Digitisation preserves historic documents

HP is helping a museum that dates back to World War II preserve its document archives while at the same time making them available to the public.

By Alison Job

Until the publication of a controversial book in 1974, The Ultra Secret by Frederick Winterbotham, very little was known about the events that took place at Bletchley Park, a hundred kilometres outside London, during the war. HP and the Bletchley Park Trust are collaborating on a five-year project that got underway in 2010 to give scholars, historians and other interested parties access to information that was kept secret for half a century. The Trust is seeking to digitally archive its extensive store of wartime paper documents such as intercepted messages, translations and maps. Once analysed, these documents will be used to raise additional funds to ensure the site stays open. To date, some 30 000 index cards have been scanned, containing detailed information about events that occurred during the war.

HP’s marketing manager, Dexter Harris explains, ‘The site has a massive archive, comprising over 1,5-million documents, that the Trust wants to digitise. This is a long laborious process that has to be carefully managed to ensure that these precious documents, which aren’t all in the best condition, aren’t damaged in any way.’

Several HP Scanjets are being used to transform the paper workflow into a digital one. A non-negotiable requirement is that the scanners’ paper handling systems be exemplary as the index cards are old and valuable. Not only should these systems not damage the documents (ie by jamming), they need to be able to process a wide variety of media. Another key element is that the file size be sufficiently small to facilitate storage yet the scan clear enough to read pencilled notes on the documents. The index cards are scanned in duplex mode (both sides simultaneously) to keep the digital asset as close to the original as possible. While the images can be cleaned up if required, the aim is to retain the authenticity of the hand-written cards.

EMEA product and business manager for HP Scanjet, Manfred Biggel identifies some of the leading trends in document scanning: ‘There’s a definite shift in demand from flatbed to sheet-fed scanning, especially in industries such as healthcare, finance and government. This is largely owing to the costs involved in handling hard copies of documents; it’s far cheaper to send, receive and store a digital file.’

Biggel goes on to say that the average organisation makes 19 copies of each document, which in turn costs $20 (US) to file, a further $120 (US) is spent trying to find misfiled documents and when one considers that one in 20 items is misfiled, taking 25 hours to recreate, the cost implications of not digitising documents cannot be ignored. Of course, when digitising documents, as with any other process, a workflow needs to be implemented that is specific to the site. At Bletchley Park, staff and volunteers are scanning the archives so a simple yet efficient system was required. The workflow implemented by HP goes from hard copy, to scan, then via OnBase software to a database archive hosted by Hyland. The Bletchley Park project represents a new partnership between HP and Hyland Software to scan and store the historical documents.

HP is supplying the scanning hardware required, including a series of scanners and PCs, as well as the underlying network. Hyland Software is providing its hosted ECM solution, OnBase Online, to archive the scanned images. The functionality contained in OnBase Online will enable the Trust to use its extensive records not only as a public research and learning resource but for essential revenue generation, with the option of potentially charging a small
James Mayhew, the regional manager of Hyland Software explains his company’s involvement in the archiving process. ‘We have a global partnership with HP whereby we provide enterprise content management and storage via the Cloud. We’re using OnBase software at Bletchley Park to catalogue its archive documents and have broken the project down into three phases. We’re currently in the digitisation and classification stage, which basically preserves the archive and provides an overview of its contents. Once this is completed, the archive will be made available via the Trust’s website, whereby historians, students and other interested parties will be able to track events during the war by accessing the information. The last stage will entail the repository becoming interactive, enabling contributions, comments and photos to be added to the archive, naturally subject to approval by the relevant experts.

‘Staff and volunteers will examine the easily accessible online documents and classify them according to dates, battles, locations, or the individuals involved. The data will then be integrated and automatically linked with other scanned content to help tell the full story of what happened. This linking of documents related to specific events has never been done before.‘

Once online and classified, the information will be available on the Bletchley Park website and eventually on information terminals around the park. The public will be able to further classify the information and comment or upload their own material, using a secure Onbase Workflow, which will be created and monitored by the Trust’s staff.

**Technology at Bletchley Park**

A series of Scanjet devices, which manage the scanning and document capture processes:
- A large format copier/printer, for maps
- A large, multi-function printer, currently in use in the archive office
- PCs and an underlying network to store the archive upon

- Hyland is delivering the underlying framework and document management solutions which will ensure the archive is agile and future-proof
- Hyland’s OnBase software allows for agility in managing the content and allowing those scanning and detailing the information to ensure everything is catalogued as is necessary, and will also be easily searchable by future generations.

Bletchley Park was occupied by the headquarters of Britain’s cryptanalytic and signals intelligence organisation by the Government Code and Cypher School between August 1939 and March 1946. Here’s some of the best brains of Britain were pitted against the enemy’s enciphered communications during the Second World War. Their success forged for Britain a decisively powerful intelligence weapon which saved countless lives and helped significantly to shorten the war.