



Sponsored article

ProQuest's *Early European Books* Project: A Collaborative Approach to the Digitisation of Rare Texts

Matt Kibble

Senior Market Development Manager, Arts and Humanities,
ProQuest (Cambridge, UK),
matt.kibble@proquest.co.uk

Project Outline

ProQuest's *Early European Books* (<http://eeb.chadwyck.com>) is an ambitious project which will build on the success of *Early English Books Online* (EEBO, <http://eebo.chadwyck.com>) 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 digitise their collections of works in all other European languages and from any location in Europe, from the era of Gutenberg, Jenson and Aldus Manutius to the end of the seventeenth 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 titles listed in Pollard & Redgrave's *Short-Title Catalogue (1475–1640)* and Wing's *Short-Title Catalogue (1641–1700)*. To date, more than 125,000 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 overarching bibliographic survey on which to base the scope of the collection, no equivalent of Pollard & Redgrave and Wing; in fact, one of the aims of the project will be to consolidate bibliographic information about printing of this period;
- 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 digitising black-and-white microfilm slides, we will be scanning every page anew in high-resolution colour, including blank pages, interleaved items, edges, spines and closures, ensuring that users get as full an impression as possible of the physical attributes of the source document.

Fig. 1: Tail edge of Niels Hemmingsen, *Postilla seu Enarratio Evangeliorum [...]*. Hafniæ: apud Christophorum Barth, 1561. Det Kongelige Bibliotek, LN 870 8° copy 1.



Publishing Model and Access

The publication of *Early European Books* has also involved an innovative approach to the digitisation 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 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.
- These files are owned by the source library.
- ProQuest provides free access to the source library's digitised 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.
- ProQuest commits to the provision of global open access to the collection at a future date.

Fig. 2: Home page of *Early European Books*, <http://eeb.chadwyck.com>.

ProQuest | **Early European Books**
printed sources to 1700

Quick Search [Need help?](#)

HOME SEARCH BROWSE INFORMATION RESOURCES MY RESEARCH HELP

Choose your language: English Dansk Italiano

Welcome to *Early European Books*

Early European Books traces the history of printing in Europe from its origins through to the close of the seventeenth century, offering full-colour, high-resolution facsimile images of rare and hard-to-access printed sources.

[Read more about *Early European Books*](#) or find out [what's new](#)

Featured content

The poems of Francesco Petrarca (Petrarch), published in Basel in 1582, with commentary by Lodovico Castelvetro, and with manuscript notes throughout in the hand of Galileo Galilei.

We aim to work with a number of major libraries concurrently, and to publish *Early European Books* as an ongoing series of collections, each one containing substantial holdings from one or more library. Customers will be able to purchase or subscribe to these collections in any combination, and they will all be cross-searchable within ProQuest's interface. The project has begun with the two relatively small 'pilot' collections outlined below, but from 2011 onwards, the production rate will be stepped up as more libraries come on board and we establish a rolling programme of larger collections of books.

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-ppi TIFF or JPEG 2000 files), together with the encoding and metadata necessary to make these images discoverable (typically in the form of a METS wrapper). The free access within the source nation can be provided by ProQuest through the *Early European Books* interface, but libraries are also free to host the files themselves in order to provide this access.

Pilot Project at Royal Library, Copenhagen

Collection One of *Early European Books* was completed in March 2010, and consists of some 500,000 pages (2,600 volumes) from the Kongelige Bibliotek in Copenhagen. For these smaller pilot projects, we have necessarily been selective in the choice of volumes to digitise, and the librarians at these institutions have played a key role in choosing particularly significant selections of works, based on bibliographically or topically defined groupings. In Copenhagen, the decision was made to digitise all items included in the standard Danish national catalogue of early books, Lauritz Nielsen's *Dansk Bibliografi 1482–1600*. This body of work includes important texts of the European Reformation by Martin Luther, Niels Hemmingsen and others, and formative contributions to the development of the Danish language, such as Christiern Pedersen's *Christian III Bible* of 1550. This was supplemented by a collection of late 16th- and early 17th-century astronomical works by the Danish astronomer Tycho Brahe and his followers, including the German mathematician Johannes Kepler's *Astronomia Nova* (1609) which revolutionised the science through its establishment of new laws of planetary motion.

Fig. 3: Thette ere thz Nøye testamenth paa danske ret effter latinen vdsatthe. Det nye Testamente paa Dansk [Christian II.s nye Testamente]. (Lybs i land til Myssen [i.e. Leipzig]): (Trøckt oc saat ... aff Melchiar Lotther), 1524. Det Kongelige Bibliotek, LN 270 8° copy 6.



Collection Two: Biblioteca Nazionale Centrale di Firenze

Collection Two of *Early European Books* was released later in 2010, and consist of a selection of 3,000 volumes from the Biblioteca Nazionale Centrale di Firenze (BNCF). Again, the selection has been carried out in consultation with the BNCF's librarians, and focuses on four collections of particular historic and bibliographic importance within the library's holdings from this period:

1. The Nencini Aldine Collection: in total, the BNCF holds approximately 1,000 copies of editions printed by the Aldine Press, mostly from the Nencini Collection, but supplemented by other collections within the library. This was one of the most significant institutions in the early history of printed books: founded by Aldus Manutius the Elder in Venice in 1495, and continued by his family until the 1590s, it played a central role in both the Renaissance revival of classical learning (by enlisting Greek scholars, editors and typesetters to publish modern editions of Aristotle, Homer, Sophocles and others) and in the course of book history (Manutius's innovations included the first use of italic type and the adoption of the smaller octavo paper size). Alongside 'pocket classics' editions of Greek and Latin authors, the collection also includes Italian literary texts such as Petrarch's lyric poetry and the first portable edition of Dante's *Divine Comedy*. The Aldine text formed the standard edition of Dante until the late 19th century, and the second Aldine edition (1515) was the first to include the famous woodcut diagramme illustrating the circles of Hell.

2. Postillati. The BNCF has a collection of around one hundred sixteenth- and seventeenth-century volumes which have been identified for the importance of the *postillati*, or marginal annotations. The collection includes editions of Euclid, Petrarch, Ariosto, Tasso and Horace which were all owned and annotated by Galileo Galilei (1564–1642), together with annotations by Michelangelo Buonarroti the Younger (nephew of Michelangelo the painter and sculptor), the playwright Lodovico Castelvetro and the poet Alessandro Tassoni. There are a number of interfoliated volumes, which the owners had rebound with alternate blank pages to allow room for annotation, including Francesco Bocchi's extensive commentaries on a Latin edition of Aristotle's *Physics*.

3. Incunabula. The collection will include 1,200 items from the BNCF's Incunabula collection, including rare first editions of Boccaccio's *Decameron* ([Naples?], c. 1470), Dante's *Divine Comedy* (Foligno, 1472) and Petrarch's *Canzoniere e Trionfi* ([Venice], 1470). Also included are 100 editions of works by the apocalyptic preacher Girolamo Savonarola from the 1490s, including many individual Florentine printings of his sermons and tracts from the era of the 'Bonfire of the Vanities'.

4. Sacred Representations. Over 600 sixteenth- and seventeenth-century volumes of *sacre rappresentazioni*, popular verse plays depicting Biblical scenes, episodes from the lives of the saints and Christian legends, which are considered by scholars to form the foundations of Italian theatre. Although many of the texts are anonymous, those by named authors include Castellano Castellani's *Figliuol prodigo* and Lorenzo de' Medici's *Rappresentazione di San Giovanni e Paolo*.

Collection Three

Launching in 2011, this will be substantially larger than the pilot collections, and will be drawn from four separate libraries. It will include 3 million pages, which we estimate could be in the region of 15–20,000 books. These will be taken from:

- Koninklijke Bibliotheek, Den Haag: in January, we announced a new agreement to digitise the National Library of the Netherlands' entire holdings of Dutch imprints up to 1700. This comprises an estimated 28,500 items, including 11,000 pamphlets, and the first batch of these digital files will be included in Collection Three.
- BNCF: after completion of the pilot project, our scanners will be staying in Florence to continue digitising the remainder of the library's enormous holdings of pre-1700 texts. Collection Three will include further treasures from their collection, including an additional 2,000 volumes of incunabula.
- Two further libraries, which we will be announcing early in 2011.

Digitisation Method and Practicalities

The digitisation 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 specific conditions of these books. The volumes come in a wide variety of sizes, including 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.

In both Copenhagen and Florence we have worked with the Hamburg-based scanning company CCS (Content Conversion Specialists GmbH, www.content-conversion.com), who have developed, tried and tested methodologies for capturing and converting historic library materials. CCS's staff work on site at the source library: they use a combination of different scanners, each of which are better suited to different volume sizes and tightness of bindings (one of the machines can scan books opened at only 90 degrees, for example), with scan operators manually turning pages, rather than the automated systems which can be used for more robust modern books. In addition, CCS have developed their own apparatus for scanning book edges, with lateral lighting to bring out the full detail of features such as gilt and embossed edges.

At both 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. CCS 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 (capturing original page numbers and the presence of page features such as coloured illustrations, manuscript marginalia, maps, portraits, or coats of arms).

Cataloguing and Collaboration

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. 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 has a team of highly experienced rare book cataloguers, who can standardise and expand on the library's own cataloguing information where necessary to give researchers the kind of detailed copy-level metadata that they have come to expect from EEBO.

In addition, we are 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 at St Andrews University (<http://www.ustc.ac.uk>), and have already begun assigning USTC numbers to corresponding editions in *Early European Books*. We have also established a partnership with the Consortium of European Research Libraries (<http://www.cerl.org>) to use data from the CERL Thesaurus, which (a) helps our indexers to map variant historical, Latin and abbreviated forms of author names, locations and printer names to a standardised form for the purposes of consistent search and retrieval, and (b) allows our users to search using modern forms of these names in their own language (e.g. Venedig, Copenhagen, Orazio, Horace) and find relevant results. Also, whenever we find variant forms which are not already listed in CERL's database, we feed this back for incorporation into the Thesaurus.

Our long-term goal is for *Early European Books* to become a 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.

For a free trial of *Early European Books* please email literature@proquest.co.uk