

ProQuest's Early European Books

ProQuest's *Early European Books* is an ambitious long-term project to make available the complete holdings of early books at major European libraries, including their rarest and most valuable items. Volumes from 1455–1700 are being digitized in high-resolution colour on site at national libraries in Florence, Copenhagen and The Hague, and at the Wellcome Library in London, with new bespoke methods being developed to accommodate the specialist nature of early books. The project aims to interoperate with other digitization and research initiatives in this field to become a focal point for research in the area and to enable new overviews, comparisons and analyses of the printed output of the early modern era. ProQuest's provision of complimentary national access means that all the books digitized at the Wellcome Library will be freely available throughout the UK from launch.



MATT KIBBLE

Senior Market Development
Manager
ProQuest

Project outline

ProQuest's *Early European Books*¹ is an ambitious project which will build on the success of *Early English Books Online (EEBO)*² by providing a single location from which scholars can study the collections of early printed sources held by libraries throughout Europe. *EEBO* is now established as the first port of call for any researcher studying early modern history or literature, but is of course limited to material printed in the British Isles, or printed elsewhere in the English language, from 1473 to around 1700. To some extent, scholarship and curricula have no doubt been skewed by the widespread availability of *EEBO* and the lack of equivalent comprehensive sources for printed works of other countries and languages. *Early European Books* will redress this balance by working with major libraries to digitize their collections of works in all other European languages and from any location in Europe, from the era of Gutenberg, Koberger and Aldus Manutius to the end of the 17th century.

EEBO has been more than 70 years in the making, beginning with Eugene Power, founder of University Microfilms, filming rare books in the British Museum in the 1930s. This established the *Early English Books* microfilm series, which has had as its aim the capturing and cataloguing of all 125,000 titles listed in Pollard and Redgrave's *Short-Title Catalogue (1475-1640)* and Wing's *Short-Title Catalogue (1641-1700)*. To date, more than

128,000 titles have been filmed, but the project is still ongoing owing to the extreme rarity of the remaining titles, thanks to active partnerships with more than 125 contributing libraries worldwide.

Early European Books is in some ways even more ambitious than *EEBO*:

- unlike *EEBO*, there is no single bibliography on which to base the scope of the collection, no equivalent of Pollard and Redgrave and Wing; in fact, one of the aims of the project will be to consolidate bibliographic information about printing of this period
- rather than working from a defined list and aiming to preserve a single example of each item, we are digitizing library holdings in their entirety, and so will often be including multiple examples of a particular edition
- the number of European printed works is so large that the project will dwarf *EEBO*: our long-term aim is for *EEBO* to become a subset of the much larger *Early European Books* database
- rather than digitizing black-and-white microfilm slides, we will be scanning every page anew in high-resolution colour, including the bindings of each volume, ensuring that users get as full an impression as possible of the physical attributes of the source document.

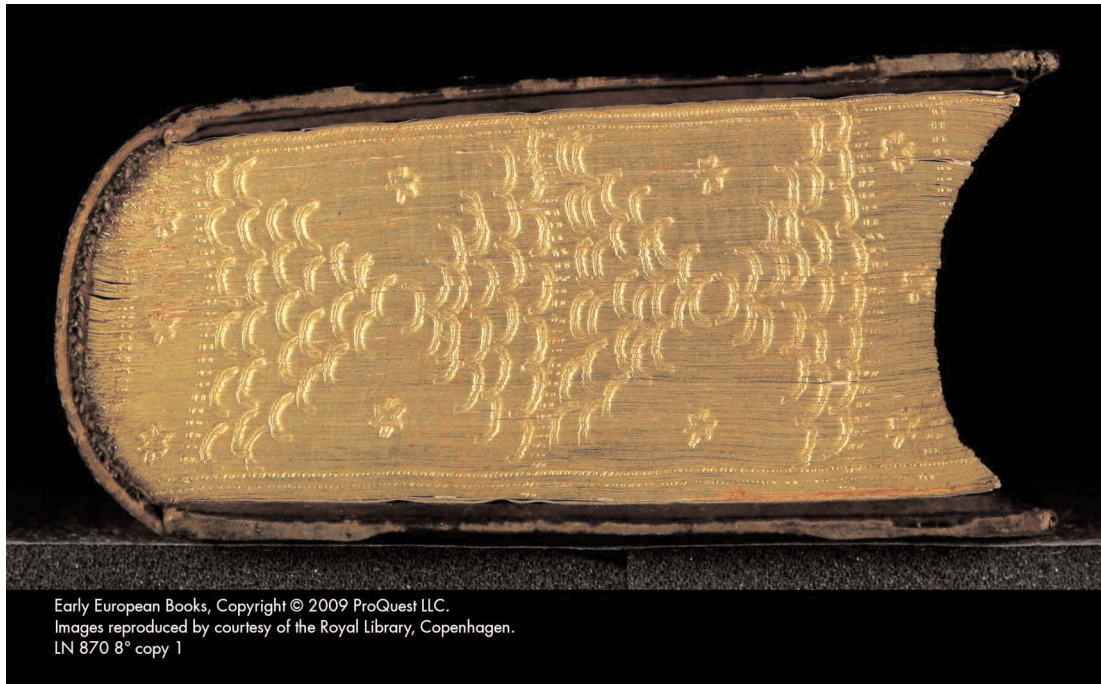


Figure 1. Tail edge with gilt decoration, from Niels Hemmingsen, *Postilla seu Enarratio Evangeliorum [...]*, Copenhagen, 1561. Det Kongelige Bibliotek, shelfmark LN 870 8° copy 1

Publishing model and access

The publication of *Early European Books* has also involved an innovative approach to the digitization of national heritage material, in terms of the models for funding and access. The standard model which we have proposed to a number of national libraries and research collections is as follows:

- ProQuest funds the scanning and creation of the digital files
- the scanning is carried out on site in the source library
- the master copies are returned to the source library for the library's use in perpetuity
- ProQuest provides free access to the source library's digitized collection within the country served by that library
- ProQuest has the rights to make the collection available commercially in all other territories, and pays the library a royalty from these sales.

We began the project with the two relatively small 'pilot' collections in Florence and Copenhagen, but this year we have stepped up the production rate and will have four scanning operations running concurrently in different libraries by the end of the year. We are publishing *Early European Books* as an ongoing series of collections, and from 2011

onwards each of those will contain books from a range of libraries. These collections are all cross-searchable within ProQuest's interface.

The publishing model allows national libraries to make some of their rarest and most fragile holdings available to their citizens. ProQuest provides the libraries with preservation-standard images (in the form of 400-dpi TIFFs or JPEG2000 files), together with the encoding and metadata necessary to make these images discoverable (typically in the form of a METS wrapper). Libraries are free to host the files themselves in order to provide the free access within the source nation, but so far all the partner libraries have chosen access via the ProQuest interface. This is provided using geographic IP recognition: so, for example, any computer within the UK will have seamless authenticated access to all the books digitized at the Wellcome Library.

Participating libraries

Our aim in selecting partner libraries has been to ensure a broad representation of the full range of Europe's national and linguistic histories. We therefore approached a wide selection of the major national libraries which have substantial holdings

of pre-1700 books, and we have also been approached by many other libraries with requests for partnerships – both national libraries and other specialist research collections. The publishing model outlined above has been arrived at through extensive discussions with library directors: we have benefited from the forums provided by such organizations as LIBER (Ligue des Bibliothèques Européennes de Recherche)³, CERL (Consortium of European Research Libraries)⁴ and CENL (Conference of European National Libraries)⁵, and are very grateful for the opportunities which these bodies have provided for open discussions with librarians and experts in the field of digitization of heritage material. The four libraries which have joined the project so far already give an excellent balance of coverage of European regions, date ranges and subject matter, and we are in close discussion with a number of other institutions whose holdings would complement these existing collections.

In order to test the feasibility of the project, we began with a pilot phase, which ran from late 2009 to mid-2011; this involved digitizing 2,600 volumes at the Kongelige Bibliotek (Royal Library) in Copenhagen and 2,750 volumes at the Biblioteca Nazionale Centrale di Firenze (BNCF; National Central Library of Florence). For these smaller pilot projects, we were necessarily selective in the choice of volumes to digitize, and limited the choice of materials to particularly significant subsets of books that were identified by the librarians at these institutions, based on bibliographically or topically defined groupings. In Copenhagen, the library's decision was to digitize all items included in the standard Danish national catalogue of early books, Lauritz Nielsen's *Dansk Bibliografi 1482–1600*, together with the library's renowned collection of works by the Danish astronomer Tycho Brahe and his followers, including the German mathematician Johannes Kepler's *Astronomia Nova* (1609) which established new laws of planetary motion.

In Florence, four distinct collections were included in the pilot project: the Nencini Aldine Collection (approximately 1,000 copies of editions printed by the Aldine Press in Venice from 1495 to the 1590s), the Postillati collection (around 60 items collected for the importance of their handwritten marginal annotations), the Sacre Rappresentazioni (over 700 rare 16th- and 17th-century editions of popular verse plays from Tuscany) and more than

1,000 items of incunabula (early printed books, from the birth of printing to the year 1500). The collection includes all of the early editions of Dante's *Divine Comedy*, up to and including the Aldine text which formed the standard Dante edition until the late 19th century, together with editions of the works of Euclid, Petrarch, Ariosto, Tasso and Horace which were all owned and annotated by Galileo Galilei (1564-1642).

With the completion of this pilot phase, we have now moved on to a full implementation phase, the scope of which will include:

- Biblioteca Nazionale Centrale di Firenze: we have extended the project to include the library's entire holdings of European books up to 1700, which is well in excess of 50,000 books
- Koninklijke Bibliotheek, Den Haag (National Library of the Netherlands): 30-60,000 books and pamphlets; scanning started in April 2011 and almost 1,500 books have gone online as of September 2011

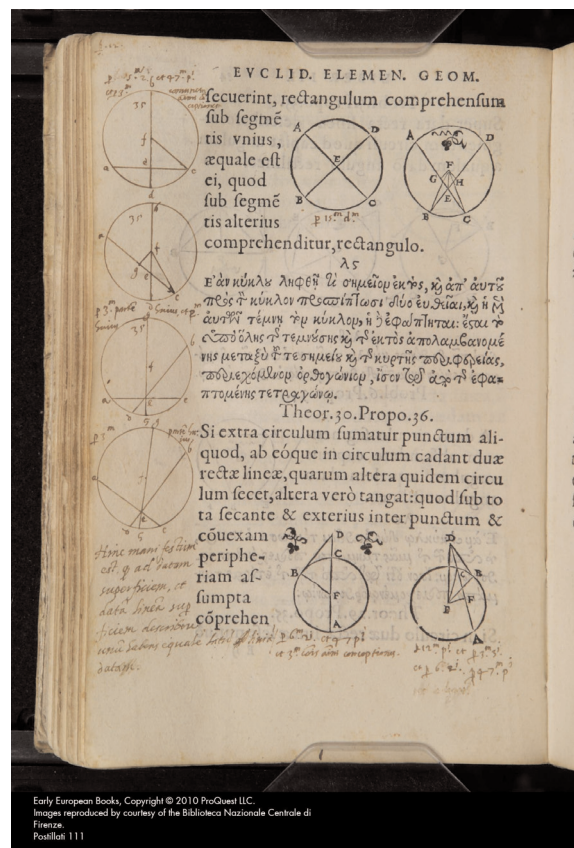


Figure 2. Euclid, 'Euclidis Elementorum Libri 15. Graece & Latine, Quibus, cum ad omnem Mathematicae scientiae partem, tum ad quamlibet Geometriae tractationem, facilis comparatur aditus', Paris, 1557. Biblioteca Nazionale Centrale di Firenze, shelfmark Postillati III. Manuscript annotations attributed to Galileo Galilei

- Wellcome Library, London: 15,500 books relating to science and medicine, ranging from alchemy to zoology; scanning started in July 2011 with the first batch of books launching in November 2011
- Royal Library, Copenhagen: we will be returning to Copenhagen in late 2011 to digitize a further 20–30,000 books, including 17th-century Danish, Norwegian and Icelandic books, and incunabula from throughout Western Europe
- more libraries to be announced shortly, to come onstream in 2012.

Digitization method and practicalities

The digitization of rare books in such large numbers brings with it a number of challenges. It requires a careful and adaptable approach in order to deal with the wide variety of sizes of volumes, and to capture non-standard elements such as fold-out maps and illustrations, interpolated slips and loose inserts, unusual bindings and closures, and multiple works bound together into single volumes. Many of the volumes are fragile, or bound too tightly to allow full opening. The bibliographic description of editions and variant copies needs to be highly rigorous to allow the

materials to be discovered and correctly interpreted by researchers. We also need to work with the library staff, fitting our needs around the workflow of the library, working together on bibliographic and cataloguing questions, and ensuring high levels of security and correct handling of materials.

We also felt that it was essential to capture every facet of these books: whereas previous microfilm and scanning projects (including *EEBO*) have tended to capture only the printed matter within each book, we are including full-colour images of all bindings, edges, endpapers, blank pages and inserts. This opens up new research possibilities for scholars who are interested in the book's history – the marks left by the book's owners and readers, and the evidence relating to printing techniques and manufacture.

We are working in collaboration with two specialist scanning companies^{6,7} who have both developed tried and tested methodologies for capturing and converting historic library materials, and are also keen to come up with innovative methods for the specific challenges these projects offer. The scanning staff always work on site at the source library to avoid the need for sending fragile and valuable items offsite. They use a combination of different scanners, each of which are better suited to different volume sizes and tightness of



Early European Books, Copyright © 2011 ProQuest LLC.
Images reproduced by courtesy of Koninklijke Bibliotheek, Den Haag.
1 C 9

Figure 3. Foreedge with clasps, from Salomon Theodotus, *'Het paradys der gheestelycke en kerckelycke lof-sangen, op de principaelste feest-daghen des gheheelen jaers gheplant door Salomonem Theodotvm'*, 's-Hertogenbosch, 1627. Koninklijke Bibliotheek, shelfmark 1 C 9

bindings, with scan operators manually turning pages, rather than the automated systems which can be used for more robust modern books. And both companies have developed bespoke approaches for scanning book edges, experimenting to find the best method for producing high-resolution images to a consistent specification which bring out the full detail of features such as gilt and embossed edges or unusual clasps and other closures.

At all four libraries, the operation has been launched with three-way meetings on site at the library between ProQuest, the scanning company and the library staff, in order to discuss practical questions such as the physical location of the operation, the assessment of the physical size of the volumes, logistics of book delivery, security, power supply, temperature of the scanning studio and so on. The scanning company set up server connections to allow for processing the image files and carrying out post-processing work such as image cropping and the addition of page-level metadata, such as original page numbering, signatures, and indexing to flag up the presence of features such as coloured illustrations, manuscript marginalia, musical notation, maps, portraits, or illustrated borders.



Figure 4. Hand-coloured illustration from Bartolommeo da li Sonetti, fl. 1485, 'Isolario', [Venice], not later than 1485. Biblioteca Nazionale Centrale di Firenze, shelfmark Magl. F.6.6

Scholarly analysis and cataloguing

As a minimum, all of the volumes in *Early European Books* are assigned the bibliographic data contained in the library's own catalogue records, including unique identifiers such as shelfmarks and reference numbers from short-title catalogues and other standard bibliographies. In some cases, the library records have not yet been converted into electronic form, and we work with the library's cataloguers to gather the information and return it to the library electronically. In addition, ProQuest adds standardized information such as modernized English-language place names and personal names, country and language, to allow consistent search results across the platform. In the longer term, our team of highly experienced rare book cataloguers are also creating more detailed copy-level records, including bibliographic notes and subject headings, to the same high standard as that which we deliver for *EEBO*.

We are also working with some of the many other scholarly and bibliographic bodies who are carrying out important work in gathering and collating information about early modern printing. We are working closely with the Universal Short Title Catalogue project (USTC) at St Andrews University⁸ and have already begun assigning USTC numbers to corresponding editions in *Early European Books*. We also established a partnership with CERL to use data from the CERL *Thesaurus*, which we have deployed to allow our users to enter standard modern forms of place names and personal names of authors or printers and find results containing historical, Latin and international variants. Our long-term goal is for *Early European Books* to interoperate with all of the major research and digitization projects which already exist in this field, and to become a much-needed focus for exactly this kind of work: not only will it allow researchers access to detailed reproductions of the holdings of major European libraries, but it will also enable new overviews, comparisons and analyses of the printed output of the early modern era.

References

1. *Early European Books*: <http://eeb.chadwyck.com> (login page accessed 6 October 2011).

2. *Early English Books Online*:
<http://eebo.chadwyck.com> (accessed 6 October 2011).
3. Ligue des Bibliothèques Européennes de Recherche (LIBER):
<http://www.libereurope.eu> (accessed 6 October 2011).
4. Consortium of European Research Libraries (CERL):
<http://www.cerl.org> (accessed 6 October 2011).
5. Conference of European National Libraries (CENL):
<http://web3.nlib.ee/cenl> (accessed 6 October 2011).
6. CCS (Content Conversion Specialists GmbH), based in Hamburg:
www.content-conversion.com (accessed 6 October 2011).
7. Diadeis Numérisation (now part of the Numen group), based in Paris:
www.numen.fr (accessed 6 October 2011).
8. Universal Short Title Catalogue project (USTC) at St Andrews University:
<http://www.ustc.ac.uk> (accessed 6 October 2011).

Article © Matt Kibble

■ **Matt Kibble**
Senior Market Development Manager
ProQuest
Cambridge, UK
E-mail: matt.kibble@proquest.co.uk

The DOI for this article is 10.1629/24255. Click here to access original article via DOI:

<http://dx.doi.org/10.1629/24255>

For a link to the full 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=24&issue=3>