

CDS Invenio Implementation at the University of Economics, Prague

Václav Šubrta

University of Economics, Prague
Centre of Information and Library Services

subrta@vse.cz

Third Seminar on Providing Access to Grey Literature
December 8, 2010

The aim of this paper is to explain the process of preparation and future implementation of the CDS Invenio institutional repository for the needs of the University of Economics, Prague. First of all, institutional goals are defined and current processes explained. An analysis of the publication activities of VŠE has been done as a part of the testing phase and its results are described. During the testing of the software several problems emerged, and the article describes them. Organizational arrangements form an integral part of the implementation of an institutional repository and therefore the article describes these as well. The paper is based on the experience gained from the preparatory/testing work on this project.

VŠE institutional repository goals

As discussed with the vice-rector for science, there are three main goals of the institutional repository at VŠE:

1. to provide means to increase citation counts of the scientific work of VŠE staff,
2. promote journals published at VŠE (there are three main journals - Acta Oeconomica Pragensia, Prague Economic Papers and Politická Ekonomie as well as several other journals),
3. provide a platform for better cooperation among various departments of all six faculties of VŠE to serve scientific purposes of the staff as well as (mainly PhD) students.

These goals indicate that some texts should be freely accessible for the whole scientific community, and that access some texts will be probably restricted to VŠE staff and students.

According to the VŠE vice-rector for science, for the first goal only peer-reviewed journal articles, articles from journals with non-zero impact factor and proceedings from international scientific conferences should be used.

The repository can and will be used not only for scientific texts, but also as a means to promote other texts collected at VŠE, like ETDs of VŠE students, currently accessible via OAI-PMH from our study system and therefore can be harvested into the institutional repository, or as a way to promote the collection that provides access to historical development of Czech economic thought among the VŠE staff (Golden Collection of Czech Economic Thought), which is a digitalisation project held by the Centre of Information and Library Services at VŠE.

There are other possibilities of how to extend the usage and usefulness of the institutional repository, mainly in two areas: working papers of PhD students and a pool for study materials currently published in the integrated study system (ISIS).

The repository can be offered for publishing working papers of PhD students at various departments of VŠE, which can serve as a means for clarification of their writings as well as the foundation of possible cooperation among departments. PhD students can place there their seminar works they are writing in the courses they attend during their studies and therefore the quality of the seminar works can increase. Other students may also benefit from reading of other students' writings.

The same can be applied to the teaching materials at present being published in the integrated study system of VŠE. The materials are provided for students of a particular course using ISIS and are linked to a specific academic year/semester. The institutional repository can offer at least the metadata for these materials, taking into account this restriction. Unfortunately, there is also a necessity to modify the study system in order to provide this kind of metadata for harvesting. Furthermore, the possibility to decide from the role of a teacher which form to be used for the materials published (full access, access limited only to an IP range of VŠE, access limited to students via LDAP authentication) must be solved. We plan to set this possible integration as one of the topics for the implementation working group (the roles and expectations from this group are described below).

Why Invenio

Invenio has been chosen among other systems in cooperation with the National Technical Library as a part of a NUŠL project (National Repository of Grey Literature). The main reasons for choosing this particular software was mainly its great functionality as well as wide range of possibilities to customize the system for the needs of a particular outcome. The system comes preconfigured as a virtual machine using Virtualbox.

Analysis of publication outcomes

As a part of the preparation phase an analysis of publication outcomes was done to determine the number of articles to be integrated into the repository. The data analysed

during this phase are accessible in the database of publication activity of VŠE at <http://pcvse.vse.cz/>.

The analysis aimed at peer-reviewed journals and journals with non-zero impact factor in the publication activity of VŠE staff and PhD students. The aim was to divide the journals into categories according to licence terms and conditions.

Analysis of the data from 2005 - 2010 revealed (among other things) that there is quite a large number of papers published in VŠE journals, which means that at least part of the publication outcomes of VŠE staff can be made accessible in the repository.

Comparing the terms and conditions, there are two main results of the analysis: (1) for the majority of journals the terms and conditions are not currently known and (2) many journals use some restrictions for the texts (either delay or access limited to the staff and students only).

For more detailed results of the analysis see the table:

	% of journals	% of papers
can be archived	9,60%	7,43%
for staff and students only	1,60%	0,16%
with delay	3,20%	0,55%
unknown terms and conditions	73,60%	61,38%
open access	0,80%	0,08%
with editorial board approval	4,80%	2,03%
the rights holder is the journal	0,80%	0,08%
journals published by VŠE	5,60%	28,30%

Basic workflow requirements

The first part of the workflow of the institutional repository will depend on the current process of storing publication outcomes of VŠE scholars. In the process there is one person at each faculty responsible for the database of the publication outcomes. This person collects the metadata from the scholars and examines whether all entries are correct. The record of metadata is then promoted from the VŠE database into the R&D IS operated by the Research and Development Council of the Czech Republic. This process works well and should remain the same. Unfortunately, this system is not integrated into the study system of VŠE and there is no current plan to integrate it. Unfortunately, the software that VŠE uses for collecting publication outcomes does not allow easy enhancement with the required functionality.

There is an ongoing discussion whether the author or the employer is responsible for the decision about making the text accessible in the institutional repository. As far as our knowledge goes, both approaches have many supporters as well as opponents. The current situation at VŠE is open to the author's decision which texts from their portfolio are

to be stored in the institutional repository. The necessary supplement of this approach is an explanatory campaign among scientists at VŠE that shows the advantages of the open access approach.

The form of eprint (preprint, postprint, ...) that can be made accessible in the repository is another issue to be solved. The discussion is even more complicated due to the fact that there is a wider range of definitions of which type of eprints means for various publishers.

Software requirements and technical issues

As mentioned above, there are several factors that influence certain setup of the software. As for our participation in the NUŠL project, we have been trying to implement these requirements into the Invenio system.

The first limitation is given by the school management - only a part of all publication outcomes will be accessible in the institutional repository (peer-reviewed journals, non-zero impact factor journals and international conferences), but the repository should not be limited just to this set of options.

Another set of limitations arises out of the terms and conditions of journals publishers:

- A number of journals allows to put the text in the repository with a delay, so there must be a way to set it up for a particular journal or conference.
- Some journals permit access to eprints for staff and students only. This can be solved by IP restrictions (texts will be accessible from the campus) and LDAP combined with EZproxy for remote access.

The next technical issue is the exchange of data between the database of publication outcomes of VŠE and the repository.

The repository should not only enable access to eprints from various journals and conferences, but also the way how to publish i.e. working papers of PhD students or their seminar works. The system should be interconnected in this regard with the study system.

There are at least two repositories that are to be harvested by the Invenio system - the ETDs repository and the digitised collection of Czech economic thought history.

In the Invenio system there is a way to set up various roles in the system - not only the super-admin, but precisely defined roles with particular rights. The issue to be solved here is how the roles are to be defined and who is to be responsible at VŠE for each task.

There is also a set of issues derived from the unified visual style of VŠE. The repository templates will be modified to be in compliance with the rules of the visual style of university webpages.

The texts published in the repository can be also advertised at other websites running at VŠE. Almost all webpages of faculties and departments use the same system (Wordpress with modifications for the needs of VŠE) and the integration of a subset of publications from the institutional repository into the webpages of a department can promote the scientific results of the department/faculty with minimal effort.

Implementation working group

The institutional repository of scientific texts to serve other purposes (as mentioned above) is quite a complicated task that influence a large part of the university. With better understanding of what the repository can offer to various parts of the university, the chance to implement the solution well increases. We propose to involve as many departments as possible into the process of implementation and the most obvious way is to create an implementation working group. There are two benefits in this approach: (1) we will be able to collect and reflect real requirements from all departments and (2) the solution will be more acceptable for all the subjects, as they will be involved in the process of institutional repository implementation. The implementation working group should consist at least of a member of school management, vice-deans for science at faculties, Computer centre staff and the Centre of information and library services staff.

The implementation working group should set the strategy, create workflow, define roles and responsibilities and manage the implementation of the institutional repository at VŠE as well as to decide how to link the repository to other systems currently running at VŠE.

Conclusion

The Invenio repository is in the preparatory/testing phase at VŠE and a lot of work has to be done to create fully functional institutional repository that meets the goals listed at the beginning of this paper. The results of the analysis show that at least part of the publication outcomes can be made accessible in the repository. The success of the repository depends on three factors: (1) the acceptance of the open access idea among scientists at VŠE, (2) the quality of technical and organizational integration of the repository into the current systems running at VŠE, and (3) the promotion of the open access idea in the wider scientific world, that can simplify the terms and conditions in current journals.