

Digital Preservation at Oxford and Cambridge

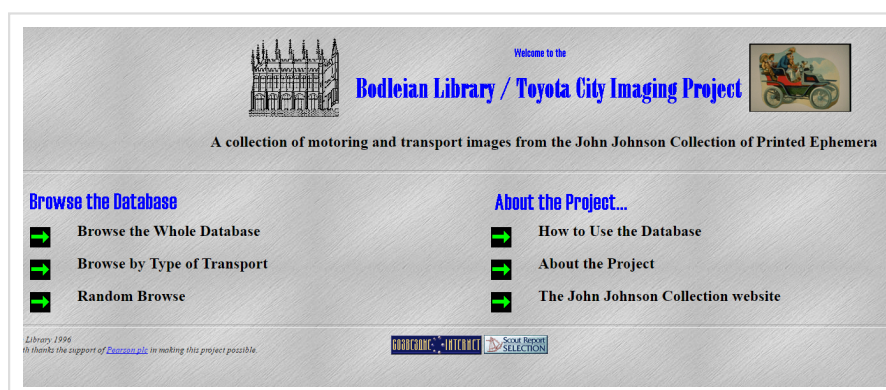
A collaborative research project to evaluate and provide sustainable recommendations for our digital preservation programmes

Over 20 years of digitization at the Bodleian Libraries

Posted on **9 May, 2017** by [ehalvarsson](#)

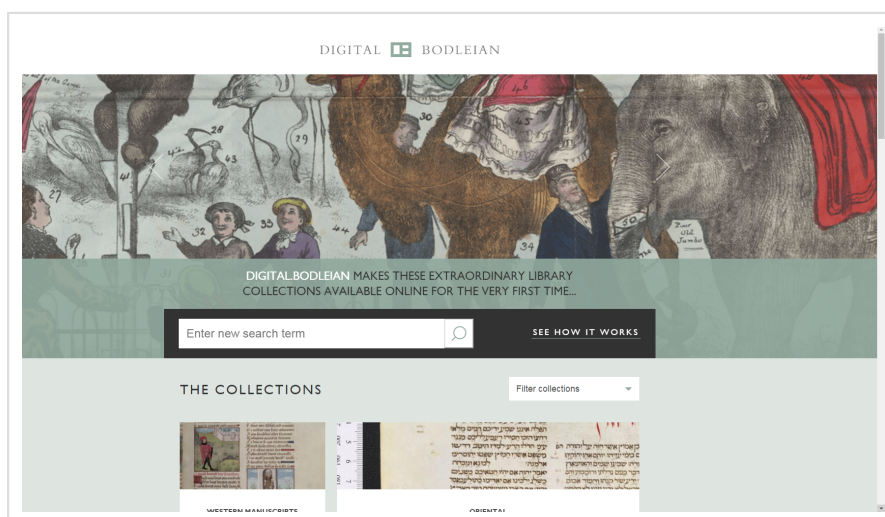
Policy and Planning Fellow Edith writes an update on some of her findings from the DPOC project's survey of digitized images at the Bodleian Libraries.

During August-December 2016 I have been collating information about Bodleian Libraries' digitized collections. As an early adopter of digitization technology, Bodleian Libraries have made digital surrogates of its collections available online since the early 1990's. A particular favourite of mine, and a landmark among the Bodleian Libraries' early digital projects, is the Toyota Transport Digitization Project (1996). [[Still up and running here](#)]



At the time of the Toyota Project, digitization was still highly specialised and the Bodleian Libraries opted to outsource the digital part to Laser Bureau London. Laser Bureau ‘digitilised’ 35mm image negatives supplied by Bodleian Libraries’ imaging studio and sent the files over on a big bundle of CDs. 1244 images all in all – which was a massive achievement at the time. It is staggering to think that we could now produce the same many times over in just a day!

Since the Toyota projects completion twenty years ago, Bodleian Libraries have continued large scale digitization activities in-house via its commercial digitization studio, outsourced to third party suppliers, and in project partnerships. With generous funding from the Polonsky Foundation the Bodleian Libraries are now set to add over half a million image surrogates of Special Collection manuscripts to its image portal – [Digital.Bodleian](#).



What happens to 20 years’ worth of digitized material? Since 1996 both Bodleian Libraries and digitization standards have changed massively. Early challenges around storage alone have meant that content inevitably has been squirreled away in odd locations and created to the varied standards of the time. Profiling our old digitized collections is the first step to figuring out how these can be brought into line with current practice and be made more visible to library users.

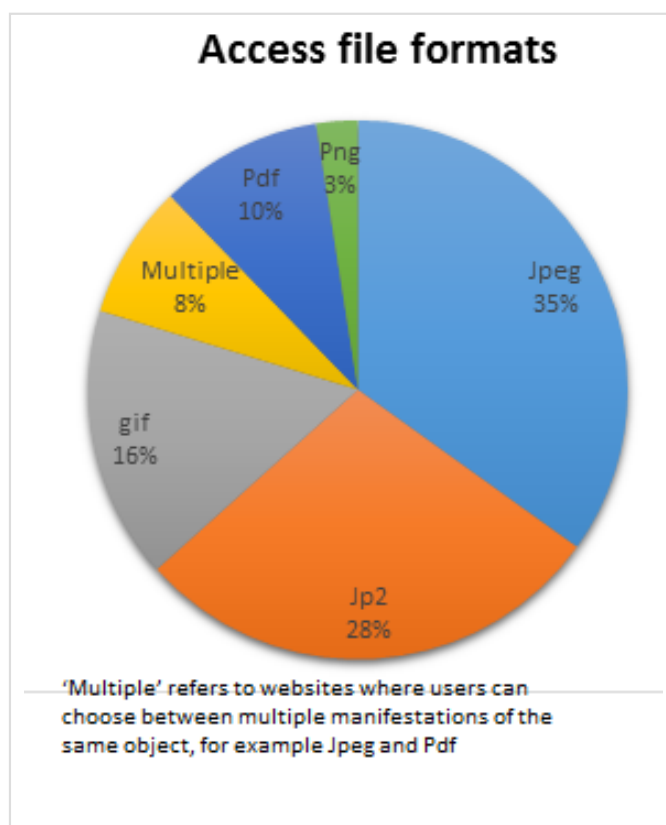
“So what is the extent of your content?”, librarians from other organisations have asked me several times over the past few months. In the hope that it will be useful for other organisations trying to profile their legacy digitized collections, I thought I would present some figures here on the DPOC blog.

When tallying up our survey data, I came to a total of approximately 134 million master images in primarily TIFF and JP2 format. From

very early digitization projects however, the idea of 'master files' was not yet developed and master and access files will, in these cases, often be one and the same.

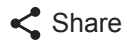
The largest proportion of content, some 127,000,000 compressed JP2s, were created as part of the [Google Books project](#) up to 2009 and are available via [Search Oxford Libraries Online](#). These add up to 45 TB of data. The library further holds three archives of 5.8million/99.4TB digitized image content primarily created by the Bodleian Libraries' in-house digitization studio in TIFF. These figures does not include back-ups – with which we start getting in to quite big numbers.

Of the remaining 7 million digitized images which are not from the Google Books project, 2,395,000 are currently made available on a Bodleian Libraries website. In total the survey examined content from 40 website applications and 24 exhibition pages. 44% of the images which are made available online were, at the time of the survey, hosted on [Digital.Bodleian](#), 4% on ODL Greenstone and 1% on [Luna](#). The latter two are currently in the processes of being moved onto Digital.Bodleian. At least 6% of content from the sample was duplicated across multiple website applications and are candidates for deduplication. Another interesting fact from the survey is that JPEG, JP2 (transformed to JPEG on delivery) and GIF are by far the most common access/derivative formats on Bodleian Libraries' website applications.



The final digitized image survey report has now been reviewed by the Digital Preservation Coalition and is being looked at internally. Stay tuned to hear more in future blog posts!

SHARE THIS:



Share

This entry was posted in [digitisation](#), [digitization](#), [project update](#) by [ehalvarsson](#). Bookmark the [permalink](#) [<http://www.dpoc.ac.uk/2017/05/09/over-20-years-of-digitization-at-the-bodleian-libraries/>].

2 THOUGHTS ON “OVER 20 YEARS OF DIGITIZATION AT THE BODLEIAN LIBRARIES”

Mazie

on **7 June, 2017 at 16:32** said:

I needed to thank you for this great read!! I definitely enjoying every little touch of it I have you bookmarked to have a look at new material you post.

Imogene

on **5 June, 2017 at 03:27** said:

Great post. I'm facing a couple of these difficulties.

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