



Provided by the author(s) and NUI Galway in accordance with publisher policies. Please cite the published version when available.

Title	The Abbey Theatre digitization project in NUI Galway
Author(s)	Keane, Aisling; Bradley, Martin
Publication Date	2015-12-17
Publication Information	Keane, A; Bradley M (2015) 'The Abbey Theatre Digitization Project in NUI Galway'. The New Review of Information Networking, 20 (1-2).
Publisher	Taylor & Francis
Link to publisher's version	http://www.tandfonline.com/doi/full/10.1080/13614576.2015.1114827
Item record	http://hdl.handle.net/10379/5574
DOI	http://dx.doi.org/10.1080/13614576.2015.1114827

Downloaded 2019-11-15T02:33:59Z

Some rights reserved. For more information, please see the item record link above.



The Abbey Theatre Digitization Project in NUI Galway

Martin Bradley

Mastermind Ireland Ltd., Dublin, Ireland

Aisling Keane

James Hardiman Library, NUI Galway, Ireland

© Martin Bradley and Aisling Keane

Address correspondence to Aisling Keane, James Hardiman Library, NUI Galway, Ireland. E-mail: aisling.keane@nuigalway.ie

ABSTRACT

National University of Ireland, Galway and the Abbey Theatre finalized a partnership to digitize the archive of the Abbey Theatre in 2012. The partnership leverages NUI Galway's position as a leader in theatre and digital humanities research and home to a range of theatre

archives. The result of the project is the creation of a major international resource for teaching and research with Irish drama, literature and history.

The project addresses two particular challenges faced by the physical archive: preservation and access. It allows unprecedented levels of access to the archive which until now, has been severely restricted due to space constraints and cataloguing backlogs.

The archive contains more than a million pages, 500 hours of video and 2500 hours of audio. The material ranges from posters, programs, photographs, minute books to lighting plans, set and costume designs, sound cues, prompt scripts and audio files, see Figure 1.

KEYWORDS

Digitization, Project Planning, Workflow, Intellectual Property, DAMS, User Experience

MASS DIGITIZATION OF ARTS ARCHIVES: THE ABBEY THEATRE DIGITIZATION PROJECT IN NUI GALWAY

Scoping, Planning and Management

Scoping a project of this scale is basically a mathematical challenge. Various sources exist for the estimation of pages by paper volume, which are a useful indicator of potential timescales involved in a digitization project (Paper and Image Estimator, published 29 August 2006, accessed 9 November 2015). Of course, what these estimators cannot predict is the extent of duplicate material that may be found that will not require digitization. According to a 2002 study by the University of California at Berkeley the average organization makes 19 copies of each document and loses 1 out of every 20 documents

(Bradley, 2002, accessed 9 November 2015). However given the differing time requirements of each format (loose pages at up to 600 pages per day, whereas photo negatives can take 6-7 minutes each) the paper image estimator is a useful tool for broadly calculating timescales with reasonable accuracy.

Equipment: A range of digitization equipment was purchased for the project, and some pre-existing equipment was found to be unsuitable. Scanners and cameras need to be able to handle the physical characteristics of each individual document type, while delivering an appropriate image resolution for each type. It was discovered that a combination of tripod mounted digital cameras with desktop studio lights, A4 and A3 flatbed scanners, A3 2400 dpi negative scanners and an A0 scanner were sufficient to meet the requirements of the paper-based elements of the project. Initial attempts were made to use a book scanner, but the irregularities of paper sizes and bindings meant this equipment was found to be unsatisfactory. Audio and Video capture required both specialist material and the outsourcing of specialized industry knowledge – particularly for the capture of ¼ inch and ½ inch audio tapes which run the risk of sticky-shed syndrome and potential damage to source material through digitization.

Staffing: It was decided that using professional qualified archivists to carry out digitization was the best approach to addressing concerns around the handling and preservation of fragile material during capture, while also ensuring a high level of attention to detail in terms of referencing over the course of a very long and repetitive process. Staff were set daily targets based on measurements of maximum capacity based on equipment types, document types and fragility of material. There has been a medium level of staff turnover during the course of the project as a very particular mindset is required to maintain concentration levels while

conducting repetitive work and many archivists have yet to adjust to the demands mass digitization work imposes, though it will undoubtedly increase in being a prominent aspect of the day-to-day work of the profession.

Data storage: It was decided to outsource data storage to Amazon S3 to allow a greater guaranteed “up” time, greatly reduced maintenance, upkeep and capital costs and a secure storage environment. Images of paper based materials were created as 600 DPI archive master TIFFS and then as PDF for delivery through the DAM. Negatives were captured at 2400 DPI, audio as 48kbps Wav and MP3 for web delivery and video as MP4. Cloud storage also serves to minimize outlay on servers and infrastructure, as it is anticipated that on completion of the 3 year project total data storage requirements will exceed 40TB. In comparison the Hubble Space Telescope Data Archive has amassed 1.2TB of data since 1993 (Borne et al, 1995, accessed 9 November 2015).

Digital Asset Management System: Aetopia Ltd., based in Belfast, was awarded to the contract to develop the user interface for the Abbey Theatre Digital Archive following a competitive process. The bespoke DAM which was developed following a lengthy process involving NUI Galway Staff and Aetopia is now available for use by researchers in the Archives Reading Room in the Hardiman Research Building and is undergoing a continuing series of improvements.

Metadata: This project was fortunate, in that the Abbey Theatre had already compiled a comprehensive relational database of all its productions, plays, authors, cast, crew and locations. It was realized early on that it would greatly aid the speed of the digitization process if items could be linked directly to this database, simply by saving them with a prefix

which denotes the document type and either a production code or a play code, where the specifics of a particular production are not known. The DAM identifies whether the code is a play or production based on the number of digits, and identifies the document type based on the code – for instance it will recognize that 4247_MPG_01 is the program for “A Pot Of Broth” performed on January 6,7 and 8th 1910, was written by W.B. Yeats, and featured the actors Máire O’Neill, J.A. O’Rourke and Arthur Sinclair, and was staged at the Abbey.

As a result, it will be possible to then search for other plays featuring the actors, writer, cast, crew or location of the performance, as well as browsing other document types that relate to that performance (for instance set designs, prompt scripts, audio, video etc).

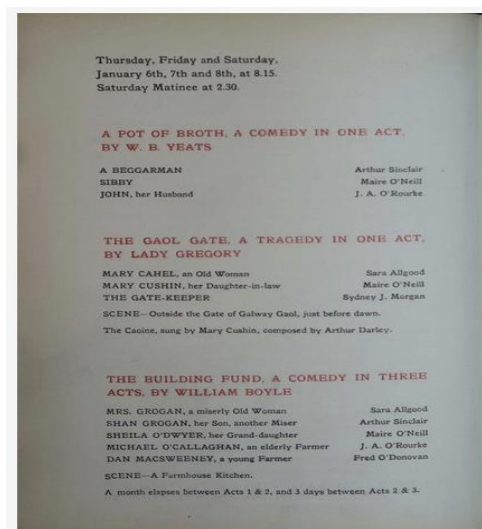


Figure 1. Abbey Theatre. *A Pot of Broth*, 06 Jan 1910 [program].

Where items can be linked directly to an Abbey Theatre play code or production code this has been done, which allows individual items to be searched by the full production history associated with each event (cast, crew, location, writer, covering dates) and associated with items which share any points in common (e.g. shared writer, set designer, costume designer etc). This allows easy and swift retrieval of items relating to specific search queries. Where it

was impossible to link items directly to Play or Production Codes new metadata has been generated along with the ability to link to Plays or Productions that may be mentioned within the body of a file, but that do not comprise the exclusive content of that file. ISAD(G) fields have also been captured for all material that is not directly linked to a Production Code. All typescript items are also subject to OCR (Optical Character Recognition) as a third layer of searchability, which allows keywords to be searched across the entire content of the archive.

Legislative Considerations

Intellectual Property: All copyright material presented as part of the Abbey Theatre Digital Archive is being made available solely on campus at NUI, Galway under the provisions of Section 50 of the Copyright Act (2000) - *Fair dealing with a literary, dramatic, musical or artistic work, sound recording, film, broadcast, cable programme, or non-electronic original database, for the purposes of research or private study, shall not infringe any copyright in the work.* (Copyright and Related Rights Act, 2000, accessed 9 November 2015).

Data Protection: A comprehensive retention schedule has been developed to allow the release of different document types according to defined release dates, based their contents. Further to this material is searched automatically, using OCR text, to find telephone numbers, addresses or other information which should be withheld under the terms of the DPA. All users are required to agree to comprehensive Terms and Conditions at each log-in to the site, which cover acceptable usage, copyright, copying, permissions and access control. A log is maintained of every user, each search carried out, each document viewed and the times of each action. In the event of any breach it will be straightforward to pinpoint an offending user.

Design of Digital Asset Management Systems: All data are hosted on the Amazon S3 cloud storage service, as are instances of the Abbey Theatre Productions Database and the bespoke Digital Asset Management (DAM) system. This maximizes system up-time as all maintenance is carried out by Amazon according to a defined service level agreement. It also serves to minimize outlay on servers and infrastructure, as it is anticipated that on completion total data storage requirements will exceed 40TB. Aetopia Ltd., based in Belfast, was awarded the contract to develop the user interface for the Abbey Theatre Digital Archive following a competitive process. The bespoke DAM which was developed following a lengthy process involving NUIG Staff and Aetopia is now available for use by researchers in the Archives Reading Room in the Hardiman Research Building.

Release Scheduling

Possibly the most complicated aspect of the project so far has been the design of the release scheduling systems built into the DAM. See Figure 2. Initially material was broken down into two categories – “Production” and “Administrative”. Production items are ones that can be linked to a specific production or play, and comprise programs, video, audio, scripts and other material that does not have a retention period or potential for DPA issues associated with it. Administrative material broadly reflects the workings of the Abbey Theatre behind the scenes, and as such contains large quantities of potentially sensitive material which cannot be released – for instance HR files – as well as material that must have set retention periods associated with it (e.g., 7, 15, 30 years) which periods were arrived at in consultation with the Abbey Theatre and taking into consideration all current Irish legislation on the retention and publication of records.

Further, it was necessary to implement a number of automated searches across most document types (with some exceptions) for phone numbers, addresses and email addresses to ensure no breaches of the Data Protection Acts (1988 and 2003) were inadvertently committed.

Use of Optical Character Recognition Software, which is applied as a dual process – once at the point of digitization material is OCR-ed using Adobe, and secondly at the point of ingress to the DAM material is OCR-ed using Tesseract – allows for not only a much greater level of referencing of the archive, allowing keyword searching of all content, but also the detection of prohibited content. As the vast majority of the collection dates from the 20th Century, and is as a result typescript, OCR has proven to be extremely effective as an indexing tool.

There is an ongoing process of teaching the DAM to ignore content based on context is also underway. For instance, phone numbers commonly appear on headed paper—so the system has been taught to recognize the address of the Abbey Theatre and ignore this and associated phone numbers. As fresh instances of commonly occurring phone numbers are identified they are built into the ignore rules.

A review and publication scheme has also been built into the system to allow users to request editorial review for material that has been held back as a result of automated detection of a potential DPA issue. This ensures a response or resolution of the issue within a set timeframe whereby material can be ‘ok-ed’ for release, partially redacted (e.g., phone number removed) or held back where the item may contain sensitive personal information.

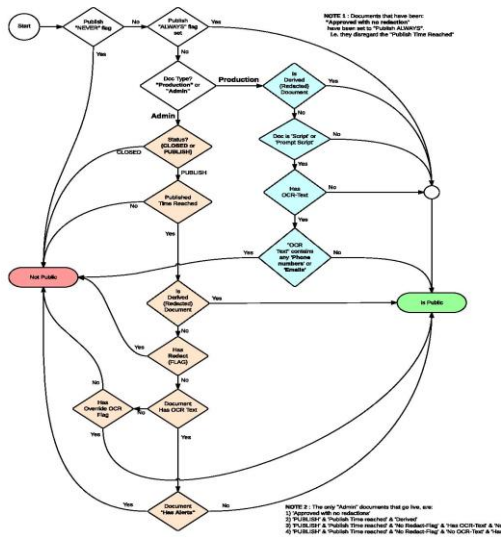


Figure 2. Release Scheduling Workflow Diagram

Digitization Progress

As of 31 December 2014, 587,397 unique pages of information, up from 308,195 at the end of 2013, had been digitized following two processes of weeding and removal of large quantities of duplicate material – firstly by the Abbey Theatre and secondly by archivists carrying out digitization work. See Table 1. Included in the available material is pre-1950 archival content from the collections of the National Library and the Programme Collection. In addition 704 audio recordings and 324 video recordings (the entire video collection) have been digitized and are available for consultation, along with 447 posters (the entire collection). To be completed in 2015 are the photograph collection, the handbills and fliers collection, the early fire-damaged stage management records from the original Abbey Theatre, lighting plans

The material captured between September 2012 and December 2014 breaks down into the following broad categories;

Table 1. Breakdown of digitized material at end of 2014

	Total at End of 2014
Programs	64,247 pages
Prompt Scripts	116,174 pages
Set Designs	1,652 pages
Scripts	116,040 pages
Lighting Designs	11 pages
Audio	704 recordings
Video	324 recordings
Venue Designs	98 pages
Posters	447 items
Photographs	1,241 items
Administrative Files	289,175 pages

User Experience

The first year of the Abbey Digitization project was devoted to establishing infrastructure and workflows to digitize over 300,000 pages. A year in, the material digitized to date was made available to researchers under supervised conditions in the Archives and Special Collections Reading Room at NUI Galway, and a greater emphasis was placed on providing and supporting access to the digital archive.

The main user groups are NUI Galway's students, researchers, and staff; external students and researchers; and academics. They vary substantially along the spectrum from absolute beginners to advanced users. The archives staff in the Reading Room provides one to one training to a new user, enabling them to understand and use the database, and become self-sufficient as their research progresses. A first time user receives a demonstration of logging into the Digital Archive, a sample search, and an overview of how to refine the content, and some of the functionality that allows them to track their research. The archivists have also given specific academic groups introductory sessions on using and searching the Abbey archive. These are most often new PhD and MA students, and on an ad hoc basis, various undergraduate groups from the College of Arts, Social Sciences, and Celtic Studies.

The access platform to the DAMS is an interface designed by Aetopia, which has to be intuitive and highly navigable in order to expedite access to the material. The interface accommodates both the 'search and explore' user who isn't looking for anything specific, and the research-led user who is conducting definitive searches.

See Figure 3 for a demonstration of how a search by play title generates over 300 results, but gives the option of refining the results by sub categories.

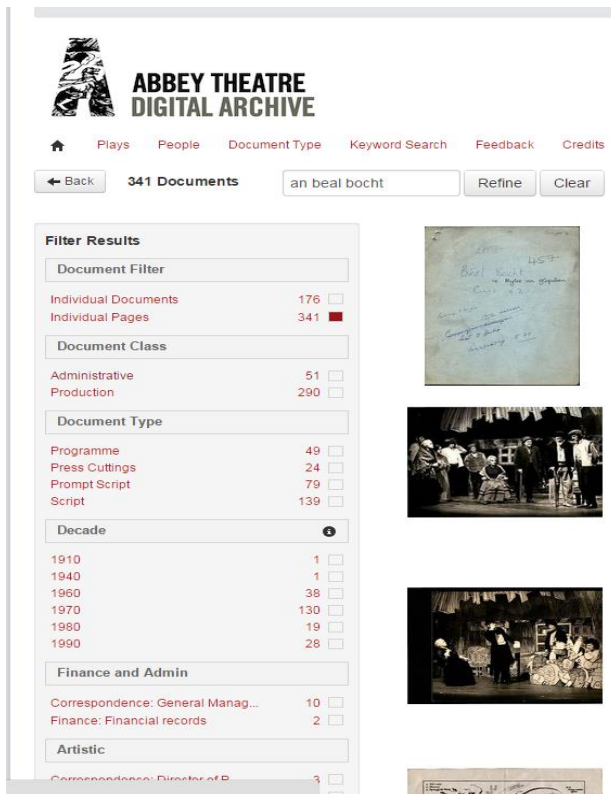


Figure 3. Screen-shot from Abbey Theatre Digital Archive, accessed 26/05/2015

Feedback received from users has been positive, with searchability, the quality of the scans, the ability to keep track of research and create a personal collection of favorite items, and the option for the user to email a formatted citation of each digital object to themselves specifically mentioned.

Each researcher can easily provide feedback through the system and suggest improvements for their experience. Some examples of issues that have been reported and resolved include improved zooming on large items, enhanced scroll-through on lengthy items, and NUI Galway users now being able to log in using their campus credentials. We have had a lot of positive and constructive engagement with users by making them part of the conversation, and mediating between them and the archive.

Ongoing suggestions that we are addressing as the project progresses include tagging individual press files and correspondence to specific productions, and transcribing written administrative material so that it can be more navigable.

One of the main challenges has been in striking a balance between addressing user's perceptions of what a digital archive should be, against our responsibility as a repository to appropriately care for, and facilitate access to material. The agreement between NUI Galway and the Abbey allows material to be made available for non-commercial university research and teaching, and is subject to the Copyright Act. A provision has been made to protect copyrighted and sensitive material, and in the case of the Abbey archive, it is available as 'view-only' under supervised conditions on campus at NUI Galway, and the archive cannot be published online.

Scholarly Engagement with the Abbey Digital Archive

With approximately one third of the total archive digitized and made available in September 2013, usage gradually grew through interacting with our users in the reading room, giving presentations, social media, word of mouth, and in late April 2014, a major public exhibition at the Hardiman Research Building. *Performing Ireland 1904-2014: a digital journey through Irish theatre*, ran from April until September 2014, and offered a taster of some of the digitized items. It included a demo of how to search the database on an interactive touchscreen, drew in connections between the Abbey Theatre and the West of Ireland, and made the digitization process part of the narrative of the exhibition.

Throughout the year, because of the rapid turnaround between material being digitized, screened, and made available, new material was constantly being uploaded to the database.

In September 2014, a module called the “Abbey Theatre Digital Archive” was made available to all taught graduate students by the Centre for Drama, Theatre, and Performance. There are several other Undergraduate and MA courses offered by the centre that require students to use the archive to complete their coursework.

To date, two scholars have been able to avail of dedicated Abbey Theatre Digital Archive PhD Fellowships, and three other funded researchers are relying on it for their work. There have been several applicants for research grants to work on the Abbey for 2015/2016 through HERA and the IRC.

Many researchers are using the Abbey digital archive alongside other Hardiman Library archive collections, such as those of the Abbey actor Siobhán McKenna, the playwright Thomas Kilroy, and institutional archives such as the Druid Theatre. This is developing cross-collection research skills in users, and is resulting in more comprehensive research, as well as increased use of our collections.

In July 2015, an international conference “Performing the Archive” will take place at the Hardiman Library. The conference is supported by the IRC and the American Society for Theatre Research, and the Abbey Theatre Digital Archive will have a dedicated panel on the main program.

A strategic advantage of this project is that it enables multi-disciplinary analysis of the archive. It has made it possible to take a wider look at trends across a huge body of material. Such analysis is, and will continue to transform scholarship of the Abbey, and challenge how we perceive it. Patrick Lonergan for instance noted that although only 3 female playwrights had new plays performed on the Abbey stage since 1940, many more have been directed,

performed, and worked on behind the scenes by women. His perception of the Abbey as a male dominated organization was challenged, and the balance shifted towards it being more female dominated. (Lonergan, 2015, accessed 9 November 2015).

Some examples of work undertaken using the archive include a Corpus of Hiberno-English Literary Dialects created as part of a PhD on Corpus Linguistics. Insight at DERI, NUI Galway's Centre for Data Analytics have also done some interesting data mining work in which they were able to establish what actors directors tended to work with. The archive as a whole is a history of production and stage management, and takes scholarship of the Abbey Theatre beyond scripts and individuals. There is great potential for fresh scholarship in the humanities and data analytics.

In time, we envisage and hope that auxiliary research such as datasets and apps could form part of the resource. Strategic and top down thinking will be essential in bringing this about.

How the Abbey Project is informing our Digital Library Strategy in terms of infrastructure and workflow, digital preservation, providing access, control, and legal compliance

The workflow and techniques employed in the Abbey Theatre Digitization Project are quite different to those in digitization projects NUI Galway had carried out in the past. The difference is mainly because of scale. Before the Abbey Project, the focus had been on creating and curating digital objects in order to open up aspects of our physical collections to all, and most often this was achieved by presenting a selection of material online. The aim was usually discoverability, though in some cases such projects also had a preservation agenda, for example to reduce demand on a 17th century pictorial map of Galway.

The Abbey Theatre's archive contains over 1 million items, and the scope was to digitize the entire collection. The project's emphasis was on mass digitization, and mindful of the fire in the Abbey Theatre in 1951, digital preservation was an inherent priority. Because the entire collection was being digitized, no appraisal work needed to be carried out, and the focus was on designing a workflow that could consistently handle a large volume of material in different formats, and allow the team full traceability and intellectual control over the archive at each stage of the transfer and digitization process. It was essential to have a suite of equipment and tools that could cope with these demands—standardize the digitization workflow, produce preservation standard images, and store them in a trustworthy repository.

The scale and nature of the Abbey work has pushed digitization activity at NUI Galway in new directions. A Digital Library Strategy Group has since been established, and an effort made to put unified digital library technology in place. In the past year, we have added some new pieces of infrastructure on which to store and work with digital images from our NUI Galway collections. One of these is a Digital Assets Management System, Islandora, and the other is Omeka which we use for digital exhibition content. At the heart of the strategy is figuring out consistent ways of managing our metadata so that we can reduce effort when it comes to extracting it from CALM and attaching it to digital objects. As well as a metadata management solution, digital preservation and strategies for coping with born digital material are all under review as part of this process.

An inevitable area that is examined here regards storage, and whether or not to set up an onsite data centre that has the capacity to grow, or opt for a cloud solution that enables us to curate and migrate objects according to our needs. The Abbey project's approach to linking

relational databases with metadata to the digital object has given us a frame of reference on how to approach this. Our aim with the Digital Library Strategy is to create systems that are agile enough to manage the different legal frameworks that exist among our collections, and to create an access platform that addresses user's needs, and hopefully exceeds their expectations in the diversity and depth of collections presented.

The experience of cross-disciplinary co-operation among a wide pool of internal and external stakeholders with the Abbey Project has also been valuable.

We are still in early stages with the Digital Library Strategy, but are working on two pilot projects to test out these pieces of the puzzle. One project is on the papers of Brendan Duddy, a collection that was listed and digitized in its entirety a few years ago. The objectives here are to test and improve the ingestion of archival metadata from CALM/EAD and the related digital objects into Islandora, test the preservation backup solution, and put systems in place that adequately control access, and can cope with restricted material. This is something we have experience in from working with the Abbey Digitization Project.

A second pilot project is on a collection that came into the Library in the last year, and is currently being listed. This is on the archive of Tim Robinson, a multi disciplinary and multi-formatted archive that lends itself to creating visualization tools that can draw in a diverse audience.

Conclusion

This final year of the digitization project is focusing on completing the retrospective digitization, making sustainable provision for ongoing content additions and cloud storage

costs, increasing use of the archive, and continuing to find ways of embedding our archival collections in academic research and curricula.

REFERENCES

[1] K. D. Borne, S. A. Baum, A. Fruchter, and K. S. Long. ‘The Hubble Space Telescope Data Archive’ on <http://adass.org/adass/proceedings/adass94/bornek.html>

[2] Martin Bradley. ‘Records Management and Irish Law’ on <http://www.archives.ie/Records%20Management%20and%20Irish%20Law.pdf>

[3] Patrick Lonergan. ‘More Thoughts (and stats) on Women Writers at the Abbey’ on *Scenes from a Bigger Picture* (January 2014)
<https://patricklonergan.wordpress.com/2014/01/22/more-thoughts-and-stats-on-women-writers-at-the-abbey/>

[4] Paper and Image Estimator on <http://www.legalscans.com/estimator.html>

[5] Government of Ireland. *Copyright and Related Rights Act*. (2000) Dublin: Stationery Office.