and accessible through the internet. Bearing in mind that other institutes as well have very similar needs, it was decided to develop the concept for an institutional repository for the whole university. In this paper the approach taken will be presented.

The remainder of the paper is organized as follows: First, the idea of digital repositories will be briefly introduced from a more conceptual perspective. Second, the approach taken by the SII will be presented and discussed. Hence, the different steps taken will be introduced followed by a brief discussion.

What is an institutional repository?

The definitions of an institutional repository vary to a great deal in the extent which artefacts are to be stored. They are ranging from every digital material created by an institution (cf. Lynch 2003, Bevan 2007) to a well defined set of materials. Within the following definition, you will find concrete types of artefacts or materials narrowly tied to an institutional

repository whereas at the same time other types are regarded as "other digital assets" not seeming to be in the immediate focus. In consequence, this forces an institution to decide, which materials are to be archived in an institutional repository.

> "An Institutional Repository is an online locus for collecting, preserving, and disseminating – in digital form – the intellectual output of an institution, particularly a research institution. For a university, this would include materials such as research journal articles, before (preprints) and after (postprints) undergoing peer review, and digital versions of theses and dissertations, but it might also include other digital assets generated by normal academic life, such as administrative documents, course notes, or learning objects." (Wikpedia 2007)

With the expression "institutional repository" goes along the idea of creating a "digital library". Jones et al. (2006) regard repositories as constituent elements of digital libraries, providing "… them with the selections of collections they present as libraries, whether institutional or disciplinary." The distinction between institutional or disciplinary is due to the scope of a repository or a digital library whether it serves a single institution or collects materials belonging to a specific discipline.

Reasons for introducing an institutional repository
Researchers and institutions benefit from institutional repositories in the same way. The most prominent reason is the increase in visibility and impact of research output (cf. Crow 2002). Building up and maintaining reputation in the scientific community is essential for academics and institutions. To measure research impact, bibliometric methods like citation analysis e.g. in the field of academic journals are used. Studies carried out (Lawrence 2001, Harnad and Brody 2004, Antelman 2004, cf. bibliography from the OpCit 2007) emphasize the importance of freely online accessible publications: It is evident they are cited more often than literature with restricted access.

The second reason is due to the change in the scholarly publication paradigm. Self-publishing scientific content and providing free access to these materials are key activities in the Open Access movement. The main conferences leading to substantial declarations are the Budapest Open Access Initiative (2001) and the Berlin conferences leading to the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003). By signing these declarations, several hundred organizations on an international level clearly state that they support the ideas of Open Access thus challenging the established, commercially oriented publisher-centric way of spreading scientific information (cf. Crow 2002). Some funding bodies even insist on having published the results produced with their financial support only by means of Open Access, like the Swiss National Fonds.

The third reason for implementing an institutional repository is based on the possible improvement of internal communication. By providing a centralized digital asset store other organizational units within a larger organization might benefit from the published materials on the one hand. But on the other hand, material not published, for instance internal technical reports, are also part of an organization's knowledge and should be accessible by everyone within this organization to allow re-use und to make cross-references which would be impossible when those digital assets are stored just locally e.g. only by a single working group.

Benefits of Open Access

There are several advantages going along with Open Access. First of all, scientific research output can be published faster without any intermediaries like a publisher. The second reason is the seemingly cost efficient way of accessing academic research output. Especially, when regarding the increasing costs of journal subscriptions leading to the so called "journal crisis" (CPB, 2000), Open Access seems to be an attractive option.

Nevertheless, publishing scientific output is not free at all. The costs for running an institutional or disciplinary repository and applying quality control mechanisms like peer review have to be taken into account. Therefore two ways of publishing materials in an Open Access manner are applied widely. In the "green road" approach, an article, for instance, besides its publication in a subscription journal, is deposited as a pre- or post-print in a repository and made freely accessibly. Some publishers put an embargo time on these materials, so that the traditional subscription models still work but for limited period of time (e.g. for one year). In the "golden road" approach, the author pays the publisher a fee and at the same time the article is published, it is made freely accessible online.

The spread of digital repositories in the European Union

To describe the current state of digital repositories in 25 countries of the European Union, a survey was carried out within the EU DRIVER-Project ("Digital Repository Infrastructure Vision for European Research") between June 2006 and February 2007. About 230 institutions with one or more digital repositories were found, of which about 50% participated in the study (n = 114) (van Eijndhoven, van der Graaf, 2007). In this study, a digital repository was defined as being institutional or thematic with research materials as content and providing Open Archive Initiative[1] (OAI) compliant access methods. The key findings can be summarized as follows:

- Mostly textual materials are included in a digital repository (about 90%)
- textual records consist for the larger part of metadata only (68%), whereas only 32% relate to full text
- The textual materials mainly consist of articles (54%). Other types play a less important role (books and book chapters: 19%, theses: 10%, proceedings and working papers 5%).
- About 20 different software tools are used to run a repository. The top three of the most frequently used packages are GNU Eprints (24%), DSpace (20%) and OPUS (11%). Locally developed software packages sum up to 17%.
- The majority (70 to 75%) of all digital repositories assigns a persistent identifier to each document for the long-term availability
- 48% use a standardized system of keywords or classification, 44% use freely assigned keywords, 8% do not use any subject or keyword indexing at all.
- Over 50% of the participating digital repositories are searchable via a general Internet search engine like Google or Yahoo.
- 95% of the full text materials are publicly accessible in the sense of Open Access; 26% is only accessible within a campus; 18% contain Open Access materials with an embargo period.

The SII approach for the introduction of an institutional repository

Taking into consideration the motives described above, the SII has decided to introduce an institutional repository. First, we had the idea to use it only for our own institute but it soon

---

[1] http://www.openarchives.org, retrieved November 14, 2007.

became very obvious that an organisation-wide solution for the whole University of Applied Sciences HTW Chur would be more advantageous. In order to establish an institutional repository we are carrying out two steps. In the first one, students participated in the form of a project course. On basis of their suggestions and findings step two has been started where a university wide approach of establishing an institutional repository has been launched.

Step 1: Involving students in a project course
As a teaching method at the SII, the form of a project course has been established. A rough project outline is given by the course instructor like the main idea and the general goals, but the project work itself and the project management is carried out by the students themselves to let them experience working in a project context. Therefore, a group of students was asked to develop a respective proposal for an institutional repository. Interviews with researchers and administrative staff in the institution were conducted in order to come up with a requirement analysis. In the next step, the students systematically evaluated software solutions for an institutional repository with the help of a criteria catalogue. This criteria catalogue was prepared on the basis of the identified requirements and distinguished between must-have and can-have criteria. Finally, the students recommended two suitable solutions: a short term solution for the institute itself and a more sophisticated one for the whole university.

Interviews to gather the institute's requirements
Interviews with the staff of the SII were carried out by the students in order to elicit the different needs and notions associated with an institutional repository. This was done as part of the requirement analysis. It was then recognized by the students, that the staff didn't share a common understanding what an institutional repository is and what functions it has to fulfil. As organizational pre-requisites, the students formulated, that
- a shared vision of an institutional repository is necessary
- the various types of documents to be stored have to be defined clearly
- all the processes surrounding the content selection and delivery must be standardized and transparent to assure a certain amount of quality control
- human and financial resources for running an institutional repository have to be taken into account
- the organizational integration on the level of the whole university has to be tackled, e.g. the technical administration or a concept for covering the costs generated by installing an institutional repository.

Within the project course, the students developed a criteria catalogue based on the requirements formulated in the interviews. The criteria catalogue used by the students consists of two main sections: general and specific criteria. In the general section they had a closer look on the software answering questions like: What are the soft- and hardware pre-requisites? What are the costs of introducing and maintaining repository software? How complex is the administration and installation procedure? In the specific section, a detailed look was taken on the features the repository software offers, notably the type of metadata and the search possibilities. Additionally, attention was paid on usability factors. Based on this criteria catalogue, the students compared four software packages for an institutional repository (MyCore[2], DSpace[3], EPrints[4] and OPUS[5]). The recommendation of the students'

---

[2] http://www.mycore.de, retrieved November 13, 2007.
[3] http://www.dspace.org, retrieved November 13, 2007.
[4] http://www.eprints.org, retrieved November 13, 2007.

group was MyCore since this software package comes in a preconfigured basis variant which needs to be adapted very little for basic purposes but still offers the capability to be thoroughly customized (e.g. by defining own metadata formats).

Comparing different software tools for running an institutional repository by listing all features and restrictions is a method which was used e.g. by Crow (2004). Of course, the features must be compared to the requirements. This was done by the student in their evaluation scheme where every criterion had the same weight and the different sections weren't weighted either. The lack of having a weighting scheme leaves room for improvement as shown in the next section.

Refinement of the evaluation methodology
A more elaborate evaluation scheme can be found in Goh et al. (2006). They designed a checklist for the evaluation of digital library software using 12 categories. The categories were determined from the literature and each category is assigned a weight which sums up to 100 percent for all categories whereas the items making up a category always sum up 10. An institution which would like to evaluate institutional repository software can adapt the checklist with its items or fine-tune the weights according to their needs. The 12 categories Goh et al. used were: content management, content acquisition, metadata, search, access control and security, report and inquiry, preservation, interoperability, user interface, standards compliance, automatic tools and support. The weights were assigned on the basis of a modified Delphi technique by four people familiar with institutional repositories. If discrepancies in the weight estimates were observed, they were resolved through discussion.

Goh et al. compared four digital library software packages (CERN CDSware[6], Greenstone[7], Fedora[8] and Eprints) and based on their evaluation scheme, they calculated a consolidated score for each software tool. The Greenstone (version 2.51) platform fulfilled most of their requirements. But at the same time they say that current open source software for institutional "… still lacks certain functionalities perceived to be important, as gathered from the literature." (Goh et al. 2006) According to their evaluation, the software tools differed a lot in the functionalities like the support for search and long term preservation capabilities.

Step 2: Formulating a concept for the university wide project implementation
Based on the experiences and results the students in the project course have worked out, the SII suggests the introduction of an institutional repository for the whole University of Applied Sciences Chur. In this article, the concept for this project is presented. The project is being split up in several phases which are influenced by elements of the evaluation process formulated by Jones et al. (2006, 72) which will be explained in more detail. Currently, we are still focussing on the first project phase which takes longer than expected since the idea of an institutional repository has to be promoted very actively and advocacy (see phase seven) is an essential factor to convince e.g. the university's administration. For this reason, emphasis is put on the role advocacy and communications management exert on the implementation phase and on the project in general.

---

[5] http://www2.bsz-bw.de/cms/digibib/opus, retrieved November 13, 2007.
[6] http://cdsware.cern.ch, retrieved November 13, 2007.
[7] http://www.greenstone.org, retrieved November 13, 2007.
[8] http://www.fedora-commons.org, retrieved November 13, 2007.

1 Ensuring financial funding
The evaluation, implementation and running of an institutional repository need to be financially funded, of course. Therefore the university's administration has to be convinced of the benefits a repository offers to the whole institution. Depending upon the willingness and financial capabilities of an academic organisation, this might be the first, but decisive, difficulty to encounter.

2 Defining requirements
On the one hand, you will find functional requirements like: Which needs do the library and the different departments have got? What types of materials will be stored? How to authorize and authenticate users for adding content to the repository? On the other hand organizational requirements must be identified like a university wide policy which content should be added to the repository and how the workflow is defined with all roles and responsibilities, e.g. who will be responsible for checking the electronic file for integrity and capturing metadata (cf. Jones et al. 2006, 85 ff.).

3 Preliminary evaluation
The aim of this phase is to select candidate software tools for a closer evaluation in the next step. In our case, we can benefit from results of the student's project course.

4 Functional evaluation by means of a comparative evaluation
During this phase, the candidate repository tools will be evaluated comparatively using the methodology as described before by Goh et al. (2006). Before, this checklist has to be adapted to the requirements elicited in phase two. The total costs of ownership of a solution have to be taken into account as well. Besides the costs of the initial development and customization of the institutional repository software, there will be technical and organizational maintenance costs: Technical maintenance costs arise from software updates, backups and the costs for hosting the repository on a server. Organizational maintenance costs can be attributed to the organizational units involved in adding an item to a repository (each author, the library as a central service provider etc.). The development of a concept for handling these issues is part of phase six.

5 Implementation
According the evaluation results, the institutional repository software will be implemented and customized to the organizations' needs.

6 Defining a maintenance concept
To ensure the enduring availability of the institutional repository and to secure the investments, a technical maintenance concept has to be developed. Part of this maintenance concept is a model for charging the costs of running the institutional repository within the different units of an academic organization, if necessary.

7 Content acquisition and communications management
Finally, when an institutional repository has been successfully installed from a mainly technical perspective, the questions arise how to enrich it with content and how to gain acceptance to make it work. Consequently, already published materials on the authors' personal home pages as well as from other sources (e.g. the issues of an electronic journal published by the institution itself) have to be integrated into the institutional repository. Corresponding workflows as well as responsible instances for adding these items to the repository have to be defined.

According to Buehrer et al. (2003) the introduction of a digital repository can be considered as a change project. Thus, we have to consider a second form of implementation beside the more technical 'implementation I'. 'Implementation II' refers to the necessary change of mindsets addressing a change of behaviour of the target groups (figure 1). Users have to be enabled to comprehend process- and organization-related changes with their group as well as within the whole university. Buehrer et al. (2003) therefore propose the communications management methodology. The distinction between Implementation I and II has been introduced by Schmid (2001).
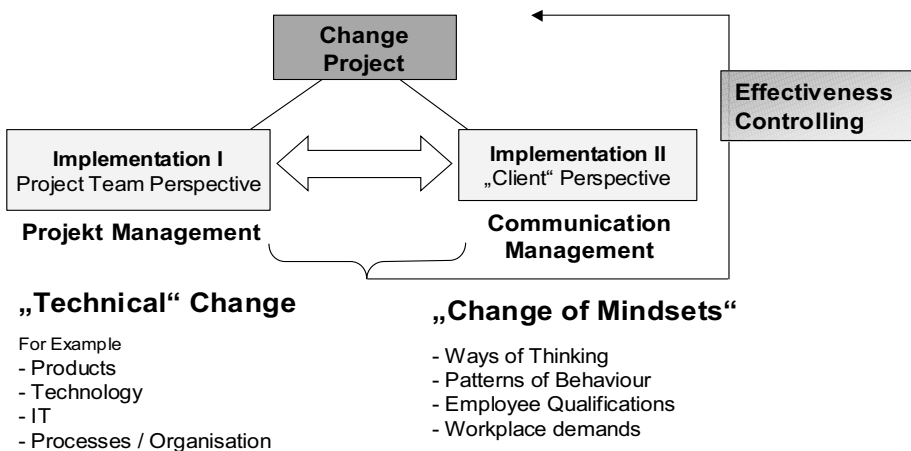

Figure 1: "Implementation I and II" (Buehrer et al. (2003).

Furthermore, appropriate marketing efforts for the whole academic staff have to be designed to promote the idea and usage of the institutional repository. This is called advocacy. Jones et al. (2006) address the issue of 'implementation II' and summarize the so called advocacy process with the help of the following figure 2.

The advocacy process is in the centre. In the layers surrounding the centre, the different activities and stages are grouped according to their logical and temporal succession. The stages in a innovation-decision process (1) are put into relation with the practical events which occur hierarchically when introducing an institutional repository (2). At the same time, the degree of acceptance within a specific social system, the so called adopters, is also integrated (3). In the knowledge step, innovative individuals start their activities around an institutional repository. The next step, named persuasion, deals with implementing a repository and promoting the idea through different communication channels and to gain interest not only on the individual level but on the level of an academic unit (e.g. institute, department) what refers directly to the "change of mindsets" mentioned before. During this phase it's crucial to win the support of opinion leaders since most of the other members of the organizational unit will follow. Jones et al. (2006) suggest that a proof-of-concept or a pilot study is useful before an academic unit finally decides to accept the repository. After having reached the implementation level, involving the university's senior management to receive top-down support is the next step. Finally, in the confirmation step, senior academic members

will recognize the repository widely and the integration into formal institutional policy is the next logical step.
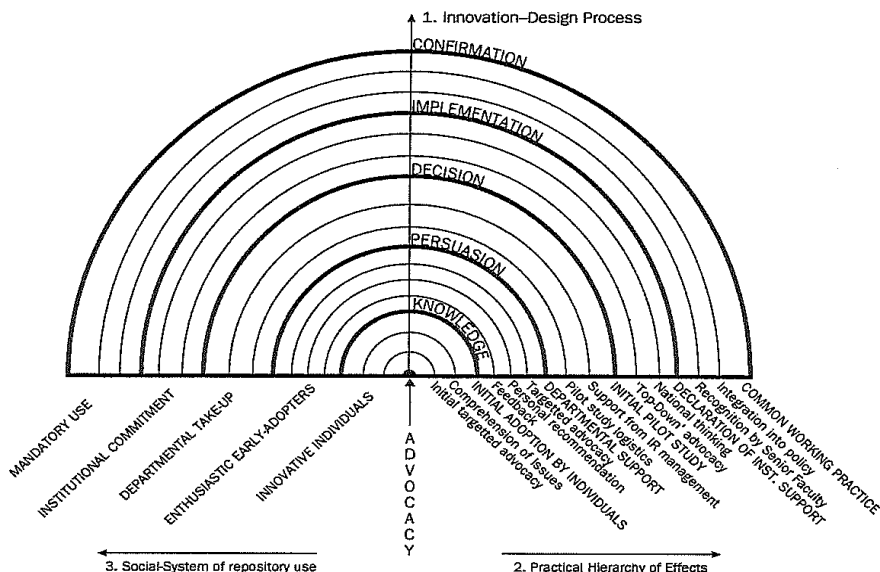

Figure 2: "Ripple-diagram showing the relationship between (1) stages in the innovation-decision process (after Rogers, 1995), (2) the practical hierarchy of effects (adapted from McGuire, 1989), and (3) the dominant social-system of repository use" (Jones et al. 2006).

Conclusion and future directives

In this paper we address the introduction of a digital repository within an university environment. As the project itself is concerned, the SII still is focussing on the first step (as of Nov. 2007) which takes longer than expected since the idea of an institutional repository has to be promoted very actively. Even if the idea of a digital repository sounds very convincing at a first glance, the concrete setup of a respective project is rather complex. As network effects apply it is obvious that the more institute within the university participate in the repository the more value will be created. On the other side each additional stakeholder adds complexity to the project. This is mainly because every institute within the university has slightly different needs because of their specific domain. Beside complexity due to different stakeholders also the issue of 'implementation II' has to be considered carefully. Only if the system fits the different needs, stakeholders will be willing to invest (a) in integrating and migrating already electronically published material into the new repository as well as (b) into learning new processes and behaviour.

References

Antelman, K. (2004). Do open-access articles have a greater research impact?" *College & Research Libraries* 65(5). pp.372-382. Retrieved October 23, 2007 from http://eprints.rclis.org/archive/00002309/.

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003). Retrieved October 25, 2007 from http://oa.mpg.de/openaccess-berlin/berlin_declaration.pdf.

Bevan, S. J. (2007): Developing an institutional repository: Cranfield QUEprints – a case study. *OCLC Systems & Services*. Vol. 23, No. 2, 2007, pp. 170-182 DOI: 10.1108/10650750710748478.

Buehrer, R., Mueller, Chr. & Zimmermann, H.-D. (2003). A Method for Systematic Communications Management in Technology-driven Change Projects. *Proceedings of the Seventh Pacific-Asia Conference on Information Systems (PACIS),* July 10-13, 2003, Adelaide, Australia. From http://ssrn.com/abstract=1018078.

Budapest Open Access Initiative (2001). Retrieved October 25, 2007 from http://www.soros.org/openaccess/read.shtml.

CPB (2000). *Tackling the journal crisis. When authors pay with money instead of copyrights*. Working Paper No. 121. The Hague: CPB Netherlands Bureau for Economic Policy Analysis. Retrieved October 25, 2007 from
http://www.cpb.nl/nl/pub/cpbreeksen/werkdoc/121/wd121.pdf.

Crow, R. (2002). *The Case for Institutional Repositories: A SPARC Position Paper. The Scholarly Publishing & Academic Resources Coalition.* Retrievef May 23, 2007 from http://www.arl.org/sparc/bm%7Edoc/ir_final_release_102.pdf.

Crow, R. (2004). *A Guide to Institutional Repository Software.* 3rd edition. Open Society Institute. Retrieved October 29, 2007 from
http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf.

Van Eijndhoven, K. & van der Graaf, M. (2007). *Inventory study into the present type and level of OAI compliant Digital Repository activities in the EU. White paper*. Version 0.9. Retrieved October 25, 2007 from
http://www.driver-support.eu/documents/DRIVER%20Inventory%20study%202007.pdf.

Goh, D.; Chua, A.; Kho, D. A.; Khoo, E. B.-H.; Mak, E. B.-T. & Ng, M. W.-M. (2006). A checklist for evaluating open source digital library software. *Online Information Review.* Vol. 30. No. 4. pp. 360-379.

Harnad, S. & Brody, T. (2004). Comparing the impact of open access (OA) versus non-OA articles in the same journals. *D-Lib Magazine*, 10 (6). Retrieved October 24, 2007 from http://www.dlib.org/dlib/june04/harnad/06harnad.html.

Jones, R.; Andrew, T. & MacColl, J. (2006). The Institutional Repository. Oxford: Chandos Publishing.

Lawrence, S. (2001). Online or invisible. Retrieved October 24, 2007 from http://citeseer.ist.psu.edu/online-nature01.

Lynch, C. (2003). Institutional Repositories: essential infrastructure for scholarship in the digital age, *ARL Bimonthly Report*, No. 226. Retrieved October 24, 2007 from http://www.arl.org/resources/pubs/br/br226/br226ir.shtml.

OpCite Project (2007). *The effect of open access and downloads ('hits') on citation impact: a bibliography of studies.* Retrieved October 25, 2007 from http://opcit.eprints.org/oacitation-biblio.html.

Schmid, B. F. (2001). What is New About the Digital Economy? *Electronic Markets*, 11 (1), 1019-6781. Retrieved from http://www.informaworld.com/10.1080/10196780122367.

Wikipedia (2007): Institutional Repository. Retrieved October 10, 2007 from http://en.wikipedia.org/wiki/Institutional_repository.