



Digitization: a successful collaborative partnership between library and publisher

While archives across the UK are recognizing the potential of digitization to unlock their treasures, the realities of a large-scale digitization project can be daunting for any institution. This article will explore how partnerships between publishers and archives can help the latter face the challenges of getting content digitized. It will take a look behind the scenes of the longstanding partnership between Adam Matthew (AM; a specialist publisher of digital primary source collections), and The National Archives at Kew (TNA; the UK government's official archive). It will explore how this kind of relationship develops; how the production process works, and how products are brought through to market.

Background

Archives Direct is a major, ongoing digital collaboration between Adam Matthew (AM; a specialist publisher of digital primary source collections)¹ and The National Archives at Kew (TNA; the UK government's official archive)². The project has involved the scanning of millions of British government documents from the 19th and 20th centuries, relating to a wide variety of global regions and focusing on British diplomatic history and international relations. The plan is to continue adding new content for the foreseeable future. AM embarked on this programme in 2009, and continues to select new material for digitization on an ongoing basis to meet demand from customers. *Archives Direct* is formed of individual modules of material that are marketed separately within a single interface, allowing customers to purchase single or multiple modules appropriate to their budgets and subject focuses. These are themed by subject or geographical region.

Archives Direct is just one element of The National Archive's strategy to digitize content through partnerships with commercial entities. TNA have established fruitful relationships with academic publishers such as Adam Matthew, Gale Cengage and ProQuest, resulting in some of their most important documents being made available for scholarly research online. In addition, major deals have been struck between TNA and genealogy sites such as Ancestry.com and Findmypast.co.uk. In all, this has resulted in a staggering 100 million document pages being digitized by TNA's commercial partners. Adding TNA's own open access digitization into the mix, you have one of the most extensive archive digitization programmes on the planet.

Such commercial partnerships have allowed TNA to rapidly digitize its collections and make them available to interested audiences, albeit for a fee. There is no doubt that TNA would have been unable to fund or manage this scale of digitization themselves; the commercial model has allowed digitization to happen at minimal cost to TNA and represents an estimated investment of c.£80m by the private sector, of which Adam Matthew are proud to have contributed a significant sum.



MARTHA FOGG

Development Director
Adam Matthew

"... one of the most extensive archive digitization programmes on the planet."

Adam Matthew and the National Archives – a brief history

The roots of *Archives Direct* lie in a longstanding relationship between Adam Matthew and The National Archives. Soon after its foundation in 1990 as an independent microfilm publisher, AM established a working relationship with what was then called the Public Record Office (later to become TNA). When AM moved into online products in the mid-2000s, an obvious partner for digitization seemed to be TNA, and it was known they were keen to work with commercial publishers.

By 2008, the publisher had focused its strategy on creating a single, cross-searchable platform which would include a broad array of TNA content – this idea later became *Archives Direct*.

Choosing content: where to begin?

The process of putting together a digital archive will be similar for most commercial partners but, at AM, a module of *Archives Direct* begins life with the editorial development team. The team faces the initial challenge of identifying content within TNA's collections that is desirable to the academic market, and then packaging it in a way that is easy and approachable for users at a range of levels, while staying true to the archival arrangements of the originals. The aim is for a lightness of touch in the editorial process so that users don't feel that editors have been over-selective or are steering their research.

“The aim is for a lightness of touch in the editorial process ...”

Selection is a pretty daunting prospect when the sheer scale of the National Archives is considered. The statistics are simply staggering: 1,000 years of history; 11 million records; 185km of shelving (see Figure 1). And despite these amazing numbers, only about 5% of the documents actually produced by government make it to TNA for permanent preservation. One way that the publisher has approached this array of choice is to focus *Archives Direct* primarily on Britain's diplomatic relations with the rest of the world. It was felt that an international audience would be most interested in how the British government has interacted with its peers across the globe, and how this relationship has altered over time as Britain's role on the world stage has changed.

So, for example, the *Confidential Print* series is based on a series of printed volumes, starting in the 1820s and running right through into the 1970s, which were created by the Foreign Office to gather together and reproduce all the significant paperwork they created. Dispatches, memoranda, correspondence, reports and even telegrams were routinely printed. Copies of the resulting volumes were then circulated to government officials. The result for the historian is an incredible primary resource which allows in-depth, but accessible, research into Britain's colonial past and its relationship with the world. Unsurprisingly, *Confidential Print* has been a major hit with researchers and lecturers.



Figure 1. Document stacks at the National Archives.

Image © The National Archives, Kew

188 But, deciding how to curate this content was still a challenge for the development team at Adam Matthew. The whole *Confidential Print* series amounts to over 10,000 volumes and digitizing all of this in one go was entirely unworkable and any resulting product would have been very difficult to market to customers. So the decision was taken to divide it up by key geographic regions which would allow customers to pick and choose what they wanted according to their subject specialisms and budgets. This also fitted well with the original arrangement of the series, which is divided up largely into subseries by region or country. To date, runs of *Confidential Print* have been published for Africa, Latin America, the Middle East and North America.

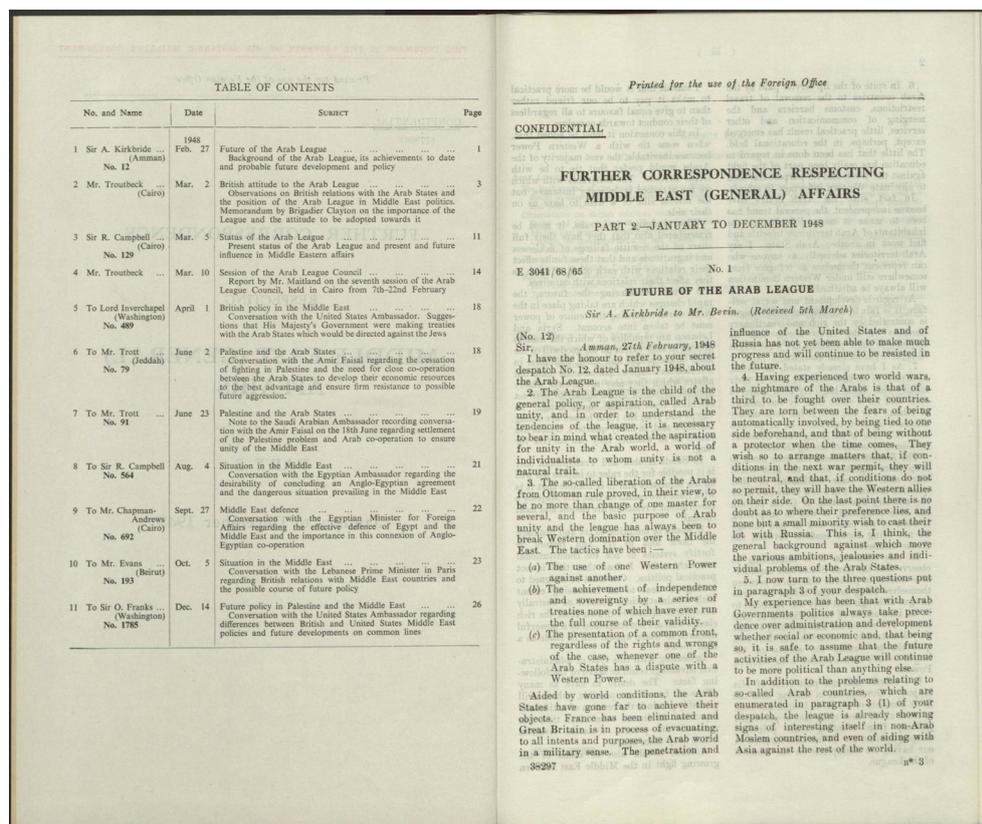


Figure 2. The *Confidential Print* series is divided up into geographic regions. The illustration shows 'Further correspondence respecting the Middle East (General): part 2, 1948.[FO 487/2:]'

Image © The National Archives, Kew

As well as the *Confidential Print* series, *Archives Direct* also includes modules such as *Foreign Office Files for China* and *Foreign Office Files for India, Pakistan and Afghanistan*. The aim is to include all files for each region and time period and not be selective in our approach. This is vital as *Archives Direct* is aimed at an academic audience, who want to know that they have all the information available.

Once the parameters of a new module are defined, a proposal is made to TNA at one of the quarterly strategy meetings; if they are happy with it, a new contract is signed and the production phase can begin.

"The aim is to include all files for each region and time period ..."

The pre-production phase

The production process begins with a thorough assessment of the material to be included in the new project, a labour-intensive task carried out by the editorial team. Each project requires numerous on-site visits over a number of weeks. TNA provide access to a large volume of material each visit, in a separate room away from the reading room where AM staff can work through large swathes of material. A detailed listing is produced with accurate image counts and digitization notes identifying issues that will impact digitization,

189 such as poor condition or the presence of large fold-out pages. It is tempting to gloss over or skip this stage of the production process, but the belief is that if time is not taken to look at the material very carefully before scanning commences, problems may result later. This approach is part of the publisher's commitment to quality and a good working relationship with source archives.

Once the editorial team have analysed the material, the first step is for material proposed for inclusion to go through TNA's rigorous 'collection care' process. Here TNA's digitization support conservators survey all the material to be digitized and assess its suitability for the process. Items marked as needing conservation are worked on by the conservators prior to scanning, a process which is funded by Adam Matthew as part of the project.

"... it is simply crucial that the scanning of documents is carried out according to carefully laid plans ..."

The main event: digitization

The setting up of a streamlined scanning operation is the lynchpin of any successful digitization project. For Adam Matthew, it is simply crucial that the scanning of documents is carried out according to carefully laid plans; any kinks in the supply chain may leave its scanners idle, which risks impacting both cost and timely delivery of products to customers. Equally, TNA want to ensure that the time taken to scan documents and have them out of the stacks is minimized. A key factor has been TNA's permission for AM to install its own scanners on site, managed by Capita, its long-term technical partners for UK digitization. This allows the publisher direct control of the scanning process and means that its projects are unaffected by the multitude of other digitization activities that TNA have under way.

Smooth running of the digitization process is a three-way collaboration between Adam Matthew, The National Archives and Capita, and all parties need to be equally invested and effective if the process is to be carried out efficiently and to sufficiently high quality. The project is managed on the publishing side by a dedicated production editor who lives and breathes the material, both digitally and through frequent visits to the archive. On TNA's side, there is a single representative responsible for managing the process and overseeing the various activities that the archive need to carry out to enable the work to progress, from putting it through collection care to pulling volumes and sending them down to the scanner operator. Finally, in the Capita corner, there is an on-site manager who oversees all the scanner operatives and ensures that scanning targets are met.

Scanning standards and practices

The relatively modern, uniform nature of government records has made digitization of *Archives Direct* a comparatively straightforward process compared to some previous projects featuring medieval manuscripts, early modern books or non-textual materials. The material digitized for *Archives Direct* is mostly either typescript, loose documents or bound volumes and dates largely from the mid-1800s to the 1980s. (Files later than this period are still closed to the public; Adam Matthew will continue to scan and add more recent documents as they are opened.)

"The editorial team then spends months quality checking every single image ..."

Images are scanned as 300dpi colour jpegs. Overhead tabletop scanners with an advanced book cradle are used, which allows for safe scanning of bound volumes without jeopardising the condition of aged spines. Despite the relatively robust condition of the material, we scan everything manually rather than using automated systems. The editorial team then spends months quality checking every single image produced by Capita, ensuring that any images not meeting its standards are rescanned and replaced.

Metadata is king

The next phase is the creation of additional metadata and full-text searchability. AM exports TNA's cataloguing as a starting point for the database, and supplements this with

190 extensive indexing covering places, people and subjects. The in-house production team at AM then makes the content more usable by applying optical character recognition (OCR) to all printed text to ensure it is full-text searchable. Finally, any chapters or 'pieces' that subdivide each document are marked up in an XML database to allow users to quickly browse across headings and get straight to the section of a document that they are interested in. These kinds of tools are essential in ensuring that scholars can be sure of finding what they want, and quickly.

Looking ahead: the future for *Archives Direct*

Although *Archives Direct* has been a flagship product on Adam Matthew's list for over five years now, there is still plenty of work to do in the future. Demand from customers for TNA material is still very strong, and the publisher will continue to bring new modules of content to market for the foreseeable future.

The digital platform will not just be preserved in aspic and merely continue to ingest more content. *Archives Direct* has just been relaunched to bring it on to AM's most recent technology platform and give it a facelift, refreshing the look and feel for users. The publisher has also taken the opportunity to gather feedback from its customers and discover what they liked and disliked about the original interface in terms of user experience and functionality. The redesign keeps the best of the original whilst enhancing and updating any outmoded technologies. (See Figure 3.)

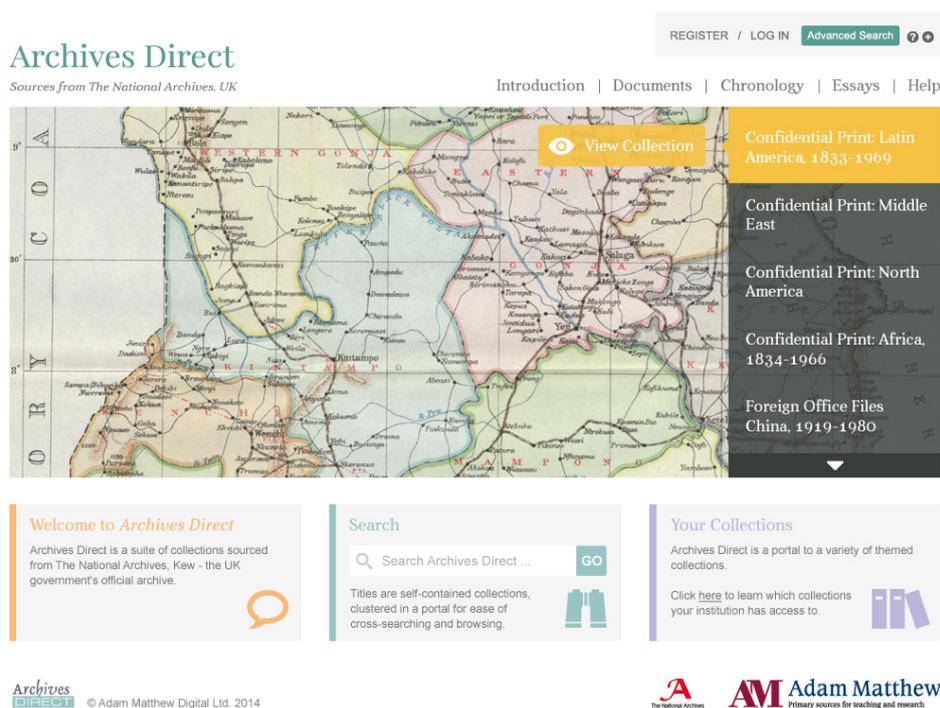


Figure 3. The redesigned *Archives Direct* platform, live to customers from July 2014.

Image © Adam Matthew

Archives Direct has proved to be one of Adam Matthew's most successful products, both commercially and in terms of the relationship with TNA, and the excellent production processes that they have established together. Undoubtedly, digital products are better when archive and publisher can establish a really positive working relationship. TNA's commitment to commercial partnerships and clearly defined digitization strategy has meant that a robust framework has been in place from the outset, with both parties understanding each other's goals and motivations. The result? Fantastic digital collections, a boon to researchers across the globe, and a partnership model that can teach us real lessons about how libraries and publishers can help each other.

References

1. Adam Matthew:
<http://www.amdigital.co.uk/>
2. The National Archives:
<http://www.nationalarchives.gov.uk/>

Article copyright: © 2014 Martha Fogg. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use and distribution provided the original author and source are credited.



Martha Fogg

Development Director

Adam Matthew, Pelham House, London Road, Marlborough SN8 2AG, UK

Tel: +44 (0) 1672 518315; Mob: +44 (0) 7789 001 429 | E-mail: Martha@amdigital.co.uk | web:
www.amdigital.co.uk | Twitter: @AdamMatthewGrp

ORCID iD: <http://orcid.org/0000-0002-3603-0119>

To cite this article:

Fogg, M, Digitization: a successful collaborative partnership between library and publisher, *Insights*, 2014, 27(2), 186–191; DOI: <http://dx.doi.org/10.1629/2048-7754.97>