

Projects without pain: demystifying digitisation with the Higher Education Digitisation Service

Joanne Lomax Smith, HEDS Consultant

The World Wide Web has provided a platform and funding initiatives have provided the means to present digitised collections to a whole new audience; to provide access to collections for a new world of interested eyes. Yet at first glance the path to a successful digitisation project can seem like a long and winding road. Answering the endless riddles posed by Technology and trying to avoid the quicksands of Copyright can mean many projects become bogged down by the technicalities of digitisation.

Yet there is help available, HEDS can provide a map, light the path and reveal the answers. From the initial consultation, through specification and production; HEDS staff can calm any fears and trample down the barriers to starting and running a successful project.

This article will look at some of the basic issues that should be considered by a project team during early planning; define some of the processes which will make a digitisation project run smoothly and also looks briefly at how using HEDS can alleviate some of the more painful processes.

Some background on HEDS

The Higher Education Digitisation Service (HEDS) was established to provide digitisation services for Higher Education and non-profit making bodies. Operating nationwide, HEDS offers expert advice, a total management package and a complete digitisation service. We will support your project development from the first feasibility assessment through to final delivery of your digitised material.

The Service is funded by the Joint Information Systems Committee of the HE Funding Councils and by the University of Hertfordshire. HEDS serves not only the HE sector, but also public libraries, museums and other non-profit making organisations.

Production as well as advice. HEDS is unique in that we are able to offer advice to our clients on how to digitise their material appropriately and also provide a full production service. The service removes the need for clients to buy equipment or to research every aspect of the digitisation process; instead projects can take advantage of HEDS' unique production service, highly competitive prices and skilled personnel. We are equipped to digitise the whole spectrum of originals from journal articles and exam papers to microfilm, glass negatives, slides and prints.

HEDS personnel have been selected not only for their technical expertise but also for their background in the library community and their understanding of its special needs. The consultants, Simon Tanner and Joanne Lomax Smith, have previously worked in national, academic and corporate libraries and bring experience of working on projects within these libraries. In particular, HEDS is familiar with the problems surrounding preservation and the handling of rare and precious assets.

Digitisation - a definition

Digitisation is the conversion of any type of original, be it paper, photographic prints or slides, 3D objects or moving film, into an electronic, digital format. The creation of a digital version of an image gives new life to the original and allows its use and manipulation by a new or wider audience.

At HEDS we work hard to make sure that the digitisation part of a project does not become more time-consuming and important than it deserves to be, freeing our clients to work on other, perhaps more important, parts of the project. After all, digitisation should be a tool and not a purpose. Using the Service in this way makes sense as it allows specialists both within the project and within HEDS to do the work they are skilled in.

Why digitise?

The main reasons for HEDS clients to digitise materials are the provision of access and the aiding of preservation of the original.

Creating a digital version of an item and then distributing it will almost undoubtedly improve access. At the most basic level it means that the item can be viewed in multiple locations by multiple users - if a core past exam paper has been hidden or stolen from the library the electronic copy will remain accessible. Digitisation can also make an item much easier to view - for example, a collection of photographic negatives or microfilm may be underused even though they contain important content because of the difficulties of viewing them - through digitisation these materials can be made more easily accessible. Digitisation can also bring new life to out-of-print items - a recent client commissioned HEDS to create a digital version of a key student text which would have been far more expensive to re-publish; the digital version also had the advantage of being searchable and comparatively cheap to reproduce for sale.

Providing remote access to digital surrogates can also ease pressure on libraries and archives by satisfying enquiries from a distance, as an example a recent enquiry to HEDS was regarding the digitisation of the standard book on the history of this particular university. The caller estimated that having the book available electronically would provide the answers to most of the general enquiries that came to the university. Digital versions can also allow an individual access to materials that would not normally be available to the public, for example special collection materials that are normally only accessible to certain researchers could be made more generally available. Furthermore, when access is improved to an item the library or archive may feel that the items are both earning their keep and improving the profile of the institution.

Digitisation may be used to aid preservation of the item, notably if the digital version can be made the only or the main method of access to the resource. This would protect the original by reducing or even completely removing everyday use, although it could be argued that this would disenfranchise users who cannot or do not use the necessary technology. The item could even be conserved and stored where it is safest rather than where it is most accessible. An example in this case would be maps in local record offices - they are a very popular resource but their size means that each viewing brings the risk of damage to the original and disruption to other users. The aim in an example such as this would be to digitise to the highest available standard so as to put off for as long as possible the need to re-digitise. We are not yet sufficiently confident to say that a resource can be digitised once and this will last forever, because we cannot know for certain the actual lifespan of a digital image.

Digitisation basics

When HEDS are contacted for advice on a digitisation project there are several key questions which require answers before the digitisation effort should go any further. Unfortunately, the convenor of a potential project often only wants an idea of how much it would cost to 'do something' with the collection and does not have the time to look into the options too carefully. However, the more time that can be spent on careful preparation the smoother the project will run and the more accurately can the budget be assessed.

The most important question is **what do you want the finished outcome to be?** This is vital - there is no point planning a project without knowing what you want the end point to be, simply knowing that you want to digitise the item is not enough. Is the item itself valuable as an artefact or purely for its content? Do you want to create a forensic copy of the original or a fast loading access version - or both? Do you want a standalone CD-ROM product, a web version or something else? Each of these issues bring decisions which will effect the cost of the finished product - and if these decisions are not faced before the funding is secured serious miscalculations can occur.

The next questions relate to the original materials. A major consideration is whether the items can be removed from the client site - if not, HEDS involvement may be to advise on setting up an in-house

scanning operation rather than doing the production work. As an example the following issues are among those that relate to photographic or paper media:

Photographic media (transparencies, prints, negatives)

- What size are the originals, are they all the same size?
- What proportion of the items have colour content? Is it important to capture the colour?
- What condition are they in - for example, are they dirty from heavy use?
- Are the slides in strips, frames or sleeves?
- Are the photographs flat or have they bowed?

Paper media

- What size are the pages, are all items the same?
- What general condition is the material in?
- Can books that are bound be stripped to loose pages for scanning?
- Is there any artwork - is it black and white, colour or line art?
- Is the text particularly small or large?

The answers to each of these questions will determine how the best digital image can be created from that item.

Factors which influence the cost of a digitisation project

It is sometimes overlooked that even when a digitisation project is undertaken in-house certain procedures can influence the eventual total cost for digitising each item. At HEDS these costs become more opaque when using the production service because each project is assessed individually and the added cost for each factor has to be taken into account when quoting the per-item production cost for a project. However, whether the scanning takes place in-house or through an agency certain factors will influence the cost - when out-sourcing the scanning this cost will be reflected in the item price, when doing the work in-house this cost will be delays and complications in the scanning effort.

The key factors are the suitability of the originals, the processes to be undertaken, whether the originals are diverse or similar and whether the scanning process can be automated or is operator-intensive.

The largest single costed factor is human intervention. Technology costs are comparatively low when compared to human costs. These costs include the time taken to manipulate an item onto the scanner; for example the item may require weighting down or placing into a frame or it may simply need to be placed into an automatic feeder. Costs also include the time taken to find the unique identifier for the item - is this on a sticker stuck to the back of the photo or on the slide frame or will the operator need to sort through lists to identify the item? Are the items sorted into batches of a similar size and colour or mixed up? - the former will produce faster results than the latter.

The key to minimising these operator costs is to **prepare the materials thoroughly** before they reach the scanning workstation; in general, the more preparation undertaken by the project team before the digitisation effort the smoother and more cost efficient the process will be. The ideal situation would be for the scanning operator to literally be able to unpack the materials as they arrive at the scanning workstation, start to scan item one and not stop until item 1001 is finished. Every time that interpretation or interaction is needed from the operator the cost will increase.

Ensuring that the original is clean is an important factor of the preparation because any dirt on the originals will transfer to the scanner and mean that the operator will need to stop and clean the scanner regularly, thereby slowing the procedure and increasing the item cost. In a project where the digitisation is handled by HEDS this cleaning is the client's responsibility because in order to prevent damage to the originals HEDS will not permit any cleaning of materials other than with an air spray.

This will remove loose dust and hairs but not fingerprints or anything that has become attached to the item. Fingerprints in particular have the potential to ruin a digital image - even a small fingerprint in the middle of a 35mm slide may be magnified by the scanning into a disfiguring mark across the digital surrogate.

Items must be fit for their purpose to get the best results; for example, 35mm slides that are fine to use as research tools and for creating images for articles and books may not be suitable for digitisation. When a 35mm slide is scanned at high resolutions - i.e. 2700dpi - any fault in the focus will be magnified on screen to become disproportionately obvious. Similarly, a microfilm that looks to be in fairly good condition even though it has been heavily used may contain minute scratches that will be magnified by the scanning process into big black lines. Likewise, with photographic prints any over or underexposure will be apparent, as will any faults in the printing process. Post-processing can be used to correct these sorts of problems but this will add to the cost. At HEDS digitisation of a representative sample of the materials is undertaken first to check that the items are suitable, that the digitisation treatment is appropriate and so that the price can be most accurately assessed. It is recommended that any in-house operation experiments in a similar way.

In general, with photographic material the best results come from originals that have been photographed with digitisation in mind. When scanning microfilm it is best to use an untouched master film, or the negative as this is unlikely to have been used.

Make sure the unique identifier is clearly marked. If the item is to receive a filename that is relevant to the original (i.e. not just sequential numbering) it is important that this is clear and apparent on the original. HEDS often recommends that a sticker with the number written on it be stuck to the slide frame or to the print's mount (not on the print itself). Ideally, the scanning operator should have as little opportunity for interpretation as possible and this means not having to search for the ID or having to choose between the various codes or mottoes that are often apparent on library and archive materials.

Handling times should be reduced by making items easier to handle and thereby lowering operator costs. For example, strips of transparencies are time consuming to place in the transparency holders which protect them on the flatbed. It may therefore prove more cost-effective to mount the strips in slide frames - which can then be easily put into transparency holders for scanning. Similarly for items packed in protective envelopes, if it is fiddly to remove the item from the envelope the handling time is increased and damage to the item may be more likely. HEDS recommends that protective coverings are made as easy to open as possible so that removing and replacing the item incurs minimal handling time and reduces the risk of damage.

Group similar items together for the scanning process. One of the most expensive parts of the scanning operation, alongside handling time, is the time taken to set the scanning equipment up. If the scanning workstation is set up to scan 35mm slides the settings will need to be changed for prints etc. Each of these changes costs time and therefore money. If the materials can be organised so that, for example, all the 35mm colour slides are scanned, followed by 35mm black and white, then colour prints and so on, the process is made as automatic as possible and therefore more efficient.

Consider using surrogates. Although the basic rule of digitisation is to scan from the original wherever possible, sometimes this is not a practical option. For example, while it is possible to scan from originals that are larger than A3 (for example posters or artwork), in reality a more cost-effective and potentially less damaging option is to use a professional photographer to create transparencies (ideally large format transparencies) which can then be scanned. This helps to protect the original from the damage that manhandling large materials onto the scanner and generally transporting them around can incur. It may also reduce the huge file sizes produced by large originals - for example, an original of around A3 size scanned in 24 bit RGB colour at 300 dpi can be up to 50 megabytes uncompressed, at 600 dpi this can rise to around 140 megabytes.

Glass plate negatives are a common part of library and archive holdings and are often a unique resource yet their extremely fragile nature makes them frustrating difficult to digitise. The extra human effort needed to manipulate each item onto the scan bed without the slightest knock or drop incurs a high price in the item cost, much higher than a plastic negative of a similar size. The need to pack and transport the negatives equally carefully also adds to the investment of time in each item. With a collection of glass plates it would be advisable to carefully compare the cost of re-photographing the negatives to create plastic negatives or positives and scanning these with the cost of scanning direct from the glass plate.

Microfilm is also a useful medium from which to digitise and has the advantage of being proven to be long-lived; unlike the digital media which still has a somewhat unknown life-span. In certain cases, it may be advisable to create a microfilm copy of the original to preservation standards and scan from this microfilm. This is often a useful way to create digital copies of bound volumes - the technique of microfilming a bound volume is older and more established than that of scanning a bound volume direct and the digital results are often equally good. The advantage of microfilming first is that the film itself will become the archive copy from which access versions of the book are made and should the digital version become outdated another version can be made from the microfilm at some point in the future.

In-house vs. out-sourcing

HEDS is in a position to be able to help projects to choose whether it would be more appropriate to scan project materials in-house on existing or specially purchased equipment or whether sending the materials to an outside agency for scanning would be more suitable and cost-effective.

The major reasons for sending materials to a bureau for digitisation are that the originals are not capable of being scanned successfully in-house (for example bound volumes) or that the intended product is particularly difficult - for example requiring advanced colour management skills. The type of equipment used for the scanning of items such as bound books or microfilms tends to be so expensive that it would be difficult for a project to justify the expenditure on such equipment, particularly given the short life-span and high maintenance costs of scanning equipment. Alternatively, the project manager may decide to use in-house resources for several reasons including that the collection cannot be moved out of the institution; that the collection is badly organised; that the digitisation needs to be phased in small amounts over a long period; or that the digitisation task is very simple. It may also be that the project can call on existing staff knowledge and equipment which would mean the project could be done in-house with limited further capital expenditure. A further reason that many projects are undertaken in-house is that the staff time, overheads and some consumables such as file storage can often be swallowed up by the institution and do not become apparent as a costed factor of the project, thus making this appear to be a cheaper option than out-sourcing. There is no easy answer to the question of whether to scan in-house or to out-source because it depends so closely on the project team, the institution and the materials.

Conclusion

In this article I have tried to stress that the key to running a successful digitisation project is preparation and a clear vision of the anticipated result. Even if a project is to out-source the digitisation effort, the preparation needed to get the materials into a state which will make digitisation efficient is still incredibly valuable.

The advice given above will hopefully help projects to know more about what to expect when embarking on a digitisation project and will maybe provide a starting point for new projects. HEDS is here to help any digitisation project, large or small, and we will try our best to answer the riddles so that your efforts can be put into the larger quest of the project as a whole.

HEDS can be contacted by telephone on 01707 286078, email as HEDS@herts.ac.uk or through our web pages at <http://heds.herts.ac.uk>.