

Teesdale Heritage Archive Access Resource

Project Report to Heart of Teesdale

1 May 2015

Introduction

A grant of £5,000 was awarded by Heart of Teesdale to the Teesdale Heritage Archive Access Resource (THAAR) project for the purchase of equipment that would enable local Heritage and History groups, as well as individuals, to form high quality digital images of documents and to convert them, where appropriate, into editable text.

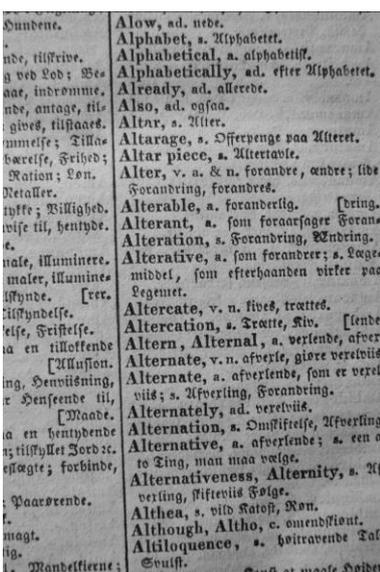
The THAAR project included volunteer time and the provision of in-kind resources both to deliver the equipment and to deliver the dissemination of information about, and training in the use, of the equipment.

The THAAR project was supported directly by The Witham Hall Ltd (The Witham) as a Restricted Fund project with The Witham providing administrative and accounting support as well as facilities for the storage and use of the equipment. The core THAAR project team consists of Roy Tranter, The Witham, and William Heyes, Teesdale Record Society. Jim Sewell, Teesdale Record Society, and Jon Smith, Barningham History Group, have both given valuable time to support the aims of the project.

Equipment Description

Appendix 1 shows a list of the equipment that was bought shortly after the award of the grant and which has been used in several applications.

The heart of the equipment is a high resolution digital camera, a Nikon D800 with a NIKOR 24-80mm zoom lens. The camera has a 36.4Mp sensor and this allows documents up to 560 x 440mm (22 x 18 inches) to be photographed at a resolution equivalent to 300 dpi. This resolution is the generally recommended resolution for converting text images into editable text using Optical Character Recognition (OCR) software.



The use of the equipment highlighted a couple of areas where additional equipment might be needed. The first was a lens capable of giving high resolution photographs of small (about 5 x 10 cm, 2 x 4 inches) document and the second was a better capability to image large numbers of unbound A4 sized documents.

The lens issue was resolved by further work with the existing lens - the macro facility on the lens was shown to produce full frame images of documents 5 x 7.5 cm in size. The image of 5 x 7.5 cm section of a page selected at random from a dictionary is shown in Figure 1. As these are at the full camera resolution, the images can be printed at much larger sizes (greater than A4) without the fuzziness usually found when blowing up lower resolution images.

Figure 1: Photograph of a 5 x 7.5cm portion of a page (shown at actual size). The image is at the full camera resolution of 4912 x 7360 pixels. It can be enlarged by a factor of about 5 and still be acceptable.

The camera has been used to photograph unbound A4 documents and although relatively quick at 5-10 seconds per page, the manual handling did introduce positioning errors and the process did

become tiring after a 100 or so pages. For this reason, agreement was obtained from Heart of Teesdale to purchase an Epson DS-510 Document Scanner with some of the grant that had not been spent during the first round of equipment purchase. This scanner has automatic sheet feed, scans double-sided A4 pages at about 20 pages per minute and produces images at 300dpi. Figure 2 shows the scanner in operation.



Figure 2: The Epson DS-510 Document Scanner in operation.

Heart of Teesdale also agreed to several small purchases such as additional batteries, power supplies and memory sticks. The full list of equipment purchased in this second round is given in Appendix 1.

Equipment usage

The first project report (1 August 2013) noted that the Teesdale Mercury Archive Project was using the equipment to photograph copies of the Teesdale Mercury for the period 1955-2005. This task was completed in August 2013 with 8 volunteers splitting the work between them and resulted in the production of some 22,500 images. Following on from this, additional work was done producing new images to replace those in the full archive (1854 - 2005) that had major problems (skewness, poor contrast, parts of the page out of frame, etc.). This produced some 7,500 images using 2 volunteers and was completed in May 2014. All of this work has allowed the Teesdale Mercury Archive project team achieve one of its objectives of transferring the fragile paper copies of the Teesdale Mercury, that were stored at The Witham, into proper conservation storage at the Durham County Records Office.

The camera system has also been used for some smaller projects:

- Several large and non-standard size historical documents relating to the property at 51 Victoria Road, Barnard Castle, were photographed.
- The Witham's Heritage Coordinator sought help to photograph some plates in a 1st edition of a Henry Witham book,.
- The Witham's Heritage Coordinator had photographs taken of selected pages from two books by Henry Witham for an exhibition about him and his role in Barnard Castle.
- Some trial photographs were taken of documents as a demonstration of the equipment to Keith Sweetmore of the North Yorkshire RO.

The Epson Document Scanner was received at the end of January 2015 and has been successfully tested using three projects:

- A ~300 page loose leaf training manual was converted into a digital document as part of a small project to create an archive of Radiological Protection documents. This proved a good test of the equipment as pages were a mix of single sided and double sided, coloured paper and white paper, gloss paper and plain.
- The Witham requested that a digital archive is created of the full accounts relating the TMAP and the THAAR projects. The scanner was used to convert invoices and bank statements from The Witham's paper archive into digital documents. This again proved to be a good test of the equipment as there were different sized pages (down to A6 size) and papers of very different thicknesses.
- A collection of about 60 magazine articles relating to scientific and computing developments and discoveries over the last 25 years was scanned. The articles had a mix of non-A series paper sizes, lightweight (60-70gsm) paper and many full colour images as well as text in a wide range of font sizes.

The User Guide to using the camera equipment has been updated to reflect the procedures that have been found to be practically useful.

A User Guide has been produced on how to convert digital images of pages into searchable text documents using Optical Character Recognition.

Other Aspects

Members of the project team have given presentations to the Middleton WI, Bishop Auckland Civic Society, Teesdale 41 Club and Barnard Castle Rotary Club about the TMA project. These presentations included a section about the THAAR equipment and its use in digitising a wide range of documents.

Members of the project team have demonstrated and discussed the use of the THAAR equipment at two meetings, one with Keith Sweetmore, North Yorkshire RO, and the other with Michael Crouch, Bideford, Devon. Both of these are in groups planning to undertake major document digitisation projects and they contacted us after reading about the work we have done.

As The Witham can no longer house the THAAR equipment, the project team has been seeking a new home for it. Agreement has been reached with The Fitzhugh Library, which is now based in Middleton-in-Teesdale, that the equipment can be housed in one of their rooms and, provided a person authorised by the Library is present, that the equipment can be used there at times when the Library is not normally open. The Library Trustees have authorised three members of the THAAR project team for this purpose.

Future Plans

A User Guide for using the Epson Document Scanner is planned following the successful trials of the equipment.

The TMA project team is currently looking at several changes that are needed to the TMA website. One of the changes will be to include a page about the THAAR equipment and its use. The site will become the main link through the internet to the THAAR project team.

The team will also take opportunities to publicise the availability of the equipment through personal contacts, presentations, occasional newspaper articles and targeted leaflet/poster distributions.

The Witham has a large paper archive of documents going back to the late 1970's. Because of pressure on storage space, consideration has been given to digitising the material and then transferring the paper documents to an appropriate off-site secure location. Should this project take place the digitisation work would be carried out by Witham volunteers over a period of several months. The THAAR project team would provide the initial training and support, as needed, as the mix of documents will require the use of the camera and document scanner systems.

Conclusion

The THAAR project team set out to specify and buy equipment that would enable large format documents, as well as standard or small format ones, to be digitised at high resolution. Such equipment has been bought, comprising a high resolution digital camera and accessories capable of photographing complex documents ranging in size from 5x7.5 cm to 44x56cm, and a document scanner optimised for the high speed duplex scanning of A6 - A4 sized documents. The equipment has been used extensively, producing well over 30,000 digital images.

The high quality of the images has enabled the use of Optical Character Recognition to convert most of the images into standard text documents that can be used in databases to provide searchable digital archives of the documents.

This success has demonstrated the capability and need for the equipment and the project team are confident that it will continue to be used effectively over the coming years.

Appendix 1

The THAAR Equipment List

Initial equipment March 2013	
Camera Equipment	
1	Nikon AF NIKKOR 24-85mm lens
2	Nikon D800 digital SLR camera
3	SanDisk 32Gb CF card (x2)
4	SanDisk 16Gb Flash Card
5	Nikon EN-EL15 camera battery (x2)
6	Expro Nikon EP-5B kit + EH5 charger
Tripods, Stands and Lamps	
7	Manfrotto 190XProB tripod and Head
8	Walimex Double screw Clamp
9	Walimex 100cm Light Stand Boom
10	Photography Light Stand
11	PRO LED lamps (x2)
12	Panasonic battery D220 for lamps (x2)
13	Battery charger
Document handling	
14	Custom made book cradle
15	Toughened glass plate
Software	
16	Nikon Camera Control Pro 2
17	Avanquest Expert PDF 8
18	ABBYY FineReader 11.0
Supplementary equipment January 2015	
19	Nikon EH-5B plus EP-5B Replacement AC Power Adapter
20	Nikon EN-EL15 Replacement Camera Battery (x2)
21	Lexar LJDS23- 64GABEU 64GB JumpDrive S23 USB 3.0 (x4)
22	Epson WorkForce DS-510 Document Scanner
23	40W 1 Output Universal Desktop Power Supply, 7.5V dc, 5.34A (x2)