

The Transfer of Knowledge from Large Organizations to Small: Experiences from a Research Project on Digitization in Wales

Clare Wood-Fisher[†], Richard Gough^{‡†}, Sarah Higgins[†], Menna Morgan[‡], Amy Staniforth^{‡†} and Lucy Tedd[†]

[†]Department of Information Studies Aberystwyth University, Aberystwyth, Wales.
clw09(at)aber.ac.uk

[‡]Department of Collection Services, National Library of Wales, Aberystwyth, Wales.

^{‡†}Centre for Performance Research, Aberystwyth University, Aberystwyth, Wales

Abstract: *This paper describes a research project being undertaken for a Masters of Philosophy degree which investigates how knowledge can be transferred from large organizations to small ones. The area being investigated involves the digitization of photographic collections held by cultural memory institutions in Wales. The researcher has spent time learning about digitization and digitizing two collections at the National Library of Wales. From this a toolkit has been developed and is being tested by digitizing a small archive of photographs at the Centre for Performance Research at Aberystwyth University. The research is being funded by the Knowledge Economy Skills Scholarship (KESS) programme, which aims to build knowledge and skills in the 'convergence' area of Wales. Supervision is being carried out by academic staff at Aberystwyth University.*

Keywords: *Digitization; KESS; Performing arts; Wales.*

I. INTRODUCTION

The three partners collaborating in this Knowledge Economy Skills Scholarships (KESS) funded work to create a strategy for the digitization for small collections whilst still upholding the highest standards are:

- National Library of Wales (NLW)
- Centre for Performance Research (CPR), Aberystwyth University
- Department of Information Studies (DIS), Aberystwyth University.

Each of these partners provided their specialist skills for the project. The NLW has provided practical expertise as well as ongoing support; CPR has provided equipment and research space; while DIS staff with relevant digitization experience have provided guidance and support throughout. DIS is a renowned department for the education of archivists, records and information managers and digital librarians from all over the world.

The researcher (Wood-Fisher) is registered for a Master of Philosophy degree at Aberystwyth University and started the research in January 2011, with a completion date scheduled for December 2011. Before undertaking this project the researcher, a qualified Records and Information Manager, had worked within the information sector in a college primarily for 16 to 18 year olds. These experiences had given some knowledge

of the technicalities of photographic processing, as well as those of picture composition, but these skills had not been combined in the creation of digital objects. She also had little knowledge of archival processes.

The researcher spent the first four months at NLW, then spent some time developing the 'toolkit' which was used to develop a digitized collection of archival photographs at the CPR. The academic aspects of the research are being supervised by staff at CPR and DIS and the researcher has undertaken Aberystwyth University's research training programme.

II. BACKGROUND TO KESS

KESS is a programme funded under the European Social Fund Convergence Programme by the Welsh European Funding Office. It runs from 2009-2014 and aims to support 400+ doctoral and research masters students from the specific 'convergence' areas (i.e. in West Wales and the South Wales valleys) of Wales. KESS aims to support the transformation of Wales' economy into one based on knowledge and skills. The KESS concept is achieved by collaborative research jointly sponsored by partners based in the convergence areas of Wales and in Welsh universities.

As stated on the website of KESS at Cardiff University (<http://www.cardiff.ac.uk/racdv/expertise/kess/index.html>) its key objectives are to:

- "increase the research capacity of organisations by linking with a PhD / Masters project;
- encourage organisations to undertake research and recruit researchers;
- prepare and train individuals to contribute to research as professionals;
- support the development of key technologies in the convergence area;
- promote higher-level skills development."

III. WORK AT NLW

The NLW is one of the primary cultural memory institutions in Wales. It houses the National Screen and Sound Archive of Wales and the National Archive of Wales as well as having a legal deposit function for books published in the UK. Broadly speaking, the NLW collects objects of cultural significance to Wales and

Welsh speakers everywhere on whatever the medium it is recorded.

The NLW's first digitization strategy was written in 2001 although there had been digitization projects taking place since 1997 (Jones, 2008). Since then it has embarked upon a continuing programme of digitization with the aim of increasing access to, and awareness of, the items within its collections. There is a specific unit to carry out the digitization work within NLW and this is viewed as being part of a whole library approach as the process interacts with all the other departments. Within the Library digitization was seen as being at the heart of the NLW's vision for the period 2008-2011 (National Library of Wales, 2009). The range of materials that have been digitized at NLW is very broad and includes manuscripts, archives, printed material, sound and video recordings, framed works of art, maps and photographs. In addition specific projects have been involved in digitizing two million pages of historic newspapers in Wales, 4,000 Welsh ballads and material on Welsh settlers in the US state of Ohio. A significant proportion of materials has been processed for Optical Character Recognition (OCR) and tagged using Text Encoding Initiative (TEI) to aid searching. Over 1 million items are publicly available via the Digital Mirror area of the NLW Website (<http://www.nlw.org.uk>).

As can be appreciated the NLW has some very large collections. For these to be managed effectively during a digitization project there is a need for specific methods to be developed to monitor and track the workflow. The workflow adopted by the NLW allows for individual digital objects as well as for the archival materials themselves to be tracked throughout. This ensures that those involved in the process know what to expect from a project and where they have reached in a particular sequence of interactions. NLW uses its own digitisation workflow and editing software, known as WOMBAT, which was developed in-house to monitor the progress of tasks and items.

Scanning is carried out in a dedicated suite where a variety of scanners and techniques are employed to create the digital objects. The selection of the technique used for particular formats of archival materials (e.g. photographs, newspapers, books or manuscripts) is determined in consultation with all the departments involved in the complete process. Included in this discussion are decisions on the equipment to be used to obtain the best possible scan without causing damage to the archival material.

The first step in the research process involved acquiring the necessary skill set to carry out a digitization project. To do this a stimulating period was spent by the researcher being trained, at NLW, in the digitization process from the beginning to the end; through scanning and adjusting images, to make the required copies, to writing and researching the metadata.

In order to demonstrate the new skills on digitization, learned by the researcher, the decision was made that she would be responsible for digitizing two collections: the H.W. Lloyd collection of approximately 400 glass negatives which include images of a prisoner-of-war camp in Bala, North Wales and the Dwywnwen Belsey collection of 180 slides, postcards and photographs of Welsh settlers in the Patagonia area of Argentina. For both of these collections the NLW had been given copyright.

Creating digital objects from the glass negatives was undertaken under the direction of staff in the Digitization Unit using a designated scanner with adjustments such as cropping the scan to the edge of the negative following on. All work had to meet the high standard required by NLW, which was not lessened for a first timer!

Having created all the digital objects specified, the process for the researcher moved to creating the metadata. This was carried out under the direction of staff in the Cataloguing Unit who gave instruction on the art of writing enough, but not too much, yet remaining accurate and concise in the description of digital objects and archival material. Consistency and succinctness was achieved using writing rules and a controlled language. This phase proved to be most rewarding as previously hidden aspects of the glass negatives that digitization had revealed could be indicated to the prospective audience in the description area of the record created. A key requirement of this skill is to be able to place yourself in the position of an NLW user and think of, and include, the search terms that might be used to access resources. Metadata is created using MARC21 and marked-up using Metadata Encoding and Transmission Standard (METS) for internet delivery and long term preservation.

It was interesting to the researcher during this phase to observe processes, which to staff at NLW are second nature, but may not be in smaller organizations. Sometimes this was easy to spot, for instance the way the whole process prioritizes the conservation of the primary source. At others it was not so easy and required considered thought to tease out, for example deciding how much to write into the description fields for the catalogue metadata.

IV. WORK AT CPR

The CPR was founded in Cardiff by Richard Gough in 1988 and now forms part of the Department of Theatre, Film and Television Studies at Aberystwyth University. As stated on its website (<http://www.thecpr.org.uk/index2.php>) the "CPR aims to develop and improve the knowledge, understanding and practice of theatre in its broadest sense, to affect change through investigation, sharing and discovery and to make this process as widely available as possible. Its programmes of work combine cultural co-operation, collaboration and exchange practical training, education and research, performance, production and promotion, documentation

and publishing, information and resource.” In contrast to the scale of NLW, the resource centre and library at CPR has a small, unique collection of materials focused on national and international contemporary theatre and performance. It has five main collections: international printed resources; the International Theatre Collection; the Performance Research Archive; the Cardiff Laboratory Theatre Archive and an archive of the Giving Voice Festivals (Centre for Performance Research, 2011). With the exception of the Giving Voice Festivals archive the archival materials collected focus on the visual aspects of performance. As such being able to see the materials is of prime importance to the user.

The concept from the start was that the researcher would research and develop a strategy, or toolkit, to digitize CPR’s photographic archive. A particular collection of photographs at CPR was identified at an early stage as being appropriate for acting as a ‘case study’ for the toolkit.

This collection, of some 80 black and white and colour prints and 150 slides, covered performances and workshops held at Blaenau Ffestiniog (in North Wales) in 1981, and known as the Blaenau Ffestiniog residency. Although the collection to be used had been pre-selected, fundamental questions regarding the aims and objectives of digitizing it still needed to be answered. Additionally copyright aspects of the project had not been fully investigated. A Digitization Policy has yet to be developed at CPR to enshrine how items for digitization will be prioritized. The Project’s starting point was to establish what was envisaged for the digitized materials, and further digitization beyond the life of the project. Having established that the eventual aim is to digitize all the archive holdings, work could begin on assessing and discussing the options open to CPR regarding the process and how they could feed into the projected future.

Copyright issues also needed to be resolved to ensure that the relevant permission was granted to digitize the photographs and slides in the collection.

Where to store the digital objects created at CPR was discussed in depth to ensure that they would be secure, useable and, if required, easily transferable in the future. It was decided that the digital objects for viewing on the Web would be stored on the University’s network and other copies off-site. This would ensure that in the case of a University failure where the digital objects could not be restored; there would be a back-up safely away from the system. Off-site storage would also be used to store a copy of the metadata on the digital objects.

Having obtained copyright and storage for the digital objects the scanning phase began with a test of the equipment that already existed within CPR. However, for various reasons this proved not to be suitable and instead it was found that excellent results could be obtained from a good domestic scanner which was then used instead.

At the same time as discussing storage for the digital objects the adoption of a metadata profile was also considered. It was decided to adopt the Dublin Core Metadata Element Set as this constitutes the minimum needed for metadata exchange. Another consideration regarding the choice of Dublin Core was the decision by CPR to become a partner of the European Collected Library of Artistic Performance (ECLAP). ECLAP is an EU-funded project that aims to promote best practice for performing arts materials online through a central resource (European Collected Library of Artistic Performance, 2011). This repository is still finalizing its metadata model but is basing its choice on Dublin Core too. The researcher attended a conference organized by ECLAP in Brussels in June 2011.

The initial decision was to capture the metadata into a database developed using Open Office software. This was dropped in favour of a spreadsheet, also developed using Open Office software, as it was felt that this would be easier for CPR staff to use and effect the transfer of data in the future. Collecting technical metadata on file size and format was undertaken using the DROID tool from the UK National Archives (National Archives, 2011). The other, more descriptive, metadata was collected by interviewing Richard Gough who had directed all the productions throughout the Blaenau Ffestiniog residency. This interview was recorded as an audio file and added to the collection as commentary on it, as it contained so much rich data that it was impossible to do it justice in a field on a form.

Access to the digitized material and its metadata will be provided initially via the CPR WebPages on the Aberystwyth University website. The digital objects and the accompanying metadata will be accessible within the CPR as soon as it is created but, eventually, when a core of digitized materials is acquired, it is hoped to install repository software to facilitate access to more of the resources rather than a selection.

V. TOOLKIT DEVELOPMENT

The development of a toolkit, to enable the staff in CPR to carry out the digitization process for themselves, has been a rolling rather than static process. Despite trialing the scanner prior to beginning the digitization at CPR it was decided very early on that there was a better method available. So the decision was made to switch to another scanner that proved to be easier to use. It was easier largely because the preview of the scan was of a reasonable size and any adjustments were shown in a preview before being accepted.

Writing the metadata was also simplified as far as possible while still containing the minimum needed for successful exchange in the future, as well as the metadata needed for preservation. The choice to use Dublin Core was taken before ECLAP decided to adopt this schema, but bore out the view already taken that it offered both the most accessible way for novices to learn and understand metadata and having viable crosswalks to most other schema.

VI. CONCLUSIONS

The research process is still in progress, but it is not too early to say that there has been a measure of success in the transfer of skills and methodology both innate and externally expressed. One of the major results to emerge most recently is that the time taken to run a large test was time well spent. This has proved to be especially so in this case where the workflow adopted was adjusted several times over the first week of the research at CPR to reflect changes in equipment.

In general the research process has been a positive one reflecting how cooperation and support from large national bodies can help the smallest of collections become more accessible for its users and gain new audiences.

REFERENCES

- [Centre for Performance Research, *CPR Collections*, Aberystwyth University \(2011\).](#)
- [European Collected Library of Artistic Performance \(ECLAP\), *European Collected Library of Artistic Performance* \(2011\).](#)
- [Jones, R. A., "A marathon not a sprint: Lessons learnt from the first decade of digitisation at the National Library of Wales", *Program: electronic library and information systems*, **42**, 97-114 \(2008\).](#)
- [National Archives, *The Technical Registry Pronom*, National Archives \(2011\).](#)
- [National Library of Wales, *Digitisation Strategy 2008/9 - 2010/11*, National Library of Wales \(2009\).](#)