Non-Text Theses as an Integrated Part of the University Repository: a Case Study of the Academy of Performing Arts in Prague

Iva Horová, Radim Chvála

Academy of Performing Arts in Prague, Czech Republic,
iva.horova@amu.cz; radim.chvala@amu.cz

Abstract

Training the professional artists of the future brings several challenges. Students at the Academy of Performing Arts in Prague (AMU), at all degree levels, are required to produce outputs including several final qualifying works. A piece of written work is mandatory, but this is usually accompanied by records of artistic performance — a graduation film, a stage role, graphic materials, a concert, a set of photographs, etc. Each of these works has its own topic, tutor, opponents, annotations and even classifications. Some qualifying works, such as films, can have multiple student input, with different graduands in the roles of screenwriter, director, producer, and so on. The preservation and discovery of and access to these works are issues of obvious importance, just as is the case with ‘traditional’ textual works.

These issues were addressed at AMU by modifications to an Open Source system, DSpace. Metadata based on the Dublin Core Standard was extended to include the relation element qualifiers. The modules for editing and displaying were modified to permit searching and viewing of the related documents. Video and audio players were integrated into the system to make the related multimedia files available directly from the primary record page. A handle server, which generates persistent identifiers, was implemented.

The automatic transfer of metadata from the AMU Study information system, KOS, into the repository was enabled, along with automated OAI-PMH harvesting into the national registry Theses.cz at Masaryk University in Brno, and exports into the AMU library system.
This paper is based on a presentation given by the lead author at the 2010 LIBER Annual General Conference, Aarhus, as part of the DART-Europe Master Class on the management of electronic theses.

Key Words: University Repository; non-text theses; metadata description; records relations

Introduction

The Situation in the Czech Republic

It is generally known that in many countries in Europe e-theses are an integrated part of the institutional repositories which have been established in recent years. In some countries they have been harvested on a national level. Finding a solution to these and all associated questions has also been very urgent in the Czech Republic. Until recently, theses were only available in fragmented sources at the local level. For these reasons, the Electronic Theses and Dissertations (ETDs) Working Group (WG) within the Association of Libraries of Czech Universities was established.

The Association was established as a common platform for Czech Academic Libraries in 2002. It aims to promote and encourage the common interests of the academic libraries. It serves as a platform to gather and exchange experiences. It supports cooperation, nationally as well as internationally. The Association is a member of LIBER, EBLIDA and IFLA. It currently has 23 institutional members, out of 25 public universities in the Czech Republic. Several Working Groups operate within the framework of the Association. The Working Groups are temporary, set up to solve special issues. The ETDs WG was one of such Working Groups.

The Electronic Theses and Dissertations Working Group within ALCU

The ETDs WG started its work in 2004 and continued until April 2010. It was the first Czech attempt to coordinate activity based on a complete analysis of the real situation in the Czech Republic. The WG had 28 members from 19 Czech universities, plus cooperators from other institutions. Its main aim was to formulate instructions and recommendations for default procedures in the
field of managing e-theses. It became very helpful for all Czech universities, because in 2006 the Higher Education Act was amended to the effect that all Czech universities are obliged to publish Bachelor’s theses, Master’s theses and Dissertations through a database.

Metadata standardisation was the first task of the ETDs WG, since the group believed that this was a prerequisite for further cooperation. The main result of these activities was the proposal of the metadata standard ‘EVSKP-MS’, later approved as a standard in the Czech Republic. The activities of the WG played an important role in raising awareness among the universities in the Czech Republic and indeed started to prepare the conditions to establish an e-theses register at the national level.

The ‘Theses.cz’ Project

Two big e-theses projects were prepared in the Czech Republic two years later. The first of them was a project of Masaryk University in Brno, which aimed primarily to offer universities a tool for tracing plagiarism. The second one was the idea of a national register, prepared by some members of the ETDs WG. The two projects were merged into one thanks to the negotiation of the ETDs WG. This ‘Theses.cz’ project was managed, programmed and launched by Masaryk University in Brno, with a grant from the Ministry of Education. The project has been operating since the end of 2008. It has delivered a system which is used to collect Czech ETDs and search them for traces of plagiarism. The system represents a nationwide, digital register of theses and all related information, captured according to the EVSKP-MS metadata standard. Today there are 29 schools participating, including AMU, and the register holds more than 80,000 records. Unfortunately, the ‘Theses.cz’ system is not open and ready to provide metadata to other aggregators. It was not conceived as an Open Access source, but rather as a detection system for anti-plagiarism.

Selected Current Issues in the Czech Republic

Legislation: Unfortunately there is disalignment between the latest amendment to the Czech Copyright Act (No. 121/2000 Coll.) and the latest amendment to the Higher Education Act (No. 120/2006) in relation to ETDs, which causes inconsistent interpretation. The obligation to make ETDs accessible
at universities is not precisely expressed. The requirement is ‘to make theses accessible via a database’. What kind of database? Full texts, or perhaps only metadata? This results in different attitudes at the schools, and various solutions are in place: free Open Access to full texts; access only on the basis of license agreements; access only via the intranet of a particular school; access only to metadata; no ETD publishing at all.

Non-text theses and access to them: The integration of non-text theses, or rather objects, into the institutional repository was solved by the Academy of Performing Arts in Prague in cooperation with the ETDs WG. It is still necessary to incorporate standards for non-text works into the ‘EVSKP-MS’ standard.

Other relevant background issues include:

- E-version collecting is not yet obligatory at schools.
- Not all universities have their own digital repository.
- Open access to repositories is not yet common. It is necessary to introduce the implementation of OAI-PMH as a standard.
- Important universities (for example Charles University, the Czech Technical University) do not yet participate in the national register ‘Theses.cz’.
- Locally, services are only partly automated (metadata is commonly handed over manually, for instance).
- DSpace is the most widely used software for repository management in the Czech Republic. There is a Czech DSpace Users Group which has 25 members and which organises e-conferences and annual seminars and which furthers the exchange of experience.

Managing e-Theses at the Academy of Performing Arts in Prague

Introducing the AMU

The Academy of Performing Arts (AMU) in Prague is a university-level school of Music, Dance, Theatre, Film, TV and Multi-media studies and the largest performing arts university-level institution in the Czech Republic. Its graduates include several prominent performing artists and world-famous
personalities, e.g. Milan Kundera, Miloš Forman, Jiří Kylián and Václav Havel. The AMU consists of three faculties: the Music and Dance Faculty, the Film and TV School, and the Theatre Faculty. Students may enrol in bachelor, master, and doctoral study programmes. The Academy has its own Library which consists of three departments located in the faculty buildings. IT support is centrally provided by the Computer Centre of AMU. Both the Library and the Computer Centre have a leading role in the implementation of new information technology at AMU.

As noted above, AMU actively participates in the ETDs WG, especially in projects relating to collection issues, metadata description and electronic access. AMU plays a key role in addressing this particular field of digital methodology in the Czech Republic. The AMU Library also closely cooperates with the Computer Centre of AMU, especially with the co-author of this paper.

AMU ETDs Specification

Due to the nature of the curriculum at AMU, students and also teachers continuously produce large amounts of projects and studies, including final bachelor’s, master’s, and doctoral theses. These works can be in very different formats: texts, audio recording, videos, photos, scores, etc. Several works are required to gain one qualification.

There is always a mandatory written work, but practical artistic works or projects are also required for each defence. These can be recorded as artistic performances, such as a graduation film, a stage role, graphic material, a concert, a set of photographs, etc. It seems appropriate to use the term ‘qualification performance’ to mark together all the qualification parts which belong to one student for his or her defence. It is important to be successful in the practical work, even more important than in the written one. Sometimes there is a collective effort, which contributes to the graduation of several students (for instance a film might be shared by a would-be film director, a screenwriter, a producer, a sound engineer, etc.).

Collecting and preserving such recorded works, easy searching, and the opportunity to preview them are issues of obvious importance, just as with ‘traditional’ textual theses.
The Starting Situation at AMU

We already solved, in a first pilot project in 2002, issues of access to selected theses at AMU. Some of our past theses are very significant and we intended to highlight them and so contribute to raising the prestige of our artistic university. They were not born-digital works. The project was one of three which aimed to create a base for the AMU digital library. We wanted to follow general trends and also to make our ‘family jewels’ more visible. One of the first works digitised was a screenplay by Václav Havel, the first Czech democratic president. The scanned texts were put on the website in a so-called ‘Electronic Reading Room’.

Two years later, in 2004, AMU began the mandatory deposit of theses in electronic form, as one of the first Czech universities to do so. We also prepared a workflow methodology and internal rules and instructions to support the whole process at AMU. The situation was changed quickly after the above-mentioned Amendment at the beginning of 2006, but thanks to our earlier activities, we were relatively ready to develop tools for making AMU’s thesis output accessible.

Legislation and Rules

The question of property rights was the first issue to be resolved. This was a very delicate task, due to the artistic character of the works. Many of them can be artistically significant. They can later become well-known. Therefore we had to decide how best to deal with copyright. The AMU solved this issue in cooperation with a team of lawyers from the Law Faculty of the Charles University in Prague. Their recommendation was that we should conclude a license agreement (LA) with all authors for accessioning from them in digital formats, and a license agreement for AMU was devised. The whole procedure is a little complicated: it involves a main licence component, which has to be signed by all students at the beginning of the study, and also partial licences for each work, each signed twice by an author and by the Dean. The LA is being generated and filled inside the AMU Study information system ‘KOS’ directly at the point of thesis submission. The author can choose from four options which define the way in which AMU may make use of each separate work. The authors are able to forbid access to works (whether in printed or electronic formats), and such outputs will
be only archived. The state of the rights is also displayed in the metadata records in the repository.

The second step at AMU was creating the internal regulations and instructions to support a theses workflow. It was necessary to give the authors guidance on what they should do and how they should do it. In close cooperation with the university’s senior management, detailed instructions were drawn up for everyone included in the workflow: administrative staff, students and teachers. All workflow participants also have ongoing support at their disposal on AMU web sites. Of course, all official university documents, including four rectors’ decisions related to theses, are available through the web, but there are also explanations, templates and manuals for starting authors, help for citing sources correctly, instructions about formats, a manual for filling the metadata into ‘KOS’, and so on. The library and IT staff have arranged a lot of briefings to explain all the details. This guidance work remains very challenging, and more staff resources are needed to deliver it effectively.

Non-Text Theses and AMU University Repository

Metadata Description and Collection

The first task to complete was setting up the methodology for describing and collecting metadata. I would like to focus on the problem of metadata description of non-text works in detail here, as they present a lot of specific problems, with implications for searching and therefore availability.

As mentioned above, almost every student of any study programme at AMU submits, besides the written ‘theoretical’ thesis, a practical, non-text part at the end of each study level. It is very desirable to capture a lot of detail about all the parts to support their unambiguous identification. For description we use the Czech metadata standard ‘EVSKP-MS’. It is based on the DC format and it has some special added elements. For more information about EVSKP-MS, see the web sites of the ETDs WG.

Originally we tried to put all the descriptive information about all parts of the ‘qualification performance’ at AMU into a single metadata
record. This is a common solution in cases when a work has conventional attachments or appendices. The problem is that these two different types of qualification work are individual and independent in topic, content, and form. Therefore, to describe them by only one record, together with the text, would be insufficient. It was clear that such a solution would lead to information chaos and that it would not be possible to identify which data belonged to which part of the ‘qualification performance’.

Therefore we tried to find another solution. We explored the possibility of the DC format to find some elements to be able to help us. In the end we decided to use the element ‘dc.relation’ and its attributes ‘has.part’ and ‘is part of’. We also decided to describe each part of the ‘qualification performance’ with a separate metadata record and then to connect each record using the relations element. This means that each student is always obliged to fill in more than one thesis metadata record. Although quite unusual, the method seems to be very effective in terms of supporting accessibility: it is possible to recognise all necessary and relevant details in the description easily and clearly. And, more importantly, we have no problems recognising the relations between works and records — which abstract describes which part, and so on (see Figure 1).

For metadata collection, we use the AMU Study information system ‘KOS’, in which tools were created especially for collecting theses metadata. These consist of separate forms for each type of work, a browser, and a procedure for generating licences. We have set a working classification for the basic types of the theses. The written work is the ‘supreme’ component, and it is marked as the main type, as type ‘A’. For the identification of all other types, separate formats are available for selection, listed alphabetically. In practice we resolve a total of 5 types: form ‘A’ for theoretical written works, form ‘B’ for other textual works like screenplays, form ‘C’ for films or audiovisual works, form ‘D’ for interpretations or performances, form ‘E’ for compositions. Each form includes some special fields (for example, form D captures the titles of all the music compositions performed, all the performers’ names and so on). This method ensures more than standard description. It is very useful in creating detailed bibliographic records for the library catalogue, and such data are also very important for archiving, thinking about the historical perspective.
Director and copyright law

In the licence agreement the author has granted the following: a) by copying the University Work in any media and any form, including electronic forms, b) by disseminating the original of the University Work in any media and any form, including electronic forms, and c) by using the original of the University Work or copying it as described in a) hereinafter, by relaying it to the public in all forms stated in the provision of Sections 18 to 23 of the Copyright Act, particularly by the live performance of the University Work and its transmission and by publishing the University Work so that anyone can have access to it in a place and time of their choice (i.e. directly on the Assignee’s websites on the Internet).
Software for Access and Archiving

The most important thing was to choose software that would be easy to adjust for our specific output and which would have some archiving functions. We would have been happy to find a tool which could be used for a quick search of the documents and an easy assessment by users of their relevance and availability. We also needed a system that would be suited for exporting and importing metadata records via XML in order to cooperate with other AMU systems. We needed to comply with the OAI-PMH protocol to support harvesting. Finally, we needed a system for the unique persistent identification of the records in web and internet space.

In the end we decided to choose DSpace as the institutional repository platform. There were several reasons for this decision:

- DSpace is used at several universities in the Czech Republic
- The existence of the Czech DSpace User Group, which offers opportunities to exchange experiences
- DSpace is Open Source, so it is relatively cheap to implement
- There are no other regular costs
- DSpace has wide usage
- It has detailed documentation
- It is easy to install, administrate, and customise
- It is not difficult to modify and maintain
- It natively supports common standards for institutional repositories

As a persistent identifier we have chosen the Handle, recommended for Dspace.

Software Modifications

We had to make some modifications to Dspace after its initial implementation. The default Dublin Core metadata set was extended to include additional elements, mainly the relation element qualifiers.

The relationship between the records could be established in two ways:

1) by creating a virtual object for each ‘qualification performance’ and linking all real objects from it;
2) by using the attribute ‘has part’ for the ‘superior’ metadata record of the textual work, and the element ‘is part of’ for the ‘subordinated’ metadata records of the non-text works.

We chose the second method. The metadata records of the textual works are designated as the main ‘superior’ records, and the other, non-text records, as the ‘subordinated’ records. Each superior record contains a link to the subordinated records, and vice versa.

The design and layout of DSpace was adapted to the AMU corporate identity. The internal organisation of faculties and departments was adapted into DSpace. The modules for editing and display were modified to permit the searching and viewing of the related documents. Video and audio players were integrated into the system, so that all related multimedia files are available directly from the primary record page. A Handle server generating persistent identifiers was implemented into DSpace. An English-language version was created.

It was also necessary to establish which e-formats would be acceptable for submission. For the written text, this is PDF, which should be created in the standard form on our server to ensure full-text searchability in the Czech language. For video records we use FLV or MP4 format (720 × 576px D1-PAL), 1500 kbps; for sound documents, MP3 format. Our intention is to provide access formats for preview; it is not possible to store native formats in the repository because that would put an enormous strain on the available storage space. In our opinion, this solution provides sufficient ‘information’ about the non-text work and, in cases of further interest, it is necessary to ask for a native record of the work at the department.

Cooperation with Other Systems

One of our goals was to transfer metadata from our repository to other databases, especially to the ‘Theses.cz’ national register. This interoperability is achieved by using the OAI-PMH protocol. DSpace has its own OAI server which enables the retrieval of metadata in Dublin Core. The Java plugin was modified to process the metadata into EVSKP-MS. Our records are now automatically harvested by Theses.cz. The national register checks our OAI server every 24 hours for updates.
Currently, we only transfer metadata about text works metadata into Theses.cz. Neither Theses.cz nor EVSKP-MS are ready to use and display the dc:relation element. We have submitted a proposal to incorporate these elements in the Czech metadata standard and in Theses.cz. Thanks to the OAI-PMH-compliance we are ready to provide data to any other harvester: it is only a question of negotiation. Access to the full version of the works is provided by links back to the institutional AMU Repository.

Metadata from DSpace are also transferred to the library catalogue. We use the T-series system, which can import in XML format. It was only necessary for us to devise templates for the transfer. The relations between the records are transferred into the AMU Library catalogue as well. The ‘main’ (superior) record represents a record on a monograph level, the others on an analytic level.

**Workflow Recapitulation**

The complete workflow model is as follows:

- a record for each thesis is placed in the Study information systems ‘KOS’, on the approval of the work, by the student’s department
- it is necessary to select the correct work-type form
- basic data about students are filled automatically by the system (level of study, programme of study etc.)
- every student supplies other data about all parts of the ‘qualification performance’
- students submit prescribed particulars, including an e-version in PDF format (on CD-R) at their departments
- a secretary has to check all the details and to load recorded works onto dedicated drive space
- metadata records are exported from ‘KOS’ in XML format (this export is not yet automated)
- the records are automatically completed with general default data (the graduation level, the name of the university and so on)
- the records from KOS are transferred in batches to the AMU Dspace Repository
- in DSpace a Handle is assigned and standard subject headings are added by the librarians
- full versions of the work are uploaded into Dspace (manually)
the metadata are transferred from DSpace via the OAI-PMH protocol to the national system Theses.cz, and also to the AMU library catalogue in bibliographic format.

Prospects

Finally I would like to mention briefly some issues that we have not as yet resolved. First, our License Agreement enables a student to set many options as to how we should treat an individual work of art, but we have no tools to create a similarly wide range of access rights within DSpace. It is necessary to simplify the administration and allocation of access rights and their roles. Second, DSpace is unable to display the hierarchy of related records. Third, we are unable to automate the loading of records into DSpace — it is necessary to do it manually. All these problems relate to software programming and modification; it is only a question of time and people. Another problem is that we have discovered many past mistakes in respect of data entry. In the Study information system KOS, some controlled language elements were set; however, perhaps this shows that more training is required for students, teachers, and administrative staff alike.

The most substantial problem is that of long-term preservation. Especially with regard to performance outputs, with their special work types and formats, it is complex. We try to keep up with standards and trends, but it is a general problem all over the world.

Now a few words about our future plans. Several school departments (for example, the Department of Photography) are interested in having their own digital archive to store their whole output. We would like to cooperate with them and offer them use of our DSpace implementation. However, the first issue that will need to be solved will again be that of metadata description. We would like to continue with our extension of metadata description and to display more details in DSpace in relation to non-text works. Much more metadata is collected about such works in KOS than can be transferred to and displayed by Dspace — yet. We intend to work with the DSpace.cz Group on these matters. We also plan to take part in a new initiative, Open Access.cz, which at the time of writing is forming within the Association of the Libraries of the Czech Universities.
If I may say so, we are ready to solve — with respect to the technology issues — practically all the problems relating to the non-text digital objects and their accessibility. Sometimes, however, we face a lack of readiness on the part of students and even teachers. Sometimes we need stronger support from our university management. Sometimes we meet a disinclination to accept new trends and methods. On the other hand, problems like this are perhaps to be expected, considering the type of school, where art is foremost and artistic individualism is encouraged. We are still in the early stages of building the AMU University Repository. The solution presented here is the first — although very important — step on the path forward. Taking the repository forward, like encouraging the acceptance of open access, is an ongoing process. The main issues seem to be human, and the main challenge to open people’s minds.

References


Websites Referred to in the Text

Notes

1 There is an e-mail conference DSpace.cz with 25 members, and the DSpace CZ website is designated for Czech users and provides: a list of installations, a list of Czech localisations, information on conferences and meetings, DSpace news, external links to registries, service, providers etc. Annually a Czech DSpace User Group Meeting is organised with the support of the Association of Libraries of Czech Universities.