

Digitization: the view from The National Archives

Licensed publication of images by commercial companies is not just a 'wider market initiative', but can be used actively to drive the digital access agenda facing archives in the 21st century. The experience of The National Archives suggests that, so long as agreements are carefully planned and structured, and content is made available at marginal cost, it is possible to meet the demands of the private sector publishers and the knowledge economy in general, whilst generating significant opportunities and benefits for the host archive.



DAN JONES
Head of Business
Development
The National Archives

In common with most archives and libraries in the United Kingdom, The National Archives, in Kew, West London, has always held great appeal both for professional researchers and for committed family historians who are willing to put in the hours required to master the various finding aids and other resources needed to unlock the records. If the wider public is aware of the Archive at all, it is probably for the New Year opening of government files under the 30-year rule. There have undoubtedly been the beginnings of a sea-change over the past few years, however. Brand recognition and awareness of The National Archives among the general public has increased sharply – to almost 70% in 2006¹ – and over 6 million people made more than 23 million visits to The National Archives' websites last year alone². In consequence, the Archive is now widely associated with online information resources, particularly in relation to family history.

Working in partnership

At the heart of this change has been a huge expansion of digital resources, from the online Catalogue and the Documents Online record delivery system to the 1901 Census. In the last few years, this has been supplemented by the appearance of co-branded online services developed with partners in the private sector under licence. This latter

programme of digitization – the Licensed Internet Associates (or LIA) programme – is particularly interesting because content licensing itself has become a strategic path, providing a means whereby the archive can achieve its core objectives, rather than a more peripheral 'wider markets' activity.

The publication of the six historic censuses (1841–1891), for example, has furthered the core objectives of the National Archives – that is to say, providing and promoting access to the records of British history – while, at the same time, fulfilling the commercial aims of the publisher (in this case, *Ancestry.co.uk*). Though archives and libraries have long permitted commercial publishers access to their collections, the products have traditionally been designed to fulfill the objectives of the publisher. The projects themselves have often been largely ancillary to the core function and operation of the institution. Of course, working with the market in this way is not entirely new, although previous private sector collaboration has generally centred on more rigid private finance initiative (PFI) investment models. Though these deals answer the capital investment issues faced by most institutions, they can create their own problems for an archive in terms of resource commitment and intensive project management. Additionally, PFI deals typically centre on fixed-cost recovery and encourage costing on a 'standalone database' basis, which can, itself, cause problems. In certain

circumstances, it can discourage the further development and enhancement of the services that are required to keep them up to date and operating to potential. It can also be prejudicial to less obviously commercial content – military records or pre-Victorian material, for example – since not everything in an archive has the commercial allure of a census or birth records.

By offering licences to publishers who are already offering substantial services in a given market sector, it is possible to harness huge economies of scale, with each project consisting mainly of incremental costs rather than infrastructure development. Royalty payments can be made on sales from the first unit, with no requirement for development cost recovery. The benchmark of what is commercially viable is lowered, enabling more marginal record series to be 'subsidized' by their more lucrative counterparts, all supported by the variety of pay-per-view and subscription payment models typically offered. Finally, this is all driven by exposure to PR, marketing resources and global reach far beyond that available to most public sector organizations and their services.

A major factor in making this possible has been the explosion of interest in genealogy in the US and UK in the last ten years – over 56 million UK census records were downloaded globally in 2006³. Co-branded licensed services to be launched during 2007 will provide access to a huge variety of archival content, including migration records⁴ and army service records⁵ among others. For several years, this has been the only established commercial market sufficiently large to justify the often huge costs of digitizing and indexing complete collections of manuscript records. However, we are now seeing a similar situation develop in other sectors – notably the development of academic research resources.

Where only two or three years ago new microfilm-based projects were common, they are now rare, and the creation of involved and comprehensive digital resources is continuing apace. Crucially, several publishers are actively developing reference products based around access to complete and unabridged record series. This is an exciting development, and we can expect to see access to a much wider-based selection of digital resources over the next few years as a result.

When a complete record series is digitized in this way, it allows the archive to provide almost all

its access via this digital resource. By securing free onsite access at the host archive as part of the licence terms, even document productions to on-site visitors are made virtually. This, in turn, allows long-term efficient storage of the original documents, and minimizes document handling, with all the resultant preservation advantages that brings. Reading rooms can be redesigned so that PC terminals providing access to a myriad of resources can gradually replace paper binders and microfilm readers. Whilst traditional learning resources which seek to select only parts of a given record series and contextualize resources around a theme are undoubtedly informative and can be very valuable, they typically do not bring with them these other benefits in terms of resource management within an institution.

Of course, The National Archives, like any public record depository, has to balance the demands of the private sector wishing to re-use its content with the interests of its end-users – UK citizens and others around the globe – and their rights to access the information. Thus it is vitally important that all deals are conducted within a rigid framework of principles. Key amongst these is that all licences granted are non-exclusive (that is, any interested party wanting to publish archive content and willing to abide by the licence terms will be afforded the right to do so). Co-branded products must include certain mandatory attributes for the end-user – notably a free search facility (the ability to interrogate the databases to a significant extent without paying a fee) and the availability of pay-per-view charging models.

It is also vital that any cost/royalty framework is standardized and applied to all equally. Typically, this may consist of standard fees for the copying and provision of the information, together with a standard flat royalty rate on any commercial revenues accrued from publication. Generally speaking, access to the records for re-use is provided through the provision of filmed copies, or digital scans, produced by the archive on a cost-recovery basis. Any interested parties thus have rights to access the material on exactly the same basis. This structure engenders competition between suppliers, precludes monopolistic exploitation, and drives innovation and value for the consumer.

Exceptionally, it may be necessary to offer a commercial partner access to the original documents in order to scan them for themselves. This applies, for example, to a particularly large record

series, where the economy of scale that the private sector can supply becomes a significant factor in the economic viability of the project. To minimize potential damage to documents, the material should be scanned only once; it is therefore vital that the rights in any digital images produced are retained by the archive, and that the same images are available for re-sale and re-use to any other interested parties on the usual terms. It is only in such an instance that granting exclusive access to the original records to one party can be said to be justified, since it can be used to further a wider access and re-use agenda with respect to archive content.

Any archive embarking on such a programme should not under-estimate the resource implications just because third parties undertake the majority of the work. Conservation and preservation is a particular area that can be affected, as already stretched resources often need to be switched from conventional conservation and preservation activities to supporting the digitization agenda through evaluation and preparation of the records for scanning. Detailed handling guidelines specific to a given format of records also need to be established and, of course, policed. Likewise, whilst facilitating multiple projects at once sounds appealing, it is, in practice, an often complex balancing act, and requires all the available resources of an archive – including document production, project management and curatorial services – to be behind the programme.

The National Archives programme

All this does not mean that The National Archives is outsourcing all new development, rather it is adapting to this new order by developing resources complementary to those offered by the market. The acclaimed 'Global Search' service using Autonomy's IDOL search technology is an example of how internal resources have been re-focused from the digitization of the records themselves to providing powerful and intuitive search tools, which can open up the records to a new generation of non-specialist researchers around the world. Over the next few years, Global Search will grow to include databases of The National Archives' material held by licensed third parties, in addition to the Archives' proprietary databases. This will mitigate one of the main disadvantages of the outsourced approach – that

of a fragmented user journey. The soon to be launched 'wiki-type' catalogue is another example. The National Archives is aiming to embrace user-attributed content to create a dynamic and powerful reference to the archives which will use the combined knowledge of staff and experienced researchers alike.

So what of the future? The fascination with history in all its form shows no sign of abating, so I hope we will see online access to an increasingly varied range of archive resources. Genealogy, particularly, seems to be becoming ever more popular. The variety of resources planned for publication over the next few years ranges from the aforementioned migration records and military archives, to parish records, war diaries and registers of death duties. The new generation of citizens will want – and expect – to relate to their archive collections digitally where possible.

As technologies improve in terms of scanning and transcription and the costs of digitization drop further, we may see other types of records and archive content become viable for e-publishing. I think we will begin to see other markets for archive content develop, to the extent that new and innovative products can be brought to market in a similarly collaborative way as for genealogy. Mapping will be an area of particular interest, and we are already seeing innovative GPS-based products being used to assist the study of First World War battlefields and the trenches of Northern France⁶, for example. One thing is certain: the demand for archival content in the e-publishing market will only increase, as will the innovative uses to which such content is put by the private sector. Archives must continue to look for ways in which this can help them achieve their own objectives.

References

1. Source: Ipsos MORI, 12–16 October 2006 (2,074 respondents).
2. Source: Webtrends/TNA, 2006.
3. Source: The National Archives, 2006.
4. BT27 UK Outward Passenger lists 1890–1960. See: <http://www.ancestorsonboard.com>
5. WO363 & 364 WW1 Army service and pension records. Circa 30 million images launching online over 12 months commencing February 2007. See: <http://www.ancestry.co.uk>

6. 'LinesMan' geo-referenced WW1 trench maps overlaid on contemporary topographical maps of Northern France to aid study and investigation of the battlefields of the Great War. See: <http://www.greatwardigital.com/>

Article © Dan Jones

■ Dan Jones
Head of Business Development
The National Archives
Kew
Richmond
Surrey TW9 4DU, UK
Tel: +44 (0)20 8392 5206
Fax: +44 (0)20 8487 1974
E-mail: daniel.jones@nationalarchives.gov.uk
www.nationalarchives.gov.uk/

To view the original copy of this article, published in *Serials*, the journal of the UKSG, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=20&issue=1&spage=16>

For a link to the table of contents for the issue of *Serials* in which this article first appeared, click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=20&issue=1>