

## 1 Introduction

Eighteenth Century Collections Online (ECCO) is an online database published by Gale-Cengage. First published in 2003, it gives access via subscribing libraries to 184,536 titles of material printed between 1700 and 1800, comprising the text and digital images of their pages. In 2020 2,092 institutions and consortia in forty-two countries subscribe; in 2019 around 7.7 million search results, images, or texts were retrieved worldwide (De Mowbray, 2020a). It is arguably the largest single online collection of specifically eighteenth-century material available via academic institutions and has had a profound impact on how researchers conduct scholarship of the period. A history of a digital resource like ECCO is important because if we are at all interested in old books, and assuming we're also aware of the exponential increase in accessing old books via resources like ECCO, then we should be interested in why digitised books look the way they do and the difference that makes, how ECCO works the way it does, and what we can – and can't – do with these books.

Eighteenth Century Collections Online is deeply rooted in a longer history of representing eighteenth-century books, the effect of which can still be traced in the way ECCO works, since it is based upon a commercial microfilm collection, *The Eighteenth Century*, the contents of which were selected on the basis of a computerised cataloguing programme, the *Eighteenth-Century Short Title Catalogue (18thC STC)*. As Sarah Werner and Matthew Kirschenbaum argue, we cannot 'posit a transcendental "digital" that somehow stands outside the historical and material legacies of other artifacts and phenomena'; rather, the digital is a '*frankly messy complex of extensions and extrusions of prior media and technologies*' (Kirschenbaum & Werner, 2014: 408; my emphasis).

This is why this book starts with a 'prehistory' in order to understand the twentieth-century contexts of earlier media technologies, the changing cultures of scholarship, and what was driving commercial academic publishing. The first section draws out two significant factors from the 18thC STC (begun in 1976). The first is that decisions had to be made about what material it would include and what material it would not – these decisions would subsequently affect the scope and nature of ECCO's content.

The second is that the catalogue was from the outset a digital project, but that presented a challenge: how would the idiosyncratic features of books produced by the hand press translate into standardised machine-readable data? This metadata (data *about* something, rather than the thing itself) would eventually shape how ECCO could be searched as well as how users apprehend the nature of old books when they are digital images. The second section moves to Research Publication's (RPI) microfilm collection, *The Eighteenth Century* (published from 1982 onwards). It sets out the scholarly and commercial contexts for the development of this new media technology from the 1930s to the end of the twentieth century, illuminating the arguments of scholars, libraries, and microfilm companies for how this new technology would enable the preservation of and enable a wider access to research materials and old books.

At this point it's worth establishing why a focus on the books within ECCO is important. My approach to the *idea* of the book is predicated on a series of axioms:

1. The form of the printed book is a particular medium for the words within (as opposed to, say, a scroll or an audiobook).
2. The meaning of a book does not solely consist of the words within; the material form and the design of the book itself have meaning.
3. Transform the book into another medium ('remediation') and you change the meanings of the book.

It's for these reasons that this book, in the chapter 'Bookishness', looks at some case studies of individual eighteenth-century books in order to exemplify the effects of the 18thC STC and the microfilming programme on how users apprehend the physicality – the material life – of hand-press books as they are presented as digital images in ECCO. Part of this is a study of how users navigate between the image of a book and its record (spoiler: there is no seamless 'fit'), but it also emphasises the effect of human agency and human decisions about technology on how old books look in ECCO. In this sense my conceptual framework for this history is indebted to the powerful arguments for the critical potential of book history D. F. McKenzie made in his lectures of 1985, in which he proposed that bibliography should concern itself with a 'sociology of texts'. I start

with his point that the term ‘texts’ should go beyond the printed text to encompass the very broadest set of human communication media and even – most vital for us – ‘computer-stored information’; there is, he argues, ‘no evading the challenge which those new forms have created’ (McKenzie, 1999: 13). The discipline of bibliography ‘studies texts as recorded forms, and the processes of their transmission, including their production and reception’ (12). Perhaps the strongest argument for my history is that a ‘sociology of texts’ should allow ‘us to describe not only the technical but the social processes of their transmission’, and it ‘directs us to consider the human motives and interactions which texts involve at every stage of their production, transmission, and consumption. It alerts us to the role of institutions, and their own complex structures, in affecting the forms of social discourse, past and present’ (McKenzie, 1999: 13, 15). Sociocultural forces, institutions, technology, and human agency all play their part in this history.

Eighteenth Century Collections Online – or any digital entity – is not a static or an unchanging entity: it *has* a history. The rapidity with which commercial publishing technology supersedes older versions of itself has meant that some circumstances of its development are now obscure and others are irrecoverable. Tellingly, a part of ECCO will become invisible from 2020 when its original interface is scheduled to be turned off: it will *literally* be history. So this book is partly an act of recovery. The third chapter, ‘Beginnings’, turns to the development of ECCO itself. In the first section, I examine the immediate contexts that shaped how ECCO was to work and to be sold. It was decisively influenced by the downward movement of the academic publishing market and the emerging so-called disruptive technologies in the 1990s (Bower & Christensen, 1995). I focus on the techno-commercial choices facing Gale by illustrating contemporary digital resources created by two of its key commercial competitors in academic publishing of the time: Chadwyck-Healey and ProQuest. In addition, some aspects of how ECCO works ‘under the hood’ are – in common with many digital products – simply invisible to the public. The chapter goes on to explain Gale’s digitisation of the microfilm collection, discussing the problems created by the use of optical character recognition (OCR) software to

automatically create text from digital images of old books, how its metadata was structured, and how ECCO's original search interface worked.

The chapter 'Interfacing' takes us forward to Gale's development of ECCO after 2010, but also returns us to the issues of access. First, it focuses on the conglomerate of deals and collaborations between Gale and various partners, all of which broadly attempted to address concerns about *who* could access ECCO, as well as *how* ECCO might be used and interrogated, including the Text Creation Partnership, Jisc, and the print-on-demand deal with BiblioLife. This chapter's last section discusses the effects and meanings of the rise of the platform: this enabled the cross-searching of aggregated digital resources in a single package, but also a new way of interfacing with data and texts that was – for Gale's platforms – influenced by their belated engagement with the scholarly field of digital humanities. However, the platform produces a crucial tension between two ways of understanding and using old books: the bibliographical (or the 'bookishness' of books) and the textual. I finish by considering the politics of how these platforms represent early print history, drawing on the insights of postcolonial digital humanities, and reminding us that the Anglocentric nature of digital resources like ECCO is a product – a partially obscured one – of human decisions made in its antecedents, the 18thC STC and the microfilm collection. Indeed, throughout the book I've tried to avoid the suggestion that technological change is the only driving force in the twentieth and twenty-first centuries; instead I hope to have demonstrated that in ECCO's history – indeed, in the history of remediating and publishing old books – technology is inextricable from culture and human decisions.

I intend my history of a digital resource to mirror the methodology of bibliography. As W. W. Greg argued:

the object of bibliographical study is, I believe, to reconstruct for each particular book the history of its life, to make it reveal in its most intimate detail the story of its birth and adventures as the material vehicle of the living word. (Greg, 1945: 27)

Like the life history of old books, I hope to reveal the intimate details of the life history of ECCO: this book is partly an argument for the application of bibliography to digital entities, conceived as a ‘material vehicle of the living word’.<sup>1</sup> Looking over my introduction I hope it’s also clear that the status of entities like ECCO and their digitised books actually challenges the notion of a linear, progressive history. First, as Bonnie Mak has noted of books that have been subject to remediation, past and present versions of the same book copy exist simultaneously; in our case, as catalogue, as record, as microfilm, as digital images, as digital text, and even as a print-on-demand copy (Mak, 2014: 1516, 1519). This is echoed in the movement between past and present when we discuss ECCO’s place within wider historical contexts. Alan Liu’s comment about how we imagine and write narratives of media is suggestive: ‘the best stories of new media encounter – emergent from messy, reversible entanglements with history, socio-politics, and subjectivity – do not go from beginning to end, and so are not really stories at all’ (Liu, 2013: 16, my emphasis).

My aim is to speak to people interested in old books, people interested in how digitised collections of books work, and people interested in the history of how new media technologies have affected academic publishing. With such a broad reader in mind, I have tried not to assume any expert knowledge of old books, technology, or academic publishing even though this means taking the odd digression to explain a microhistory of file formats, or microfilm publishing, or some bibliographical terminology. The challenge of my history is to trace the digressive reverberation of ideas and debates that surrounded how we access and what we do with old books, and the chronological messiness of books whose lives have been subject to constant change. But this history is more than that; it is also an argument that we should better grasp the nature of something students and scholars rely upon for their understanding of eighteenth-century print and an argument for recovering, reading, and researching digital resources critically.

<sup>1</sup> This book’s methodology is indebted to the interface between book history and digital humanities; in addition to those cited here see, for example, the work by Ryan Cordell, Johanna Drucker, Alan Galey, Jerome McGann, and Whitney Trettien.

## 2 Prehistory

### *Cataloguing the Eighteenth Century*

Eighteenth Century Collections Online is based on a microfilm collection produced between 1982 and the early 2000s. This film collection itself was based on a catalogue of books begun in 1976 called the Eighteenth-Century Short Title Catalogue, under the editorship of Robin Alston, consultant at the British Library, co-edited with Henry Snyder in the United States. From 1987 this project expanded to include material printed before 1700 and was eventually renamed the English Short Title Catalogue (ESTC). In its current form the ESTC is an online catalogue of printed material published from the fifteenth century to the end of the eighteenth century. It's difficult to capture the sheer scale of the ESTC and its ambition: looking back over the project from 2003, Thomas Tanselle reaches for numbers: the 'file (achieved at a cost of about 30 million dollars) consists of some 435,000 records, indicating the location of over 2,000,000 copies in 1,600 libraries around the world' (in Snyder & Smith, 2003: xi). Currently it comprises more than 480,000 records from 2,000 libraries worldwide.<sup>2</sup> However, the project was initially confined to material printed between 1700 and 1800, and because it is this that shaped the underlying nature of ECCO, this history concentrates on the catalogue project before 1987.<sup>3</sup>

The first discussions about the possibility of a catalogue that would cover the eighteenth century began as early as 1962 amongst members of the Bibliographical Society, and such a catalogue was perceived as the logical next step from the two Short Title Catalogues (STCs) covering material printed between 1475 and 1640 (Pollard and Redgrave) and 1641 and 1700 (Wing). However, it was from 1975 that the catalogue got the necessary backing to start. More significantly, the discussions and plans by the leading editors for the 18thC STC emphasised the necessity that it be produced as

<sup>2</sup> British Library – Projects – English Short Title Catalogue: [www.bl.uk/projects/english-short-title-catalogue?\\_ga=2.29691066.42946434.1568626674-1896136714.1415013819](http://www.bl.uk/projects/english-short-title-catalogue?_ga=2.29691066.42946434.1568626674-1896136714.1415013819) [accessed 27 May 2020].

<sup>3</sup> I refer to the project between 1976 and 1987 as the 18thC STC, and the project as it currently exists as the ESTC.

an electronic file so that it could be managed and queried by computers (Alston & Janetta, 1978; Crump, 2003: 54; Snyder, 2003: 21–30; Alston, 2004). One of the most significant aspects of the 18thC STC is that it was deeply influenced by developments in computer-aided library cataloguing and online networked computer systems in the 1960s. The technology of machine-readable catalogue records would be an essential aspect of how metadata (that is, information *about* an object, distinct from the object itself) works alongside images and text of old books in digital archives and collections of the twenty-first century.

Korshin's 1976 grant application to get the 18thC STC off the ground included the participation of Hank Epstein, the director of the Stanford University computing team (Alston, 2004). Notably, the Stanford Research Institute was the first, in 1963, to demonstrate an 'online bibliographical search system', an 'online full-text search system', and systems that could be used remotely over long distances, and it was the first to use a screen display for interaction between human and computer (Bourne & Hahn, 2003: 14–15). In 1978 the Stanford-based Research Libraries Group developed an online networked database called the Research Libraries Information Network (RLIN).<sup>4</sup> In 1980 the 18thC STC at the British Library formed an important collaboration with the group, and by 1985 US and UK teams of the 18thC STC were able to edit the same file interactively online (Crump, 2003: 55).

Technological solutions to managing information had been the subject of both visionary projects and practical application since the end of the nineteenth century; two particular figures are often cited as seminal thinkers in this field. One is Paul Otlet (1868–1944) who, with Henri La Fontaine, designed a huge card catalogue in the 1890s entitled the 'Universal Bibliographic Repertory'. Otlet worked with Robert Goldschmidt on microfilming in the 1920s and 1930s, after which Otlet published a collection of his essays on the future of information science, *Traité de Documentation*, in 1934. The other is Vannevar Bush (1890–1974): in 1945 he proposed – but never built – a machine called 'Memex' that would

<sup>4</sup> 'Online' before the birth of the Internet and the Web merely means that two or more computers were directly connected via a closed network.

display microfilm and that also had the capability to apply, search, and retrieve keywords about the information on the microfilms (Deegan & Sutherland, 2009: 125–6; Borsuk, 2018: 209–13).

However, more significant than either of these men for the future of library information was Henriette Avram (1919–2006). Her career as one of the first computer programmers led eventually to designing library information systems at the Library of Congress in 1965. It was here that she developed the first – and what would become the standard – computerised library cataloguing system throughout the world: MACHine Readable Catalogue (MARC) (Rather & Wiggins, 1989). Before the advent of computer-aided catalogues libraries used card catalogues; the details about a book, for instance, would be recorded on a three-by-five card. The system for ensuring all libraries had access to and could update library catalogue records involved card-copying services and transporting duplicate card records by mail. By contrast a record that a machine can read can be disseminated and centralised much more easily. A MARC record is divided up into coded fields each of which contains a designated type of information, such as author, title, library location, subject, and many more. Avram's pioneering work necessitated a thorough understanding of computing *and* the principles of bibliographical cataloguing. The MACHine Readable Catalogue, then, was not only about designing a record to be parsed by a computer; it also set the standard for bibliographical records that libraries across the world would follow. The eventual result is the kind of human-readable record you can see on the online ESTC catalogue entry for the 1789 issue of Patrick Browne's *The Civil and Natural History of Jamaica* (Figure 1). Later I use this book to explore the relationship between the physical book and its record.

In fact, Alston discussed the 18thC STC with Henriette Avram in 1977 (Alston, 2004). In the earliest discussions the catalogue was to adopt the principles of the seminal catalogue of early print: Pollard and Redgrave's 1926 *A Short-Title Catalogue of Books Printed in England, Scotland, & Ireland and of English Books Printed Abroad, 1475–1640*. But it was clear that the 18thC STC would be a computerised catalogue and that therefore records would follow the much more detailed standards required by MARC (Alston & Janetta, 1978: 24–6). In his 1981 lecture 'Computers and Bibliography' Alston was adamant that computing would enable more



ESTC System No.	006359980
ESTC Citation No.	T89758
Author - personal	● Browne, Patrick, 1720?-1790.
Title	● <b>The civil and natural history of Jamaica. Containing I. An accurate description of that island, its situation, and soil; with a brief account of its former and present state, government, revenues, produce, and trade. II. An history of the natural productions, including the various sorts of native fossils; perfect and imperfect vegetables; quadrupeds, birds, fishes, reptiles, and insects; with their properties and uses in mechanics, diet, and physic.</b> By Patrick Browne, M.D. illustrated with forty-nine copper plates; in which the most curious productions are represented of their natural sizes, and delineated immediately from the objects, by George Dionysius Ehret. There are now added complete Linnaean indexes, and a large and accurate map of the island.
Publisher/year	● London : sold by B. White and Son, at Horace's Head, Fleet-Street, M,DCC,LXXXIX. [1789]
Physical descr.	[6],viii,503,[47]p.,plates : map ; 2°.
General note	With a half-title and four additional indexes. With a list of subscribers. A reissue of the sheets of the 1756 edition, with a new titlepage. Part of the 1756 edition was printed by William Bowyer; his records show 738 copies printed on crown and 12 on demy; John Hart may also have printed part of the edition.
Citation/references	Goldsmiths'. 13843 Maslen and Lancaster. Bowyer ledgers, 4023
Surrogates	Microfilm. Woodbridge, Conn. : Primary Source Microfilm, an imprint of Gale Group, 2002. 1 reel ; 35 mm. Unit 361. (The Eighteenth Century ; reel 12612 , no. 02). Microfilm. New Haven, CT Research Publications, Inc., 1978. 1 reel ; 35mm. (Goldsmiths'-Kress Library of Economic Literature ; reel 1412, no. 13843).
Subject	● Tropical medicine -- Early works to 1800. ● Medical climatology -- Early works to 1800. ● Natural history -- Jamaica -- Early works to 1800.
Copies - Brit. Isles	● Brighton Central Library ● British Library ● Cambridge University Library (includes Sir Geoffrey Keynes Collection, British & Foreign Bible Society, & Peterborough Cathedral) ● Glasgow University Library ● National Library of Ireland ● National Library of Scotland

Figure 1 ESTC record, Patrick Browne, *The Civil and Natural History of Jamaica*, 1789 (screenshot, 22 May 2020)

powerful scholarship. Alongside the inventions of writing and printing as technologies of knowledge, Alston noted, ‘we have now added a third (and by comparison with the former two) quite remarkable one: the storage, and virtually instant retrieval, of information about the present and the past in electro-mechanical form, and the mechanical aids available to assist in this

process are now formidable – more powerful than anything we have known’ (Alston, 1981a: 379).

Using machine-readable records meant that users could perform much more sophisticated searches than were possible with card catalogues (Zeeman, 1980: 4). In addition, cataloguers, ‘keen to facilitate greater access to the online records, tended subconsciously (perhaps) to transcribe long titles’, enabling users to conduct complex keyword searches (Crump, 1988: 5).

However, using MARC for the 18thC STC would not be a seamless fit. It’s worth reminding ourselves that each book in this period was handmade, the product of a series of processes which depended ‘upon a complex sequence of events, all of which were determined by humans capable of fallibility, stupidity, laziness, inconsistency, disobedience’ (Alston, 1981a: 372). These included the making of paper, ink, and metal type, as well as composing type into sentences, locking those sentences into a frame, using the press to make each sheet, proofreading, and compiling it all into a book (Figures 2 & 3).



Figure 2 Metal type