

The British History Online digital library: a model for sustainability?

by Jonathan Blaney and Jane Winters
Institute of Historical Research, University of London

About British History Online

British History Online (BHO) (www.british-history.ac.uk) was established with private charitable funding from The Andrew W. Mellon Foundation of New York (as part of its scholarly communications programme). It began with a one-year pilot project in 2002, which had three main phases: planning (including extensive consultation); establishment (including the digitisation of sample datasets and the development of the information architecture); and benchmarking (of both existing research practice and researchers' expectations). The pilot project determined the most academically and technically effective structure for the two major establishment phases of BHO, from August 2004 to July 2008, which focused on the development of content and sustainability respectively. In 2007 we also secured additional funding from the Arts and Humanities Research Council in the UK to add content and develop an online annotation tool encouraging scholars to comment on, link between and correct material on the site.

The British History Online digital library provides access to the core printed primary sources for the history of Britain, from the 11th to the 19th century (the 20th century has largely been excluded because of the copyright problems surrounding the digitisation of 20th-century works; it is also the case that a degree of 'completeness' is much easier to achieve for the earlier periods).

Areas of core strength include:

- Local and regional history – notably the material in the volumes of the Victoria County History of England
- Historical geography – mainly the Epoch 1 Ordnance Survey maps for the entire country (begun in the 1840s)
- Urban and metropolitan history – including occupational, tax, population and court records
- Parliamentary history – the Journals of the Houses of Commons and Lords, parliamentary diaries, the Statutes of the Realm and so on

- Ecclesiastical and religious history – monastic, cathedral and parochial records
- Administrative history – most significantly The National Archives, calendars, including the State Papers Domestic, Colonial and in Foreign Archives, the Close Rolls, the Treasury Books and Papers, the Cecil Papers and so on.

All material is double-keyed into XML to a minimum 99.9% accuracy and checked by a subject specialist, ensuring that searches reveal the maximum possible results and that no material is ‘hidden’. Confidence in the accuracy and authority of search results is essential to academic researchers and is crucial to their take-up of digital resources.

There are a number of routes into this content, most of them available from the home page. The main navigation at the top of the page is by place, subject, period or type of source, supplemented by direct access to the most popular material on the site – the Ordnance Survey maps and the State Papers. In the left-hand navigation menu some of these options have been further refined, with users able to browse by ‘top source’ or by region. A full text search is also available.

The material on the site is presented within a standard framework. The header gives information about the ‘sponsor’ or owner of the publication, for example the History of Parliament Trust or the London Record Society, the name of the volume, the date of publication, the pages that are being viewed, citation information (which can be saved to the user’s personal bookshelf), the option to add a comment and the facility to search within the page (highlighting particular words or terms). The structure of the original printed sources is not replicated, but where layout is essential for sense this can be accommodated (as for example with the complex *Calendar of Treasury Books and Papers*, which contains nested columns of figures).

Finally some facts and figures about the project:

- 961 volumes, of varying length and complexity
- 1.48 million page views (210,000 visits) in January 2009
- Core staffing of 2.4, with support from an academic team in the IHR
- 40,701 registered users.

Planning for sustainability

Let us turn now to the question of sustainability – in technical, financial and academic contexts. Funding for digital research in the humanities in the UK has tended to be responsive mode and project-based – that is, individual researchers (more rarely teams of researchers) send project proposals to funding bodies on an ad hoc basis, outside any overarching strategic framework. Funding is most commonly for three years, after

which responsibility for sustaining a particular website or online database reverts to the researcher or his or her host institution, with varying degrees of success. Before BHO, the IHR had hosted several projects of this kind – some, like our online journal *Reviews in History*, were successfully absorbed within core staffing, others simply had to end when the funding ran out.

From the outset, we wanted the British History Online digital library to be something different. The first decision was to identify a project that would build on existing institutional strengths, rather than reflecting the interests of an individual researcher. It would involve all departments of the IHR, bringing together academic, library and technical staff in a way that had not happened before.

The Institute of Historical Research began to develop its web presence very early, with its IHR-Info website launched in 1993. It also has the best open-access collection of printed historical sources in the UK. BHO was designed to develop and support these areas of core activity, and this alignment with wider institutional strategy has been essential for the project's long-term success and sustainability.

The degree of planning afforded by the staged project process, and the pilot phase in particular, meant that we had more time than usual to think about the long-term sustainability of BHO after we could no longer rely on either state or private funding. We also benefited enormously from working with a funder, The Andrew W. Mellon Foundation, which had experience of supporting and developing not-for-profit services such as JSTOR – and crucially was prepared to consider allowing us to introduce an element of charging.

Technical sustainability

From the beginning at British History Online we have tried to design the information architecture with two complementary aims in mind:

- That it should be robust and scalable enough to grow exponentially
- That it should be flexible enough to take account of our growing knowledge of user needs.

The overall arrangement on the site must be fixed and therefore the overall architecture must be correct from the beginning. Our lowest-level unit of organisation – a group – may contain one volume or dozens, but they will all be strongly thematically linked, as for example the group 'Assize Records', which gathers together four related sources.

Once a group of texts is created with a unique URL, for example <http://www.british-history.ac.uk/catalogue.aspx?type=2&gid=53>, it is clearly not possible to move publications from this group into another – users may have cited the URL for this group, and persistent URLs are crucial

in encouraging citation of digital resources. In other words groups can grow but they cannot be disassembled.

However as the number of groups on the site grows (we have 150 at the time of writing) then it becomes increasingly important to organise access to them in ways that allow users to find what they are looking for. And, equally importantly from a scholarly perspective, to be confident that they have found all the resources relevant to their query.

The categories which can be used for access have evolved over time. Originally there was only a top-level list which contained basic categories such as primary, secondary and so on. We have now added a second layer.

Currently there are seven subject categories:

- Administrative and legal history
- Ecclesiastical and religious history
- Economic history
- Intellectual, scientific and cultural history
- Local history
- Parliamentary history
- Urban and metropolitan history.

Subjects are also now weighted by relevance, so that the most important volumes in a particular category are offered to users at the top of the list when they are browsing a category. This is a key benefit of BHO and one that cannot be replicated by automated search functions: we can offer the considered evaluation of an experienced editor rather than the outcome of a search algorithm.

Indeed, we have refined these search options further. A search can be confined to a single series or even to a single volume. This improvement was added early on, as the weight of material on the site began to increase, because it was foreseen that common search terms would soon return too much information. For example, a search such as ‘Exeter’, at the time of writing, gives 8,020 results across a diverse range of sources. More recently we made it possible for people to mix and match browses from a menu of checkboxes divided into *Subjects*, *Regions* and *Periods*.

It is possible to make multiple – or zero – selections from any of these three categories. So that by selecting, for example, ‘North’, ‘Intellectual+Urban’ and ‘18th century’ you find the seven publications which match those various criteria. Again, these are weighted so that the most directly useful volume appears at the top. Using JavaScript, the display automatically refreshes as soon as a box is checked or unchecked, giving the browsing experience a more intuitive and exploratory feel (naturally, in compliance with accessibility guidelines, the feature will still work without JavaScript being enabled but it is then necessary to press the Submit button to refresh the display). Additionally this new function gives the user an immediate sense of the areas of strength on the site.

To be able to reflect users' needs as the site evolved we carried out user surveys and also analysed typical activity on the site. For example, the latter showed that there was a requirement for browsing by date, and so we applied a date range to each volume, indicating how much of that volume covered the specified period, and also categorised volumes according to how chronologically their information was structured. This means that a user looking for, for example, the 17th century, will not only get a prioritised list of volumes which cover that period, but chronological volumes will sort to the top, ahead of something like a history of London which is organised geographically.

As the volume of traffic grew sufficiently high we could gain important information from the list of referring sources. For example, a list provided by Google Analytics for March 2009 shows that search engines and Wikipedia were the biggest referrers, but that these users viewed an average number of about four pages per visit. By contrast referrals from history sites such as the Victoria County History and the Institute of Historical Research's own website, although fewer in number, showed an average of about 13.5 page views per visit.

An important point to stress here is that many of our site changes are deliberately reactive. We do not try to determine what the user needs will be and then anticipate them, so much as try to respond to needs that arise.

We are now planning another major addition to the structure of the website, as content continues to grow. We aim to add a new, third layer to our taxonomies to accommodate the growing number of texts (at the time of writing we have 961). By adding a 'parent group' we can aid navigation. For example under the current structure we have:

Primary Sources > Guides and Calendars > State Papers, Milan
Primary Sources > Guides and Calendars > State Papers, Venice
Primary Sources > Guides and Calendars > State Papers, France

While keeping the groups as they are, we might consider adding a parent group called 'Papers in Foreign Archives':

Primary Sources > Guides and Calendars > Papers in Foreign Archives > State Papers, Milan

In this case there will be the added benefit of emphasising that, unlike other material in the State Papers series – which are also in the Guides and Calendars section – these are not English state papers relating to these foreign powers but the papers of those powers which relate to England. But the benefit for the site architecture is that it will aid our plans to introduce new navigation functions such as tree views.

This change needs to be planned carefully to ensure that it, too, is sustainable into the future, and also that it is reasonably intuitive to use. We will engage in user testing of the structure and of the site when a

proposed new taxonomy is employed. A useful exercise to carry out during this process is a card-sorting task: in this a number of cards containing categories or areas of the site are given to people and they are asked to arrange them in a structure that seems natural to them. After repeating the process with various people then a structure that makes sense to a majority begins to emerge.

We also altered the technology itself, as the site saw more traffic. All functions that were originally written in Classic ASP were replaced with new forms written in ASP.NET. This software provides a richer development environment.

In the summer of 2007 BHO employed, after a tendering process, a professional design company to produce a set of templates which could be easily applied to the existing structure, to aid navigation and to promote content of particular importance. The earlier version of the site was very text-based and had no images or background colours. The redesign introduced plenty of images on the home page – for example the image in the top left of the page cycles through a gallery of appropriate pictures, and the six initial main categories (Places, Subjects, Periods, Sources, Maps, Text Search) all have an eye-catching image associated with them. A colour scheme within the BHO green livery was used for navigation: dark green boxes indicate site announcements and general information, and light green boxes indicate navigational jumping-off points.

We have continually added new features to the site, and spend a lot of time thinking about what would be genuinely useful. Among the things we have added are:

- A bookshelf facility, allowing users to save favourite searches and publications
- Two RSS feeds, one to the journal and one to newly-uploaded volumes
- A journal announcing publications and other news on the home page
- A blog (with BHO posting as part of the Digital Publications department group blog)
- An auto-twitter of the above blog posts using Twitterfeed
- Bookmarking facilities on social networking sites
- Provision of tag clouds of most popular searches
- An annotation feature, which will be discussed later in this paper.

Financial sustainability

Financial sustainability is not traditionally something with which academics in universities have had to engage, but the availability of funding for the creation of digital resources has changed all of that. In the past, a research project might have led to the publication of a monograph and one or two scholarly articles. The publishing process was handled by

a university press, which dealt with editing, production and dissemination. The book and/or articles would be purchased by individual researchers and universities, as well as being deposited in each of the UK's copyright libraries. The author would make very little, if any, money, but the products of his or her research would always be available for consultation.

Digital resources pose other challenges. They will never be complete in the way that a book is complete. Even if a discrete collection of material is digitised and published online, and no further academic input is needed, the underlying hardware and software will require ongoing maintenance and support. Someone has to take responsibility for this, and unfortunately this involves time, expertise and ultimately funding.

With BHO, we decided that we would attempt to move towards self-sustainability on a not-for-profit basis. The aim was to fulfil the IHR's remit to provide services for the wider historical community, while at the same time generating sufficient revenue to maintain a core staffing and meet other central office costs. Should we be fortunate enough to achieve a profit in any particular year, this will be ploughed back into the project, for example supporting additional digitisation.

Extensive market research and consultation in Phase I of the project indicated that no single revenue stream would be sufficient to cover BHO's operational costs. Long-term development and sustainability was consequently dependent upon the establishment of a range of options, no one of which would individually be sufficient to maintain the central office but which together would offer a guaranteed minimum level of income annually.

The first decision that we took was to continue to make the bulk of the material on the site freely available to users both within and outside higher education – this has remained the case for well over 80% of all sources on the site. Then – and this was completely new territory for us – we identified a number of possible sources of finance that could be developed and trialled during the last phase of our funding from The Andrew W. Mellon Foundation:

- Advertising, whether performance-based (referral fees are earned through links to affiliate marketing sites) or period-limited display (allowing advertisers to reach particular demographic and special interest groups)
- An advertising opt-out, whereby users are given the option to sign up for advertisement-free browsing of the site, for an annual fee
- Donations, with users offered the opportunity to support scholarship by becoming 'friends' of British History Online
- Print on demand is offered for those users who would still like to have hard copy of the material that they are reading online. They can pay for and download entire volumes, and in some cases a greater degree of personalisation is available. For example, it is possible to download a

single parish history from the larger Victoria County History volumes, with an appropriate cover design, contents page etc.

- Consultancy and publishing services, primarily the digitisation of material for third parties, which might be published through BHO or returned to the sponsor for publication elsewhere
- Research grant income, building new funding applications around our existing infrastructure
- And finally, and most significantly, institutional and individual subscriptions.

Some of these ideas and approaches have proven to be more successful than others. When we first proposed including advertisements, we were worried both that it would have a negative impact on the perceived quality and authority of the site, and that our users would find it too intrusive. Consequently, we decided that we would keep any advertising outside the frame of the web page, so that it did not interfere with the text or draw the eye too much. There is also no advertising present on the home page. We are able to select the type of adverts that we are happy to display, and Google allows them to be closely related to the content of the site – so there might be adverts for historical maps, UK census records and so on.

As it turns out, our concerns were completely misplaced, and nobody has ever complained about the presence of adverts on the site. The disadvantage of this, albeit a minor one, is that we quickly abandoned the advertising opt-out facility as it was never used. The donations facility proved similarly unpopular. Despite the apparent unobtrusiveness of the advertising, however, it brings in significant income for the project – around £1,000 per month – and has become an important element of our sustainability strategy.

As BHO has developed and become more widely known, demand for consultancy and publishing services has increased. This has been notably marked since the removal of funding from the UK's Arts and Humanities Data Service in 2008, which left a number of resource creators committed to publishing and preserving the outputs of their projects but no longer with any means of doing so (the AHDS had previously acted as a preservation service for the outputs of publicly-funded research).

The collaboration takes two main forms, either a partnership to digitise material through BHO (often involving a joint funding application); or BHO essentially acting as a digitisation contractor and/or publisher. An example of the first type of activity is our work with the Survey of London. A joint application was developed by staff of BHO and the Survey, which was submitted to English Heritage for funding, and the digitisation has been a genuinely collaborative process. The 45 volumes of the Survey have added an interesting architectural history element to the site, as they include detailed plans, drawings and photographs of the most significant of London's buildings.

In addition to these partnerships and collaborations, BHO is also well placed to access research council and other government funding in the UK. In the past couple of years there has been a reduction in the amount of money that is available to support digital projects. The Arts and Humanities Research Council, for example, cancelled its dedicated digital research programme in 2008 and the Joint Information Systems Committee (JISC) announced that it will not be funding any large-scale digitisation in the medium term. However, there is still support available for the development and enhancement of existing resources, whether adding new content or new functionality. An example of this type of project, which would not have been an option without the existing BHO infrastructure, was the award of funding from the AHRC mentioned earlier. This allowed us to complete the entire set of calendars of state papers, which had only been partially digitised with funding from The Andrew W. Mellon Foundation.

The last element of the financial model was the implementation of some form of subscription service, which would support the ongoing free delivery of most of the material on the site. Following wide consultation with academic users, higher education information services specialists and academic publishers, The National Archives' calendars (specifically the Calendar of State Papers, Domestic) were identified as 'premium content' around which a subscription-based service could be developed.

BHO 'Premium Content' was launched in the autumn of 2007, with a variety of options open to subscribers. UK universities are able to subscribe via the Joint Information Systems Committee, a government-funded body which licenses content for higher education institutions. JISC applies a banding system, which allows smaller institutions to sign up at a significantly lower rate than the larger, research intensive universities. The maximum annual charge is £1,000, the minimum £400. We have devised a similar banding scheme for universities and colleges in the US, based on the publicly available Carnegie classification system, and a lower flat rate applies in the rest of the world. The subscription rate is fixed for at least the next three years, and new content has already been added at no extra charge, for example the full run of the Calendar of Close Rolls a few months ago. The Parliament Rolls of Medieval England will similarly be made available to subscribers later this year.

The implementation of a charging regime of this kind has not been straightforward. There were considerable technical and customer service issues to be overcome – a secure payments system had to be developed, for example, and subscribers are entitled to expect a level of support which would never be available from a free, university-based website. An advertising strategy also had to be devised, with email and print campaigns, special offers and so on, and the team certainly now knows more about library budgets and purchasing than we did three years ago.

However, we are starting to see successful results, and have easily exceeded our targets for the first year as a standalone service. At the time of writing, 30 institutions and 624 individuals have subscribed to the Premium Content. Most of the institutions are in the UK – a total of 19 – but there are 10 US subscribers, one in Ireland and one in Australia. A handful of European universities have requested free trials, but as yet none has signed up. Obviously, the economic climate is now very different from that in late 2007, and advertising revenue at least is starting to show signs of levelling off. But we are as confident as we can be that the project is financially sustainable in the medium to long term, without compromising the Institute's status as a provider of free resources for historians.

Academic sustainability

The final aspect of our sustainability strategy relates to the academic integrity of the project, and here we are experimenting with adapting the community elements of Web 2.0 for scholarly application.

Many of the calendars of state papers which have been digitised by BHO were produced in the 19th century, a time of rather different editorial assumptions and emphases. Research carried out subsequently may modify the narrative presented in the calendars in significant ways. Furthermore the accuracy of some of the calendaring has been questioned.

Technological advances mean that a great deal of material can be, and has already been, electronically aggregated. This seemed to be the appropriate time, therefore, to examine whether the academic community is willing to engage with Web 2.0 technology and annotate the calendars on BHO.

We have introduced an annotation feature which allows users to comment on calendar entries. The top of each activated page has a link called 'Comment on this article'. From there the user is taken to a list of categories, into one of which their annotation must fall:

- Correct factual error
- Transcription error
- Add document transcription
- Link to corresponding resource
- Annotate an entry
- Add background.

Users must further state whether their primary source for this annotation is manuscript, print, or both. By using these as radio buttons we hope to steer users towards scholarly annotations: the list of types of annotations makes it clear that only certain categories are expected, and the inclusion of a primary source with a citation is mandatory.

As a final check on the validity of the annotations, all are moderated before publication. We have included a function that allows the moder-

ator to message the person who left the annotation and ask him or her to modify the text or to add more detail. Additionally there is a facility – unlikely to be used, to be sure – to block any malicious user from posting, either temporarily or permanently.

We have found that it is even useful to add the errata for a volume (or group of volumes) as annotations, even though these already occur in a separate page at the beginning or end of the digital version, because users are not likely – as is also the case with print versions – to check them.

Annotations added even at this early stage have led to important corrections. For example this entry, from the Letters and Papers of Henry VIII, Volume 5, concerns a letter from Jean du Bellay to Montmorency:

1187. John Du Bellay, Bishop of Bayonne, to Montmorency. The king of England is much pleased with Montmorency for negotiating the interview between him and the king of France ...

(<http://www.british-history.ac.uk/report.aspx?compid=77489>)

However Rémy Scheurer, in his three-volume scholarly edition of du Bellay's letters, claims that both the addressee and the recipient of this letter are incorrectly ascribed in this calendar entry. Scheurer gives Gilles de La Pommeraye as author, and Anne de Montmorency as recipient: 'Cette lettre, connue par une copie du XVI^e siècle conservée à la Bibl. nat., fr. 3003, fol. 23, ne comporte ni adresse ni signature. Elle a été attribuée à du Bellay par ... Gairdner, *Letters and Papers* ... En vérité, elle émane de l'ambassadeur de France à Londres, Gilles de La Pommeraye.' *Correspondance du Cardinal Jean du Bellay* (ed. Rémy Scheurer, Volume 1, p. 280n., Paris, 1969).

The addition of this information as an annotation has several important effects. It registers the fact that the author and recipient are editorial conjectures (something nowhere indicated in the calendaring) and that there is a scholarly debate concerning the identity of the writer and recipient. Although experienced scholars will be well aware that all such calendaring is not free of error, this kind of annotation is also a salutary reminder to beginning researchers and to members of the public that all such apparently impartial and rigorously assembled sources need to be treated with due caution and scepticism.

It is hoped ultimately that something very like new 'editions' of the calendars may emerge over time, with scholars becoming invested in the improvement and development of core sources for their research.

Conclusion

Clearly, this strategy is still at a very early stage of development. And, of course, the success or otherwise of plans for sustainability will only become clear in the years to come. It is also the case that this strategy is not fixed. The challenges of the digital environment are constantly evolving, and academic research projects will similarly need to evolve if they are to survive. However, in considering the interrelated questions of technical, financial and academic sustainability from the outset, we have at least given British History Online a good chance.