

York Case Study | March – June 2010

Contents

Summary	Page 2
Borthwick Institute on-demand digitisation service	
Process	Page 3
Outcomes	Page 3
Recommendations	Page 6
Electronic Key Texts	
Process	Page 7
Outcomes	Page 7
Recommendations	Page 14
Digitisation Service	
Process	Page 14
Outcomes	Page 14
Recommendations	Page 15
Appendix 1 – Summaries of archives and special collections on-demand digitisation services	Page 16
Appendix 2 – Summaries of EKT services at other higher education institutions	Page 23
Appendix 3 – Technical requirements for archival master images from the Borthwick on-demand digitisation service for YODL	Page 28
Appendix 4 – Standards and best practice guidelines for creating digital archival master images	Page 33
Appendix 5 – Model for providing an online course reading service	Page 37
Appendix 6 – Electronic Key Texts Funding Models	Page 47
Appendix 7 – Electronic delivery of files for the Borthwick Institute via YODL – proposal	Page 53

Summary

The York case study looked at on-demand digitisation in two areas of the University of York Library and Archives: the on-demand digitisation service at the Borthwick Institute for Archives and the Library's Electronic Key Texts (EKT) pilot project. It focussed on key areas which were identified as needing improvement, or where more information was needed to inform service development.

Borthwick Institute on-demand digitisation service

The focus of the Borthwick case study was on improvements to workflow in three areas: capturing requests; long term storage of digital images produced by the service; and delivery of digital files to customers

- An investigation of similar archives/special collections on-demand services revealed that the Borthwick service was on a par with other comparable services (with the exception of The National Archives) in the first and third of the three areas identified above, but lagged behind some services in the long term storage of images in a repository.
- A web form was created to capture image requests.
- Technical standards and a model for long-term storage of images in YODL were drawn up.
- A model for delivery of images to customers using YODL was drawn up.

Electronic Key Texts

The case study explored a number of key areas for developing the existing EKT pilot project into a full service offered to the whole University.

- An investigation of other EKT services provided information for subsequent stages of the case study.
- A service model was drawn up, listing different options for each stage of the lifecycle of the digital reading, with possible advantages and disadvantages of each option.
- Predicting the likely demand is extremely difficult, but some 'educated guesses' based on the experiences of other services, particularly Leeds and Sheffield, are possible. It seems likely that demand for the service will be high and continue increasing over the first three years and probably beyond. It seems likely that demand will reach at least 1000 readings per year within three years and an eventual figure of 5000 does not seem impossible, based on the number of modules run by the University.
- Based on timings obtained from the EKT pilot study, we calculated costs of £10.79-£14.85 per reading. If the levels of demand suggested are anywhere near correct, this could mean costs for the service of between £10790 and £74250 a year. These figures exclude liaison time, which is very variable.

- Different funding models were drawn up, exploring the idea that academic departments might be asked to subsidise the service in some way. It is recommended that all readings be paid for by departments, but subsidised by the Library.

Digitisation Service

Julie Allinson has drawn up an initial proposal for a Digitisation Service, which would draw together a number of strands of digitisation activity carried out by the University of York Library and Archives, including the Borthwick on-demand service and EKT. To support the development of this idea, an initial meeting was organised as part of the case study with key members of staff employed in each of the digitisation activities to discuss and compare workflows (process engineering). A number of areas of crossovers emerged, which will be discussed at further meetings.

Borthwick Institute on-demand digitisation service case study

Process

The Borthwick case study focussed on three areas identified by them as needing work: receipt of digitisation requests; uploading digitised documents to YODL; and delivery of digital files to customers. The study consisted of a number of stages:

- Investigation of similar services: a small sample of other archives/special collections on-demand digitisation services were contacted and investigated in detail. This resulted in a set of rich descriptive accounts which informed subsequent stages (Appendix 1).
- Receipt of digitisation requests: a web form was created to capture digital requests.
- Uploading of digitised documents to YODL for long-term storage and access:
 - Technical standards for digital archival master images, including file formats and metadata, were identified and collated. Best practice guidelines were then created from this.
 - Existing Borthwick scanners were assessed for fitness for purpose on the basis of the technical standards.
 - Models developed for storing digitised archive documents in YODL
- Delivery of digital files to customers: models were produced for a more efficient way of delivering digital files to customers over the internet and for storing images in YODL.

Outcomes

Investigation of other archives/special collections on-demand digitisation services

(See Appendix 1 for detailed descriptions of the services investigated).

With the exception of The National Archives, none of the archives/special collections on-demand digitisation services looked at was significantly further ahead of the Borthwick in terms of receipt of requests and delivery of files to customers. The most common methods for capturing requests were in-person visits, telephone calls, emails and online PDF forms which could be printed and sent by post/fax. None, apart from TNA, used a web form to submit requests over the internet. The most common methods for delivering files to customers were by post (on a CD/DVD), email and FTP (e.g. using services like YouSendit). TNA has a bespoke software system to manage and send requests.

A number of the services looked at (excluding TNA, which adds documents to their DocumentsOnline service on a project basis) add images created as part of their on-demand service to their institutional repositories. This places them ahead of the Borthwick service, which does not currently do this.

Receipt of digitisation requests

After investigating the other archives' approaches to the receipt of digitisation requests, it was decided to create a simple web form. This will make the process of receiving requests more efficient and reduce the amount of correspondence. It is also possible for the Borthwick staff to enter details into the form for telephone and in-person orders, thus creating a single logging system for all requests. The requests can then be managed using Microsoft Outlook or similar tools. The form was created by the Web Office with fields generated in consultation with Sara Slinn and tested by Borthwick staff. It became live for customers in September 2010 and is now receiving requests.

Uploading of digitised documents to YODL for long-term storage and access

(See Appendices 3, 4 and 7 for information on technical standards and storage).

A proposal was created for permanently keeping those images generated as part of the on-demand service, which are judged to be of wider interest than just to a single customer, and making them available via YODL (see Appendix 7). The key recommendation is that a decision is taken at the point of receiving the digitisation request as to whether the images will be worth keeping or not. If they are to be kept in YODL, then it needs to be digitised to the standards recommended below. Images for permanent storage would be uploaded to a collection in YODL with full metadata in VRA Core 4.0 format. These images would be available to University of York members, with only the metadata available to the public. This would enable images to be sold to the public, if this option was desired. Sharing within the White Rose Consortium could happen in the future: the main technical challenge which would allow this is integrating YODL with Shibboleth, something which is on the YODL project plan. The image files which are actually sent to the customer would be derived from the master file and uploaded to a temporary delivery area of YODL, as described below and in Appendix 7.

A large number of technical standards for the creation of digital archival master images were consulted and are listed in Appendix 4. The principle ones used were *US National Archives and Records Administration (NARA) Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images* (Puglia, Reed and Rhodes, 2004) and the Cornell Quality Index (QI) (Cornell University Library, 2003). The NARA guidelines are the most detailed and comprehensive guidelines consulted and ones on which a number of other guidelines, especially of US digital libraries, are based. The

Cornell QI provides a method for judging the scanning resolution needed, depending on the size of the smallest significant character, or the width of the narrowest stroke, in a textual document.

The Cornell QI was applied to sample documents from the Borthwick Institute and it was found that, for the majority of cases, the current practice of scanning at 300dpi was also sufficient for long-term storage. For textual documents with 1.37mm as the height of the smallest character, a QI of 8 (excellent) is achieved at 300dpi. The majority of the material scanned by the service is hand-written, with characters significantly greater than 1.37mm in height. However, resolution would need increasing for printed documents with very small type. The Cornell method also recommends that the thinnest stroke width is represented by at least two pixels across its width. The sample documents passed this test at 300dpi.

The NARA guidelines recommend that photographic prints should be scanned so that the digital file has at least 3000 pixels along the longer edge (minimum guideline), but the ideal guideline is 4000 pixels (for prints 10 inches or less on the longer side – which most of the prints held by the Borthwick are). This means that 300dpi is inadequate for most photographic prints, even if the minimum guideline is followed. Some prints need to be scanned at a higher resolution than the 600dpi maximum which the Borthwick's book scanner is capable of (those less than about 6 inches long for the full guideline and those less than 5 inches long for the minimum guideline). The Library Photographer scans photographic prints at a resolution sufficient to achieve 4000-5000 pixels on the longer edge using an A4 flatbed scanner which is capable of much higher resolutions. The recommendation is, therefore, to follow the full NARA guideline of 4000 pixels on the longer side and use a flatbed scanner for scans which need to be over 600dpi.

The YODL *Content Model for Images* (Allinson *et al.*, 2009), was also consulted. The main recommendations from this were that material for YODL should be scanned as TIFF, which is in accordance with all of the published standards consulted, and catalogued using the VRA Core 4.0 metadata schema (for which YODL has an editor).

The Borthwick currently use two scanners: a Scan2Net SMA21 book scanner, which scans up to A2 in size, and a Konika Minolta MS6000 MkII microfilm scanner. Since microfilm scans would not be added to YODL, only the book scanner has been considered. The book scanner has a maximum resolution of 600dpi up to A4, 400dpi up to A3 and 300dpi up to A2. In addition to the problem with photographic prints (which can be met with a relatively inexpensive A4 flatbed scanner), the only problem which this presents in terms of resolution is with items over A3 in size and text less than 1-1.5mm high. This problem is at present theoretical; as such documents are rarely, if ever, encountered by the Borthwick. Such documents, if encountered should be photographed by the Library Photographer, using the Hasselblad digital camera. (The Borthwick are aware of a need to digitise large items, over A2, which, at present, need to be removed from the Borthwick and taken to the Library photography studio, which is impermissible for some types of material. They are also aware of conservation issues affecting the Scan2Net SMA21 book scanner, due to the fact that the scanner lightly touches documents – making it unsuitable for documents with seals, ones that are not perfectly flat and very delicate ones – and that the book cradle is only capable of handling books which are able to open to 180 degrees).

The Borthwick recently acquired a third scanner, mainly for walk-in use by the public, but which may sometimes be used for the staffed on-demand service. It solves the problem of delicate and not flat documents, as there camera part does not come into contact with the document, but only takes pictures

which are roughly equivalent of 300dpi, so doesn't change the other recommendations. It is a Book2Net Kiosk (<http://www.microfilm.com/php/display.php?ItemID=379>).

Delivery of digitised files to customers

(See Appendix 7 for full proposal).

A number of possible solutions presented themselves for electronic delivery of files to customers. It was felt that the nearest to hand and most flexible solution would be to deliver via YODL. Since secure delivery was not a requirement, this would be fairly easy to implement. A proposal was drawn up to create a publically-accessible collection in YODL to which images for delivery to customers could be uploaded, once payment has been received. The customer would be sent an email with a link to the record in YODL, with instructions as to how to download their files.

This model has undergone some testing and a small number of technical problems identified (e.g. multiple files downloaded with the same filename and over-wrote each other), but these should be easy to solve (most have been fixed). A more serious issue is that the YODL interface, to which customers would be directed to download their files, has a complex appearance, which may confuse users. It is proposed that a dedicated and much simplified interface be developed as part of the planned new YODL interface later in the year, just for delivery of files to customers.

Concurrently with work on this model, work has been done by the YODL Team in the area of Fedora disseminators and content models, two features of Fedora which could allow a single, more sophisticated, workflow which would be the same for both types of material (permanent storage and delivery to customer), but handle them differently for customers. (Technically, a disseminator can be used to deliver content differently depending on its content model).

Recommendations

- Implement web form when the CMS migration is complete.
- Scan all images for long-term storage in YODL as TIFF files
- Scan textual documents for long-term storage in YODL using the book scanner at 300dpi, unless they contain characters less than circa 1-1.5mm high, for which increase the resolution to 400dpi
- Digitise textual documents for long-term storage in YODL which are over A3 size with characters less than circa 1-1.5mm high using the Hasselblad digital camera
- Scan photographic prints for long-term storage in YODL to an image size of 4000 pixels along the longer side (adjusting the resolution accordingly). Use a flatbed scanner, such as the one in the Library photographic studio, for resolutions over 600dpi.
- Identify material for permanent storage in YODL at the point of capture, so that it can be captured at appropriate standards for YODL (see section on technical standards) and sufficient metadata captured. Derive the version sent to the customer from the archival version. Material identified as not for inclusion in YODL can be captured to the same standards as at present.

- Ingest material for permanent storage to sub-collections of the 'Borthwick Institute' collection in YODL, catalogue using VRA Core 4.0 and set access control so that the public can see metadata only and University of York (or White Rose) can see the images.
- Once all identified technical problems have been fixed, set up a small pilot project to test the use of YODL to send digital files to customers (temporary storage).
- Create a simplified interface (Fedora object view page) in the new YODL interface for customers.

Electronic Key Texts case study

Process

The work for this case study focussed on the transformation of the pilot service into a full, University-wide service. The study consisted of the following pieces of work:

- Investigation of other EKT services: a small sample of similar services at other HEIs were contacted and investigated in detail. This resulted in a series of descriptive accounts, which informed subsequent steps. A published study into EKT services (Hedges and Secker, 2010) and a spreadsheet of their raw data (used with permission) also provided valuable information for this stage. Some case studies from the NoWAL consortium were also consulted (NoWAL, n.d.).
- Service models: information from the investigation of other EKT services was used to develop a service model to inform future developments at York and provide information for institutions setting up similar services.
- Estimate of potential take up: information from the investigation of other EKT services was used to make a tentative estimate of the potential take-up for the York service when it goes University-wide.
- Costings for University-wide service: costings for offering a full EKT service to the whole University were established based on the estimate of potential take up and the time-and-motion study
- Funding models: possible funding models for the EKT service were drawn up, based on discussions with Elizabeth Harbord and Julie Allinson. The previous stages assumed full library funding of the service (the dominant model identified in the investigation of other services), but this exercise considered ways in which academic departments might be asked to contribute financially to the provision of readings (which they already do by creating paper or digital course packs internally).

Outcomes

Investigation of other EKT services

(See Appendix 2 for full details of other services).

The investigation of other EKT services resulted in the following conclusions.

- Demand for online course readings has tended to increase, even for established services, by up to 100% in one year. All of the institutions looked at in detail have seen an increase.
- All of the services looked at in detail use one of two methods for obtaining the majority of their online course readings: British Library’s Higher Education Scanning Service (HESS) and in-house scanning. Ingenta Heron is more likely to be used as a secondary method, as it is the most expensive method. One institution also outsources to a reprographics company outside of term time.
- The CLA licence is the main means of copyright clearance used. Again, Heron is used either as a secondary means, for material not covered by the CLA licence, or isn’t used at all. Only one institution does a significant amount of clearance direct with publishers, but they also increasingly use the CLA licence.
- Staffing models and the number of FTEs employed vary. Hedges and Secker found that the number of staff ranged from 0.5 FTE or less to 5 FTE or more, with the average being 2. Sheffield has 6 FTE. Some HEIs have dedicated teams. Others split the workflow across more than one team. Another model is to have a very small core team (e.g. one single member of staff), but draw in other people at busy times. Staffing models do not seem to be connected to institution size (e.g. Leeds and Sheffield are both much larger than York, but Leeds only has a single member of staff working all the time on EKT).
- Record management, yearly renewals and the CLA return are a significant burden as they are extremely time-consuming. Solutions include Heron PackTracker, Excel or Access databases and bespoke applications.

Service models

(See Appendix 5 for a more detailed version of the service models).

Information from the investigation of other EKT services was used to develop a model which breaks the lifecycle of an electronic reading down into stages informed by the LIFE model (<http://www.life.ac.uk>). Different possible options for each stage are presented in a table, with advantages and disadvantages for each. This was presented at the online course reading Exchange of Experience event held by LIFE-SHARE in Leeds in April 2010 and the experiences of delegates fed directly back into it. It thus forms a synthesis of a wide range of experiences of running EKT services.

The model and the experience of talking with people about it both seem to strongly confirm the direction already embarked on with the EKT pilot. No major possibilities or factors were identified which have not already been considered. The following table presents the options already decided on for the EKT pilot project, with key points to back up the choice from the model.

Lifecycle stage	Method(s)	Reasons
Selection – how are readings selected for digitisation?	Recommendation from lecturers	Provides a direct service to academics and likely to be the most useful readings to students.

Creation or purchase	British Library HESS (Higher Education Scanning Service) for most readings, with in-house scanning for readings not available via HESS	HESS is quicker and more cost-effective than Ingenta Heron. In-house scanning can provide readings which are not available in the British Library (under the CLA licence the institution has to own a copy of the book, so a copy will always be available in stock).
IPR/licensing	CLA in most cases, with clearance via Heron for cases not covered by the CLA	The CLA licence offers the most cost-effective and easiest way of clearing readings. The University already pays for the subscription and there is no additional per-reading charge. Heron is more expensive and slower, but may be able to obtain permissions for material not covered by the CLA licence. Direct clearance with publishers is possible, but has a high cost in terms of administration.
Administration – time taken to administer the service (e.g. compiling the CLA return and weeding material) emerged as a key point from the investigation of other EKT services	Heron PackTracker	Heron PackTracker offers powerful functionality (e.g. it can automatically generate the CLA return) and avoids the need to create and test in-house solutions.
Metadata	Filenames and links (e.g. link in online reading list)	The general consensus at the Exchange of Experience meeting was that metadata needs to be kept minimal – enough to identify a reading and administer it. Readings are not kept permanently and are not worth detailed cataloguing.
Storage and access	VLE	Different institutions have come up with different solutions to storage and access – it depends a lot on the local set up (e.g. whether the institution has a

		<p>VLE). From looking at the experiences of other institutions there seems to be no reason not to follow the earlier recommendation at York that the VLE should be used. This seems to be the preferred method at institutions where it is available (some institutions had EKT services before they had VLEs).</p>
--	--	---

Estimation of likely demand

Estimating the likely demand for the EKT service once it changes from being a pilot study to a University-wide service proved very difficult. An investigation into whether EARL and PEDRO could be used to estimate the total number of eligible readings currently on reading lists at the University revealed that this was problematic and it was therefore decided to use other methods to estimate demand.

Looking at the experiences of other EKT services, and the results of a published survey of EKT services (Hedges, J. and Secker, J., 2010), the following very broad points can be made.

- EKT services can expect to see a large increase in demand year on year. Hedges and Secker (2010) found an average increase of 41% between figures for 2007-8 and projected figures for 2008-9. Sheffield reported an increase of 50% over the previous year in 2008-9 and 25% again 2009-10. Leeds predicts an approximately 60% increase from 2009-10 to 2010-11.
- There does not seem to be any sign of demand levelling off, even for long-established services.
- Currently, established services at larger institutions like Leeds and Sheffield are offering readings numbering in the low thousands annually. The largest number of readings any of Hedges and Secker's correspondents expected to offer in 2008-9 was 3200. Leeds provided 2760 in 2009-10 (of which 1859 were new acquisitions and 901 reused from previous years) and expects to provide around 4500 in 2010-11. Sheffield made 4174 readings available in 2009-10, of which around 2036 were new acquisitions. It is possible to scale these figures down to 'York-size', but difficult to base valid predictions on when York would reach a particular level.
- Even established services with high levels of demand are at this point in time offering a smaller number of readings than the total number of modules running in their institutions. For example, Sheffield made around 3200 readings available in 2008-9, but in the same year 4920 modules ran at Sheffield. Even given that not all modules will ever make use of the service, it still suggests that demand is still far from exhausted at these institutions.

- One important factor which affects the level of demand is how well the service is promoted to academics and departments. Several services report being cautious in how they promote their services, for fear of not being able to cope with demand.
- Some services limit the number of readings they offer to each module. Limits encountered range from five to forty. If York limited the number of readings they were prepared to offer to five per module, the service would have a maximum theoretical demand of 11815!
- At the EKT Exchange of Experience event the consensus was that services would have to start charging for, or limiting the number of, readings they offered, as the current model of offering unlimited numbers free of charge to academic departments was unsustainable.
- The eventual impact of ebooks is a factor which might lead to the levelling off of demand, but it is difficult to predict how quickly this will happen.

All of this means that making projections for York over the next three years is very difficult indeed. It does not seem unreasonable that the service should catch up with current rates of scanning at Leeds and Sheffield within three years or less, leading to figures of around 1000 readings a year at that point, if scaled down to York's size. Predicting eventual levels of demand and the point at which demand will level off is more difficult still, but the size of the potential 'market' (2363 modules ran at York in 2008/9) suggests that demand might rise much higher than 1000 readings a year. Although a corpus of scanned readings will be built up, which can be reused from year to year, it is likely that much of the growth in demand before the service 'matures' will be from new modules using the service, so it is unlikely that the body of already scanned material will dent the demand for new material until the total demand for readings begins to level off. Even if only half of the modules in the University each had five readings, that would mean over 5000 readings. (Also, experience from the History of Art Image Service shows that, even though History of Art has maintained an image collection for years, the constant development and changing of modules means that there is a constant need for new material). It seems likely that initial demand at York will be high, as there has been a lot of pressure from academic departments for the service to be introduced.

Time and motion study and costings

Timings were obtained for processing a sample of book chapters requested from the EKT pilot project where the British Library was not able to provide scans. (This included retrieving them from the library shelves, scanning them, tidying the files in Paint Shop Pro if necessary, converting the JPEG files to PDF and OCR-ing them). These timings were passed to Lucy Jaques, who produced a workflow diagram, with timings, for the whole process, including for ordering readings from HESS (see EKT Phase 2 final report).

The following table details costs per reading for each stage, based on Lucy's timings and model. It assumes most of the process being carried out by a grade 4 member of staff, with in-house scanning carried out by a grade 3 assistant. It does not include liaison time, which can be very variable (and time-consuming). Nor does it include the purchase and ongoing costs of the scanning equipment, which are difficult to calculate per reading.

1 FTE service co-ordinator grade 4 spine point 17 (roughly mid grade) = £24666.90 inc NI + pensions
 = £12.82 ph = 21p per min

1 FTE assistant grade 3 spine point 13 (mid grade) = £21925.39

= £11.40 ph = 19p per min

Add to track list & re-name file using PackTracker details. 2-5mins. Grade 4		£0.42 - £1.05	
Eligibility checks. 2-6mins. Grade 4.		£0.42 - £1.26	
Info entered into PackTracker & recorded for CLA return; ordered from BL using PackTracker. 2-5mins. Grade 4		£0.42 - £1.05	
If ordered from BL		If scanned in-house	
PDF downloaded from BL; PDF cover added & OCR-ed. 4-6mins. Grade 4.	£0.84 - £1.26	Retrieve, scan, OCR and add PDF cover. 30-90mins. Grade 3	£5.70 - £17.10
Cost of order from BL (assuming under CLA)	£8.35	Upload to VLE CMS; give read/manage rights. 1-3mins. Grade 4.	£0.42 - £1.05
Upload to VLE CMS; give read/manage rights. 1-3mins. Grade 4.	£0.21 - £0.63	Inform academic, including link to material & info on how to use. 2 mins. Grade 4.	£0.42
Inform academic, including link to material & info on how to use. 2 mins. Grade 4.	£0.42	Inform Bibliographic Services of new item available digitally. 2mins. Grade 4.	£0.42
Total (per reading):	£11.08 -£14.02		£8.22 - £22.35

Assuming 90% will be BL and 10% will be in-house, as at Leeds, the average cost for a reading, would be £10.79-£14.85.

Cost for 1000 readings = £10790-£14850

Cost for 5000 readings = £53950-£74250

A surprise, based on these timings, is that in-house scanning appears potentially cheaper than using BL HESS. Lucy adjusted my timings to allow for eventualities such as books being difficult to find on the shelves, so I suspect that the lower figure might often be closer to the mark. My timings were based on a small sample, however.

Funding models

(See Appendix 6 for detailed funding models).

A number of possible funding models were drawn up, based on discussions with Elizabeth Harbord and Julie Allinson. The models are summarised below, but the full document in Appendix 6 also includes pros and cons for each.

1. Service staffed and fully funded by the Library. Library offers all academic departments unlimited free digital readings. This is the model currently followed by many services, such as those run by Leeds and Sheffield University Libraries. The Library absorbs the cost of external services such as HESS and Heron.
2. Service staffed and partially funded by the Library – limited free readings plus paid-for readings. Library offers all academic departments a limited quota of free digital readings per module and then offers further readings at cost price (staff + equipment + external services e.g. HESS).
3. Service staffed and partially funded by the Library – limited free readings plus no paid-for service. Library offers all academic departments a limited quota of free digital readings per module. Further readings are scanned by academic departments themselves. This model seems to be followed by some services in other HEIs.
4. Service staffed and partially funded by the Library – all readings part subsidised by the Library and part paid for by departments. Library offers all academic departments an unlimited number of digital readings, but all readings are paid for by departments. The Library subsidises all readings, probably by paying for staffing costs and equipment and passing the additional costs of using services such as HESS and Heron to academic departments.
5. Service staffed by the Library, but fully paid for by academic departments. Library offers unlimited course readings to academic departments, but all readings are paid for by departments at full cost price (staff + equipment + external services).
6. Self-service – academic departments do their own scanning using equipment provided by the Library and under Library guidance. This model is based on a suggestion that the Library could provide some kind of digitisation suite, or self-service scanners, which departments could use to do their own scanning. The Library would also offer training and guidance, but, essentially, copyright and CLA compliance would be the responsibility of the departments.
7. A combination of models 3 and 6. Library offers a limited quota of readings free to departments. The Library also offers a digitisation suite within the Library building for departments to do any extra scanning themselves.

The best of these options seems to be option 4. It seems likely that offering all readings free of charge to departments will eventually become too expensive, ruling out option 1. The final two options, which both involve setting up a self-service digitisation suite, would be highly problematic to implement and don't offer departments enough benefits. Option 4 is fairer than options 2 and 3, as these options penalise departments which make heavy use of the service. It would also be less burdensome to implement than options 2 and 3, as there would be no need to keep a record of how many free readings had been done for a particular department.

Recommendations

- The experiences of other EKT services and discussions at the online course reading Exchange of Experience event suggest that the service model chosen for the EKT pilot project is the best one for York to take forward into the University-wide service.
- Consider limiting or phasing promotion of the service initially, until levels of demand are known, especially if the service is to be fully funded by the Library.
- Monitor how much time is spent on in-house scanning in order to see if this could be cheaper than using BL HESS
- The Library should offer all academic departments an unlimited number of digital readings, but all readings should be paid for by departments. The Library should subsidise all readings, probably by paying for staffing costs and equipment and passing the additional costs of using services such as HESS and Heron to academic departments.

Digitisation Service

Process

The idea of a single digitisation service is to develop a service which will focus on image digitisation and share workflow, staff and resources across CCS, YODL and the Borthwick. Part of the value of the case study into EKT and the Borthwick service is that has started a process of thinking about how to do things better across the board.

To support the development of this idea, a process review exercise was initiated as part of the case study. This to date has consisted of a meeting, in August 2010, involving key members of staff involved in digitisation activities. The key processes involved in the various digitisation activities were mapped out. The next stage is to organise a further meeting to examine crossovers and similarities between the processes.

Outcomes

Some of the areas which emerge from the exercise are:

- Retrieval of library material from shelves and reshelving (EKT and Photocopying. History of Art very occasionally).
- Electronic delivery of digital files (Photography, Borthwick, Photocopying).
- Storage of copyright declaration forms (Photocopying, Borthwick).
- Storage of digitised files and the ability to check previous digitisation (Photocopying, Borthwick, EKT, Photography)
- Electronic receipt of requests (EKT, Borthwick, Photocopying? History of Art?)

- Completing the CLA return (EKT, History of Art and Photocopying – use PackTracker for Photocopying?)
- Scanning of library materials (EKT, History of Art and Photocopying)
- Scanning special collections/archive material (Borthwick, Photography)
- Taking payments (Photocopying, Borthwick, Photography, EKT?)
- Price setting (Photocopying, Borthwick, Photography? EKT?)
- Service standards (all)
- Copying agreements – what can be done with copies? (All)
- Permissions to publish (Borthwick, Photography, Photocopying? History of Art?)
- Awareness of regulations and licences (all).
- Business service continuity (all).

Recommendations

- Organise a follow-up meeting to the digitisation process meeting to explore crossovers.

References

Allinson, J, et al. (2009) *York Digital Library content model for images*. [Online]. Available at: https://vle.york.ac.uk/bbcswebdav/xid-448730_3 (Accessed 15/07/2010).

Cornell University Library (2003), *Moving theory into practice: digital imaging tutorial*. [Online]. Available at: <http://www.library.cornell.edu/preservation/tutorial/conversion/conversion-04.html> (accessed 15/07/2010).

Hedges, J. and Secker, J. (2010), *Survey of digitisation of core readings in UK higher education*. [Online]. Available at: http://clt.lse.ac.uk/Projects/Survey_finalreport.pdf (accessed 15/07/2010).

NoWAL (n.d.) *Digitisation: case studies*. [Online]. Available at: http://www.nowal.ac.uk/digitisation.php?page_id=3 (Accessed 15/07/2010).

Puglia, S., Reed, J. and Rhodes, E. (2004) *Technical guidelines for digitizing archival materials for electronic Access: creation of production master files – raster images*. [Online]. Available at: <http://www.archives.gov/preservation/technical/guidelines.pdf> (accessed 15.07/2010).

Appendix 1. Anonymised summaries of archives and special collections on-demand digitisation services

Institution A

Institution A handles a very large volume of requests for copied documents from the public – about 2000-2500 per month. They offer digital files, prints, microfilm and microfiche. They have a fully automated software system to manage requests, which they developed themselves. This includes a web form for requests. The system changed slightly in Spring 2010. The old system involved the customer submitting their request details via a web form. Staff would then retrieve the document, count the pages and decide on the copy process to be used (based on size). This information would be fed into the software, which would generate an estimate of the price and automatically email this to the customer. The customer would then have to pay online and the order would be fulfilled.

The new system automatically generates the estimate without the staff having to retrieve the document. The previous system was not very cost effective, as less than 44% of requests for an estimate resulted in an order. Sometimes, people used the estimate system to make sure that the document was available when they came in person to the archive. If a document had to be sent from a remote site, this could be expensive. In the new system, the estimate is based on the average number of pages for the series that the requested document is from. The software maps the page number average against a price banding system and emails the customer with the price. There are three bands: A=£10; B=£20; C=£30. Thus, there is now a minimum charge of £10. As well as being more cost effective, the new system will reduce the handling of records.

There are three types of delivery: print, disc and email. They are priced in order to encourage people to use the method that uses the least consumables – partly for cost and partly for environmental considerations. Thus, email will be the cheapest method. In this method, the software will send the file to an FTP site and the customer will be emailed a link to download it. This will work for 21 days. A copy of the file will be kept for three months on file store, in case of queries about it. No files are kept beyond this point. Certain records have been microfilmed. If a customer requests a record which is microfilmed, they automatically get a scan from the microfilm, unless they order colour. Many popular documents and series have been digitised and added to the repository service, where they can be downloaded for a fee.

All of the scanning is done with overhead book scanners. They have different sized ones, between A3 and A0. The smallest size that they will do is A3 – this is the appropriate size for foolscap documents, which are longer than A4. They are currently changing their standard dpi from 150 to 300 – this change is driven by improvements in technology which allow scans at the higher resolution to be done at the same speed as lower ones could previously be done. All scans will now be JPEG, unless the customer pays extra for TIFF. They don't do any post processing of images. Their commitment is only to research (i.e. legible) quality. There is a commercial service, which is separate (for publishers etc).

The software allows customers to track the status of their requests by logging into the website. The turnaround time for requests ranges from 7 to 14 working days, depending on the type of material being digitised and the number of pages. The online order form asks for the following information:

- media (paper, digital, microfilm, microfiche)
- colour or monochrome
- delivery method (post, person, download)
- CD-ROM or DVD (for digital files by post)
- microfilm process
- Details of up to 5 documents
 - Catalogue reference
 - Copy all pages?
 - Details – large text box
- Personal details
- Personal details
 - email
 - address
 - telephone
 - reader ticket no. (optional)

The service has a dedicated team.

Institution B

Institution B does not have an online ordering service at the moment, but they are hoping to put one in place, hopefully later this year. At the moment, orders for digitisation are taken via the Reader Services department, either from readers in person, or by phone, mail etc. Institution B does not at present have an ecommerce system, which limits what they can do, although they are hoping to get one this year. Payments can only be taken by card from users in person in the library, or by cheque, postal order etc remotely. They are developing an online form which will come directly to the Reprographics Department, and charge the customer directly, cutting out the Reader Services Department. They are currently working on creating a copyright declaration form with a digital signature.

Customer profile is very varied: students, family history researchers, academics, publishers etc. etc.

The volume of on-demand digitisation that they do is fairly low – just under 210 requests a month. They do not have a large team. There is only one person working full-time on on-demand digitisation

They use two medium format single shot digital cameras with Phase 1 P45+ and P65 camera backs. These are mounted on stands from Icam. The P45+ will do up to A3 one-to-one scale and the P65 up to A2. They also have a colour track scanner; a contact scanner for digitising maps up to A0 and a Nikon medium format slide scanner. The latter is mostly used for internal projects or doing jobs for other libraries and individuals who are concerned about scan quality and the conservation of their originals

Since orders come via the curators in the Reader Services department, the curators will decide whether the image is to be kept in the Digital Object Database. If it is fragmentary (e.g. only one page of a long report), or there are copyright issues, then the image will be captured as JPEG and it will simply be sent to the customer, then deleted (kept for about one month only). If it is to go in the Digital Object Database, then it is digitised to 400dpi as uncompressed TIFF. The technical standards are based on advice from TASI/JISC Digital Media and Cornell University.

Recently, Institution B simplified their price list. They used to offer different qualities/file sizes, but found that customers and Reader Services staff did not understand these options very well and it was time-consuming to explain. Now they just have one price and offer the highest quality (it saves work as well, as it avoids needing to convert TIFF to JPEG). For items which are to go in the Digital Object Database, they will use a software package to create a JPEG surrogate as well.

Scans are delivered on CD/DVD, email or by FTP – the same charge applies for each method. Usually, the curators in Reader Services recommend to the customer which is the best for them.

They are hoping to get self-service scanners in the library and have made a business case for this, but are awaiting the money.

Institution B still creates microfilms, usually for project work and at the request of publishers who want to make out of copyright books available as online resources. It is cheaper for the latter to order a microfilm and then digitise the microfilm themselves. Institution B gets a royalty from this work. The advantage of microfilm is its longevity and there is no need to migrate the image. However, microfilming is relatively low volume and will eventually be replaced by digitisation.

Institution C

The Imaging Service at Institution C produces digital copies of items from the collection. Paper copies are made by reading room staff using a conservation copier. Requests for digital images are received via a PDF ordering form which is available online and which the customer prints off and sends in by post or fax. A lot of people still make first contact by email, phone etc, but are still asked to fill in a paper form. They would like to develop online ordering.

Once a request form has been received, the details are entered in a spreadsheet and given a job number by the administrative assistant. Once payment is received via credit card or cheque, or via Institution C's invoicing system, the job is then passed to the photographer. The photographer retrieves the document from the stack and photographs it. Some post-processing is done by the photographer. Appropriate conservation work is carried out on the physical document prior to an after photographing. Conservation staff may refuse digitisation on the basis of the condition of the document. The following formats are

offered to the customer on the order form: JPEG (72dpi), small TIFF (300dpi), or large TIFF (600dpi). The images are then sent to the customer, either by email (JPEG, small TIFF, small number of images) or burned to CD/DVD or sent via FTP (large TIFF or a large number of images). Turnover time is 5 working days for the express service and 2-4 weeks for the normal service (though large jobs can take longer). Most customers are from outside the Institution and many from outside the UK. Institution C does not have an online payment system, so payments have to go through the till.

The service uses the following equipment:

- two Phase One 645AF III camera bodies
- one Mamiya 645AFII camera body
- Phase One P25+, P45 and P65+ digital camera backs
- Mamiya 45mm, 80mm and 120mm macro lenses
- Apple Mac Pro Quad Core and Mac Book Pro computers
- Copy stand with two Multiblitz Xenolux 1000 flash heads
- Traveller's Conservation Cradle, with 24 daylight LED lights
- Bowens Lighting kit

The Service has two fully-equipped imaging studios.

Digital images are also catalogued and uploaded to Institution C's repository. All images are made public, unless IPR prevents this. Currently, there is a backlog of images to add, so items do not get uploaded to the repository immediately. The metadata for images in the repository is based on VRA Core 3. The physical archives are catalogued with a mixture of paper and electronic lists. There is currently no electronic linking between images in the repository and the electronic archival finding aids, but Institution C is working on a service which will cross search both.

The Imaging Service does not have a formal team as such. There are two full-time members of staff: the service manager and a cataloguer/administrator. They currently have one half-time photographer, who is on a fixed-term contract and who does all of the on-demand photography. Beyond this, they have staff employed on a project basis. The photographers come to the Service with training and experience as professional photographers, but are offered additional in-house training in the handling of special collections/archive materials by conservation staff. Project staff are also offered in-house cataloguing training.

Institution D

Institution D runs an on-demand special collections digitisation service. Requests are received from readers in person in the special collections reading room, or by phone and email. There is currently no online form for ordering digitised documents/images, although Institution D's repository software is capable of doing

this. They offer photocopying and digitisation services, but reading room staff will often advise readers that, if the document they want photocopying is too delicate to place on a photocopier, digital photography is the only option. All customers making a request have to fill in a form, even if the initial request is by email. Customers are either invoiced or pay when they collect their disc from the reading room.

Scanning is done either within Special Collections, or by Institution D's institution-wide design and reprographics service. In Special Collections they have one flatbed A3 scanner capable of scanning up to 1200dpi – although the highest resolution they have scanned to is 600dpi and the standard is 300dpi – and one digital camera on a stand. The camera is not a high-end camera and isn't able to capture enough detail at larger sizes, making it of limited use. Images are not good enough for publication. It is used to avoid placing delicate items on the flatbed scanner. All images for customers are captured as JPEGs, but TIFF may be used for project work.

Anything which Special Collections isn't able to do, or which is for publication, is sent to the institution-wide design and reprographics service. Sometimes the item is sent off-site to the design and reprographics service, but a photographer will come to Special Collections for material which cannot leave Special Collections.

Pricing is at the same rate, regardless of whether digitisation was carried out in-house or by the design and reprographics service. Pricing is per page (with a reduced rate for over 11 pages). A one-off handling charge and an extra charge for difficult material are also levied. Users are allowed to take their own digital photographs in the reading room free of charge, but subject to restrictions and to signing a form.

Delivery of digitised images is usually by post on a disc. The service will email files to customers, but prefer not to. The customers are diverse: academics, researchers (e.g. for TV companies); private readers etc. Digitised images are also added to Institution D's repository with metadata, where allowed by copyright restrictions.

The Special Collections team consists of the Head of Special Collections, a service manager, four part-time front-of-house staff and one member of staff engaged on projects. On-demand digitisation is mostly done by the front-of-house staff. One of the difficulties that the service has faced is the difficulty of quantifying work in order to price the service. There has been an increasing complexity to the methods used – once it was just photocopying and film photography. Digital equipment requires a higher skill level than traditional photocopying and there are more possible options to explain and understand. The service has seen an increase in both staffed digitisation and users taking their own digital images. The increase in users taking their own pictures does not seem to have reduced demand for the staffed service.

Institution E

Institution E runs an on-demand digitisation service for special collections and archive material, which is a recent addition to their photocopying service. The service is fairly undeveloped as yet and is not currently advertised, as they are not confident that they can meet a large demand. They currently meet perhaps no more than one request a week. The photocopying service continues to run alongside the digitisation service. They will digitise instead of photocopying for several reasons:

- preservation – less damaging to originals to use a camera than a photocopier
- demand – sometimes customers ask for digital image
- easier to send to remote customer by email.
- publishing – publishers want high quality digital images

They have been delayed in offering a fuller service by building works – a room to house a digitisation suite has been delayed by six months, but it is now ready. It will house an A3 flatbed scanner, a digital camera with a copy stand and equipment for digitising audio and video material. At the moment, most digitisation is done using a handheld digital camera (which is ‘good’, but not top-end), sometimes with larger items laid on the floor, or with staff standing on chairs to get distance. The largest items they have digitised are slightly larger than A3. Individual images are emailed to customers as JPEGs, but multiple files are copied onto CDs, which are posted to customers. Images are also saved on institutional filestore, with no formal metadata, as Institution E does not currently have a digital repository (though there are plans to acquire one).

There is no formal ordering procedure for digitisation. Requests come either by phone/email, or from users of the reading room. Institution E is currently looking at a pricing policy. About 50% of the customers are from within the Institution. Demand for imaging – both for photocopying and digitisation – is low. This is probably because most users come to the archive to do research *in situ* rather than coming with a very specific request.

There are four members of staff (some full-time and some part-time) engaged in the digitisation service, but this forms only a small part of their work. Lack of staff time has been part of the reason why the service has not been advertised. They hope to formalise the service more when the digitisation suite is up and running. Institution E also has a large collection of audio visual material, which they have already had requests to digitise from. They have some basic equipment for digitising audio visual material, but are acquiring better equipment.

Institution F

Institution F is a small subject-specific archive hosted by a university. The archive collections consist of books and journals, photographs, ephemera, newspapers and audio visual material. Most of the digitisation activity relates to the photographs, newspapers and ephemera. All of the photographs held by the archive (circa 80,000) were digitised about ten years ago, but the standard of the digitisation was not good enough by today’s standards. In addition to this, the master files were stored on CDs, some of which are now unreadable. Low resolution versions of all of the images are on the archive’s website, available to the public.

The customers of the digitisation service fall into two categories: publishers and general customers. For publishers, or general customers requiring a high quality image, Institution F will redigitise photographs to a higher standard, but, for most users, the original digital files are used. Publishers are sent a high-quality digital file, but other requests are met with a digital print. They do not give digital files to general customers, as they have had too many incidents of images being reused without permission or commercially exploited. One issue with this approach is that the digital prints have limited light-fastness. Prints are offered in two qualities, with the higher quality costing £5 for A4 and £7 for A3. There is a £5 administration charge and an additional charge for more complex searches or research.

For general customers, there is no formal ordering procedure or form to fill in. Most requests are received in person or via email or telephone. With publishers, a contract will be signed, but this is handled by the Institution F's Director, rather than front-of-house staff.

Equipment used consists of an A4 flatbed scanner with a light hood for transparencies; an A3 flatbed scanner and a Canon digital camera. The service will share the digitisation suite which the host university is setting up and this equipment will be shared. Anything larger than A4 is digitised by being laid on the floor and photographed using the hand-held camera. Ephemera, including posters, are photographed in this way. They used to have a digital camera on a stand, but the focal depth didn't allow large originals to be photographed. They are hoping to acquire a mounted camera which will allow them to photograph larger things to a higher quality. Minimal post processing is currently done, but they may consider doing more if they get a mounted camera.

Digital files from ephemera are not systematically kept, as this would require too much manual editing of finding aids (many of which are MS Word documents), to record that the item has been digitised. If an item is in poor condition, a note may be made that it has been digitised and the file kept, to avoid having to handle the material any more. This means that the same items may be digitised repeatedly, but this is less effort than storing files and updating finding aids (though this may have to be reconsidered, as repeat digitisation is happening more and more). Institution F does not have any server space – digital files are kept on DVDs (previously on CDs). Audiovisual material is not currently digitised as part of the service.

Institution F has an online finding aid. This consists of static web pages with MS Word documents for individual collection lists. The collection is catalogued (mostly to item level) by collection and by format (e.g. posters, notebooks, flyers, photographs etc). This means that the same item is separately (but identically) catalogued on more than one list. The finding aid has only recently been created – before that Institution F had no formal record of its holdings.

Institution F is looking for funding for a repository and federated search service which will bring together library holdings with digitised photographs and ephemera. This would mean digitising the entire holdings of ephemera. They hope to include Web 2.0 features such as login, RSS, saved searches, personalisation and user-contributed metadata. Decisions such as metadata profiles and a platform are still to be made

Institution F has two full-time staff, plus volunteers. There is a Research Director, who spends most of her time on outreach activities; a Collections Manager who is responsible for day-to-day running of the Archive; an archive assistant three days a week; a cataloguer three days a week; and a book cataloguer one day a week.

The demand for digitisation is increasing (currently about twenty requests per week), but will disappear altogether if everything is digitised and made available online. It is envisaged that the repository will give access to medium quality images, but that high quality images will still not be given out to general customers, even for a fee (because of the problem of unauthorised use).

Appendix 2. Anonymised summaries of EKT services at other higher education institutions

Institution G

Institution G has been running an online course reading service for four years. Readings are digitised in-house from paper copies of books and journals held in the library, using the library's existing reprographics service. Copyright clearance is done under a CLA license. Ingenta Heron is used for copyright clearance of materials not covered by the CLA license and for materials where the library's copy is in too poor condition to digitise. The British Library's HESS is not used at all.

Institution G saw a slow initial take up of the service, but this was probably due to the fact that, prior to the Institution signing up to the CLA license, Heron was used and the cost passed on to departments. (Another possible reason was lack of knowledge of the service, as the service was started on the initiative of the library, rather than departments). Since the introduction of the CLA license the service has been free to departments (except where Heron is used) and the service has grown quickly. After the third year, there were 930 digitised readings in the database, but in the last year this number has more than doubled to about 2000.

Digitised readings are stored on the University Portal, where they can either be accessed directly by students, or accessed via links embedded in VLE modules. The only metadata created is a filename containing the name of the author, data and module code. Readings are not catalogued on the OPAC, but a staff note on the catalogue records for the print copies records that a digital copy has been taken, to prevent the print copy from being disposed of.

The service does not have a dedicated team as such. There is one full-time assistant who carries out most of the day-to-day tasks apart from scanning. Scanning is carried out by the reprographics team (1 FTE). No additional staff have been recruited to the reprographics team in support of the service; the additional workload is probably counterbalanced by a reduction in other sorts of copying carried out by the team, due in part to the availability of online resources and printers (and, presumably, the reduction of ad-hoc scanning of course readings). Overall management of the service is carried out by a member of staff who dedicates part of her working time to this task (hours depending on demand). Two further members of staff are trained and able to help out at busy times.

The workflow for a request is as follows:

- Requests for scanning received from lecturers via an online form on the University Portal. This is picked up via a shared email account
- Request is checked against the Library catalogue to see if the book/journal is in stock in a paper version
- Request is checked against the Library catalogue to see if it is available electronically. If it is, it will not be scanned.

- If it is from a book, staff will check to see if it is available as an ebook and, if it is, they will ask the lecturer whether they want the Library to simply purchase the ebook.
- Request is checked against the CLA list of excluded works.
- If the request is still to be scanned, Information Management staff retrieve the printed book/journal from the shelf. If it is on loan, it is requested from the user. This is one cause of delay in fulfilling requests.
- The details of the request are entered in an Access database. This was developed in house and is able to generate CLA returns and coversheets. A coversheet is printed and sent, with the book/journal and a slip detailing the pages to be scanned, to the Reprographics Team.
- The Reprographics Team scan the requested pages. The first page to be scanned is the coversheet and the whole is saved as a PDF file on a shared drive. The scans are not submitted to any optical character recognition process, so the text is not searchable. The quality of the scans is not as good as Heron scans, but the only complaints that they have had have related to intricate figures, where detail has been lost.
- The PDF file is then uploaded to the Portal using CMS software and a link sent to the lecturer, who may then embed the link in a VLE module.

Management of the service is done mainly using the Access database and Microsoft Outlook. The Library does not have a subscription to PackTracker – they would have considered it but already had a successful in-house solution. An audit of the readings in the database is carried out annually. This time-consuming task entails contacting lecturers to find out which modules are continuing and which readings are still needed. The readings for any module which is not running during the following semester have their links in the Portal removed (though the actual PDF files are retained until they are needed again). The same sort of procedure is gone through for defunct modules and for readings which the lecturers flag up as no longer needed.

Readings can be requested at any time during the year, but there are deadlines set some time before the beginning of each term, in order for the Library to guarantee that readings will be available before the start of term. The busiest time of year is late summer (during the summer course reading scanning is a large part of the work of the Reprographics Team). It can take less than a day for a request to be completed, if the book/journal is in stock and the team is not too busy. However, lecturers are asked to allow at least four weeks. While most of the scanning carried out is in response to lecturers' requests, there has also been a programme of replacing popular paper photocopies from the short loan collection with digital copies, though this takes a lower priority than requested readings.

Institution H

Institution H has been running an online course reading service since 1997. The service was begun at the initiative of an academic faculty, but subsequently taken over by the library. In the first few years, demand

doubled every year. It is still increasing, but at a slower rate. The service currently serves 302 modules. There are about 6000 readings in the database.

All course readings are scanned in house. Readings are requested by lecturers using a form available on the library website, which may be printed off and sent to the library. Physical books and journals from the library's collection are first photocopied and then the photocopy is scanned using Epson scanners (though they are changing to Plustek OpticBook scanners, which allow scanning right up to the page gutter). Scans are converted to text using OmniPage optical character recognition software. Turnaround time can be as little as two days, but lecturers are asked to allow six weeks. There is a deadline in June for obtaining material before the start of the academic year in September. The service is free to departments (the service has its own budget), but each module is limited to 24 readings.

Institution H has built up experience of in-house copyright clearance directly with publishers, but also uses the CLA scanning and photocopying license and Heron. Their experience suggests that in-house clearance is a viable option. It is time consuming and requires a lot of chasing, but the key is organisation. One advantage of direct clearance is that the permission applies to the whole university, so there is no need to restrict access to modules, as with the CLA license. Fees from publishers vary from free up to about £200 for a chapter. They don't very often get refusals and sometimes get permission to use extra material, especially from the USA, where copyright laws are more permissive for educational purposes. They have also cleared material from companies such as Ikea and Next. They work on a costing model in which they will pay up to 8p per page per student on the module, or up to £100 for a chapter (as the former might work out more than the latter for a large course). If it comes to more than that, they will ask the lecturer how important it is and will sometimes pay up to £150. Permissions usually last for one academic year and need to be renewed each year. An annual 'housekeeping' process is needed to work out which material needs renewing and to renew it.

Readings are stored on Institution H's VLE (with a list of links to them on the library's website, so that lecturers can see what other lecturers have requested). Access control is handled by the VLE software.

Institution H has a small dedicated team, consisting of a Team Leader (0.5 FTE) and four administration assistants (1.7 FTE). The scanning and day-to-day administration is carried out by the administration assistants. Institution H has developed an Oracle-based tool for managing their service, which is tailored to their needs. The Team Leader also acts as copyright adviser for the institution and handles non-routine questions which arise.

Institution I

Institution I provides online course reading by using the British Library's Higher Education Scanning Service (HESS), which provides scanned documents under the CLA license. Material provided by HESS is OCR-ed using Adobe Acrobat, but this is not completely accurate and Institution I recognises that this is not acceptable for students with visual impairment. Accordingly, if a course has a visually impaired student on it, they will request materials from Heron, who OCR scans to a much higher level of accuracy. If HESS is unable to supply a reading (because the item is not held by the British Library), Institution I will scan in-house – this accounts for about 10% of readings.

Institution I have tried using a local printer to digitise material and found that they provided a good quality and value service. However, this method meant that physical material was kept out of circulation for longer while it was at the printer's and so they only do this during vacation times.

The online course reading service at Institution I is administered by one full-time member of staff, who requests scans from the HESS via their Document Supply Service. Other members of staff and casual student workers can be called upon at busy times, but they have had problems with training and continuity. They have also had issues with the service being 'tacked onto' the rest of the library's services and not very well embedded. However, they feel that it can be an advantage not having too many links in the chain. The service has gathered popularity since it was introduced and staff at Institution I have warned that a new online course reading should be introduced in phases, with sufficient staff in place to cope with demand. They have found that time has been a greater concern than cost.

Institution I used to use Heron's PackTracker software to manage their service, but have now developed their own in-house tool, which saves them a lot of time.

Readings are stored as PDF files in a folder in the VLE and accessed via VLE modules. The only metadata consists of a filename with the name of the module and the author. Access control is handled by the VLE software and is limited by module.

The service is not heavily promoted amongst academic departments, as they fear that they could not cope with greater demand. The service is free to departments and is funded from the library budget, with extra funding from an alumni fund. If the latter were removed, they would have to reconsider their strategy (and possibly start charging departments).

Institution J

Institution J mainly scans course readings in-house, under the CLA photocopying and scanning license and from both the original source and copyright-cleared photocopies. Heron is used for material not available under the CLA license. With non-CLA permissions, Institution J have been able to gain permission to digitise more than one chapter/article from publications, but usually no more than two or three, as publishers will object. Scanning is carried out using Plustek Opticbook 3600 scanners. Scans are OCR-ed using Adobe Acrobat Professional and saved as TIFFs. Minimal cleanup of the images is done. Readings are branded with the library's logo, so that users know that it is a library resource.

Institution J has a dedicated team (6 FTE) and they deal with reading lists, paper and digital course packs and scanning (with the majority of the time being spent on digital readings). The service is very strongly supported by the Librarian, hence the investment in staff. The team is divided up into four subject-area teams, which handle requests from those areas. The team is also responsible for reading lists, but spends around 80% of its time on course readings. One of the advantages of having a larger team which also has other tasks is that the service is better embedded in the library and staff in the team are more aware of the larger library service. The service began life as a pilot to five academic departments. There is more demand for the service from humanities departments than science departments and some of the heaviest users were part of the initial pilot phase. Institution J do not currently limit the number of readings which can be

digitised for a course, but may have to in future. The service made 2156 readings available to students in 2007-8 and projected that they would make about 3100 available the following year. In the event, they experienced an increase in demand of 50% in 2008-9 and 25% in 2009-10.

Readings are stored on a secure server and linked to in reading list and VLE software. Heron PackTracker is used for maintaining a database of readings and for generating the CLA return. The service is paid for from a dedicated library budget. Readings available in ebooks are not scanned. However, this creates an issue, as, if the ebook is purchased, it comes out of the small departmental budget rather than out of the online course reading budget. Course packs have a twelve-month 'lifespan' and have a renewal date of 90 days before the end of term.

The service is promoted to departments by Academic Liaison Librarians, but the team is also planning to do some of their own promotion, as Institution J only has a small team of Academic Liaison Librarians. They are working on leaflets advertising the service. Emails sent to teaching staff notifying of deadlines for submitting requests also increase awareness of the service. However, they worry that, if they promote the service too widely, they won't be able to cope with the demand. The service is currently free to departments. The service has been targeted at departments which have fared badly in the National Student Survey, and this has been a way of demonstrating the service's usefulness and obtaining funding from the library.

Appendix 3. Technical requirements for archival master images from the Borthwick on-demand digitisation service for YODL

Overview

This document will look at the technical requirements needed for digitising material from the Borthwick Institute to create archival master images¹ for storage and delivery via YODL. The Borthwick Institute feel that their existing practices are satisfactory for customer needs (although they are responsive to the occasional customer who may need images at a higher quality). However, they do feel the need to examine the requirements for publishing images in YODL and for long term preservation. The key requirement for textual documents is legibility. Slightly different considerations apply to photographic prints, which are also scanned as part of the service (negative and other transparencies are not currently scanned as part of the service).

Requirements are derived from the YODL *Image Content Model*, which defines how image files are handled in YODL, and from a survey of published standards, which are presented in Appendix 4. Digital archival master images should be saved as uncompressed TIFFs, with 24 bit colour, RGB colour space, at full size and with a resolution of at least 300dpi. A longer discussion on resolution follows, as this is one area where a single recommendation is not possible. The Cornell University Quality Index and the NARA guidelines from Appendix 4 are used to work out guidelines for resolution for textual documents and photographic prints.

Current practice and equipment

The Borthwick have two scanners: a Scan2Net SMA21 book scanner and a Konika Minolta MS6000 MkII microfilm scanner. The book scanner is capable of scanning originals up to A2 in size. For material which cannot be scanned by either of these, they use the Library Photographer. Documents larger than A2 can be photographed by the Photographer, using a Hasselblad H3D digital camera, provided that they are able to be removed from the Borthwick and taken to the photographic studio, located in the adjacent JB Morrell Library. Images scanned from microfilm would not be archived in YODL, so this document will consider the book scanner. For reference, the technical specifications for both scanners are as follows:

	Scan2Net SMA21 book scanner	Konika Minolta MS6000 MkII microfilm scanner
File formats	JPEG, TIFF, PNG (Current	JPEG

1

The NARA *Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images*, on which some of the recommendations here presented are based, distinguishes between *production master files*, which are used to generate surrogates for access, and *preservation master files*, which are used for preservation purposes. The NARA guidelines relate to the former. In this document the term ‘digital archival master’ is used essentially to cover both, as a full preservation policy for YODL has not yet been developed and NARA guidelines covering preservation master files do not seem to have been produced.

	practice is to output as JPEG)	
Resolution	300dpi up to A2; 400 dpi up to A3; 600dpi up to A4 (although they only currently scan to 300dpi).	200, 300, 400, 600, 800 dpi (PC mode), or 400, 600 dpi with MSP 3000 printer attached
Bit depth	24 bit colour	1-bit bitonal; 8 bit greyscale
Colour space	RGB	N/A

For original documents scanned with the book scanner, current practice is to scan in colour (24 bit, RGB) at 300dpi and output the file as a JPEG, which is then sent to the customer. Occasionally, they scan as greyscale, usually at the request of the customer, but, in future, they will remove this option, except for high quality scans for publication. More recently, they have experimented with converting JPEG files to PDF, so that multiple images (e.g. of a document with several pages) can be bundled together, but so far this has only been done for scans from the microfilm scanner (though with a view to extending it to the book scanner).

Requirements for YODL

YODL's *Image Content Model* (https://vle.york.ac.uk/bbcswebdav/xid-448730_3) defines requirements for images in YODL. YODL supports a range of file formats, including TIFF, JPEG, PNG, GIF and JP2 (partial). PDF is also supported. TIFF images are treated as archival master copies. This is in agreement with the published standards in Appendix 4. When a TIFF file is ingested into YODL, three JPEG versions are automatically created for dissemination: a full sized JPEG for download; a medium sized JPEG for screen preview; and a thumbnail. From this and from published standards and best practice guidelines (see Appendix 4) the following requirements emerge:-

	Archival master in YODL	Version sent to customer*
File format	Uncompressed TIFF	JPEG/PDF
Colour space	RGB	RGB
Bit depth	24 bit colour	24 bit colour (occasionally 8 bit greyscale)
Size	1:1	1:1
Resolution	At least 300dpi, but see discussion below	300dpi

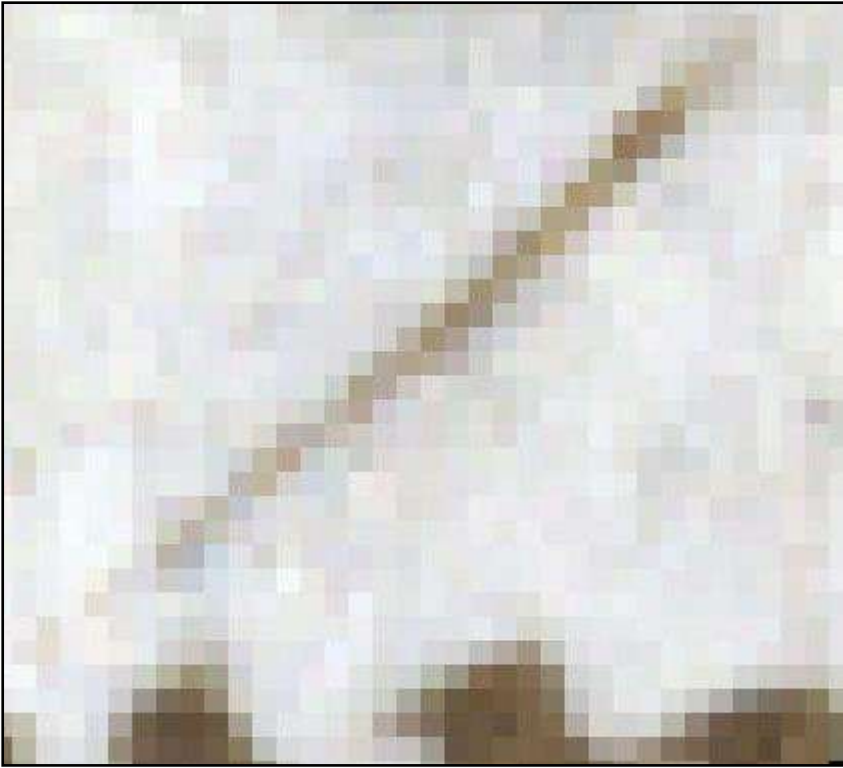
* Represents current practice.

Resolution for textual documents

Resolution, measured in dots per inch (dpi) or, more accurately, pixels per inch (ppi), determines the amount of information contained in a digital image of a given size. This is one area in which the guidelines and standards consulted vary (although 300dpi emerges as a minimum for many purposes). The reason for this is that the resolution needed depends on the nature of the material being digitised (e.g. the size of characters in the text); whether it is being digitised as bitonal (black and white), greyscale or colour; and user needs. For the Borthwick Institute, the main requirement in terms of user needs is legibility. This includes being able to discern (but not necessarily examine in detail) fine strokes of the pen; being able to distinguish creases and tears in the support of the document from pen marks; and being able to see very faint pen marks. The Borthwick feel that their current practice of scanning documents to 300dpi is sufficient for this and several examples of colour scans (viewed as TIFFs or Photoshop PSD files) seem to bear this out.

One way of checking this conclusion is to apply Cornell University's Quality Index (QI) to the scans (see Appendix 4). This provides formulae for calculating the resolution needed to capture the smallest significant detail in a document at different levels of quality, ranging from QI of 3 or less (barely legible) to QI of 8 (excellent). This can be calculated based on the height of the smallest character in a textual document, or the narrowest stroke or line width in a document. ('Character' seems to refer to things of letter size, not punctuation or other small marks). Scanning at 300dpi, the smallest character that could be captured at QI 8 (excellent) would be 1.37mm high. Most of the scanning done by the service is of handwritten documents and it is more or less impossible that a handwritten document would have letters this small. Printed books from the Library's special collections also form a small part of the scanning done and there is the possibility of printed archive material being requested, so it is not outside the bounds of possibility that documents with text as small as 1-2mm may be scanned. However, this is unlikely to form a very large part of the service and may have to be dealt with as a special case, when it arises, by increasing the resolution. (The only case in which this would be difficult would be for items larger than A3, which would probably have to be digitised by the photographer). In the example scans, which were captured at 300dpi, an individual character would pixelate heavily if enlarged to fill the computer screen, but be perfectly legible (including fine strokes) if viewed at a comfortable reading size.

Stroke or line widths of less than 1mm are difficult to measure accurately, but the Cornell method recommends that no less than 2 pixels are used to represent the finest line or stroke or line. The finest 'sharp' (i.e. not faded or broken) pen strokes in the example image files are about two or three pixels wide. Some fine strokes are clearly discernable when the digital image is viewed at full size, but resemble an example of an inadequately rendered line on the Cornell QI website when viewed greatly enlarged. However, this seems to be because the lines are faded or not very crisp and are broken (as when a pen skips over the grain of the support, or when a mark has been worn by abrasion). See illustrations for examples.



Example of a very fine pen stroke from one of the example scans, enlarged to show how many pixels form the line.



Example of a line which is much more faint and broken, but still discernable as a pen stroke.

Colour is important for discerning faint or unclear pen strokes and for distinguishing pen strokes from creases and tears.

Resolution for photographic prints

The Borthwick on-demand service digitises a range of different types of photographic prints (not negatives or transparencies, as they are not equipped for this). Standard practice is to scan at 300dpi, as with other document types, and they feel that this is adequate for most customers. However, there is a need to examine standards for archival master images.

Determining the resolution at which to scan photographic prints (and other continuous tone images) for archival purposes is difficult, as there is no fixed element which can provide a metric (such as a character or line width). The NARA guidelines (see Appendix 4) have a *minimum* recommendation of 3000 pixels along the longest edge of the image. To calculate the resolution from this, divide 3000 by the length of the longest edge of the print (in inches) so, for example for a 4" x 6" print: $3000/6=500\text{dpi}$. Following this

model, the resolution will be lower the larger the size of the print. The full recommendations of the NARA guidelines are summarised in the following table.

Print size	Pixels along longest edge	Resolution
8" x 10" or smaller	4000	approx. 400dpi-approx. 800dpi
larger than 8" x 10" and up to 11" x 14"	6000	approx. 600dpi-approx. 430dpi
larger than 11" x 14"	8000	approx. 570dpi and lower

The Borthwick's current practice of scanning at 300dpi is inadequate for most photographic prints, even if the minimum guideline is followed. Some prints need to be scanned at a higher resolution than the 600dpi maximum which the Borthwick's book scanner is capable of (those less than about 6 inches long for the full guideline and those less than 5 inches long for the minimum guideline). The Library Photographer scans photographic prints at a resolution sufficient to achieve 4000-5000 pixels on the longer edge using an A4 flatbed scanner which is capable of much higher resolutions. The recommendation is, therefore, to follow the full NARA guideline of 4000 pixels on the longer side and use a flatbed scanner for scans which need to be over 600dpi.

Metadata

YODL's Image Content Model (https://vle.york.ac.uk/bbcswebdav/xid-448730_3) defines a profile of the VRA Core 4.0 metadata schema for images. This was devised for describing images of artwork and cultural objects, but has already proved able to adequately describe materials such as manuscripts and playbills. The schema is based on the distinction between a 'work', meaning the original artefact being digitised, and the 'image', meaning any image or surrogate of it. In the case of a scanned original manuscript, the 'work' section would describe the manuscript and the 'image' section would describe the digital image of it. An editor for VRA Core has been created for the YODL interface and has been in use for some time.

Recommendations

- Scan all images for YODL as TIFF files. JPEGs for delivery to customers can be derived from the TIFF version.
- Scan textual documents using the book scanner at 300dpi, unless they contain characters less than circa 1-1.5mm high, for which increase the resolution to 400dpi.
- Digitise textual documents over A3 size with characters less than circa 1-1.5mm high using the Hasselblad digital camera
- Scan photographic prints to an image size of 4000 pixels along the longer side (adjusting the resolution accordingly). Use a flatbed scanner, such as the one in the Library photographic studio, for resolutions over 600dpi.

Appendix 4. Standards and best practice guidelines for creating digital archival master images

US National Archives and Records Administration (NARA) Technical Guidelines for Digitizing Archival Materials for Electronic Access: Creation of Production Master Files – Raster Images

By Steven Puglia, Jeffrey Reed, and Erin Rhodes, 2004

URL: <http://www.archives.gov/preservation/technical/guidelines.pdf> [accessed 5th May 2010]

The NARA guidelines are an authoritative and informative source of information and advice on technical standards for digitisation, aimed at those planning and carrying out digitisation projects. A number of other standards (e.g. Indiana University DLP and California Digital Library) are based on them. They provide specific guidelines for different kinds of material, including different kinds of textual documents; different formats of photographic material; and oversized material. The bulk of the document discusses standards for ‘production master files’ (i.e. master files from which derivative images can be created for access purposes such as screen display and printing), but an appendix discusses requirements for ‘preservation master files’. It seems like there are yet to be agreed standards for this.

Indiana University Digital Library Program (IU DLP) Use of Digital Imaging Standards and Best Practices

By Jenn Riley, 2004

URL: <http://www.dlib.indiana.edu/education/workshops/Ista04/handout9.pdf> [accessed 5th May 2010]

Indiana University guidelines are based on early published recommendations from NARA, but seem to roughly reflect the current alternative minimum standard from the NARA guidelines on many points. For archival documents, they scan master images at 300dpi, 24-bit colour or 8-bit greyscale and save as uncompressed TIFF files. For photographic materials, they scan to a long side pixel dimension of 3000 and adjust the resolution accordingly. Printed texts are scanned at 600dpi as 1-bit bitonal images, which are saved as compressed TIFFs (using CCITT group 4 compression) and subject to optical character recognition.

California Digital Library Guidelines for Digital Images (CDL GDI)

2009

URL: http://www.cdlib.org/services/dsc/tools/docs/cdl_gdi_v2.pdf [accessed 5th May 2010]

CDL GDI are substantially based on the NARA guidelines, with some differences because of local needs and requirements. However, the guidelines are worth referring to, as they re-present the tables of specifications for different material types from the NARA guidelines in a slightly simpler form.

Library of Congress Digital Formats for Content Reproductions

1998

URL: <http://memory.loc.gov/ammem/formats.html>

These guidelines are now rather old, but present useful standards for a number of material types. For some printed textual material (e.g. late 19th and 20th century), they recommend scanning as bitonal images at 300dpi and compressing them as TIFFs. However, it is unlikely that this approach would be suitable for the type of material which the Borthwick deals with. For textual material reproduced as image (i.e. not scanned as a bitonal image and then converted to text) they allow for some 'routine' material (i.e. not 'treasures') to be saved as JPEG/JFIF, compressed to a ratio of 5:1. Otherwise, they recommend 8-bit greyscale or 24-bit colour uncompressed TIFFs at 300dpi. For photographic materials, they recommend 3000-5000 pixels for the longer edge.

JISC Digital Media advice papers

URL: <http://www.jiscdigitalmedia.ac.uk/stillimages/> [Accessed 5th May 2010]

JISC Digital Media provide a range of training and advice on digitising still images. Advice papers include basic guidelines, choosing a file format, colour management, advice on equipment, workflow etc. JISC Digital Media recommend uncompressed TIFF files for archival master images.

The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials

By the Humanities Advanced Technology and Information Institute (HATII), University of Glasgow, and the National Initiative for a Networked Cultural Heritage (NINCH), 2002.

URL: <http://www.nyu.edu/its/humanities/ninchguide/> [Accessed 5th May 2010]

A useful best practice guide to planning a digitisation project, including image capture.

Cornell University Quality Index (QI)

2000-2003

URL: <http://www.library.cornell.edu/preservation/tutorial/conversion/conversion-01.html> [Accessed 5th May 2010]

Digital Imaging for Archival Preservation and Online Presentation: Best Practices. Working Paper

MATRIX: The Center for Humane Arts, Letters and Social Sciences Online at Michigan State University, 2001

URL: http://www.historicalvoices.org/papers/image_digitization2.pdf [Accessed 5th May 2010]

Guides to Quality in Visual Resource Imaging

Digital Library Federation, 2000

URL: <http://www.diglib.org/pubs/dlf091/> [Accessed 5th May 2010]

Preparing Digital Surrogates for RLG Cultural Materials

OCLC Research Libraries Group, date unknown

URL: <http://www.oclc.org/research/activities/past/rlg/culturalmaterials/surrogates.htm#recs> [Accessed 10th May 2010]

Contains useful recommendations for digitising a range of materials, including text, images and audio.

Creating Digital Resources for the Visual Arts: Standards and Good Practice

Visual Arts Data Service (VADS), date unknown

URL: http://vads.ahds.ac.uk/guides/creating_guide/contents.html [Accessed 10th May 2010]

An introduction and practical guide to creating digital resources for the visual arts. Topics include IPR, creating digital resources, metadata, collection management, delivery, storage, preservation and introducing specialised digital formats.

Session on care and handling of library materials for digitisation: In-House Course Handout

LIBRARY OF CONGRESS NATIONAL DIGITAL LIBRARY PROGRAM and the CONSERVATION DIVISION, date unknown

URL: <http://memory.loc.gov/ammem/techdocs/conserv83199b.pdf> [Accessed 10th May 2010]

Contains advice on the handling of special collections/archive material.

Technical notes: by type of material. General comments on digital reproductions of textual materials for American Memory

Library of Congress, date unknown

URL: <http://memory.loc.gov/ammem/dli2/html/document.html> [Accessed 10th May 2010]

Appendix 5. Model for providing an online course reading service

This model was drawn up based on experience at Leeds and Sheffield and a few other university libraries to inform the York service. The model is broken down into several stages in the lifecycle of a digital course reading: selection, creation/purchase, IPR/licensing, administration, metadata, and storage/access. It suggests a number of alternatives for each stage and attempts to evaluate each. Due to the complex, changing nature of this area, and differences in institutional context, it is not always possible to make absolute judgements, but it is hoped that the following points may be useful as starting points to inform discussion.

Selection

	Advantages	Disadvantages
In response to academics' requests	<ul style="list-style-type: none"> • Provides a direct service to academics. • Likely to be the most useful texts to students. • Can directly query the academic (e.g. to see if it is worth paying the extra money to copyright clear a reading via Heron). • Gets academic buy-in (e.g. they may recommend the service to other academics). • Enables library input in the creation of reading lists (e.g. it may be better to buy a reading as an ebook instead). 	<ul style="list-style-type: none"> • No control over demand (unless limits on the number of readings per module are applied). • Hard to manage expectations, as academics may not understand the copyright limitations etc.
from checking reading lists	<ul style="list-style-type: none"> • May be used to free physical copies of books from being placed in key texts/short loan collections • Ensuring delivery of core reading • May be a way of 	<ul style="list-style-type: none"> • Academics may feel you're 'interfering' • Time consuming to do. • May be difficult to work out eligibility of material for digitisation (e.g. if lists are split up by week or seminar, it may be difficult

	managing demand	to work out whether only one extract from a particular book is wanted, or whether different extracts of the same book are listed under different weeks).
recommendations from librarians	<ul style="list-style-type: none"> • Helps embed service in library as a whole to have librarian involvement 	<ul style="list-style-type: none"> • Based on interpretation – librarians may not know whether it is a core reading or not. • Dependent on personality of librarian – can lead to a bias towards those librarians who are keen on the service. • Can be over keen! This can lead to large number of requests. • Can lead to duplicated effort, if librarians put time into searching through reading lists.
replacing paper photocopy collections	<ul style="list-style-type: none"> • May be used to create space in physical library • Good way of taking stock 	<ul style="list-style-type: none"> • Usually only possible to do partially • Need to do a lot of checking e.g. module is current?

Creation or purchase

Many HEI online course reading services use a mixture of the following approaches to either creating or acquiring digitised readings. For example, if HESS is used as the default means of obtaining readings, this could be backed up by in-house scanning for material which is covered by the CLA licence but not held by the British Library and Heron used for material which is not covered by the CLA licence.

	Advantages	Disadvantages
In-house scanning. This can be done from the original books, from photocopies or from copyright cleared copies	<ul style="list-style-type: none"> • May utilise existing photocopying staff – extra work may be offset by reduction in other forms of reprographic work (e.g. due to increasing use of home/office printers). • Cheaper than Heron. • Work can be done to agreed standards • Potentially for fast turnaround and flexibility (e.g. for urgent requests). Can be quicker than Heron. • Not relying on third-party suppliers – greater control over turnaround time. • Developing in-house skills and capabilities 	<ul style="list-style-type: none"> • Need the capacity to do it, in terms of staff, expertise, space and equipment. • Material has to be taken out of circulation in order to be scanned. • Material may be out on loan when it is need for scanning, delaying delivery • Material has to be retrieved from the library shelves – this has to be factored into timings/costings for in-house scanning. For example, it may be more of a barrier for institutions which have multiple libraries, as material may need to be retrieved from a different site.
British Library HESS (Higher Education Scanning Service) http://www.bl.uk/reshelp/atyourdesk/docsupply/	<ul style="list-style-type: none"> • Faster service than Heron. • Cheaper than Heron. 	<ul style="list-style-type: none"> • Not as high quality as Heron. • Not OCR-ed.

<p>productsservices/hescanning/index.html</p>	<ul style="list-style-type: none"> • No need to retrieve items from shelves • Can get copyright fee paid 	
<p>Ingenta Heron</p> <p>http://heronweb.ingenta.com/heron/</p>	<ul style="list-style-type: none"> • High quality. • OCR-ed. • Good for materials not covered by the CLA license. 	<ul style="list-style-type: none"> • More expensive than HESS • Slower than HESS
<p>Outsourcing to a reprographics company</p>	<ul style="list-style-type: none"> • Can be good value for money and high quality. • No need to invest in equipment, space, training and staff. Reprographics companies are likely to have higher quality equipment, and more expertise than a library is able to afford. They also have established workflows and can benefit from economies of scale which may be passed back to the customer. • Many universities have reprographic companies on campus, so may 	<ul style="list-style-type: none"> • Need to send books offsite, taking them out of circulation for longer. • You still need to check the quality of material scanned. • Not developing in-house expertise and capability, although someone in the institution needs to be knowledgeable about digitisation to be able to liaise with the company*.

	<p>be convenient.</p> <ul style="list-style-type: none"> • They can OCR it. • No ongoing subscription. • May charge per page, not flat fee – so can end up cheaper. 	
--	--	--

* More advantages and disadvantages of outsourcing versus in-house scanning can be found in the JISC Digital Media article *To outsource or to digitise in-house?* at:
<http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/to-outsource-or-to-digitise-in-house/>

IPR/licensing

	Advantages	Disadvantages
<p>CLA licence</p> <p>http://www.cla.co.uk/#h</p>	<ul style="list-style-type: none"> • Single fee for institutional membership – no charge for each reading . • Less onerous than clearing direct with publishers. • Everyone has it – it’s there anyway! 	<ul style="list-style-type: none"> • Can only scan one chapter/article per book/journal per module. • Reporting can be onerous. • Some publishers/material is excluded. • Material published in some countries is excluded. • Readings only cleared for specified module(s). • Can’t do digital to digital copying if you don’t have the comprehensive licence

<p>Clearance direct with publishers</p>	<ul style="list-style-type: none"> • Material can be cleared for the whole institution, not just for a course (though this may not be a great advantage, as material may not necessarily have wide subject relevance). • More than one chapter/article per book can be cleared (though maybe only two or three chapters/articles). • Sometimes permission is granted free of charge, or at low cost. 	<ul style="list-style-type: none"> • Time consuming and complex to administer. • Fees can be expensive (up to £200 asked for a single chapter/article). • Permissions need renewing (and paying for) year on year.
<p>Clearance via Heron</p> <p>http://heronweb.ingenta.com/heron/</p>	<ul style="list-style-type: none"> • More than one chapter/article can be cleared (though not usually more than two or three). • Heron can be used for clearance for items which are digitised in-house. • Heron sometimes has deals with publishers which can make it cheaper than direct clearance 	<ul style="list-style-type: none"> • Cost – you have to pay per reading cleared • Can take quite a long time for readings to be cleared. • Locked files [?] • Permissions need renewing and paying for year on year.
<p>Clearance via Copyright Clearance Centre</p> <p>http://www.copyright.com/</p>	<ul style="list-style-type: none"> • Immediate clearance • Easy interface • No subscription charge • Used by Heron for 	<ul style="list-style-type: none"> • Cost depends on Stirling/Dollar exchange rate, as based in the US • Bias towards US publishers • Not able to offer

	some clearance work	clearance for everything <ul style="list-style-type: none"> • Price is per student per page – can be very expensive • Permission needs renewing every year
--	---------------------	--

Administration

	Advantages	Disadvantages
Heron PackTracker	<ul style="list-style-type: none"> • Has powerful functionality e.g. can automatically generate CLA return • No need to invest time and effort in developing an in-house solution 	<ul style="list-style-type: none"> • Subscription is expensive
In-house Excel, Access or other database	<ul style="list-style-type: none"> • Can be created to meet specific local needs 	<ul style="list-style-type: none"> • Time consuming and expensive to create • Requires expertise in database construction to create • May not be as heavily automated as PackTracker (e.g. may not be able to generate CLA returns)

Metadata

	Advantages	Disadvantages
Filenames and link texts (e.g. the title and author can form a link to the PDF in a VLE or portal)	<ul style="list-style-type: none"> • Quick and easy to create. • Filename stays connected to file even when 	<ul style="list-style-type: none"> • Limited information can be captured

	downloaded.	
Administrative database e.g. Access db	<ul style="list-style-type: none"> • Flexible and customisable way of administering readings 	<ul style="list-style-type: none"> • May be difficult to join up with other services and require duplicated effort to keep up to date. • Metadata may not be public – just for admin purposes
OPAC/MARC 21	<ul style="list-style-type: none"> • Existing interface which is already visited by users. • Existing bibliographic metadata standards (e.g. MARC) and expertise can be used. 	<ul style="list-style-type: none"> • May not be possible to suppress records – visible to people who don't have access
Repository/Dublin Core, MODS and other formats	<ul style="list-style-type: none"> • Access may be controlled to metadata, so not exposed to users who don't have access to content. 	<ul style="list-style-type: none"> • May not really be necessary for items which aren't going to be stored / preserved long term

Storage and Access

	Advantages	Disadvantages
VLE	<ul style="list-style-type: none"> • Access control may be already set, or be easy to set. • Readings are embedded directly in students' workspace – convenient and visible • Natural place to put it – focused on learning materials 	<ul style="list-style-type: none"> • Proprietary system – may be not possible to adapt to local needs • Not possible to record metadata (though this may not necessarily matter, as it may be judged unnecessary to record much metadata for readings which are not going to be kept over the long term)

	<ul style="list-style-type: none"> • Academic has control of material for their own course, which may encourage buy-in on the part of academics. 	<ul style="list-style-type: none"> • Devolves control of material to academics: fully centralised control is an advantage from a legal standpoint
Reading list software	<ul style="list-style-type: none"> • Convenient to students • May link to content held anywhere else (e.g. in a VLE CMS) 	<ul style="list-style-type: none"> • Linking only – need to store elsewhere
Web portal	<ul style="list-style-type: none"> • Technically simple to do • May be a good option in the absence of a VLE or other infrastructure 	<ul style="list-style-type: none"> • May not be as visible and convenient as in a VLE
Library catalogue	<ul style="list-style-type: none"> • Links the digitised reading to the physical book/journal in the library collection, which may be useful, for example, for alerting library staff not to withdraw a book from which a reading is being held (if licensing is under the CLA license, which requires that a print copy is held). • Record can be suppressed and used for administration purposes 	<ul style="list-style-type: none"> • The CLA licence prohibits making readings searchable, so this option may not be allowed if readings are provided under this licence • Linking only – need to store elsewhere
Digital repository	<ul style="list-style-type: none"> • Provides a strong infrastructure for storing and providing access to material • Good provision for metadata (though this not be crucial) 	<ul style="list-style-type: none"> • May be more focussed on storage/preservation • The CLA licence prohibits making readings searchable, so this option may not be allowed if readings are provided under this licence

Appendix 6. Electronic Key Texts Funding Models

Some general points

- At present, academic departments are responsible for doing their own scanning of course reading and their own CLA reporting. The Library is taking on responsibility for the CLA licence and an advisory role for copyright within the University, so anything which improves the University's adherence to the CLA licence and copyright law could be seen as supporting this aspect of the Library's duty. Centralising scanning for course readings and the associated CLA reporting within the Library could definitely be a way of doing this. This is an advantage for all of the staffed models (1-5) to the extent that the service is used.
- Regardless of who pays for it, offering a staffed, centralised service in the Library could offer efficiency savings and be more convenient to departments than doing it themselves (assuming that the service was easy to use, responsive and had a reasonable turnaround time etc.).
- All of the models, apart from model 6, depend on costings being worked out for both in-house scanning and for ordering digital material from external services (e.g. British Library's HESS and Ingenta Heron).
- All of the models depend to some degree on projections of the total number of readings likely to be requested during the coming academic year.
- Even established EKT services such as Leeds and Sheffield are currently providing a smaller number of readings per year than their institutions run modules – hence the potential demand is huge.

1) Service staffed and fully funded by the Library

Library offers all academic departments unlimited free digital readings. This is the model currently followed by many services, such as those run by Leeds and Sheffield university libraries. The Library absorbs the cost of external services such as HESS and Heron.

Pros

- Will be popular with departments and will undoubtedly produce the greatest take-up of the service of any of the models.

Cons

- The most expensive option.
- No way of managing the cost to the Library of running the service.
- Probably unsustainable in terms of cost – other university libraries which are running fully funded services are coming to this conclusion.
- No way of managing demand – demand may outstrip capacity.

- A free service may encourage academics to order readings that are not strictly needed.

2) Service staffed and partially funded by the Library – limited free readings plus paid-for readings

Library offers all academic departments a limited quota of free digital readings per module and then offers further readings at cost price (staff + equipment + external services e.g. HESS).

Pros

- More financially sustainable. Enables the Library to manage the cost to itself of running the service – the number of readings offered free could be calculated by dividing the service's budget by known costs and the projected number of modules using the service. Any readings offered beyond the free quota would be paid for by academic departments.
- Demand may be more manageable – the quota may discourage departments from ordering readings that they don't really need. Also, the paid-for readings would provide an income to pay for the additional staff time and equipment to cope with the extra demand.

Cons

- May seem unfair: the experience of other institutions shows that academic departments vary in how heavily they rely on digitised course readings, so some departments may not use their full quotas and others may want greatly in excess of their quotas. There may be some way of matching the size of the quota to some projection of the likely uptake of the service per department/module. This would solve the issue of unused capacity, but could still be seen as unfair.
- Administrative burden of recording how many readings have been provided for each module.
- Administrative burden of receiving and recording payments from academic departments.
- Calculations as to how many readings the Library can afford to offer free per module should probably take into account the fact that the service is likely to grow, so that for the first few years, the service will not be used for every module in the University. This would have to rely on projections of uptake, which are difficult to make. Unless it is accepted that the number of readings offered free per module would go down as the number of modules served increased, the service's budget would have to increase as the service grows.
- Academic departments may prefer to do scanning themselves once their quota is reached. Departments would have to pay their own staffing and equipment costs anyway, but, as they are already doing this, they may just continue doing so. The only incentive to use the paid-for service would be convenience and any efficiency savings that the Library is able to offer.
- Income from paid-for readings may be unpredictable (at least while the service gets established), making it difficult to judge staffing levels needed to manage the extra demand.

3) Service staffed and partially funded by the Library – limited free readings plus no paid-for service

Library offers all academic departments a limited quota of free digital readings per module. Further readings are scanned by academic departments themselves. This model seems to be followed by some services in other HEIs.

Pros

- It is difficult to see any advantages of this model over model 2, except that it removes the problem of having to put in place the staffing to cope with the paid-for service before knowing what the income from the service will be and the administrative burden of receiving payments from academic departments. The main attraction may just be that it is a simple model.

Cons

- Has all of the cons of model 2, except the two noted above.
- Would force academic departments to carry out any scanning they want beyond their quotas themselves.

4) Service staffed and partially funded by the Library – all readings part subsidised by the Library and part paid for by departments

Library offers all academic departments an unlimited number of digital readings, but all readings are paid for by departments. The Library subsidises all readings, probably by paying for staffing costs and equipment and passing the additional costs of using services such as HESS and Heron to academic departments.

Pros

- Fairer than models 2 and 3.
- No unused capacity, as in models 2 and 3 if modules fail to use up their quota.
- Provides an income to offset at least some of the costs – more financially sustainable than model 1.
- May encourage academic departments to use the service for their entire course reading scanning needs, since the only costs being passed to them are those which they would incur if they were doing it themselves (e.g. cost of ordering readings from HESS). Additionally, it would save departments the staffing and equipment cost of doing it themselves.
- No administrative burden associated with keeping a tally of how many free readings done for a module, as with models 2 and 3.
- Would discourage unnecessary requests, as all requests would have some cost (unlike model 1).

Cons

- May not be easy to fix a budget for the service in advance. Even though certain costs are being recouped, the service would still cost the Library more the higher the demand is, without providing extra income to pay for increasing demand. One answer might be to fix a budget and then fix the

amount of Library subsidy per reading depending on projected demand (so that the amount of subsidy is not linked to the staffing and equipment costs as such, but is calculated as a percentage of the total cost of providing a reading). This might make the cost per reading to academic departments higher or lower than it would be if only the costs of external services are passed to them, depending on what the budget and demand projection are. The projection of demand on which this idea is based is difficult to make, but would be needed to judge staffing levels for this model to work anyway (see below).

- Staffing levels need to be based on projected demand, since there is no cap on demand, as with model 3. Demand could still outstrip capacity, without providing extra income, as in model 2.
- Administrative burden of receiving and recording payments from departments.
- Readings scanned in-house would end up being free to departments, if only external costs are passed on. The pilot service is using HESS as the preferred method, with in-house scanning used for material not held by the British Library. It could be that the additional staff-time cost of in-house scanning compared to ordering via HESS could be calculated and charged to departments, or that departments are charged the price of a HESS order for in-house scanning.

5) Service staffed by the Library, but fully paid for by academic departments.

Library offers unlimited course readings to academic departments, but all readings are paid for by departments at full cost price (staff + equipment + external services).

Pros

- Fully self-funding, from the Library's point of view. Would require no Library funding at all.
- Would not be any more expensive to departments than doing it themselves, and may be slightly cheaper, if the Library is able to do it more efficiently than individual departments.
- May be more convenient to departments than doing it themselves.

Cons

- The cost/convenience value of having the Library handle digital course readings may not be great enough to persuade departments to use the service. Since the costs of departments doing it themselves are less visible to them than the fee charged by the Library, it may be difficult to persuade departments that it is offering them any value.
- It would not reflect well on the Library politically not to subsidise the service at all, if it is able to.

6) Self-service – academic departments do their own scanning using equipment provided by the Library and under Library guidance

This model is based on a suggestion that the Library could provide some kind of digitisation suite, or self-service scanners, which departments could use to do their own scanning. The Library would also offer training and guidance, but, essentially, copyright and CLA compliance would be the responsibility of the departments.

Pros

- Training and advice service would encourage departments to scan responsibly and increase the technical skills of departments.
- Would save departments equipment costs – would be a more efficient use of equipment for the University as a whole, as it would essentially be a shared equipment scheme, funded by the Library.
- Could be cheaper to the Library than offering a staffed, but not fully paid-for, service (though it would include the costs of staffing the advice/training service, maintenance of equipment and the space).

Cons

- No way of preventing breaches of copyright law and the CLA licence (other than by providing guidance). Since departments would be breaking the law using Library equipment and on the Library premises, this would put the Library in a fairly intolerable position.
- Academic departments may still prefer to scan in-house. The only advantage offered to departments would be free equipment, but this would be offset by the inconvenience of having to come to the Library to do scanning.
- Access to the digitisation suite would have to be controlled somehow, to prevent it being used by general members of the University or public.
- There would be no way of controlling the kind of scanning done by departments, even if access to the suite could be controlled. It would probably end up being seen as a free all-purpose scanning facility for academics.
- It is not clear how the advisory/training part of the service would work. Would there be a member of Library staff permanently in the digitisation suite, or would staff from academic departments be able to book a training session from Library staff? Who would staff this? Would any attempt be made to make staff from departments to undergo some form of training or induction before allowing them to use the facility?
- Need to provide a space, probably a room within the Library. The staffed models also have space requirements, but this model requires a dedicated, semi-public space, rather than office space which could be shared with other Library staff.
- Even if training is provided, it is still more difficult to prevent accidental damage to scanners if a large number of different people are using them (e.g. scanners can be scratched by rings with stones – would all departmental users be as careful to remove/cover rings as a small team of

dedicated staff?). Lack of ownership often leads to lack of care. Equipment maintenance costs would probably be higher for this model than staffed models.

- Assumes that the principle way of obtaining digital readings is in-house scanning – whereas HESS is a cost-effective alternative. It is difficult to see how this model could incorporate this. Encourage departments to order their own readings from HESS? Do it for them? But this is going further towards the other models.
- Difficult to maintain a central database of readings – departments could at best be encouraged to contribute details of their scans to a central registry, but could not necessarily be relied upon to. This would make re-use of readings for different modules/departments (and sharing readings with other institutions difficult (both allowed under the CLA licence).

7) A combination of models 3 and 6

Library offers a limited quota of readings free to departments. The Library also offers a digitisation suite within the Library building for departments to do any extra scanning themselves.

Pros

- The only obvious advantages are relative to models 3 and 6 themselves: it makes up for some of the disadvantages of each of these ideas, but it is difficult to see any overall advantage.

Cons

- Still falls foul of all of the disadvantages of model 6.
- Would probably be as expensive as models 3 and 6 put together.

Conclusions

- Option 1 (the current model for the pilot study) probably rules itself out in terms of financial sustainability.
- Options 6 and 7, which involve the creation of a self-service digitisation ‘suite’, would be too problematic to carry out.
- Of the remaining options, option 4 (all readings part subsidised by the Library and part paid for by departments) is probably the fairest and easiest to implement. This option seems to be the best one.

Appendix 7. Electronic delivery of files for the Borthwick Institute via YODL – proposal

General

Some material which is requested by customers of the on-demand digitisation service may be of wider interest and so is a candidate for inclusion in YODL. Other material is of such narrow interest that it is likely only to be of interest to the customer. In this proposal, a decision needs to be made at the point of digitisation which of these categories the request falls into. If it is decided that the material should also be added to YODL, as well as being sent to the customer, it should be digitised to the standards proposed in the *Technical standards for archival images from the Borthwick Institute on-demand digitisation service for YODL* document. For example, this would mean saving to TIFF rather than JPEG. It might include deciding to digitise the whole of a multi-page document, even if the customer only wants part of the document. Files can then be converted to an appropriate format for delivery to the customer (e.g. TIFF file converted to JPEG or PDF). If the request is deemed of interest only to the customer, then it should be digitised to the current digitisation service standards (e.g. scanned as JPEG, not TIFF) and ultimately deleted (once delivered to the customer).

Delivery to customers

- Create a collection in ‘Library and Archives’ called ‘Digitisation Service Admin’ with a sub-collection called ‘BIA-delivery’. Both collections should be public.
- Once payment has been received from the customer, the file is submitted to BIA-delivery, using the Dublin Core metadata editor
 - For image files (JPEG) the document type is selected as ‘image’ at the start of the workflow
 - For PDF files the document type is selected as ‘unsupported’
- In the Dublin Core metadata editor, only the title field is filled in with a title in the following format: ‘date (yyyy-mm-dd) – space – order number’ e.g. ‘2010-06-01 56748’ (where ‘56748’ is a completely made up order number).
- The record is saved and published.
- If there are multiple files for one document, after the record has been submitted (with the first image attached), the rest of the images would need to be added using the ‘edit datastreams’ link in the object view page.
- A link to the object view page with instructions on how to download is emailed to the customer.
- At regular intervals (monthly?) someone visits the ‘BIA-delivery’ collection and deletes all records over a month old (or whatever is agreed). Since they should sort in the collection browse page in title order and since the title begins with the date, there should be no problem identifying records to delete. It would be possible to move the records to a separate holding collection, accessible only

to Library and Archives staff, instead of deleting them. This could be a longer term, but still temporary, holding area, in case customers don't download their images within the time they were in the BIA-delivery collection. However, it would be problematic to see this as a way of holding onto files prior to eventually ingesting them into YODL, for the following reasons: (a) they may not have been digitised to the right standard; (b) if they have, they are likely already to have been ingested into YODL; and (c) no metadata is attached to them enabling them to be identified or described.

Ingest into YODL for permanent storage and delivery

- Images to be added to YODL for long term storage should be uploaded to sub-collections of the 'Borthwick' collection
- Metadata created in VRA Core 4.0 format, using YODL's metadata editor. At least some metadata needs to be captured at the point of digitisation, even if it is just the archival reference number. It would be better if the images were uploaded to YODL straight away, with minimal metadata than for them to be stored somewhere awaiting fuller cataloguing. Some form of metadata needs capturing in any case, otherwise it will be impossible to identify the images at a later date. Minimal catalogue records could be upgraded at a later date. Only three fields would be needed for a minimal VRA record: worktype (manuscript, photograph, or whatever); title (could be the archival reference number); image type (digital photograph or image – referring to the digital file).
- For records where multiple TIFF files need to be uploaded (e.g. for a multi-page document) there needs to be a way of distinguishing them. One solution might be to enable datastream labels to be edited ('page 1', 'page 2' etc.), but this would require development work. Another solution would be to create a collection for each multi-page document and upload each page to a separate Fedora object (this is the way manuscripts have already been represented in YODL).
- Access control set so that University members can access all files but the public can only see the metadata.
- An embargo may be applied for images which have been digitised as part of the on-demand service.

LIFE-SHARE Project |

Beccy Shipman | Matthew Herring | Ned Potter
life-share@leeds.ac.uk | www.leeds.ac.uk/library/projects/lifeshare

