



Responding to feedback on Jisc Historic Books

03 September 2014



INTRODUCTION	3
FEEDBACK	3
PDF DOWNLOADS	3
SEARCH OPTIONS	5
RESULTS RELEVANCY	6
IMAGE VIEWING	7
SPEED	8
VOLUME IDENTIFICATION	9
NO BROWSE OPTIONS	10
FUTURE DEVELOPMENT	11

Introduction

While the content in Jisc Historic Books (Early English Books Online, Eighteenth Century Collections Online and 65,000 recently digitised titles from the British Library 19th century collection) is valued highly by the community, access to it through the Jisc Historic Books interface did not fully meet the needs of the community. Based on feedback from the community, as well as a need to ensure financial sustainability for the future by bringing the maintenance and development in house at Mimas, the decision was taken to build a new platform to replace Jisc Historic Books.

To help to ensure the new platform met the needs of the community more fully, the team firstly reviewed all of the feedback received on Jisc Historic Books since its launch in August 2011 and identified the key issues to be addressed. Feedback was collated from a variety of sources including discussions with the Advisory Board, user satisfaction survey responses, conversations and correspondence with academics and librarians as well as enquiries to the Mimas helpdesk.

The new platform, Historical Texts, was developed by Knowledge Integration, Gooii, Sero Consulting, Mimas and Jisc Collections in collaboration with the Jisc Historic Books Advisory Board and User Group. It was released on 23 June 2014 and replaced Jisc Historic Books. This document explains how Historical Texts addresses the key areas where improvements were requested for Jisc Historic Books.

Historical Texts: <http://historicaltexts.jisc.ac.uk>

Feedback

The main areas where feedback had been received from the community and improvements had been requested are listed below, along with how these issues have been addressed with Historical Texts.

PDF downloads

Problem in Jisc Historic Books

You could download individual pages, a range of pages or a whole publication as a PDF but the PDFs had to be generated on the fly as you requested them. This meant you would have to wait for a PDF download to be processed and this could be a matter of seconds, minutes, hours or, if there was a backlog of larger requests, sometimes longer than a day. The time it took depended on the number of pages requested and the number of requests ahead of yours in the server queue. In some cases, where there were technical or data problems, they failed to generate at all.

When you requested a download you would be required to enter your email address and the request would go into the interface queue first, then get passed to the PDF server queues. There were two PDF

servers, one processed requests over 200 pages and the other requests under 200 pages. All requests were handled one at a time in each server queue. Once your request reached the top of the queue the system would locate all of the individual images on the server, PDF each one then merge them into one document.

As part of the PDF generation the images were also processed using optical character recognition (OCR) software to create searchable PDFs. Although requests of less than 50 pages often only took a few seconds or minutes to process, works with several hundred pages could take hours. If any of the images requested were too poor quality the software would be unable to OCR it and the whole PDF job would fail. The PDF generation could also fail due to temporary network connection issues between the servers.

Once the PDF had been generated, or had failed to generate, an email would be sent to the user. It would contain a link to download the PDF if it had been successful or a request to contact the helpdesk for assistance if it hadn't. The PDF was kept in a temporary cache for a short time then deleted, so another researcher requesting the same text would usually have to wait for it to be generated again.

The download feature also used a non standard network port (8282) and this port was not always open on an institutions firewall. If it wasn't then you would get a connection error message and not be able to request a PDF download or, for PDFs you had already requested, you wouldn't be able to follow the link to download it. The port could be opened by the institutions local IT team, but that was at their discretion. This often caused frustration for users who couldn't understand why they could sometimes access downloads but at other times couldn't, depending on where they were connecting from, and assumed that there was a problem with the service itself.

PDF download problems caused a great deal of frustration, especially when a text was needed quickly for a teaching class or work that on that day. Download related issues accounted for over half of the helpdesk calls. The Mimas team either had to download the PDF again if it had failed due to a temporary network issue, locate all of the images on the server and PDF them manually if there was an OCR failure, or liaise with library and IT staff locally over network port configuration.

Resolved in Historical Texts

All of the PDFs have been pre-generated and stored on the server. The British Library have provided the PDFs for their 19th century collection and the Historical Texts development team have created the PDFs for EEBO and ECCO.

This means that when you select the "Download publication as a PDF" now the PDF will download instantly. Depending on the browser you are using it will either download automatically or open in the browser so you can then save it to your local machine. Individual images can be downloaded instantly as well in JPEG format.

The non standard network port 8282 is not required for the download feature anymore, avoiding any download access problems.



Search options

Problem in Jisc Historic Books

The native interfaces for EEBO and ECCO have additional search options that Jisc Historic Books did not provide. Researchers asked for options such as subject and genre searching and variant spelling support. They also wanted to be able to search within the Thomason Tracts in EEBO specifically and search for EEBO periodicals using issue level criteria such as the day and month of publication.

The bibliographic reference search is a key method researchers use to locate texts but due to the variable data in the bibliographic reference number field it was difficult to get results to return, even if the publication was in the resource. If you entered only part of the identifier sometimes you would get no results and in other cases unrelated results, as it returned matches that were similar as well as those that were an exact match.

Resolved in Historical Texts

A wide variety of additional search options are available in Historical Texts including searching by language, illustration description, subject and genre. You can also specify whether the search terms are found in the bibliographic information or the full text, whether the results should contain illustrated or not illustrated works and choose sub collections to search such as Thomason Tracts in EEBO and ECCOII in ECCO. The bibliographic information has been carefully reviewed and identifiers extracted and indexed better and a variant spelling feature has been developed.

Providing the ability to search using issue level criteria such as day and month was dependent on data from ProQuest which Jisc Collections and Mimas did not previously have. This data was provided in August 2014 and the team will now be able to develop this functionality as part of the next phase of development.

The screenshot shows the 'Historical Texts' search interface. On the left, a sidebar lists search categories: All, Title, Author, Printer/Publisher, Place of publication, Bibliographic Number, Language, Illustration description, Subject, and Genre. The main search area includes a search bar with 'Fuzzy search' and 'Variant spelling' checkboxes, and a date range selector ('Including undated') with a dropdown menu showing 'e.g. 1600, 1620-1650, 1654, 1712-1738'. Below this are collection selection checkboxes for EEBO (1473-1700), Thomason Tracts, ECCO (1701-1800), ECCO II, and BL 19th Century (1789-1914). There are also search within options for 'Description' and 'Full Text', and a filter for 'Illustrated' or 'Not Illustrated' works. At the bottom, there's a 'MUST include' dropdown set to 'Title', a search field, a 'Fuzzy search' checkbox, a 'Remove' button, and a '+ Add row' button. A note at the bottom says 'New features coming soon! Keep an eye on our [development roadmap](#) to see what we're working on.'

Results relevancy

Problem in Jisc Historic Books

The service was built using Autonomy IDOL software, which provides conceptual searching rather than traditional keyword searching, but some users questioned the relevancy of the results. Researchers sometimes expected to see different results at the top of their results list and questioned why results that seemed a closer match for their search terms were sometimes lower down. Also, as it was not using

keyword searching, you would not necessarily see your search terms in some of the results, so it could be difficult to work out why some results had been returned at all.

Autonomy create and licence the IDOL software, but they do not provide much information on how it works, so it was difficult for the Jisc Historic Books team to work out why particular results were returned for searches as well.

The sheer volume of results for some searches could also be overwhelming. Both individual pages and book results were displayed in the same results list in the default "book" display view, with the page results just displayed lower down the list. This meant often large numbers of results were returned and there could be confusion when interpreting these numbers.

Resolved in Historical Texts

The data in Historical Texts has been indexed using different search software to Jisc Historic Books. It uses Elasticsearch software (<http://www.elasticsearch.org>). As the software is open source, they way it works is transparent and documented so our technical team are able to understand and explain why certain results are returned for a particular search and how the relevancy is worked out. This also means the team can work out the best way to enhance how the system works in response to feedback from the community.

It uses keyword rather than conceptual searching so you should be able to identify where your terms match in the text and so why the results were returned. Only whole publications are returned in the results list, not individual pages, so you won't get hundreds of results if your terms match multiple pages of a single text. You will see the individual page matches when you open up the publication.

Image viewing

Problem in Jisc Historic Books

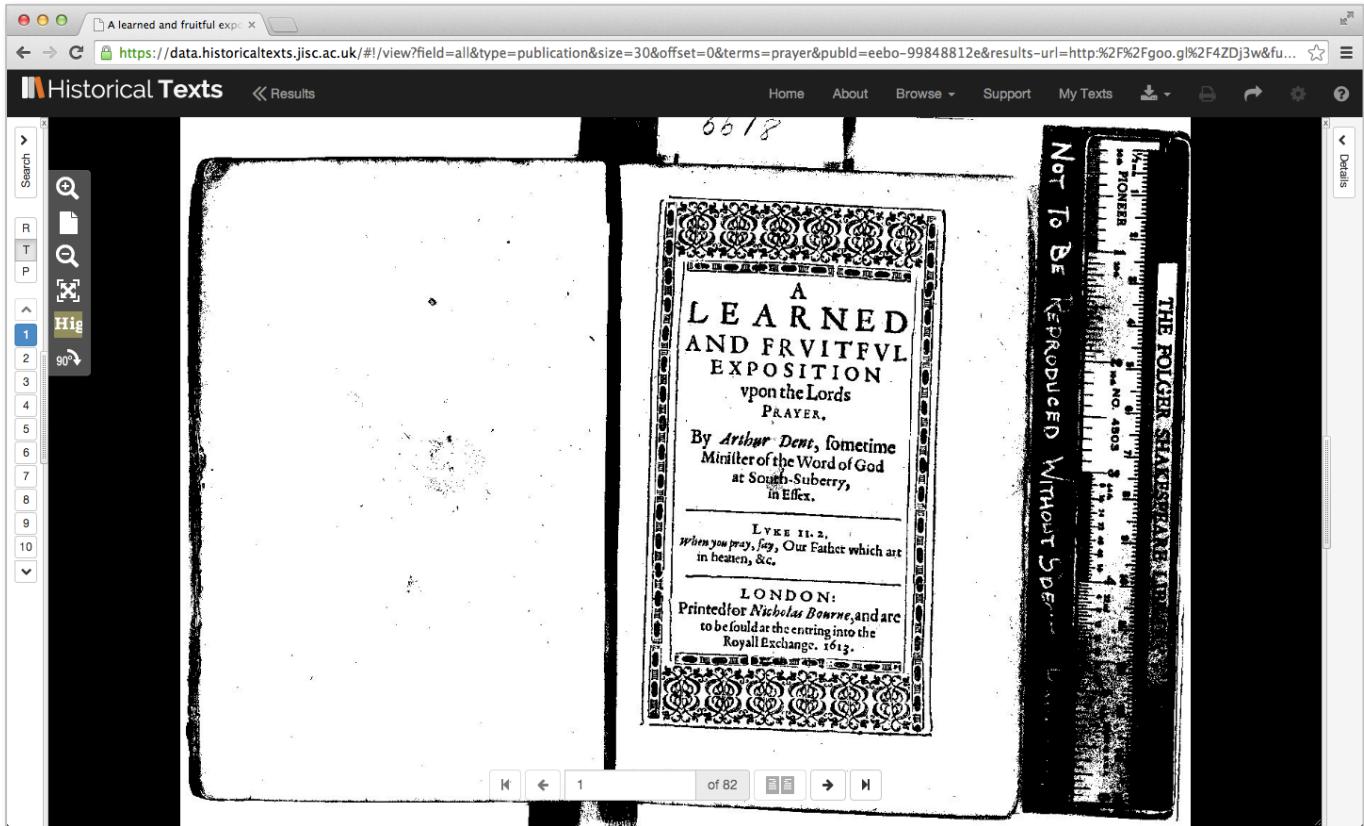
The viewer was displayed as a modal window which was overlaid on the results page. Even at its maximum size it was smaller than the on screen space available. A full screen view was provided which hid the side search panel and the top toolbar within the viewer, but it was still only the full screen of the modal window not the full screen of the screen itself. The buttons to zoom in and out didn't allow granular change to the zoom and researchers found them clunky.

Resolved in Historical Texts

The viewer in Historical Texts is a lot more flexible. It has three panels with the left and right side panels closed by default to maximize the space available to view the images.

The toolbar on the left hand side of the central panel enables you to zoom in and out but you can also use a mouse scroll wheel or a trackpad for more control. The deep zoom is provided by the open source OpenSeadragon software. You can also rotate the images, making it easier to view illustrations printed on their side for example, which are quite common in the BL collection.

It also provides a full screen button so you can view the images without any of the viewer and browser visible at all, and you can zoom in and out in full screen view using your mouse scroll wheel or a trackpad as well.



Speed

Problem in Jisc Historic Books

There were several areas where the system could be slow at times. The UK Federation Service Provider IdP was shared between Jisc Historic Books, Jisc Journal Archives and Jisc Content. This meant that the login process was complicated by additional re-directs depending on which resource you had chosen and this had an impact on the time it took to login.

Once logged in, search results could take many seconds to return sometimes and once you had opened one, the images could take several seconds to load. When navigating through a publication image by

image, or a set of thumbnails at a time, you would sometimes see a “please wait” message for several seconds or longer while the system was trying to load the images. It made trying to quickly navigate through a text difficult at times.

Resolved in Historical Texts

A new UK Federation Service Provider IdP has been created specifically for Historical Texts and logging in now usually only takes a matter of seconds.

Historical Texts runs in the cloud (we are using Amazon Web Services) and uses Elasticsearch open source software. Results are loaded quickly and navigation through images for a publication is fast. Thumbnails load continuously as you scroll down the list, so you can quickly scan through works of hundreds of pages. You can also switch off viewing the results page thumbnails using the on/off toggle button to make the page load faster, particularly useful when viewing large numbers of results per page. Additionally, you can close the side panels in the viewer, so you do not need to wait for the thumbnails or full text to load to make it even quicker.

Volume identification

Problem in Jisc Historic Books

There were two issues with volume identification. Firstly, the volume number was not displayed on the results page. This meant that if you had searched for a title with 8 volumes, they would all be listed on the results page with the same title and no way of knowing which was which until you opened each one in the viewer and checked the title page.

The second issue was specific to the British Library 19th Century Collection. There was no volume information in the expected bibliographic field in the data, so this information wasn't indexed and meant that all of the volumes for a title were merged into one large publication.

Resolved in Historical Texts

The volume information is displayed clearly on the results page for publications with multiple volumes for all three collections.

The team also carefully reviewed the bibliographic information and identified and extracted the volume information which was hidden in part of the filenames for the British Library 19th Century collection.

Relevance Author Printer/Publisher Volume Date

First < Previous 1 2 3 Next > Last Results per page: 30 100 500 1000 2000

	Dictionary of National Biography. Edited by Leslie Stephen. vol. 1-21. (By Leslie Stephen and Sidney Lee.) vol...
1	Author: Stephen, Leslie K.C.B. Publication date Publication place Smith, Elder & Co.: London, 1885-1900 Bibliographic number/Identifier 474 pages Volume: 43 BL Full Text ▾ More Information ▾
	Dictionary of National Biography. Edited by Leslie Stephen. vol. 1-21. (By Leslie Stephen and Sidney Lee.) vol...
2	Author: Stephen, Leslie K.C.B. Publication date Publication place Smith, Elder & Co.: London, 1885-1900 Bibliographic number/Identifier 442 pages Volume: 16 BL Full Text ▾ More Information ▾
	Dictionary of National Biography. Edited by Leslie Stephen. vol. 1-21. (By Leslie Stephen and Sidney Lee.) vol...
3	Author: Stephen, Leslie K.C.B. Publication date Publication place Smith, Elder & Co.: London, 1885-1900 Bibliographic number/Identifier 476 pages Volume: 25 BL Full Text ▾ More Information ▾

No browse options

Problem in Jisc Historic Books

There were no browse options and some researchers had advised they preferred to navigate the collections through a browse list rather than entering information in a search box. Specifically, they wanted to be able to browse a list of authors as well as browse the Thomason Tracts by volume. Identifying publications from the Thomason Tracts collection in EEBO in particular was difficult when searching.

Resolved in Historical Texts

There are two ways to browse the content now. You can browse everything included in the resource by author, printer/publisher, date, Thomason Tracts and Thomason Tracts Broadsides. You can also browse your particular search results by relevance, author, printer/publisher and date.

The screenshot shows a web browser window for 'Historical Texts' at <https://historicaltexts.jisc.ac.uk/#/results?field=all&type=publication&size=30&terms=%20&mode=by-volume>. The search bar contains the term 'declaration from York'. The results page displays two items:

- A declaration from York** by Wortley, Francis Sir (1591-1652). It was printed by A.N. for Thomas Warren in 1642, London. The item has 5 pages and is part of the Thomason Tracts series, Volume 153, Issue 1. It includes links to EEEBO and more information.
- The Queenes Majesties propositions to the states of Holland, concerning the differences betwixt His Majestie a...** by Pym, John (1584-1643). It was printed for I.H. and T.R. in 1642, June, 30, London. The item has 5 pages and is part of the Thomason Tracts series, Volume 153, Issue 1. It includes a link to EEEBO.

On the left, there are filter options for 'Printer/Publisher', 'Author', and 'Date'. The 'Volume' tab is selected in the navigation bar. The total number of hits is 400817.

Future development

The service is in continual development and the latest features we are working on can be found on our development roadmap:

<http://historicaltexts.jisc.ac.uk/developmentroadmap/>