200 Manuscripts in 200 days: high throughput digitization of the Advocates' manuscript collection

Gavin Willshaw; National Library of Scotland; Edinburgh, UK

Abstract

The National Library of Scotland is pioneering a new approach to the digitization of medieval manuscripts, using the Book2Net Dragon system to image over 200 fragile and unique volumes in just one year. This paper provides an overview of the project to date, highlighting the balance the Library has been able to reach between maintaining quality and high throughput while minimizing damage made to the material. It discusses why the Library purchased the Dragon system, summarizes its key features and provides an honest assessment of the Library's experience to date. Furthermore, the paper outlines how this new work strand has been integrated into the wider digitization program and the strategic aim to have a third of collections in digital format by 2025.

Introduction

I am the Mass Digitization Service Manager at the National Library of Scotland and in my role I oversee the Library's high throughput digitization activity, managing a team of ten photographers and digitization operators who image collections material in pursuit of the Library's One Third Digital strategic aim.

In this presentation I will focus on our current project to digitize approximately 200 medieval manuscript volumes in just one year, a speed and scale far higher than the Library's traditional approach to manuscript volume capture. I will focus on the imaging system we used for this project – the Book2Net Dragon – and discuss how far it has been able to meet our project requirements. In sharing our experience, I hope to provide delegates with a better understanding of whether this system would be suitable for use in other projects involving the digitization of rare and fragile material.

The project

The project aims to digitize all 225 medieval manuscript volumes in the Advocate's collection, a unique series of books that were donated from the Faculty of Advocates when the Library was founded in 1925. While being old, rare and fragile, the collection does lend itself well to high throughput digitization because it is relatively standardized in terms of size, most volumes are fairly robust and there are very few serious conservation issues. We were fortunate that one of the Library's donors was willing to fund all aspects of the project including staffing, equipment, cataloguing and conservation; while these elements make up the wider project, in this talk I will focus mainly on the image capture element.

Capture requirements

Before beginning the capture process, we identified our requirements for a suitable digitization system and considered whether any of our existing equipment could be used on the project. Through this process we established the five main requirements below.

1. **High throughput**: we wanted a system that would allow us to capture volumes at a faster pace than our usual approach to manuscript digitization. In practical terms, we committed to having the collection online by the end of 2020, meaning a high throughput rate of approximately one volume per day.

2. **Good quality images**: we digitize manuscript material at 400 ppi in full colour and output TIFF files in the AdobeRGB colour space, cropping outside the book edge in order to give the viewer an accurate representation of what the original item looked like. Any system we used would need to meet these requirements.

3. **Minimal damage**: while we accept that digitization is a contact process, and there will always be some impact on the physical item, we wanted a system that would cause as little damage as possible to the manuscript volumes.

4. **Gold leaf**: many of the books contain illustrations and illuminations in gold leaf. Our system should be able to accurately represent the nature of this in the digital images.

5. **Reliable, with low barriers to entry**: our focus is on the capture of good quality images at high throughput, meaning any system we use for this work had to be reliable and able to cope with a high workload. As several staff were likely to be working on the system, it had to be relatively straightforward to use and not require in-depth photography expertise to operate.

We knew quite quickly that none of our existing equipment would be suitable for this project: our Copibook V-Shapes were one option, but their fixed 120° cradle and heavy glass plate ruled them out because of the damage they might do to the manuscripts volumes, while our single Phase One camera with copy stand was not appropriate because of the low throughput on that system, which would require digitization in the conventional method of capturing all the recto pages, rotating the book, capturing the verso pages then combining the images into one set.

This meant undertaking a procurement process, during which it became apparent early on that there was only one system on the market that appeared to meet all our requirements: the Book2Net Dragon V-shape. The only other option that we considered was the Conservation Cradle, which uses a vacuum to hold pages down during imaging, but

through discussions with colleagues it became clear that this system would also be too slow for our project requirements.

The Dragon

We purchased the Dragon system from Microbox GmbH and it was installed at the Library in late August 2019. The Dragon is a V-shape imaging system with two 71 MPS cameras and it appealed to us for this project because of its flexibility: the cradle, lighting and camera positions are all highly adjustable, meaning the system can more easily be adapted to the capture requirements of different books. It comes with a removable glass plate but also has a separate, less damaging, system for holding book pages flat, a feature which will be outlined in more detail below.

The system has been installed at the Library for over seven months, and we are now in a good position to reflect on the extent to which it has met our requirements, and therefore how suitable if is for the high throughput imaging of rare and fragile heritage materials.

Requirement 1: High throughput

Our minimum project requirement was for a system which could digitize one full 200-400 page manuscript volume per day, while also allowing enough time for the operator to fetch the volume from the stacks, prepare the system, capture the entire book from cover to cover, undertake an initial QA and return the book to the shelf.

The Dragon is capable of capturing several hundred pages of a standard modern book per hour, so we were confident that it could meet our requirements even when taking into account the older and more fragile nature of our material. As it turns out, aside from a period of teething problems at the start of the project, the system has largely been able to deliver on our requirement of one manuscript volume per day: at the end of February 2020, our digitization operator had digitized 86 volumes of the 96 volume target for that period. Although we are ten volumes behind schedule, the initial period included set up, training and other disruptions and we are now imaging above the anticipated monthly rate.

Requirement 2: Good quality images

We wanted a system that could deliver high quality images, by which we mean sharp images at 400 ppi which accurately represent the look of the original item. The Dragon's cameras are able to deliver these requirements and provide an image quality that matches our other imaging systems. Much of this is quite subjective, but viewing the image at 1:1 reveals the quality and sharpness that the system can produce.

Requirement 3: Minimal damage

As important as the image quality, we also required a system that could digitize books in a safe way, preventing or at least minimising damage done to the original items. The Dragon meets this requirement in two ways: through its Butterfly System plexi finger page holding solution and the adjustable nature of its book cradle.

The Butterfly System employs four user-operated mechanical plexi fingers which hold the volume's pages down near the gutter, thereby flattening them without applying stress or pressure to the spine, boards or joints. The operator can adjust the height, length and position of the fingers,

meaning they can be used for multiple book sizes and types, while the lack of a glass plate removes the danger of paint or other substances being removed from the volumes during imaging. We have found this system to be very effective for preventing damage to books: none of the volumes imaged to date have suffered from the stress damage that would be caused by using a conventional V-shape glass plate.

The Dragon has a flexible book cradle that can be adjusted to angles ranging from 80° to 120°, meaning it can support books that do not fully open to the default angle of 110°. The system's cameras can also be adjusted to the same angle, meaning they will always remain parallel to the page of the book being captured. To date we have not yet used this feature because the first hundred volumes can be imaged safely at 110°, although it was demonstrated during the procurement process and the operator has received training. Our plan to implement this feature in late Spring 2020 has been delayed by the coronavirus crisis.

Requirement 4: Gold leaf

Many manuscript items in the collection contain illuminations and pages or illustrations painted in gold leaf, so our hope was than any system we used would be able to accurately represent the nature of this in the digital images. The Dragon's adjustable lighting means the angle at which the light hits the page can be increased giving a better representation of gold detail. This process was demonstrated during procurement but we have not yet been able to accurately replicate it on our own collections material, so we don't have a workflow for integrating pages captured on the gold leaf settings with the rest of the book.

Requirement 5: Reliable, with low barriers to entry

Our final requirement was that any system we used would be reliable, meaning it could be used continuously for 37 hours per week over a multiple year period with minimal technical issues or disruptions, and have low barriers to entry, meaning a typical digitization operator could learn to use it effectively after a day's training. Our experience so far is that aside from the teething issues explored below, the system has been reliable enough for us to capture volumes at our required rate, and the team from Microbox have been quick at responding to and resolving issues as they arise. Our operators received one day's training followed up by remote support, and this has been sufficient for our purposes.

Issues and Challenges

As has been outlined above, the Dragon system has largely been able to meet our project requirements and we are on track to complete our project to deadline. However, that is not to say that our experience has been completely free of issues: I will now outline some of the bigger issues and challenges we have faced to date.

Challenge 1: Teething issues

While we appreciate that teething issues are a normal and expected part of installing any new system, we were disappointed to lose a lot of time in the first three months of the project with several issues which were not fully resolved until the end of 2019. Without going into too much detail, these included a yellow-green colour appearing in the gutter on one page side, the appearance of recurrent white

dots in several images, and an issue with focus that meant sharpness was not even across the two cameras. We would estimate that as a result of these issues we lost approximately one third of our capture time in this period, although this was offset as Microbox were able to deliver the system four weeks earlier than we had originally hoped. Since the start of 2020 we have not suffered any more issues of this nature and we have been able to increase our throughput to anticipated levels.

Challenge 2: Plexi fingers in images

The Butterfly System allows for the safe capture of fragile material, but using this system does require acceptance of a few compromises. The most obvious of these is that a plexi finger appears in every image captured on the system and on some occasions, it is impossible to avoid text being wholly or partially covered by it. The fingers are adjustable, but they need to be as close to the book's gutter as possible in order to minimize curvature: we attempt to assuage this wherever possible by positioning the plexi finger over a complete word.

Challenge 3: Size limits

Another issue with the Butterfly system is that it takes up a lot of space on the cradle, thereby reducing the Dragon's capture area from A2 to around A3+ per side. This limitation has meant that three of the volumes in the collection had to be photographed by our colleagues in the Imaging Service team, using Phase One cameras on a traditional copy stand set up but imaging at a much slower rate. This team also capture some of the books' bindings because there is no way to image them on the Dragon system.

Challenge 4: Colour Charts

The Library's previous approach to digitizing manuscripts involved including a colour chart in every image but, without the frame of a glass plate to attach these to, it has proved difficult to include them in our digital images. Instead, we capture the colour charts before imaging begins and then again after the book is digitized: these images are then included as part of the image pack when we publish the book online. Post digitization, we check the RGB values on the Black and White patches and investigate any major discrepancies. To date this has not been an issue, and the RGB values have remained stable at around 240 on the white patch and 50 on the black.

Challenge 5: Adjustability

The final challenge that we have faced is that while the system is very flexible, making these adjustments is by no means an easy, quick or straightforward task. We had, perhaps naively, hoped that the angle of the cradle and cameras could be adjusted relatively quickly as each book was imaged, but we now understand that this process can take the best part of a day as it requires a lot of small tweaks and adjustments to get the calibration of the system right. As a result, we grouped the collection into clusters of books requiring similar settings so that we only have to make these adjustments once or twice during the whole life of the project.

Conclusion

Since receiving the Dragon at the National Library of Scotland in August 2019 we have been able to successfully pursue our project to digitize over 200 medieval manuscript volumes in just one year and, prior to the coronavirus crisis, were on track to have all volumes digitized and online by the end of 2020. This success is largely down to the flexibility offered by the Book2Net Dragon, whose adjustable features, and, in particular, the Butterfly System, has enabled us to image rare and unique material at scale without causing damage to any of the volumes. If you accept the presence of plexi in images and understand that there will be teething issues and other minor inconveniences, then we strongly recommend the system for anyone else wishing to undertake a similar digitization project.

Author Biography

Gavin Willshaw is the Mass Digitization Service Manager at the National Library of Scotland, where he oversees the Library's digitization program. He is a chartered member of CILIP, holds a PGDip in Library and Information Studies from Robert Gordon University and has over five years' experience of managing digitization at cultural heritage organizations in the UK.