

# Digital Preservation at Oxford and Cambridge

A collaborative research project to evaluate and provide sustainable recommendations for our digital preservation programmes

## Guest post: The 6-million-image gap

Posted on **16 October, 2017** by **Emma Stanford**

*Bodleian Digital Library Systems and Services' Digital Curator, Emma Stanford, guest blogs for the DPOC project this week. Emma writes about what she is doing to close some of the 6-million-image gap between what's in our tape archive and what's available online at [Digital.Bodleian](#). It's no small task, but sometimes Emma finds some real gems just waiting to be made available to researchers. She also raises some good questions about what metadata we should make available to researchers to interpret our digitized image. Read more from Emma below.*

Thanks to Edith's hard work, we now know that the Bodleian Imaging Services image archive contains about 5.8 million unique images. This is in addition to various images held on hard drives and other locations around the Bodleian, which bring the total up to almost 7 million. [Digital.Bodleian](#), however, our flagship digital image platform, contains only about 710,000 unique images—a mere tenth of our total image archive. What gives?

That 6-million-image gap consists of two main categories:

**Images that are online elsewhere** (aka the migration backlog). In the decades before Digital.Bodleian, we tried a number of other image delivery platforms that remain with us today: Early

Manuscripts at Oxford University, the Toyota City Imaging Project, the Oxford Digital Library, Luna, etc., etc. Edith has estimated that the non-Digital.Bodleian content comprises about 1.4 million images. Some of these images don't belong in Digital.Bodleian, either because we don't have rights to the images (for example, [Queen Victoria's Journals](#)) or because they are incomplete selections rather than full image sets (for example, the images in the [Bodleian Treasures](#) exhibition). Our goal is to migrate all the content we can to Digital.Bodleian and eventually shut down most of the old sites. We've been chipping away at this task very slowly, but there is a lot left to do.

**Images that have never been online.** Much of Imaging Services' work is commercial orders: shooting images for researchers, publishers, journalists, etc. We currently store all these images on tape, and we have a database that records the shelfmark, number of images, and list of captured pages, along with information about when and how the images were captured. Searching through this archive for Digital.Bodleian-appropriate images is a difficult task, though. Shelfmark notation isn't standardized at all, so there are lots of duplicate records. Also, in many cases, just a few pages from a book or manuscript were captured, or the images were captured in black-and-white or grayscale; either way, not suitable for Digital.Bodleian, where we aim to publish fully-digitized works in full colour.

I'm working on extracting a list of complete, full-colour image sets from this database. In the meantime, we've started approaching the problem from the other direction: creating a list of items that we'd like to have on Digital.Bodleian, and then searching the archive for images of them. To do this, we asked the Bodleian's manuscript and rare book curators to share with us their lists of "greatest hits": the Bodleian's most valuable, interesting, and/or fragile holdings, which would benefit most from online surrogates. We then began going through this list searching for the shelfmarks in the image archive. Mostly, we've found only a few images for each shelfmark, but occasionally we hit the jackpot: a complete, full-colour image set of a 13th-century bestiary or a first edition of a Shakespeare play.

Going through the archives in this way has underlined for me just how much the Bodleian's imaging standards have changed in the last two decades. File size has increased, of course, as higher-resolution digital scanning backs have become available; but changes in lighting equipment, book cradles, processing software, rulers and colour charts have all made their mark on our images

too. For me, this has raised the question of whether the technical metadata we're preserving in our archives, about when and how the images were captured, should also be made available to researchers in some way, so that they can make an informed choice about how to interpret the images they encounter on sites like Digital.Bodleian.

In the meantime, here are some of the image sets we've pulled out of the archive and digitized so far:

[Jane Austen's juvenilia](#)  
[a 13th-century bestiary](#)  
[the Oxford Catullus](#)



— MS. Bodl. 764, fol. 2r (detail)

SHARE THIS:



This entry was posted in [digitisation](#), [digitization](#), [standards](#) by [Emma Stanford](#). Bookmark the [permalink](#) [<http://www.dpoc.ac.uk/2017/10/16/1039/>] .

This site uses Akismet to reduce spam. [Learn how your comment data is processed](#).