Examining the Early Modern Canon: *The English Short Title Catalogue* and Large-Scale Patterns of Cultural Production

Mikko Tolonen, Mark J. Hill, Ali Zeeshan Ijaz, Ville Vaara, and Leo Lahti

> *But where are such critics to be found? By what marks are they to be known? How distinguish them from pretenders?*
> —David Hume, *"Of the Standard of Taste"* (1757)

CANON-MAKING VERSUS LARGE-SCALE PATTERNS OF CULTURAL PRODUCTION

This chapter offers a data-driven approach to constructing and examining the English canon (ca. 1500–1800). This is part of a long-term project started in 2013, which involves extracting data from *The English Short Title Catalogue* (ESTC), a database that contains records of works published between 1473 and 1800 in Britain and its former colonies. The

M. Tolonen (✉) • M. J. Hill • A. Z. Ijaz • V. Vaara
University of Helsinki, Helsinki, Finland
e-mail: mikko.tolonen@helsinki.fi; mark.hill@helsinki.fi; ali.ijaz@helsinki.fi; ville.vaara@helsinki.fi

project has been undertaken by the Helsinki Computational History Group (COMHIS) to which the authors of this chapter belong. COMHIS is an integrated multidisciplinary team that combines big data approaches with expert subject knowledge in intellectual and book history to study the early modern period.[1]

For this chapter, we have quantitatively constructed a canon of works that were published most often, most frequently, and for the longest period of time by making use of a processed version of the ESTC and analyzing it in terms of time, people, places, and materiality.[2] Importantly, our aim has not been to curate a canon but to extract one based on a systematic quantitative investigation of publishing patterns. Therefore, this chapter provides both methodological insights into such a task and historical insights into the history of (mainly English) printed works. One crucial aim of this study is to demonstrate the enormous analytical potential of harmonized metadata catalogs. For us, this study functions first and foremost as a proof of concept[3] and lays out groundwork for a series of case studies developed with the ESTC.

Defining the Canon: A Brief History

When speaking of written works, the canon commonly refers to "lists of approved authors" of literature.[4] In his important monograph on the

[1] For more information about the group and the authors' backgrounds, see https://www.helsinki.fi/en/researchgroups/computational-history

[2] For a technical description of our canon index, see footnote 45. The code for reproducing the figures and tables in the chapter is to be found at https://doi.org/10.5281/zenodo.4003898

[3] We are fully aware that some of our statistics, especially with respect to particulars, are bound to change as we harmonize and further enrich the ESTC data, as well as improve the overall data quality. We do not see this as a weakness, however, but as a possibility to generate transparent and reproducible inferences that are based on the most reliable and comprehensive data sets currently available. These inferences can be further refined over time, leading to more precise evidence. At the same time, we have put forward research questions that can be answered with our current data and that are focused on large-scale qualitative trends. We expect that the conclusions we have drawn will not be significantly revised by future improvements in data quality.

[4] Wendell V. Harris, "Canonicity," *PMLA* 106, no. 1 (1991): 116, doi: https://doi.org/10.2307/462827. For the etymology of the canon and its different uses, see also Trevor

L. Lahti
University of Turku, Turku, Finland
e-mail: leo.lahti@utu.fi

making of the English canon in the eighteenth century, Jonathan Kramnick calls it "a pantheon of high-cultural works from the past."[5] What works are to be included in such a pantheon, however, is far from unambiguous. Samuel Johnson believed that canonical works of literature are able to communicate timeless human experiences due to the uniformity of human nature; these are works which transcend the particular.[6] Similarly, David Hume believed in a "universal" form of art which is superior to lesser forms, such as those rooted in the mundane and the particular.[7] It is from this perspective that the literary critic emerges as a moderator of the canon, or, as Ernst Robert Curtius put it, "the artificial safeguard of a tradition."[8] Conversely, there are those who dismiss the idea of objective canonicity, and argue, instead, that the canon is a reflection of contemporary or past values.[9] While our age, 250 years after Johnson's time, is "as much post-canonical as post-modern,"[10] this is not a recent view. In late eighteenth century, Isaac Disraeli wrote that "different times … are regulated by different tastes. What makes a strong impression on the public at one time, ceases to interest it at another … and every age of modern literature might, perhaps, admit of a new classification, by dividing it into its periods of *fashionable literature*."[11] Other thinkers, such as Jean Jacques Rousseau, went even further, arguing that the very idea of culturally celebrated works

Thornton Ross, *The Making of the English Literary Canon: From the Middle Ages to the Late Eighteenth Century* (Montreal: McGill-Queens University Press, 1998), 23–25. A canon need not be literary only; a cultural canon can also include, for example, objects, such as the Notre Dame Cathedral, UNESCO sites of world heritage, or even cultural trends, such as tulipomania in Holland. For details, see Frits van Oostrom, *A Key to Dutch History: The Cultural Canon of The Netherlands* (Amsterdam: Amsterdam University Press, 2007).

[5] Jonathan Brody Kramnick, *Making the English Canon: Print-Capitalism and the Cultural Past, 1700–1770* (Cambridge: Cambridge University Press, 1999), 1.

[6] Ross, *The Making of the English Literary Canon*, 267–92. About the nature of literature, see also Terry Eagleton, "The Subject of Literature," *Cultural Critique* no. 2 (1985): 95–104, https://doi.org/10.2307/1354202

[7] David Hume, "Of the Standard of Taste," in *Essays Moral, Political, Literary,* ed. Eugene F. Miller (Indianapolis: Liberty Fund, 1987), 226–50. For an insightful interpretation of this text, see Noel Carroll, "Hume's Standard of Taste," *The Journal of Aesthetics and Art Criticism* 43, no. 2 (1984): 181–94, https://doi.org/10.2307/429992

[8] Ernst Robert Curtius, *European Literature and the Latin Middle Ages* (Princeton, NJ: Princeton University Press, 2013), 253.

[9] Charles Altieri, "An Idea and Ideal of a Literary Canon," *Critical Inquiry* 10, no. 1 (1983): 37–60, https://www.jstor.org/stable/1343405

[10] Peter Robinson, introduction to *Textual Scholarship and the Canon*, ed. Hans Walter Gabler, Peter Robinson, and Paulius Subačius (Leiden: Brill, 2008), 1.

[11] Isaac Disraeli, "Literary Fashions," in *Curiosities of Literature*, 7th ed. (London: John Murray, 1823), 4:171–72.

being universally good is a dangerous fiction.[12] More recently, Richard Proudfoot has argued that "'canons' were abhorred as restrictive practices by critics and theorists [of Shakespeare's time] who wanted to impose other kinds of restriction, of their own choosing, on the study of literature."[13] It is unsurprising, then, that defenses of particular values, such as those found in Harold Bloom's *The Western Canon*, continue to raise questions of inclusion and exclusion.[14] While, for Johnson and Hume, it was the task of the literary critic to identify canon-worthy authors and works, for critical theorists and post-colonial advocates building canons upon individual judgments is open to, and perhaps deserving of, criticism.[15]

It is this ambiguity around the very idea of canonicity that this chapter aims to overcome through a quantitative analysis of patterns of large-scale cultural production. By moving away from debates around style, subjectivity, objectivity, and universality, and turning, instead, to availability, this chapter hopes to sidestep this debate. While most scholarship on early modern and eighteenth-century British canon formation approaches this subject from the perspective of canon-makers like Samuel Johnson or

---

[12] Jean Jacques Rousseau, *Politics and the Arts: Letter to M. D'Alembert on the Theatre*, trans. Allan Bloom (Ithaca, NY: Cornell University Press, 1968).

[13] Richard Proudfoot, *Shakespeare: Text, Stage & Canon* (London: Arden Shakespeare, 2001), 38.

[14] Harold Bloom, *The Western Canon: The Books and School of the Ages* (Harcourt: Brace, 1994). See also Lillian Robinson, *In the Canon's Mouth: Dispatches from the Culture Wars* (Bloomington: Indiana University Press, 1997). On different perspectives of inclusion and exclusion from a post-colonial perspective see Leslie A. Fielder and Houston A. Baker, eds., *English Literature: Opening Up the Canon* (Baltimore, MA: Johns Hopkins University Press, 1979); Gail Low and Marion Wynne-Davies, eds., *A Black British Canon?* (Hampshire: Palgrave Macmillan, 2006); and Ankhi Mukherjee, *What Is a Classic? Postcolonial Rewriting and Invention of the Canon* (Stanford, CA: Stanford University Press, 2013). For aspects of gender, see Paul Salzman, ed., *Expanding the Canon of Early Modern Women's Writing* (Newcastle upon Tyne: Cambridge Scholars Publishing, 2010).

[15] Importantly, however, both of these perspectives are also visible in the more pragmatic aspect of canon formation: the syllabus. Once a work enters the teaching curriculum, canonicity is naturally enforced until a revisionist wave of canon formation emerges. In this way, even the critics of the canon find themselves acting as judges of values represented by these works. For details, see Jan Gorak, ed., *Canon* vs. *Culture: Reflections on the Current Debate* (New York: Garland, 2001); John Guillory, *Cultural Capital: The Problem of Literary Canon Formation* (Chicago: University of Chicago Press, 1993); Harry Levin, "Core, Canon, Curriculum," *College English* 43, no. 4 (1981): 352–62, https://doi.org/10.2307/377120; Richard Bradford, *Is Shakespeare Any Good? And Other Questions on How to Evaluate Literature* (Malden: Wiley-Blackwell, 2015), 166–93; and Blair Hoxby, *What Was Tragedy? Theory and the Early Modern Canon* (Oxford: Oxford University Press, 2015).

John Dryden, a key aim of this chapter is to introduce a systematic method for studying the canon and the process of canon formation which does not focus on particular authors or literary critics.[16] To achieve this goal and move beyond ontological debates on what the canon actually is, we have also expanded the temporal limits of canon formation.

Much of the debate around the early modern English canon focuses on whether literature, in its modern sense, was born in the eighteenth century or earlier.[17] As Kramnick argues, "the decisive reception of the English literary past was settled during the mid-eighteenth century. Years of critical discussion coalesced then into a durable model of literary history and aesthetic value."[18] However, in a similar vein to the individual-centered

[16] On the one hand, this emphasis on individual canon-makers can be seen in the debate about "when literature was invented," culminating in Richard Terry's study, "Literature, Aesthetics, and Canonicity in the Eighteenth Century," *Eighteenth-Century Life* 21, no. 1 (1997): 80–101, https://www.muse.jhu.edu/article/10407. Even when scholars of the early modern canon are aware of theories of cultural production and pay attention to wider social processes, they still emphasize the role of individual canon-makers. Jeremy Lopez, for example, in *Constructing the Canon of Early Modern Drama* (Cambridge: Cambridge University Press, 2014) examines the history of canon-making through anthologies, thus focusing on selections by individuals. On the other hand, both Kramnick and Ross base their take on the canon on Pierre Bourdieu's distinction between canon as a set of values reflected in literature, and canon as a result of large-scale cultural production. Cf. Pierre Bourdieu and Jean-Claude Passeron, *Reproduction in Education, Society and Culture*, trans. Richard Nice (London: Sage Publications, 1977); Pierre Bourdieu, "The Market of Symbolic Goods," *Poetics* 14, nos. 1–2 (1985): 13–44, https://doi.org/10.1016/0304-422X(85)90003-8; and Pierre Bourdieu, "The Field of Cultural Production, or the Economic World Reversed," *Poetics* 12, nos. 4–5 (1983): 331–56, https://doi.org/10.1016/0304-422X(83)90012-8. For a traditionally constructed historical case of pamphlets and the marketplace, see Alexandra Halasz, *The Marketplace of Print: Pamphlets and the Public Sphere in Early Modern England* (Cambridge: Cambridge University Press, 1997).

[17] In his analysis, Douglas Lane Patey emphasizes the importance of the eighteenth century in the process of canon formation, given its focus on aesthetics; see "The Eighteenth Century Invents the Canon," *Modern Language Studies* 18, no. 1 (1988): 17–37, https://doi.org/10.2307/3194698. Following the debate started by Thomas P. Miller, Clifford Siskin, Howard Weinbrot, Barbara M. Benedict, Robert Crawford, J. Paul Hunter, Thomas Bonnell, and Jonathan Brody Kramnick in the same issue of *Modern Language Studies*, Richard Terry emphasized the importance of earlier times in the process of canon formation; for details, see "Literature, Aesthetics, and Canonicity in the Eighteenth Century." For a more recent discussion and a similar emphasis, see David Fairer, "Historical Criticism and the English Canon: A Spenserian Dispute in the 1750s," *Eighteenth-Century Life* 24, no. 2 (2000): 43–64, https://doi.org/10.1215/00982601-24-2-43

[18] Kramnick, *Making the English Canon*, 1. While there are several works that challenge the formative importance of the eighteenth century, all of them focus on the role of individual agents in the process of canon formation. See Ross's *Making of the English Literary Canon*, for instance, which considers cultural attitudes toward literature, and Jane Spencer's *Literary*

approach, this assessment starts from the position that, before one can define the canon, one must define its contents. To avoid this question, we take the preliminary position that printing itself was the starting point for a work to be potentially included in the canon. While we do not deny that the eighteenth century, and particular types of written works, are key to the canon, we think that defining the canon by these features is to put the cart before the horse.

Since Elizabeth Eisenstein's seminal study, *The Printing Press as an Agent of Change*, it has been clear that the history of the book is an integral part of the formation of early modern civil society.[19] On the whole, however, the role of the printing industry in the process of canon formation during the early modern period has remained understudied, especially from a quantitative perspective, despite scholars from outside of book history highlighting the benefits of such an approach. Benedict Anderson, for example, coined the term "print capitalism" to describe the complex dialectics of modernization through literacy and commerce tied to the book trade.[20] Similarly, Jürgen Habermas identified the print and its growing mass consumption as key to the emergence of the public sphere.[21] In the same tradition, this chapter acknowledges that large-scale cultural production is related to the formation of common values, but it starts from the premise that to engage with the canon at this level requires a wider view of what the canon *could* be.

To come to understand what the canon could be, we turn to Alastair Fowler. Fowler's taxonomy identifies six kinds of canon: the potential, the accessible, the selective, the official, the personal, and the critical.[22] While

Relations: Kinship and the Canon 1660–1830 (Oxford: Oxford University Press, 2005), which focuses on author networks.

[19] Elizabeth Eisenstein, *The Printing Press as an Agent of Change: Communications and Cultural Transformations in Early Modern Europe* (Cambridge: Cambridge University Press, 1979); see also Lucien Febvre and Henri-Jean Martin, *The Coming of the Book: The Impact of Printing, 1450–1800* (New York: Verso, 1976). For more recent literature, see Thomas Munch, *Conflict and Enlightenment: Print and Political Culture in Europe, 1635–1795* (Cambridge: Cambridge University Press, 2019).

[20] Benedict Anderson, *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (New York: Verso, 1983).

[21] Jürgen Habermas, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society* (Cambridge: Polity Press, 2003). See also Halasz, *The Marketplace of Print*, 162–204.

[22] Alastair Fowler, "Genre and the Literary Canon," *New Literary History* 11, no. 1 (1979): 97–119, https://doi.org/10.2307/468873. Wendell Harris has suggested that

the selective, official, personal, and critical canons are not directly within the scope of our study of large-scale cultural production, the other two get closer to what we aim to engage with here. The potential canon "comprises the entire written corpus, together with all surviving oral literature"[23]; that is, an ideal canon is made up of the totality of the written corpus, including what has not survived, as well as oral literature which is potentially lost and often scattered when not. Therefore, for our purposes, we turn to the "accessible" canon, or the portion of the potential canon which is available at a given time. An important upshot of this approach to the canon is that "creative writings [are] not dissociated from referential ones such as history, oratory, letter writing, and preaching."[24] To be printed is to be potentially canonical (we could call this the *commercial canon*) and, importantly, we have a robust record of these works in library catalogs, such as the ESTC.

This study takes the ESTC and treats it as a record of cultural production, and thus as the "accessible" canon. Its contents (which we will return to in more detail shortly) are taken as data points which can be quantified and tracked over time. Based on these records, some works emerge as having been printed more frequently than others; some subjects and topics come into and fall out of fashion; and some authors gain or lose audiences at different times. When approached from this perspective, decisions about who should and can be canonical are no longer made by us, but by the cultural environment they emerge from. We disconnect ourselves from individual canon-makers and focus, instead, on large-scale cultural production. Thus, Shakespeare and cookbooks are treated as equals in their potentiality to be part of the canon. Our underlying principle is that these are works which particular people decided that should be printed at particular moments in time.

By focusing on products that were printed over an extended period of time, our approach to the canon could run the risk of being exploratory, mundane, intrinsic, and perhaps even positivist. This should not be regarded as a weakness, however. Instead, it is, perhaps ironically, in line with the tradition initiated by Quentin Skinner and the Cambridge School: it allows us to create a canon which is born out of its own historical context, similar to the one that would have been accessible to those living at a

other kinds of canon can also be considered, such as the pedagogical canon. See Harris, "Canonicity," 113.

[23] Fowler, "Genre and the Literary Canon," 98.

[24] Terry, "Literature, Aesthetics, and Canonicity in the Eighteenth Century," 98.

given time.[25] Moreover, by looking at all works in the ESTC, and comparing them across subjects and genres, a new way of seeing how cultural capital is formed emerges. That is, by comparing the popularity of aesthetically valuable literature with that of more mundane works of domestic economy, we may be able to recognize things about literature that definitions of aesthetic value miss. As Alex Thomson suggestively put it, "the very idea of literature might be a function of the way that we look at the past. What has been seen as literary in the past has often been treated dismissively by subsequent generations, so it seems perfectly reasonable to say that a book can be literary at one time and not at another."[26]

Defining the Canon: Available Data

The ESTC is a comprehensive union catalog listing early modern books, serials, newspapers, pamphlets, broadsides, and other ephemera printed between 1475 and 1801.[27] Covering over 480,000 documents held by

[25] See Quentin Skinner, *Visions of Politics. Volume I: Regarding Method* (Cambridge: Cambridge University Press, 2002), and Mark J. Hill, "Invisible Interpretations: Reflections on the Digital Humanities and Intellectual History," *Global Intellectual History* 1, no. 2 (2016): 130–50, https://doi.org/10.1080/23801883.2017.1304162. It is worth noting that the concept of "canon" had particular meanings at specific historical moments. While we acknowledge that these meanings informed decisions regarding which works deserved reprints or critical editions, we do not aim to uncover them here, but the body of works which were available to readers at particular moments in time. For more on the historical meaning of "canon," see Jan Gorak, *The Making of the Modern Canon: Genesis and Crisis of a Literary Idea* (London: Bloomsbury, 2013), 47.

[26] Alex Thomson, "What Is Literature?" in *The Edinburgh Introduction to Studying English Literature*, ed. Dermot Cavanagh, Alan Gillis, Michelle Keown, James Loxley, and Randall Stevenson, 2nd ed. (Edinburgh: Edinburgh University Press, 2014), 3–16.

[27] For a discussion of the ESTC and its possibilities and limitations, see Michael F. Suarez, "Towards a Bibliometric Analysis of the Surviving Record, 1701–1800," in *The Cambridge History of the Book in Britain. Volume V: 1695–1830*, ed. Michael F. Suarez and M. L. Turner (Cambridge: Cambridge University Press, 2009), 37–65; Michael F. Suarez, "Book History from Descriptive Bibliographies," in *The Cambridge Companion to the History of the Book*, ed. Leslie Howsam (Cambridge: Cambridge University Press, 2015), 199–219; Stephen Karian, "The Limitations and Possibilities of the ESTC," in *The Age of Johnson: A Scholarly Annual. Volume 21*, ed. Jack Lynch (New York: AMS Press, 2011), 283–97; Peter Stallybrass, "The Library and Material Texts," *PMLA* 119, no. 5 (2004):1347–52, https://doi.org/10.1632/003081204X17914; Alex Weedon, "The Uses of Quantification," in *A Companion to the History of the Book*, ed. Simon Eliot and Jonathan Rose (Malden: Wiley Blackwell, 2009), 33–49; Leo Lahti, Niko Ilomäki, and Mikko Tolonen, "A Quantitative Study of History in the English Short-Title Catalogue (ESTC), 1470–1800," *Liber Quarterly*

more than 2000 libraries, it is an essential record of early English print culture. However, like other library catalogs, it is also "a greatly underestimated source of knowledge."[28]

It is important to note that the original purpose of an analytical bibliography is not to support quantitative research, but to preserve as much original information regarding the printed material as possible. However, at the same time, each ESTC record represents a unique printed document, edition, reprint, or variant. Theoretically, every known variant of a work should have its own distinct record in the ESTC. When we combine this fact with the realities of the hand-press printing period, a technology which remained remarkably stable until the nineteenth century, it becomes possible, through careful harmonization,[29] to treat these records as comparable units. For example, estimating the popularity of a particular work based on the number of its editions and reprints makes sense due to the relative stability of print run counts (something which, following technological innovations in the print industry during the nineteenth century, becomes more difficult due to the larger variations between print runs). Thus, the ESTC allows us to examine the canon from a data-driven perspective.[30]

25, no. 2 (2015): 87–116, https://doi.org/10.18352/lq.10112; and Maureen Bell and John Barnard, "Provisional Count of STC Titles, 1475–1640," *Publishing History* 31 (1992): 47–66.

[28] Mikko Tolonen, Leo Lahti, Hege Roivainen, and Jani Marjanen, "A Quantitative Approach to Book-Printing in Sweden and Finland, 1640–1828," *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52, no. 1 (2019): 57, https://doi.org/10.1080/01615440.2018.1526657

[29] The original, unprocessed bibliographic metadata is typically not directly comparable across the catalog due to varying naming conventions, spelling errors, missing entries, and other technicalities. The data quality and comparability can be substantially enhanced by automated harmonization procedures, as described in more detail in Leo Lahti, Jani Marjanen, Hege Roivainen, and Mikko Tolonen, "Bibliographic Data Science and the History of the Book (c. 1500–1800)," *Cataloging & Classification Quarterly* 57, no. 1 (2018): 5–23, https://doi.org/10.1080/01639374.2018.1543747

[30] Our methodological approach was described in Lahti, Marjanen, Roivainen, and Tolonen, "Bibliographic Data Science and the History of the Book (c. 1500–1800)," and it was later used in Tolonen, Lahti, Roivainen, and Marjanen, "A Quantitative Approach to Book-Printing in Sweden and Finland, 1640–1828." See also Lahti, Ilomäki, and Tolonen, "A Quantitative Study of History in the English Short-Title Catalogue (ESTC), 1470–1800." Philip Gaskell's idea of "London average" (750–1500) as a reliable estimate of early modern British print runs was supported, for example, by Richard Sher, in *The Enlightenment and the Book. Scottish Authors and Their Publishers in Eighteenth-Century Britain, Ireland, and*

It is also important to understand that, even though an immense amount of work has gone into processing the ESTC, it remains imperfect both as a catalog and a dataset.[31] Like other early modern catalogs, it cannot be considered as comprehensive because it does not contain information about every existing early modern British publication. Some works have been lost while others have been collected with concern for posterity.[32] Thus, while a quantitative analysis of the early modern period is exciting, we must recognize the limits of the data we have access to. Although there are statistical approaches to estimating lost works, this chapter will focus, instead, on one aspect of the dataset which we can more confidently explore: the most popular works recorded in the ESTC.[33] These are works which are both historically more likely to survive and quantitatively more representative of the overall publication output.[34] With these reservations in mind, one of the strengths of the ESTC's records lies in their robustness: they contain as many as 420 fields, each with its own attributes, ranging from a work's physical features to information on the libraries holding copies of the work. Their weakness, however, lies in the significant effort needed to extract this information at scale. While the ESTC records use the Machine-Readable Cataloging (MARC) 21 standard, making use of

---

America (Chicago: University of Chicago Press, 2007), 86–87. See also Sarah Werner, *Studying Early Printed Books, 1450–1800: A Practical Guide* (Malden: Wiley Blackwell, 2019), 23–25.

[31] Cf. Stephen Karian, "The Limitations and Possibilities of the ESTC."

[32] See, for example, the impact of the Civil War on the number of pamphlets published or the damage caused by the Great Fire of London in 1666 in terms of lost works (Fig. 3.1). There are several other issues that complicate these historical anomalies, such as Thomason's tracts causing a peak in the number of variants during the Civil War, duplicated titles when Wing and STC catalogs are combined, approximated publication years causing a peak in publication numbers at five-year intervals starting from 1500, and other issues specific to the data collection and cataloging process.

[33] For analyses and estimates of the surviving record and the ESTC, see Flavia Bruni and Andrew Pettegree, eds., *Lost Books: Reconstructing the Print World of Pre-Industrial Europe* (Leiden: Brill, 2016).

[34] It is evident that the survival rate of frequently published canonical works is higher than that of single-sheet publications and other ephemera. On broadsheets, see Andrew Pettegree, ed., *Broadsheets: Single-Sheet Publishing in the First Age of Print* (Leiden: Brill, 2017).

this data requires extensive processing.[35] A brief overview of the process clearly shows this challenge.[36]

Actors

The data for actors connected to entries in the ESTC are contained in a wide variety of MARC fields.[37] The extraction process varies depending on the type of actor (author, printer, publisher, editor, etc.) and the amount of information about that actor documented in the metadata. In an ideal case, we are provided with the actors' names, years of birth and death, and their role in a particular printed work; in less ideal cases, we may only be provided with a set of initials or a verbatim repeat of a work's imprint.[38] As

[35] Lahti, Marjanen, Roivainen, and Tolonen, "Bibliographic Data Science and the History of the Book," 5–10. The Library of Congress provides an extensive overview of the MARC 21 standard at https://www.loc.gov/marc/

[36] The end result of processing the MARC catalog is a dataset comparable to linked data. Distinct entities, such as actors connected to the titles documented in the original MARC catalog, are separated out and assigned unique identifiers; these identifiers are, in turn, used to link various entities. A conscious decision was made to keep the infrastructure around the dataset as light and as simple as possible. The whole processed dataset is stored in a collection of CSV tables of entities; these entities include works, titles, actors, and linking tables that document these connections, when needed. These tables, in turn, are stored in a thoroughly documented Git repository. From a modelling point of view, this resembles an SQL database and can be easily exported to other formats. For details, see Leo Lahti, Ville Vaara, Jani Marjanen, and Mikko Tolonen, "Best Practices in Bibliographic Data Science," i*n Proceedings of the Research Data in the Humanities 2019 Conference: Data, Methods, and Tools*, ed. Jarmo Harri Jantunen, Sisko Bruni, Niina, Kunnas, Santeri Palviainen, and Katja Västi (Oulu: Faculty of Humanities Linnanmaa, 2019), 57–65.

[37] The process of unifying actors, and some initial research making use of this data, are detailed in Mark J. Hill, Ville Vaara, Tanja Säily, Leo Lahti, and Mikko Tolonen, "Reconstructing Intellectual Networks: From the ESTC's Bibliographic Metadata to Historical Material," in *Proceedings of the Digital Humanities in the Nordic Countries 4th Conference: Copenhagen, Denmark, March 5–8, 2019*, ed. Constanza Navarretta, Manex Agirrezabal, and Bente Maegaard (Aachen: CEUR-WS.org, 2019), 201–19.

[38] Information on book trade actors was to a large part distilled from a field (260b) described in the MARC standard as "Name of publisher, distributor, etc." In reality, however, it most often repeated faithfully the publisher's statement of a particular title. Examples vary widely in length, level of detail, and style. For instance, they can vary from brief statements, such as "Sold by J. Newton and R. Bland," to detailed descriptions, such as "Printed by T. Bensley, Bolt-Court, Fleet-Street. Sold at Providence chapel on Monday and Wednesday Evenings, and at Monkwell Street Meeting on Tuesday Evenings; by W. Baynes, No. 54, Paternoster-Row; T. Green, No. 93, and J. Baker, No. 226, Oxford-Street; J. Cobbin, No. 14, Hertford-Street, Fitzroy-Square; at the Chapel in the Cliff, Lewes, Sussex; T. Barston, Castlegate, Grantham, Lincolnshire; and by A. Batten, Sen. Wellwyn, Herts."

each record is distinct and there is no explicit information linking the actors or the works recorded, a large amount of cleaning and unification is needed. Strings representing data points must be extracted, cleaned, corrected, compared (internally and externally, by making use of other databases), unified (when appropriate), and finally harmonized. Incomplete or varying spellings make the task even more difficult as identifying and grouping mentions of the same historical actors requires extensive post-processing. We have currently compiled a database of 144,399 unique actors (of which 56,693 are primarily book trade actors and 67,924 primarily authors), with 1,107,777 links to titles documented in the ESTC.[39]

Works

An additional issue with the ESTC is the lack of unifying links between multiple editions of the same work. If one searches for *Romeo and Juliet*, for instance, there may be 80 results based on exact string matching, but no explicit links among these works. Moreover, when one accounts for title variations, repetitions, or commentaries, connecting the correct items becomes a very difficult task. To address this issue, we have used a work-field dataset as the foundation of our process. Since this is an integral part of the workflow that enabled us to start discussing the canon based on ESTC records, it is necessary to explain this process in more detail.

Our aim was to draw out relations between discrete records and link them as single works; in other words, to create a relational model. The fields we chose for matching different records included, first, the edition statement, which provides a specific edition number for each record. Currently, only a small subset of the whole ESTC contains information in this field (around 44,000 records). Second, we used the title and title remainder fields to identify the complete title of a record, while the title uniform field provided a representative title for the work. Lastly, we used the publication year of the record to provide a chronological ordering of the various editions. Distinguishing information, such as publication date or edition number (if available), was then used to determine more precisely what a particular edition is and organize all the editions of a work in chronological order.

The dataset was created through a multi-stage harmonization process, which began with an initial cleaning of the various fields needed for this task. Any unwanted characters and stop words were removed, and the text

---

[39] Other roles include translator, dedicatee, and unknown.

was converted to lowercase. An initial dataset was then created by combining the actors responsible for works with harmonized titles; these actors were used to differentiate among different works with the same title. This provided a unique work-field identifier for each record. The dataset was then segmented and harmonized on a per-actor basis.[40] This was performed by specialized algorithms, which were able to determine the representative work-fields of these editions. This allowed for an effective grouping of editions despite differences in title lengths and spelling. Finally, as some works have multiple actors attached to them (often at different times), another harmonization step was necessary to properly collate these duplicated works into single, unified work-fields.

Figure 3.1 shows a comparison among the original record counts from the ESTC (including first editions and reprints), the processed unique works as derived from the work-field dataset (where the reprints have been removed), and the first prints of the works included in the canon that we extracted. The original records include prints without any actor information, which raises the number of records to over 480,000. However, by processing unique works information, we effectively normalized the whole ESTC dataset, which resulted in a significantly lower number of works for each decade. As the canon dataset is based on the work-field dataset, it follows a similar trend.

It is important to note that the work-field harmonization and dataset creation is an ongoing process, which is iteratively and continuously being improved for more accurate grouping of editions. At the moment, however, the work-field dataset consists of 200,378 works covering a total of 361,245 records.

Subject-Topics

It was important for this study to include information about the subject-topics (such as "Religion," "Literature," and "History and Geography") of the works examined.[41] While other approaches may see this as a necessary step for determining what to include in the canon (i.e., which

---

[40] We have chosen only those records where the actor attached to them had a specific role. The list of accepted actor roles included author, corporate author, translator, and attributed name. All the other records either did not have any actors attached to them or had actors of other roles than the ones selected, so they were not used in the creation of our dataset.

[41] Fowler has underscored the relevance of the information related to genre, when thinking about the canon, in "Genre and the Literary Canon," 100.
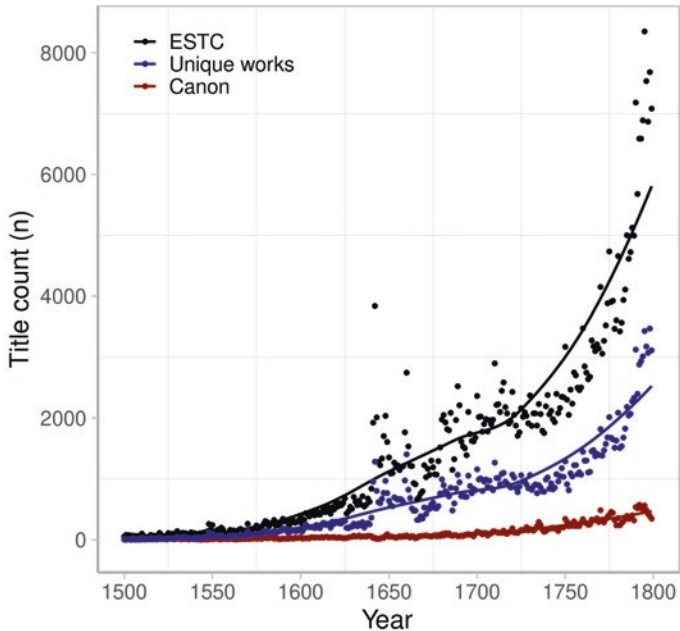
Fig. 3.1 Total documents, works, and canon items in the ESTC per year (1500–1800)

subject-topics are more aesthetically or culturally valuable and, therefore, worthy of being included), for our purposes it allowed for further qualitative reflections, such as temporal changes within the canon. This information, however, has not been recorded in the ESTC comprehensively (roughly only half, or 266,207 documents in the ESTC, have subject headings) or systematically (there are 12,553 unique subjects).[42] It was, therefore, necessary for us to modify and enrich this data.

There are numerous models proposed for categorizing subject-topics, and they range from classification systems developed by ancient authors, to early modern attempts to revise such systems, to current efforts to

---

[42] To complicate things even more, many subjects recorded in the ESTC are questionable. For example, John Arbuthnot's political pamphlet on John Bull is tagged in the ESTC under the subject heading "Bulls."

create appropriate models for quantitative analysis.[43] Because the approach in this study is computational, a hierarchical classification system by subject was aimed for. To this end, we chose the Dewey Decimal Classification (DDC) system.[44] As classifying every single work in the ESTC was not an option, we chose, instead, to hand-classify a selection of works and, then, used this information with the existing subject headings in the ESTC to create a conversion table which was used to further extrapolate the data. That is, each entry with subject-topics in the ESTC was compared with the manually entered data, and this training dataset was used to find the most common equivalent ESTC subject-topic for each DDC category. The resulting DDC-to-ESTC translation table was then used to assign Dewey-style subject-topics to non-hand-categorized ESTC entries. In total, we hand-classified 1153 works, which represent a total of 47,041 individual documents. From these, we were able to classify another 62,342 documents with the conversion table.

We should note that the catalog has many unique and rare topics: 7957 topics were used in only one instance and they range from individual psalms to specific years to "Granby (Race horse)." In the end, we identified 53,683 works with a subject-topic in the original ESTC but with no equivalent in the DDC. This is a typical example of the diminishing returns of manual work in digital humanities: the remaining untranslated subject-topics are increasingly rare, so the payoff for each additional manual entry decreases.

Defining the Canon: Methods

As stated, our aim in this study is to construct a canon using a data-driven approach and analyze the works contained within it. As this is a data-driven approach, there is no qualitative judgment made at the curation stage. Instead, we aim to construct a list of canonical works that is born out of historic publication records. To do so, it is necessary to define a set of features which could be considered representative of canonical works

[43] For a quantitative example, see David. L. Gants, "A Quantitative Analysis of the London Book Trade, 1614–1618," *Studies in Bibliography* 55, no. 1 (2002): 185–213, https://doi.org/10.1353/sib.2005.0004

[44] We have used a two-level system. The general categories are: "Information & General Works," "Religion," "Social Sciences," "The Arts," "Literature," "Philosophy," "Natural Science & Mathematics," "Applied Science," "History & Geography," and "Language." These further break down to a total of 94 subcategories.

and examine the entirety of the ESTC for these features. The features we have chosen include the total publication count and publication frequency, which have been further normalized by the overall publishing activity during the same period. Based on this canon index,[45] we identified the top 1000 works that were printed in relatively high numbers, relatively frequently, and over a relatively long period of time.[46]

This method has resulted in the inclusion of works that qualitatively would never be considered as canonical but, nevertheless, have been included in the data-driven index as they have identical data profiles to works that we would expect to be included. For example, while almanacs may be considered less historically important to many, their frequent and

[45] We have defined the canon index as $C_w = T_w \times (N_w/N)$, where C_w is the canon index for work W, T_w is the publication frequency (total number of distinct publication years) for work W, N_w is the total publication count for work W, and N is the total number of all works published between the first edition of work W and the year 1800. The ratio (N_w/N) indicates the overall share of the given work in all publications within the time period that starts from the first publication of work W and ends in 1800. The normalization of the covered time span improves the comparability between earlier and later publications. The canon index increases with the total publication count and publication frequency. We have tested multiple metrics and methods to create a reliable index of "canonical works," making use of historical expectations and ensuring a qualitative balance among features. We have chosen the above-mentioned features for several reasons. First, while the number of unique versions (or editions) of a given work is certainly an indicator of a work's importance, this feature alone is not an indication of canonical importance. Works could be printed in very high numbers over very short periods of time and then quickly forgotten (see, for example, political pamphlets during the English Civil War). While these works are interesting in their own right, they are more likely indicative of topicality than canonicity. Thus, the longevity of a work, i.e., the frequency and total period during which it was reprinted, was deemed a key feature. However, this metric also needs to be treated as a relative feature, allowing for works to be measured against their temporal peers, as works printed earlier had a much higher longevity potential. Thus, each work was also measured as a proportion of its potential contribution to the publication output for the period in which it was published. That is, a work published in 1750 has been measured as a proportion of all works published between 1750 and 1800 rather than against all works in the ESTC. It is important to note that works first printed at the very end of the years covered by the ESTC cannot have their future impact measured as there is no data post-1800. There are, therefore, fewer new works marked as canonical in the last decade of the eighteenth century. The identification of the top canonical works was relatively robust to the choice of index; we have chosen to use an index that is intuitive and easy to calculate.

[46] Although 1000 may appear to be an arbitrary number, it meets two useful criteria: first, it is a number that a human is still able to engage with when qualitatively examining results and assigning subject-topic classifications; second, it is a number after which the total number of reprints for a work begins to drop below 20.

stable reprinting patterns mean that they feature in our list. Their relevance, of course, depends on the context of publication. While an almanac lacks literary significance, from a commercial standpoint it is very significant, a point which can also be made about commercial catalogs that emerged in the latter part of the eighteenth century. Similarly, while grammar books played a key cultural role at the time of their publication, some may argue for their exclusion from lists of literary bestsellers.[47] Indeed, for the most part, these works do not fit the category of printed cultural material we are interested in. For this reason, printed versions of laws and political reports, liturgies and local church documents, catalogs, almanacs, curricula, periodicals not published as collected works, and annual reports from various associations have been excluded from our canon,[48] leaving us with 856 works.[49]

The Canon: Works, Time, and Subject-Topics

When treated as one continuous historical canon, the top 20 works we have extracted based on the current data are displayed in Table 3.1.[50]

---

[47] Sigfrid H. Steinberg, *Five Hundred Years of Printing* (London: Penguin, 1955), 99–105.

[48] Of the 154 works considered not relevant to the canon, the top ten were Church of England liturgies, general public acts (one set authored by "Britain" and one by the Parliament), Irish proclamations, Connecticut laws, Catholic church liturgies, *Rider's British Merlin*, Quaker yearly epistles, miscellaneous official documents by King Charles I, and reports from the Court of Chancery.

[49] It should be noted that the earlier part of the period covered by the ESTC is slightly overrepresented, as we can see in the long tail. There are at least two reasons for this. First, there is a back catalog of written material which had existed for much longer, yet it is only recorded from the start of the ESTC. This catalog includes authors who lived prior to the sixteenth century, in particular ancient authors, who can only be recorded at the beginning of the ESTC records. Second, printing itself was a much more specialized industry early on. This means that what was chosen to be printed may already have met some subjective criteria established by printers, which made it more likely to be a work with longevity. At the time, the differences between early printing and traditional manuscript production were minimal.

[50] It should be noted that the publication years of these works have been extracted from the ESTC following the method laid out previously. We do not claim that this data includes every edition of a work; the many changes in the way in which a work has been recorded in the ESTC makes the detection of all editions difficult. As harmonization of public documents, such as general acts, continues, they will be mapped together in the future based on this additional information. Information about the full canon can be found in the online code and data supplement at https://doi.org/10.5281/zenodo.4003898

Table 3.1   Top 20 canonical works (1500–1800)

No.	Author	Work	Eds.	No.	Author	Work	Eds.
1	Thomas Sternhold and John Hopkins	*Book of Psalms* (1553–1762)	592	11	François Fenelon	*Telemachus* (1699–1800)	171
2		*The Book of Common Prayer* (1549–1795)	373	12	Mathurin Cordier	*Colloquies* (1651–1800)	140
3	William Lily	*Short Introduction to Grammar* (1543–1800)	241	13	Alexander Pope	*Essay on Man* (1733–1800)	172
4	Aesop	*Fables* (1484–1800)	226	14	Desiderius Erasmus	*Colloquies* (1519–1800)	124
5	John Bunyan	*Pilgrim's Progress* (1678–1800)	237	15	Homer	*The Iliad* (1581–1796)	142
6	Richard Allestree	*The Whole Duty of Man* (1658–1800)	181	16	Lewis Bayly	*The Practice of Piety* (1612–1792)	128
7	John Milton	*Paradise Lost* (1667–1800)	211	17	Robert Dodsley	*The Oeconomy of Human Life* (1750–1800)	185
8	Ovid	*Metamorphoses* (1552–1800)	153	18	Isaac Watts	*Hymns* (1707–1800)	152
9	Daniel Defoe	*Robinson Crusoe* (1719–1800)	213	19	William Vicker	*A Companion to the Altar* (1707–1800)	131
10	John Wesley	*Hymns* (1740–1798)	203	20	Edward Young	*Night Thoughts* (1742–1800)	168

One can immediately notice the diversity of these works. Included are the expected works of poetry and fiction, but also devotional literature and language grammars—works which would normally be excluded from a literary canon. However, as we have stated, our aim is not to curate a canon but to extract one. It is possible to purge works *post hoc*, but before such a step is taken, it is worth noting and investigating what does make the list. Questioning the works and authors who make up the extracted canon will be a recurring theme throughout this chapter, but to give one example, we will turn now to William Vicker's *A Companion to the Altar*.

A relatively obscure work today, this short text was written to spiritually prepare readers for holy communion. During the eighteenth century and

beyond, the work was often offered as a gift to those preparing for confirmation and was frequently printed along with *The Book of Common Prayer*. This practice can be considered as an explanation for its 131 records, and perhaps be seen as a good reason to exclude this work from our canon. However, the strength of our approach lies in confronting and reflecting on works such as these. To have been published in such large numbers means it reached a large audience, and it is, therefore, worth reflecting on the number of people who would have been familiar with its contents. We know, for example, that it was one of the 20 volumes owned by Jane Austen and that "she made constant use of the devotions contained in it."[51] Thus, while the work itself may be seen as distinctly uncanonical by other definitions, its relationship to both the canon and the historical-cultural moment of its publication should encourage further reflection.

Of course, raw statistics offer only one view of many, especially if we are looking at the canon atemporally. For example, when one looks at the frequency of publication for works included in the ESTC, the emphasis will be on the latter part of the eighteenth century, when most printing activity took place. However, by constructing a data-driven canon which takes into consideration the relative longevity and publication frequency of these works, we can also provide a temporally representative selection. While this temporality can be seen in the overall distribution of these works (Fig. 3.2), this approach also allows us to make more specific analyses. For example, in Fig. 3.3, we can also see the works which were most frequently printed per decade.

As previously noted, we have also assigned subject categories to many of the works in the ESTC, as well as to all the works which make up our canon. By categorizing these works, we can also examine the canon by subject. For example, the top-ten literary works in each category can be seen in Table 3.2.[52]

This data also allows us to examine the changes in the distribution of subject-topics in the canon over time and recognize that subject-topics emerge and subside at particular historical moments. These shifts are not entirely surprising. As Fowler noted, "the complete range of genres is by no means equally, let alone fully, available in any one period … Moreover,

---

[51] Irene Collins, "The Rev. Henry Tilney, Rector of Woodston," *Persuasions* no. 20 (1998): 156.

[52] Although in the top-ten, the collected works of Swift, Pope, Virgil, Horace, Milton, and Shakespeare are not included.
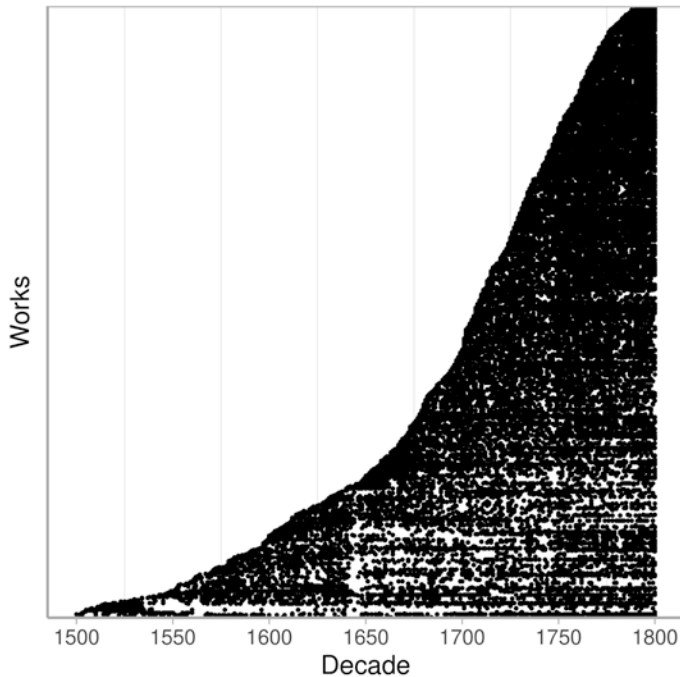
Fig. 3.2 The Full Canon (1500–1800). These canonical works have been sorted by the first publication year. Individual dots indicate the publishing year for the initial publication and all subsequent reprints

each age makes new deletions from the potential repertoire."[53] We can see this in Fig. 3.4: by including more than strictly literary works in the canon, we can see the emergence of new types of printed documents from the mid-seventeenth century onward.

Religion and Literature

By combining data covering top-works, temporality, and subject-topics, we can begin to construct more complicated versions of the English canon. One can see, for example, the importance of grammar books during early printing era (under the category "language"), and then

---

[53] Fowler, "Genre and the Literary Canon," 110.

Table 3.2 Top 10 canonical literary works (1500–1800)

| Fiction	Poetry	Drama
Aesop's Fables	Milton, *Paradise Lost*	Addison, *Cato, a Tragedy*
Bunyan, *Pilgrim's Progress*	Ovid, *Metamorphoses*	Gay, *The Beggar's Opera*
Defoe, *Robinson Crusoe*	Pope, *Essay on Man*	Ramsay, *The Gentle Shepherd*
Homer, *The Iliad*	Young, *Night-Thoughts*	Shakespeare, *Hamlet*
Goldsmith, *The Vicar of Wakefield*	Thomson, *The Seasons*	Lillo, *The London Merchant*
Lesage, *Giles Blas*	Gay, *Fables*	Otway, *The Orphan*
Cervantes, *Don Quixote*	Juvenal, *Satires*	Shakespeare, *Romeo and Juliet*
Sterne, *A Sentimental Journey*	Butler, *Hudibras*	Rowe, *The Fair Penitent*
The Fables of Phaedrus	Blair, *The Grave*	Otway, *Venice Preserv'd*
Sterne, *Tristram Shandy*	Gay, *Acis and Galatea*	Rowe, *Jane Shore*
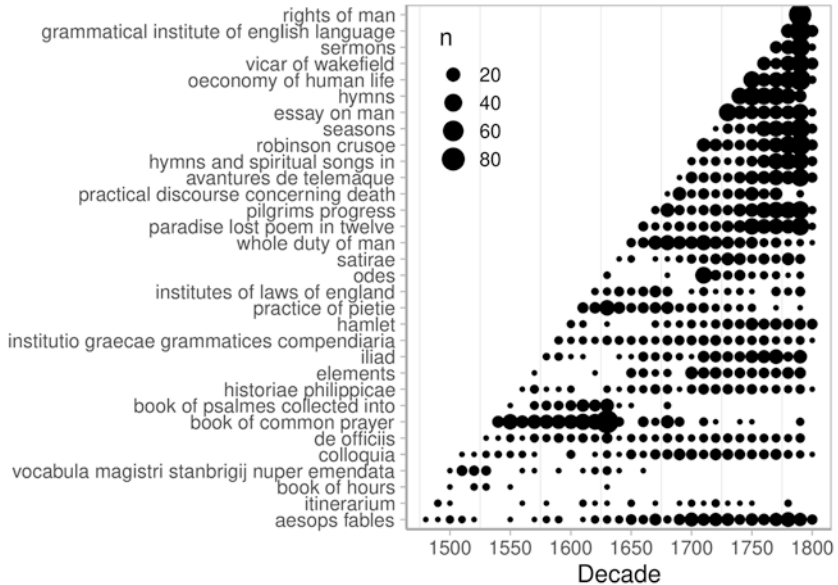
Fig. 3.3 Most frequently printed works (1500–1800). The point size indicates the number of reprints for each work (rows) during a given decade (columns)
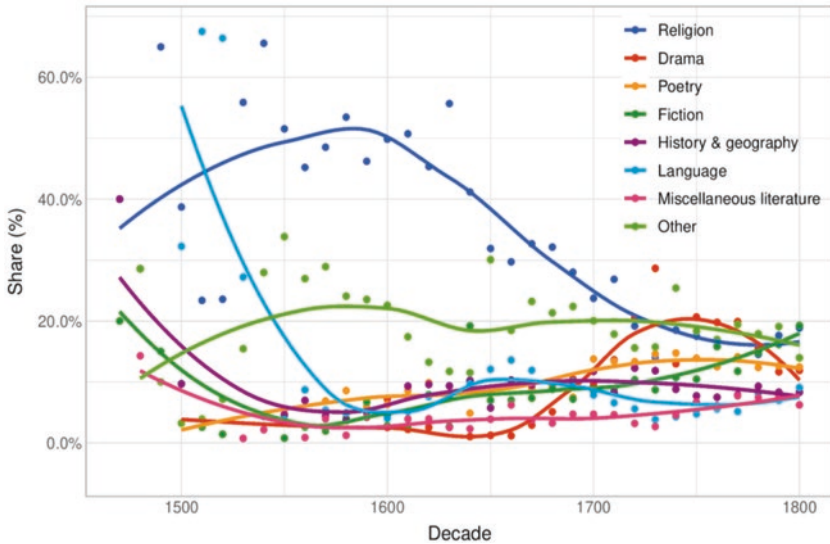
Fig. 3.4 Temporal variation in the relative frequency of the most common subject-topics in the canon (1500–1800). The relative variability in publishing frequency is higher in earlier time periods due to the lower number of total published works

recognize the dominance of religious works up to the late seventeenth century, followed by the rise of literary genres (drama, fiction, and poetry) at the beginning of the eighteenth century.[54] On the whole, it becomes possible to recognize epistemological shifts between the Renaissance and the eighteenth century.[55] In particular, frequently printed late Renaissance works are largely religious tracts, classical works, and grammar books. While the classics, especially *Aesop's Fables* and Cicero's *On Duties*, continue to be printed until the end of the eighteenth century, religious works, such as *The Book of Hours*, *The Book of Common Prayer*, and *The*

---

[54] For early religious publishing, see Patrick Collinson, Arnold Hunt, and Alexandra Walsham, "Religious Publishing in England 1557–1640," and Ian Green and Kate Peters, "Religious Publishing in England c.1640–1695," *The Cambridge History of the Book in Britain. Volume IV: 1557–1695*, ed. John Barnard, D. F. McKenzie, and Maureen Bell (Cambridge: Cambridge University Press, 2002), 29–66, and 67–94, respectively.

[55] It is, of course, true that detecting an epistemological shift cannot be based solely on reprint records. Publication records do, however, offer one perspective.

Book of Psalms, become relatively less apparent. When looking at the eighteenth century, however, works that are more traditionally considered canonical do emerge—particularly with respect to literature. Thus, while literature clearly had a central place in the canon, its volume in terms of printed works compared to religious works is cemented by 1700. When analyzing the subject-topic distribution in the canon before and after the eighteenth century, the decline of religious and grammar books, and the rise of literary genres, especially drama, becomes apparent. This pattern holds true for all the works in the ESTC that include subject-topic information (although, overall, religion holds its place better than history and geography, for example).

Thus, confirming some scholars' historical expectations, the eighteenth century emerges as a key moment in the history of the canon. We would hesitate to state that this makes the eighteenth century representative of the canon, however. We would, instead, note that there is an epistemological shift regarding what is considered canonical at this point in time. This does indicate that, depending on one's analytical aims, the year 1700 is a potentially useful marker.[56] Other potential markers do exist, however, as we will see below.

Donaldson v. Becket and the Importance of 1774

It has been claimed that, "[o]n 22 February 1774, literature in its modern sense began."[57] With Donaldson v. Becket, the House of Lords ended the perpetual copyright and, consequently, London's monopoly over print in Britain, allowing for a back catalog of cultural goods to enter the public domain. The impact of this event on the print industry was profound. As we have noted in previous quantitative research into the ESTC, the relationship between the London printers and publishers in the 1770s changed radically.[58] Our concern in this chapter, however, is to identify what was offered to the public at a large scale after this act. While there is clear

---

[56] Depending on one's aims, other years may be more appropriate. For example, one may also want to look at the post-1666 print industry as it was rebuilt following the Great Fire of London. See Hill, Vaara, Säily, Lahti, and Tolonen, "Reconstructing Intellectual Networks," 206–208.

[57] Ross, *Making of the English Literary Canon*, 297.

[58] See Hill, Vaara, Säily, Lahti, and Tolonen, "Reconstructing Intellectual Networks," 201–19.

evidence that Ross's thesis of radical change should be visible in our data, the question is whether our canon diverges from this expectation or not.

To answer this question, we created a post-1774 canon which was compared with the complete data-driven model. We extracted the 1000 most printed works originally published before 1746 (thus making them potentially public domain), reprinted after 1774, and printed in Great Britain (where the law applied).[59] We then compared these works with those included in the larger canon, looking for substantial differences. It must be noted, however, that the limits placed on the post-1774 canon mean that it is a substantially smaller subset. While the entire processed ESTC has currently 361,245 harmonized recorded documents (representing 200,378 works), only 8925 of those documents (1997 works) meet the post-1774 criteria. Therefore, to compare the two directly is not entirely meaningful as a substantial number of works in the data-driven canon are missing. However, we can still examine how the top-works in the post-1774 canon are distributed in our data-driven canon.

The results show that almost a quarter (23.2%) of all post-1774 works can still be found in the larger data-driven canon. Moreover, the overall distribution of the highest ranking post-1774 works (i.e., most printed) overwhelmingly falls at the top end of the larger canon. This is not entirely surprising, though, as printing frequency is one of the defining features of the data-driven canon. This comparison does, however, verify that this data-driven approach recovers works similar to those in the post-1774 canon with substantial coverage and accuracy. What is more important to this study, however, is what is *not* captured in the post-1774 canon.

Amongst the top 500 post-1774 works, only 41 are not in the data-driven canon. Of these, 16 are works of literature. The types of works not included in these 41 titles are spread across subject-topics, but the most common type of work is drama (eight titles). On the other hand, amongst the top-500 data-driven canonical works which were in the public domain, 59 are not in the post-1774 list. Of these, 21 are works of literature. Importantly, the works which did not make the larger data-driven canon but are in the post-1774 list are generally printed less frequently. When

---

[59] While there is evidence that printers in Scotland had accepted the end of perpetual copyright as early as the 1740s, that is, before the publishers in London, it is not clear that this was universally the case. As Scotland is important culturally and as a center for reprinting works, it was included in our analysis. See John Feather, *Publishing, Piracy and Politics: An Historical Study of Copyright in Britain* (London: MansFigell, 1995), 81.

including all prints (not just those in Britain) the mean number of prints in the data-driven canon is 19, while the number of works in the post-1774 canon is 7. If we only look at prints in Britain, 90% of post-1774 works have 10 or fewer reprints.

Interestingly, a total of 145 literary works found in the larger data-driven canon are missing from the post-1774 canon. A significant number of these works (50) are works of fiction. These include works which were still covered by copyright in 1774, such as Henry Fielding's *Tom Jones* (1749), Samuel Johnson's *Rasselas* (1759), and Oliver Goldsmith's *The Vicar of Wakefield* (1766), as well as older works which, while still published in the eighteenth century, did not reach the same edition count as some newer works, such as Boccaccio's *Decameron* (first recorded in the ESTC in 1562), Robert Greene's *Pandosto* (1588), and Philip Sidney's *Arcadia* (first recorded in the ESTC in 1590).[60] Additionally, 34 works of poetry, 34 works of drama, and 27 miscellaneous literary works are not included in the post-1774 canon. Authors of works that did not make the post-1774 cut include Jonathan Swift, Hannah More, Jean-Jacques Rousseau, Voltaire, Michel de Montaigne, Alexander Pope, Ned Ward, John Dryden, Benjamin Franklin, Thomas Paine, and Richard Steele.[61]

Overall, while there is clear evidence that the 1774 legal changes did have a substantial impact on publishing, especially of works that one would consider canonical, this is an event which takes place so late in our dataset that its impact is likely to be seen more in the following periods not covered by our dataset. This includes works which, for various reasons, fell out of favor toward the end of the eighteenth century, as well as works which were still under copyright in 1774. As we are interested in the canon as it developed and was available over the entirety of the period covered by the ESTC, the actions of a temporally specific group of people should not be over-represented. However, it is worth noting one important upshot of the comparison between the two datasets: the general overlap of coverage and the opportunity to understand the causes of the missing works provides further verification of the methods used to construct our canon.

[60] For more on the shift in the works which make up the canon, see Ross, *The Making of the English Literary Canon*, 210–52.

[61] It should be noted that many of these works may have found their way into the emerging "collected works" genre. However, new collections of works by an author are not treated as continuations of previous works in our data, but as new editions. While this is methodologically correct for our purposes, it does mean that there is potential for under-representation at the author level.

People: Authors

When discussing the people responsible for the canon (with particular emphasis on authors), we must acknowledge that the importance we place on them today is radically different from the importance their contemporaries had placed on these authors in their time. As Adam Rounce has noted with regard to Samuel Johnson, "[t]he desire to be recognized as an author and to profit from it coexists with an awareness that much work in the burgeoning world of print and literary journalism was not especially intended to be handed down to posterity; the sparsity of works with Johnson's name on the title page in his lifetime indicate his pragmatism."[62] In fact, it was only toward the end of the eighteenth century that authorship began to take the form that we recognize as typical today and, therefore, for the majority of the era covered by the ESTC the notion of authorship was quite different.

One of the clearest examples of this is anonymity. Up until the eighteenth century it was quite common for authors not to be credited, by choice or by practice, for their work. For instance, between 1679 and 1800, the ESTC has 239 records with authorship attributed to a "Lady." Many similar examples can be found in our canon: multiple works by Defoe were initially published without any attributed authorship, Philip Francis' criticisms of George III's government were penned under the name "Junius" for obvious reasons, and Richard Steele used the *nom de plume* Isaac Bikerstaff in *The Tatler* (which was itself borrowed from Jonathan Swift).[63] Bikerstaff is indicative of another aspect related to authorship: the practice of collaborative writing. While Steele can be credited with most issues of *The Tatler*, he was not the sole author: Joseph Addison and Jonathan Swift also contributed pieces to the periodical, a practice which would be taken further with *The Spectator*. For many, in fact, authorship was not something that was held in any regard. The "hack" author, paid by the page or word, willing to put his or her skills to use for any topic or patron, was a common presence during the eighteenth century and was successfully immortalized in Pope's *Dunciad* (1728, 1729, 1743). In other words, when thinking about the canon, it is

---

[62] Adam Rounce, "Authorship in the Eighteenth Century," *Oxford Handbooks Online*, 2015, https://doi.org/10.1093/oxfordhb/9780199935338.013.38

[63] Other pseudonyms included in the canon are: "Countess of Huntingdon's Connexion," "Gentleman in the Country," "Gentleman of Oxford," "Lover of their Precious Souls," "Person of Quality," and "Protestant."

important, on the one hand, not to put undue weight on authorship, and, on the other, to recognize the exceptionality of those who were able to step into the authorial spotlight.

Based on our records, 833 works in the canon have a person or organization as author and 556 of these are unique. There are different ways of speaking about who the top authors were or may have been: one can count the authors who published the most canonical works, the authors who published the most editions of these works, or the authors who have the most records in the ESTC. In each case, a different picture emerges, although we see repetition (Table 3.3).

Table 3.3 Top authors and works, canon editions, and total works recorded in the ESTC (1500–1800)

Author	Works	Author	Canon eds.	Author	ESTC recds.
Shakespeare, William (1564–1616)	20	Shakespeare, William (1564–1616)	779	Wesley, John (1703–1791)	1217
Watts, Isaac (1674–1748)	9	Watts, Isaac (1674–1748)	570	Shakespeare, William (1564–1616)	1085
Bunyan, John (1628–1688)	8	Bunyan, John (1628–1688)	480	Defoe, Daniel (c. 1660–1731)	1082
Goldsmith, Oliver (1728–1774)	8	Defoe, Daniel (c. 1660–1731)	416	Watts, Isaac (1674–1748)	1059
Wesley, John (1703–1791)	8	Wesley, John (1703–1791)	377	Swift, Jonathan (1667–1745)	770
Defoe, Daniel (c. 1660–1731)	7	Lily, William (c. 1468–1522)	375	Pope, Alexander (1688–1744)	664
Dryden, John (1631–1700)	7	Goldsmith, Oliver (1728–1774)	366	Voltaire (1694–1778)	648
Rowe, Nicholas (1674–1718)	7	Milton, John (1608–1674)	356	Bunyan, John (1628–1688)	641
Virgil (70 BCE–21 BCE)	7	Gay, John (1685–1732)	347	Christie, Mr. (James) 1730–1803[a]	616
Voltaire (1694–1778)	7	Pope, Alexander (1688–1744)	343	Dryden, John (1631–1700)	548

[a]The inclusion of James Christie is worth mentioning: founder of the auction house Christie's, his attributed works are catalogs of sale. While certainly not the type of work one would generally consider canonical, his inclusion is a point of historical interest as it reflects both changes in the print industry and the rise of commerce. Additionally, it draws attention, again, to the relationship between the canon and large-scale cultural production
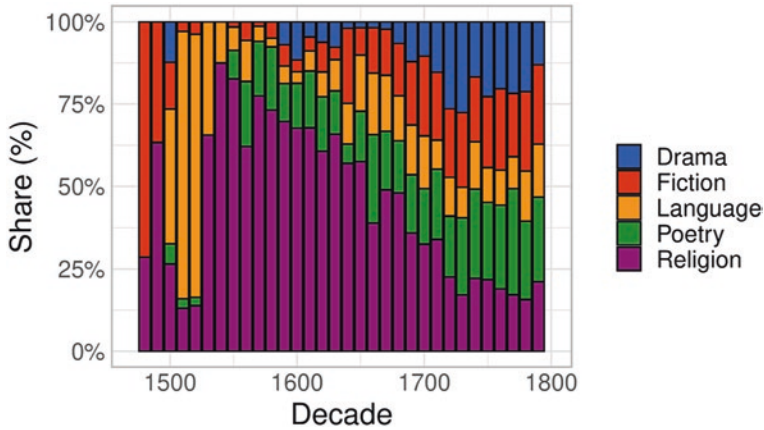
Fig. 3.5 The most popular subject-topics for the ten most printed works in each decade from 1500 to 1800

When we look at the most published works per decade (Fig. 3.3), we can generate a complementary list that attempts to estimate which works were "bestsellers." In this case, no author who makes the top-ten-per-decade list has more than four works included. Additionally, the list emphasizes early authors, such as John Stanbridge (1463–1510), Robert Whittington (approximately 1560), Richard Allestree (1619–1681), John Brinsley (active 1581–1624), and Edward Coke (1552–1634), none of whom make the general list. This is likely due both to the lack of competition when printing first emerged and to the high number of reprints of grammars (see also Fig. 3.5). In contrast, authors like Isaac Watts and Daniel Defoe have only two works that make the top-ten-per-decade list. This is an important insight: it shows that there are various ways to integrate the data that allow for both contextual and *longue durée* insights.

When examining the subject-topics by top-ten canon authors (Table 3.4) literary works dominate the list, although there are a few outliers, such as Defoe's familial advice, *The Family Instructor* (1715), the conduct piece, *Religious Courtship* (1722), and Wesley's medicinal textbook, *The Primitive Physick* (1747). While temporality plays an obvious role in these results, overall there is a decent spread of authors, covering various time periods and genres (Fig. 3.7).

Table 3.4 Distribution of subject-topics among works by top authors (1500–1800)

Category	Count
Drama	34
Poetry	14
Religion	13
Fiction	10
History & geography	6
Philosophy	4
Sermons	4
Domestic economy	1
Education & manners	1
Medicine & health	1
Miscellaneous literature	1

In addition to identifying canonical authors, we have also generated some general statistics regarding the lives of the authors working during the ESTC era. By looking at the authors whose birth and death years are available, we notice that the average age of authors when publishing a first work is quite high (over 40). Additionally, many authors, and especially canonical authors, continue to be published after their death, and, with the advent of the public domain, more and more frequently (Fig. 3.6).

Analysis of posthumous publication frequency indicates that being published after death is more common for authors during the early modern period, ancient authors excluded (Fig. 3.6). According to our data, the median number of years to the first publication after death is one; however, for 2.7% of the 1455 authors who were included in this analysis, the first posthumous publication appears over 100 years after their death. The frequency of posthumous publications within the first 50 years after death shows a steadily declining pattern over time.[64] For the first half of the sample (until the 1650s) the percentage is higher, but we should remember that at the time it was more difficult to be printed due to limitations in the print industry. Thus, existing resources were most likely

[64] We have used the 50-year window after an author's death because it removes the bias associated with the fact that later authors have fewer years for republishing (the bias remains for the last 50 years but this is a side issue as most data is directly comparable and the declining trend is clear). In principle, the declining trend could also be explained by increasing intervals of republishing but, in our data, the average publishing time after death is getting systematically smaller, not larger, so this is not a likely explanation.
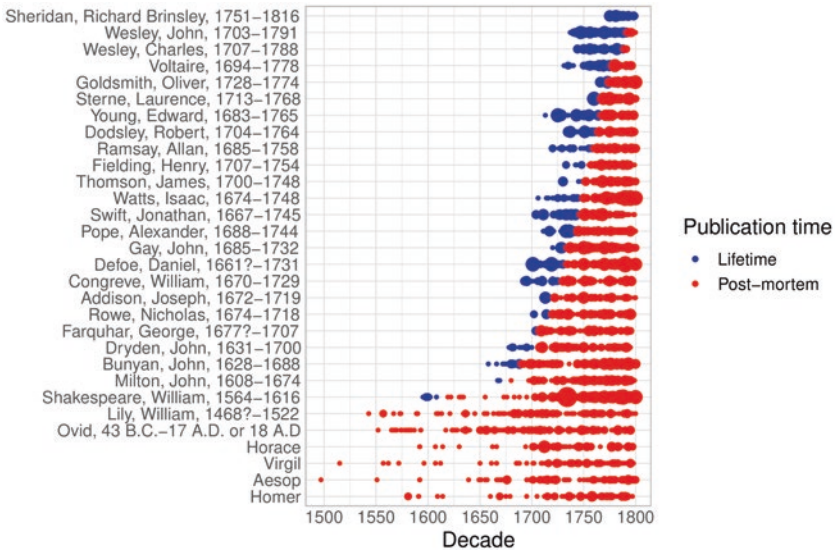
Fig. 3.6 Top authors (1500–1800). The point size indicates the number of publications for each author, including reprints (rows), per year (columns). The color indicates publication before (red) and after (blue) death, respectively. The authors have been sorted by their death year

directed toward printing canonical works, or works by established authors (Fig. 3.7).

As noted previously, while most authors we have extracted from the ESTC can uncontroversially be labeled as canonical, there are some whose inclusion could be contentious. However, as was the case with William Vicker, we may want to reflect on their inclusion before deciding to purge them from the canon. William Lily, for example, is worth noting. Strictly in terms of publication counts, it is understandable why he ranks so high: his *Latin Grammar* (although written by many hands, including Erasmus) was granted a royal monopoly as the only Latin textbook to be used in schools from 1540 onward. However, while Lily was mainly known as a schoolteacher and grammarian, his contribution to humanist education should not be overlooked.[65] Instead, we should acknowledge that his

[65] Mary Beth Stewart, "William Lily's Contribution to Classical Study," *The Classical Journal* 33, no. 4 (1938): 217–25, https://www.jstor.org/stable/3291195
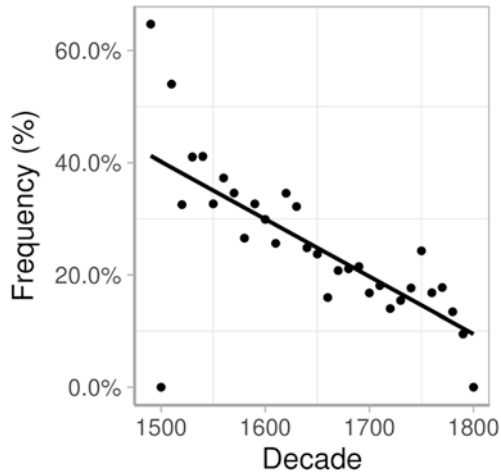
Fig. 3.7 Posthumous publication frequency. The percentage of authors in the ESTC with posthumous publications during the first 50 years after death, grouped by decade (1470–1800). This analysis includes the 1544 authors whose lifetime data is available for the investigated period with one or more posthumous publications

work was a cultural constant amongst all "upper-class British males" for over two centuries, and that its contents, including hundreds of quotes from Roman writers, were known by most educated readers by heart.[66] This fact was disapprovingly recorded by John Locke in *Some Thoughts Concerning Education* (1692): "Custom serves for reason, and has, to those who take it for reason, so consecrated this method, that it is almost religiously observed by them, and they stick to it, as if their children had scarce an orthodox education unless they learned Lilly's grammar."[67] If one of the aims of a data-driven approach to canon formation is to high-light works that were essential to the cultural space of the time, Lily's inclusion is an important one—the impact of his grammar was profound, influencing a host of canonical authors, including John Lyly, Ben Jonson,

[66] Nancy A. Mace, "The History of the Grammar Patent from 1620 to 1800 and the Forms of Lily's Latin Grammar," *The Papers of the Bibliographical Society of America* 100, no. 2 (2006): 177, https://www.jstor.org/stable/24293669

[67] John Locke, "Some Thoughts Concerning Education" in *The Works of John Locke, Esq.; in Three Volumes* (London: Printed for John Churchill and Sam. Manship, 1714), 3:73.
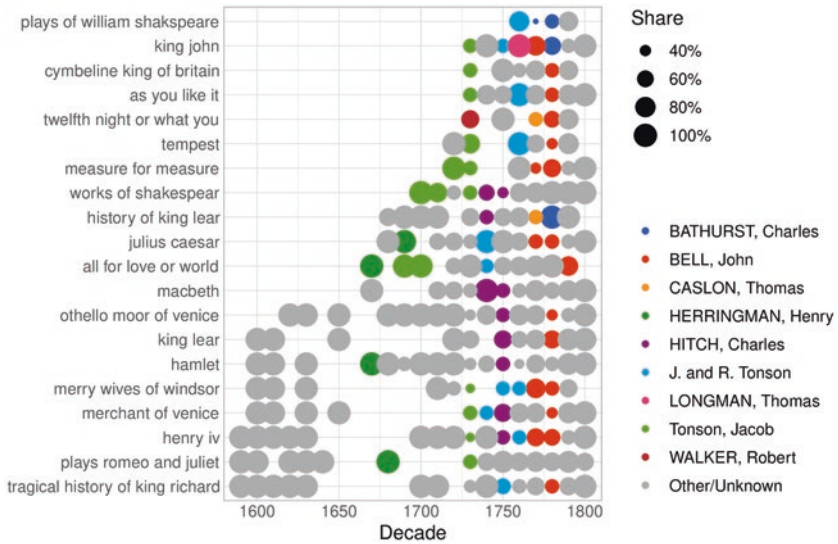
Fig. 3.8 Timeline of Shakespeare's publications included in canon. The point size indicates the share of the publisher with most prints of the indicated work (rows) per decade (columns)

Thomas Fuller, George Borrow, Charles Lamb, Edgar Allan Poe, and, of course, William Shakespeare (Fig. 3.8).[68]

Finally, it is worth turning to, perhaps, a more conservative canonical author who emerges from the above analysis. One benefit of a data-driven approach to the canon is that it enables us to shed new light on the printing and publishing history of specific authors. If we focus on Shakespeare's publications, for instance, a few points of interest emerge. For instance, while popular during his lifetime and continuously published throughout the seventeenth century, Shakespeare was printed less in the midseventeenth century. This was partially caused by censorship during the English Civil War and Interregnum years (1642–1659), which was not the most propitious time for printing and performing plays.[69] However, the

[68] R. Smith, "Lily, William (1468?–1522/3), Grammarian and Schoolmaster," *Oxford Dictionary of National Biography*, 2008, https://doi.org/10.1093/ref:odnb/16665

[69] Emma Depledge, "Shakespeare in the Civil War and Interregnum Years, 1642–1659," in *Shakespeare's Rise to Cultural Prominence: Politics, Print and Alteration, 1642–1700* (Cambridge: Cambridge University Press, 2018), 13–38.

data also shows that Shakespeare reemerged most strongly in the eighteenth century. This is directly related to the impact of several publishers who, by printing individual plays by Shakespeare, encouraged a growth in the popularity of his works and initiated a newly invented canon-making business (in which Pope, Dryden, Johnson, and others were an important part). Thus, we can see in our data the effect that the publishing efforts of Robert Walker, Jacob Tonson, John Bell, Edward Harding, and others, had on Shakespeare's canonization. It is, therefore, important that we also discuss these actors.

People: Printers and Publishers

Developments in the literary canon went hand in hand with developments in the book trade itself. This trade was transformed and driven by the economic expansion that occurred during the seventeenth and eighteenth centuries, when the scale and volume of all kinds of printings greatly expanded. By the end of the eighteenth century, this led to significant changes in the structure of the book trade, one of which was the increased specialization of the people involved.[70]

Within the existing data regarding book trade actors (printers, publishers, and booksellers), the data on booksellers is the most sporadic. Canonical works, however, provide better information on publishers than the dataset as a whole (Fig. 3.9). Of all the titles cataloged in the ESTC, 27% do not mention any book trade actors; in the subset of titles included in the data-driven canon, this number is 13%. Out of the three categories of book trade actors, publishers are the best represented in the data while booksellers are the least represented, with 85% of titles not mentioning them. The printer information is also missing from 64% of titles. Overall, the number of specialized roles linked to canonical publications increases toward the end of the period, as expected.

An observation relating to data quality should be made here. The ESTC joins two major catalogs, STC and Wing, with the dividing line between the original catalogs running at the year 1640. The pre-1640 period seems to have more carefully documented metadata but, at the same time, the number of entries in the database immediately shoots up after the divide. Part of this can be attributed to an increase in the publication activity, but

---

[70] See, for example, James Raven, *Publishing Business in Eighteenth-Century England* (Woodbridge: Boydell Press, 2014), 40.
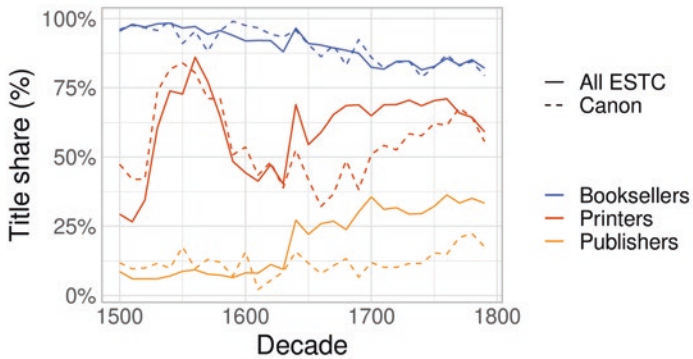
Fig. 3.9 Titles with missing book trade actor data. This figure charts out the coverage of the book trade actor data, as found in the catalog records (1500–1800)

this is also partially due to the better coverage of the published material in the latter half of the catalog. The number of unique book trade actors mentioned in the catalog follows a more consistent curve compared to the absolute number of titles, which can be taken as an indication that the individual actors involved have been detected reasonably well.

Looking at the publications linked to individual book trade actors also reveals a wide variety of profiles. A publisher's output can vary from a few to hundreds of titles, so it makes sense to create a rough categorization of the publishers based on this variable.

To explore this aspect of the book trade from a data-driven perspective, we divided the publishers into percentiles according to their publication output. The publishers were ranked yearly by their output, and as expected, the highest quantiles of the book trade dominate the data. What is immediately apparent is that the top 1–5% publishers account for over half of all publications (where the publishers are known), with no major variance over time (see Fig. 3.10 for a closer look at the first percentile's share). This matches historical expectations: the discussion around the 1774 Donaldson v. Beckett decision (and the propositions made to the Parliament by both established London booksellers and the "little low Stall Booksellers in Middle Row Holbourn"[71]) emphasizes the division

---

[71] Thomas F. Bonnell, *The Most Disreputable Trade: Publishing the Classics of English Poetry 1765–1810* (Oxford: Oxford University Press, 2008), 35–36.
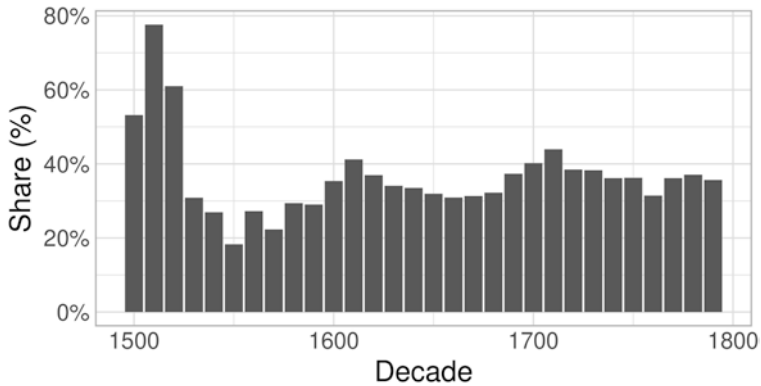
Fig. 3.10 Share of publications by the largest publishers (top 1%)

between the economic leaders of the book trade and the numerous, but less established, latter actors. The dividing line runs between the top London publishers, who owned many of the more lucrative and valuable copyrights, and members of the London trade who were not part of this select circle, as well as the less central publishers in Scotland and provincial England.[72]

The copyright battles of the eighteenth century were not over the right to print in general, but rather over the possession of intellectual property rights that had greater financial potential, such as canonical works. Here, too, the leading publishers had higher percentiles. Their share of the data-driven canon is proportionally higher, as illustrated in Fig. 3.11.

It has been claimed that the end of the eighteenth century was pivotal in changing the nature of the publishing business, with publishers starting to rely less on profits from "safe" reprints and taking on a more modern, entrepreneurial character.[73] The increase in mentions of booksellers in the publishers' statements could be taken as a confirmation of this claim, as it reflects the commercialization of the print industry in more than one way (see the overview of the actor data in Fig. 3.9). With increasing specialization within the trade, and a growth in the market for books, an increased need for advertising followed. Indeed, publisher statements often include advertising-like language, and provide practical details, such as bookshop

[72] Feather, *Publishing, Piracy and Politics*, 94.
[73] Ibid.

Fig. 3.11 Share of the canon by largest publishers (top 1%) for unique works, as derived from the work-field dataset
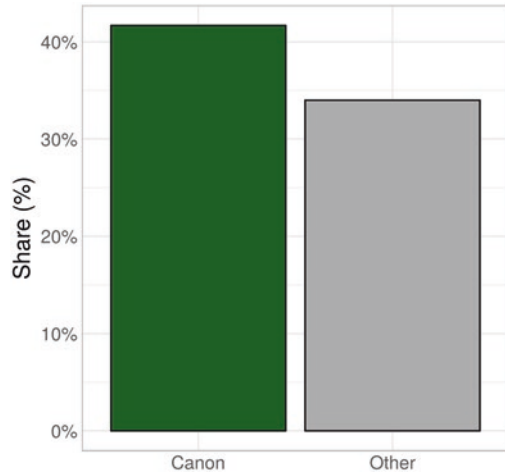
locations, lists of other works available, and so on. On the other hand, however, the distribution of canonical and non-canonical works does not significantly change during the eighteenth century. The canon stayed in relatively few hands, even after 1774. While there was a gradual increase in the "fluidity" of the publication business during the eighteenth century, that is, works tended to change hands more often, this trend is relatively temperate, with no sudden hike in the 1770s. Additionally, while the relative number of new works and reprints by new actors compared to reprints by established actors does increase (Fig. 3.12), the changes are less dramatic than it must have appeared to the worried copyright-owning printing elite of the time.[74] In fact, consumers' demand for books increased dramatically toward the end of the eighteenth century, which can be seen in the number of printings documented in the ESTC for that period.[75] Thus, even if the established elite was challenged by those seeking to make inroads into an expanding market, they were well positioned to defend their hegemony and exploit the new opportunities created by the public's increased demand for books.

[74] Bonnell, *The Most Disreputable Trade*, 169.

[75] See, for example, Raven, *Publishing Business in Eighteenth-Century England*, 214–18, and Karen O'Brien, "The History Market in Eighteenth-Century England," in *Books and Their Readers in Eighteenth-Century England: New Essays*, ed. Isabel Rivers (London: Continuum, 2001), esp. 123–30.
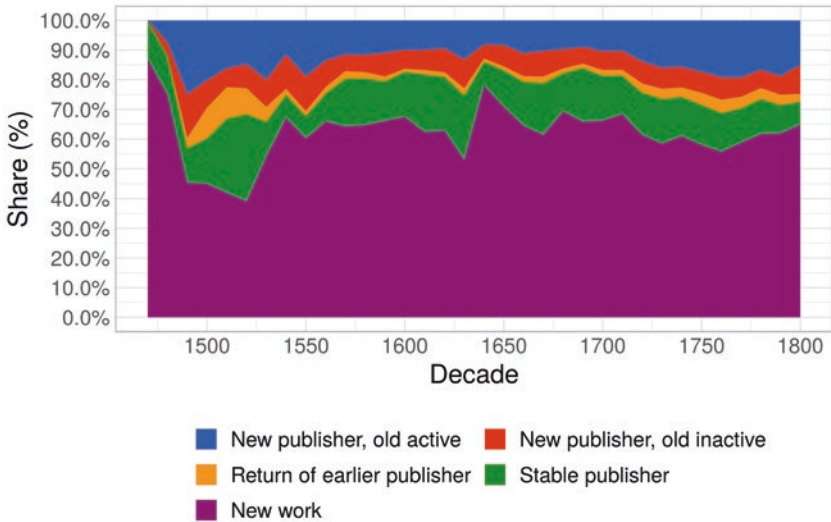
Fig. 3.12 Publishing and reprint patterns by publisher role in the printing sequence. This figure indicates changes in the reprint publishing patterns by different publishers over time and charts out the "fluidity" of the book trade. Each work had its publishers explored in chronological sequence to find out how often the publications changed hands. New publications ("New work") and new editions of the same works by the same publishers ("Stable publisher") were traced. Publications changing hands were traced both in the cases where the previous publisher disappeared from the book trade ("New publisher, old inactive") and where the publication changed hands, but the previous publisher stayed active ("New publisher, old active"). Cases of the publication returning to the hands of a previous owner are relatively rare ("Return of earlier publisher")

It is also known that, after 1774, other publishing monopolies, such as those covering almanacs, grammars, and law books, were under attack.[76] This highlights the importance of specialization within the trade and the significance of printed materials outside the subject-topics generally seen as canonical. When looking at the division between these categories, we can see that many publishers specialized in relatively limited subject-topics (Fig. 3.13). At the same time, new subject-topics were clearly part of a broader portfolio of publications. This means that we have different types

---

[76] Feather, *Publishing, Piracy and Politics*, 94–95.
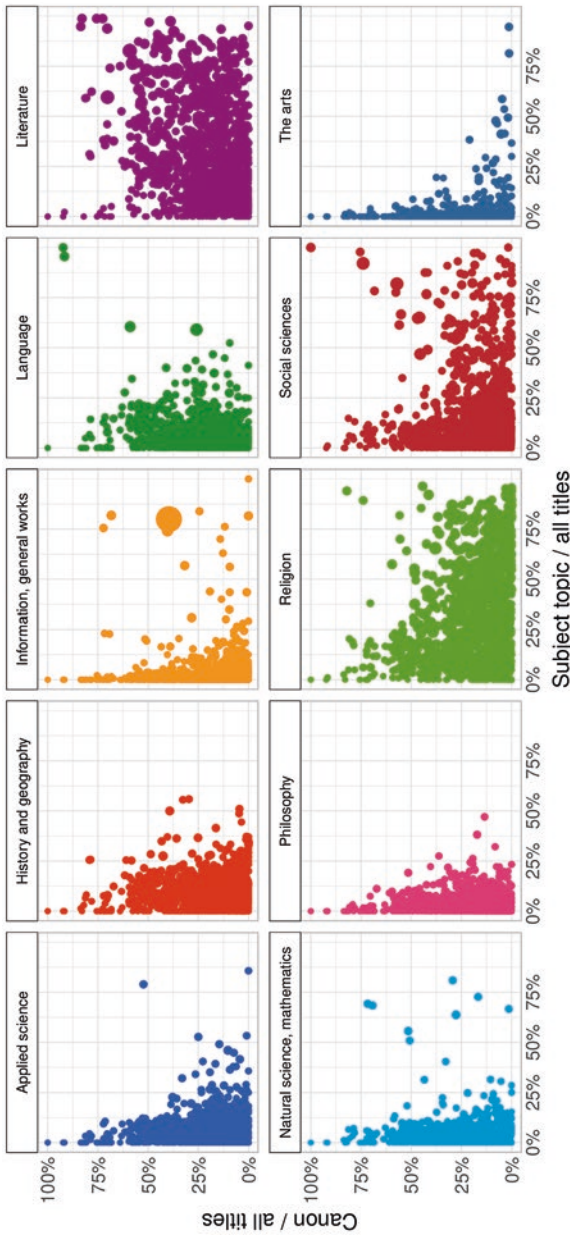
Fig. 3.13 Publisher subject-topic specialization and canon share. This figure illustrates the differences and similarities in the publishing landscapes of the identified subject-topics (individual scatterplots). Each dot represents an individual publisher, the horizontal axis indicates the publisher's topic specialization, the vertical axis indicates the portion of all publications by a publisher included in the canon, and the size of the dot indicates the publication volume of a publishers in a particular field. The large dot in the center right in "Information, general works" represents the Stationers' Company

of publishers throughout the early modern era: those who specialized and those who published in response to demand. Literature, religion, social sciences, and, to some extent, information and general works (almanacs and the like) appear to be fields where the market was large enough to accommodate specialization, and a limited group of well-established publishers were occasionally able to dominate these markets based on monopoly rights. A good example of this phenomenon is the single most voluminous publisher of the time, the Stationers' Company of London, which dominated the information and general works publishing market of eighteenth-century London.

Gender and the Book Trade

While it would be exciting to claim that our data-driven approach has allowed for a reassessment of the gender imbalance (amongst other imbalances) in the history of the canon, this is not the case. As seen in Fig. 3.14, most canonical authors were male. In fact, only 1 in every 32 authors is female, and men have on average 38 reprints per work compared with only 31 by women. Within the book trade as a whole the gender imbalance is not as great, although it is still significant: 1 in roughly 14 book trade actors in our data is female.

Fig. 3.14 Works by female authors in the data-driven canon per decade
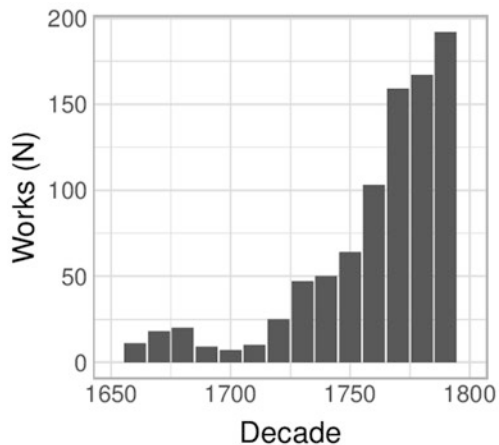
In total, 21 authors in the canon are female, and only three of them have multiple works included: Susanna Centlivre (1667?–1723) with three, and Elizabeth Singer Rowe (1674–1737) and Hannah More (1745–1833) with two each. The remaining authors have only one work each: Hannah Glasse (1708–1770), Lady Mary Wortley Montagu (1689–1762), Anne Fisher (1719?-1778), Mary Collyer (?-1763), Madame Marie-Catherine d'Aulnoy (1650 or 1651–1705), Elizabeth Raffald (1733–1781), Madame Jeanne-Marie Leprince de Beaumont (1711–1780), Hester Chapone (1727–1801), Catherine Talbot (1721–1770), Frances Brooke (1724?–1789), Eliza Smith (died cca.1732), Frances Burney (1752–1840), Madame de Gomez (1684–1770), Madame de Graffigny (1695–1758), Elizabeth Moxon (life years not available), Mary Brook (cca. 1726–1782), Anna Letitia Barbauld (1743–1825), and Elizabeth Grey, Countess of Kent (1581–1651). There is, however, a clear growth in the number of female authors and in the number of works by female authors that make up the canon over time. In fact, if we only look at the eighteenth century—which is relevant in this case as there is only one female in the canon prior to this period, Elizabeth Grey, Countess of Kent—the split is less severe, albeit still unequal: the disparity between reprints drops by nearly two works, with one female author for every 18 male authors.

The subjects covered by female authors include domestic economy, drama, education and manners, fiction, language, miscellaneous literature, and religion. Since the number of female-authored works is quite small, it is somewhat difficult to compare their subject coverage with that of the larger number of male-authored works. However, we do see that, in general, more women write fiction, and fewer women write educational and religious works.[77]

Places

The centrality of London (and of the Stationer's Company) to the early modern English book trade is well documented.[78] This reality is visible in

[77] These findings confirm and supplement the findings of William Underwood, David Bamman, and Sabrina Lee related to the nineteenth-century authorship. For details, see "The Transformation of Gender in English-Language Fiction," *Journal of Cultural Analytics* (2018): 1–25, https://doi.org/10.22148/16.019

[78] For more on London's place in the history of the book trade, see James Raven, *The Business of Books: Booksellers and the English Book Trade 1450–1850* (New Haven, CT: Yale
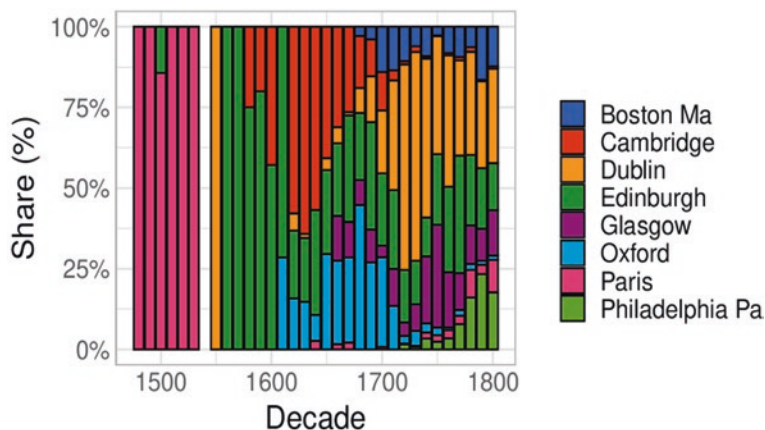
Fig. 3.15 Fraction of publications by place for the top publication places excluding London (1500–1800)

the publication records extracted from the ESTC as well, with London publishing more works by a factor of ten to its closest rival, Edinburgh. What is more, the non-London-based publication industry only begins to mature toward the end of the seventeenth century (and even then, works printed in London continue to dominate local markets).[79] As we see in Fig. 3.15, the early prints outside London are coming almost entirely from Paris, with Edinburgh, Cambridge, and Oxford dominating the seventeenth century, before Dublin and North American publishers begin to emerge in the eighteenth century.

When it comes to the canon, however, a different geographical picture emerges, as seen in Table 3.5. There are at least two findings revealed by this data: first, the importance of particular areas as producers of canonical works, and, second, the importance of particular areas as producers of reprints of popular works.

Regarding the former, London unsurprisingly dominates the print market. In fact, of the 30 cities which are recorded as sources of first editions

University Press, 2007). See also the works referenced in Hill, Vaara, Säily, Lahti, and Tolonen, "Reconstructing Intellectual Networks."

[79] James Green, "The British Book in North America," in Suarez and Turner, *The Cambridge History of the Book in Britain. Volume V*, 544–59.

Table 3.5 Top 10 printing locations in the whole ESTC and in the canon (1500–1800)

ESTC Records		Canon Total Prints	
Location	Count	Location	Count
London	318,708	London	21,273
Edinburgh	32,146	Dublin	2668
Dublin	26,273	Edinburgh	1778
Boston, MA	10,665	Glasgow	1071
Philadelphia, PA	10,282	Boston, MA	665
Oxford	7418	Philadelphia, PA	646
Glasgow	5358	Oxford	331
New York, NY	4716	Paris	289
Cambridge	4371	Cambridge	251
Newcastle	2726	New York, NY	232

of canonical works, London is responsible for 606, or roughly 84% of all titles. With regard to the movement of individual prints of canonical works, this means that more than ten times as many works originally printed in London were subsequently printed elsewhere than the other way around. Of course, London was the political, financial, and cultural capital of Britain, so these findings may not be surprising. However, what is surprising is the magnitude of this imbalance, with London dwarfing the rest of the canon. Edinburgh, the center of the Scottish Enlightenment, comes a distant second as the source for first editions that will become canonical, with 38 titles (Table 3.6).[80]

On the other hand, there are cities that are centers for reprinting editions of canonical works which were not originally printed there. Edinburgh fits this case, with the second highest number of reprints (288), followed by its Scottish compatriots Glasgow (252) and Aberdeen (57). The key player in reprints of canonical works, however, was Dublin. While Dublin did have a number of first editions (18), these are disproportionate to the number of subsequent editions it printed (427, the most of any city). This is almost certainly due to Dublin's privileged legal and cultural position:

[80] It should be noted, however, that the place of publication for 125 first editions is unknown due to the lack of detail in the ESTC about first editions, or due to multiple cities publishing editions of the same work in the same year. Also, our counts are by work and do not include reprints or editions beyond the first. Therefore, the highest number any one city can achieve is the total number of works counted in our canon, that is, 847.

Table 3.6 Locations for the first printed editions and for subsequent editions of canonical works (1500–1800)

Canon: First editions		Canon: Non-local subsequent editions	
Location	*Count*	*Location*	*Canon*
London	599	Dublin	427
Edinburgh	37	Edinburgh	288
Dublin	18	Glasgow	252
Oxford	12	Philadelphia, PA	148
Cambridge	8	Boston, MA	139
Pairs	5	London	95
Boston, MA	4	New York, NY	74
Philadelphia, PA	3	Belfast	71
Amsterdam	2	Oxford	62
Manchester	2	Aberdeen	57

being beyond the reach of the British law and the Stationers' grasp, and having a large English-speaking population, meant that Dublin's printers were in a privileged position. And while there was certainly a local market for these works, it is also clear that Dublin was not the only intended market for these works, and many were exported to Britain and beyond. Thus, while we know that Dublin was an importer of canonical works first printed elsewhere, it was also an important exporter of these works.

A similar, yet smaller, pattern is also seen in the colonies, with Philadelphia, Boston, and New York reprinting numerous non-local works. Overall, the movement of works between Europe and the colonies remains largely unidirectional, according to the ESTC records (Fig. 3.16).[81] On the whole, however, there is a remarkable movement of various editions of these works among various locations. While books obviously traveled as any other material object would, the works themselves as less tangible things were recreated in various locations throughout the world, being reprinted by local book trade actors for commercial and intellectual profit.

[81] Base Mercator projection map used for visualization created by Geordie Bosanko and shared under the CC BY-SA 3.0 license at https://commons.wikimedia.org/wiki/File:Mercator_Projection.svg. This base map was further edited by Adrienne Hawkes.

Fig. 3.16 Movement of canonical editions from original print location

Population must also be taken into consideration when looking at publication counts. When one considers the size of London compared to that of other cities at the time, its dominance may not be surprising. However, size does not appear to be the key contributing factor to the number of prints, relative to population size, that emerge from a city.

As seen in Fig. 3.17, smaller cities were capable of producing many more works than larger cities in both absolute and relative terms. The reasons for this vary. University cities like Oxford and Cambridge printed for specific niche markets, while colonial cities developed their own markets, which were less reliant on imports.[82] Dublin and Glasgow were also important reprint centers, although we can see a decline in their print output compared to London following the liberalization of the market in 1774. While we have already touched upon Dublin's relationship to

[82] Bristol's inclusion in Fig. 3.17 can be attributed almost exclusively to John Wesley, a prodigious writer who had an important impact on both British and colonial print markets in the early modern era. Interestingly, the publication place of over 300 of his works (more than 20%) was Bristol (comparatively, only Swift had more works published outside of London, i.e., in his hometown, Dublin). Moreover, when looking at Bristol's publishing record, Wesley accounts for almost 20% of all publications from Bristol in the ESTC. When John's brother, Charles, is included, this number increases to over 25% of all works from Bristol in the ESTC.
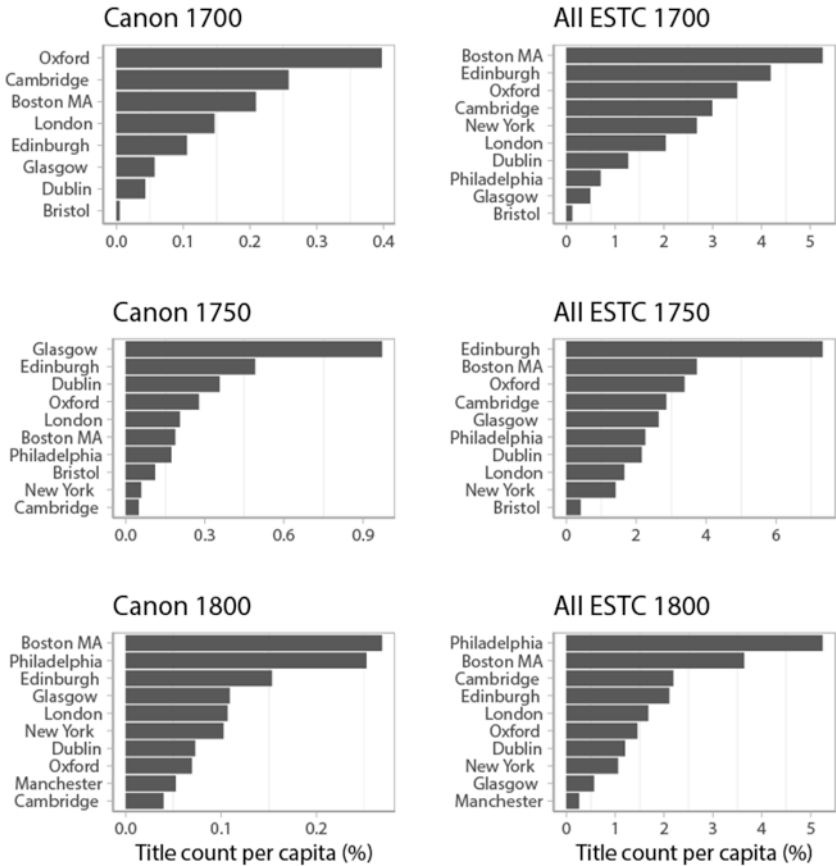
Fig. 3.17   Number of prints per capita (1700–1800)

reprints, Glasgow is also worth noting given that its print history is argu-
ably tied to the production of editions of "classics."[83]

Glasgow's dominance in canonical printing (Figs. 3.17 and 3.18) can
be tied to a specific printing house run by the brothers Robert and Andrew
Foulis.[84] Initially elected as printers to the University of Glasgow in 1743

---

[83] Thomas F. Bonnell, "The Elzevirs of Glasgow: Robert and Andrew Foulis," in *The Most Disreputable Trade*, 39–67.

[84] For more on the impact of the Foulis brothers on the eighteenth-century Scottish intellectual life, see Steinberg, *Five Hundred Years of Printing*, 139–41.
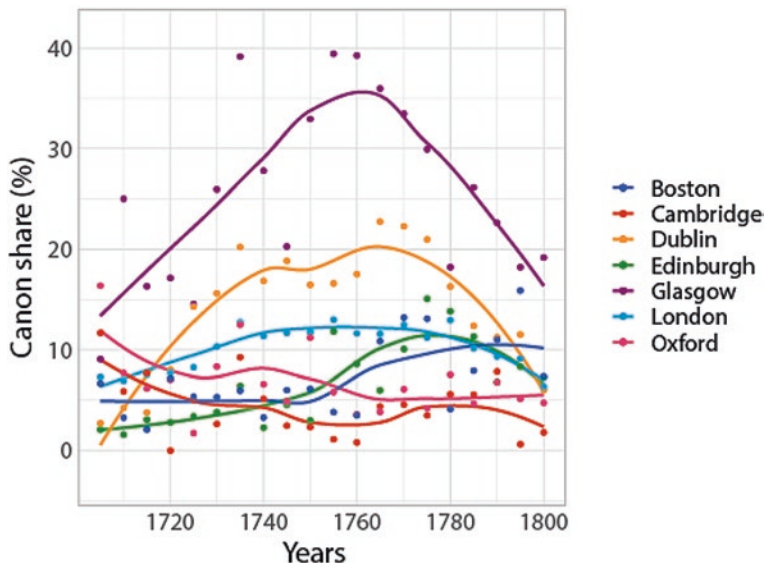
Fig. 3.18 The fraction of canonical editions compared to all editions per city (1700–1800)

and charged with printing works by ancient authors, by the 1750s the Foulis brothers had turned to printing Elzeviresque editions of English classics by Milton and Gray, before engaging in mass market reprints of dubious legality. With the deaths of Andrew in 1775 and Robert in 1776, the period of Glaswegian dominance ends, although Robert's son, also Andrew, was able to revive the business to some extent from the 1780s on, with a renewed focus on exporting books to the American market.

Importantly, their genius was found not only in whom they printed but also in how they printed. While their editions of the ancients were often expensive folios or quartos, their reprints of contemporary English poets were duodecimos: they were cheaper to print and thus more affordable to a broader audience. As Thomas F. Bonnell notes, "they had crossed a divide from an old world of monumental scholarly and typographical ventures devoted to ancient Greek and Latin texts into a new world of selling multi-volume collections of modern vernacular classics to a larger and

more diverse readership."[85] It is this transformation of the print trade which we turn to next.

Materiality

There is one further aspect of print that is worth taking into consideration when thinking about the canon: materiality, or the composition of a printed work, such as page size, format, number of pages, and print area.[86] This is, again, an aspect of print which is often overlooked when one engages with historic works quantitatively.[87] However, the material attributes of a printed work have meaning to a reader, as made explicit by Joseph Addison in issue 529 of *The Spectator*: "I have observed that the Author of a Folio, in all Companies and Conversations, sets himself above the Author of a Quarto; the Author of a Quarto above the Author of an Octavo; and so on, by a gradual Descent and Subordination, to an Author in Twenty Fours."[88] In other words, within the materiality of a work, there were ingrained literary and social signifiers that contemporaries would have been familiar with.

When evaluating the most dominant book formats attached to subject-topics over time, the significance of the smaller book format emerges. In particular, we notice the growth of the octavo and duodecimo formats (Fig. 3.19), which were more portable, easily fitting in one's pocket, and thus more easily perused or read at different occasions. While the trend can be seen across all subject-topics, it is particularly visible in the case of

---

[85] Bonnell, *The Most Disreputable Trade*, 62.

[86] Print area quantifies the paper consumption in sheets for a unique copy of a document; the combined print area across different documents in a given time period can be used to quantify the breadth of the printing activity.

[87] For more work on materiality, see Tolonen, Lahti, Roivainen, and Marjanen, "A Quantitative Approach to Book-Printing in Sweden and Finland, 1640–1828," 57–78; Lahti, Ilomäki, and Tolonen, "A Quantitative Study of History in the English Short-Title Catalogue (ESTC), 1470–1800," 87–116; Lahti, Marjanen, Roivainen, and Tolonen, "Bibliographic Data Science and the History of the Book (c. 1500–1800)," 10–23; and Eetu Mäkelä, Mikko Tolonen, Jani Marjanen, Antti Kanner, Ville Vaara, and Leo Lahti, "Exploring the Material Development of Newspapers," *TwinTalks at DHN 2018 Understanding Collaboration in Digital Humanities*, https://cst.dk/DHN2019Pro/TwinTalks WorkshopProceedings.pdf

[88] The Spectator No. 529 Thursday, November 6, 1712. *Oxford Scholarly Editions Online*, 2014, https://www.oxfordscholarlyeditions.com/view/10.1093/actrade/9780198186137. book.1/actrade-9780198186137-div1-104
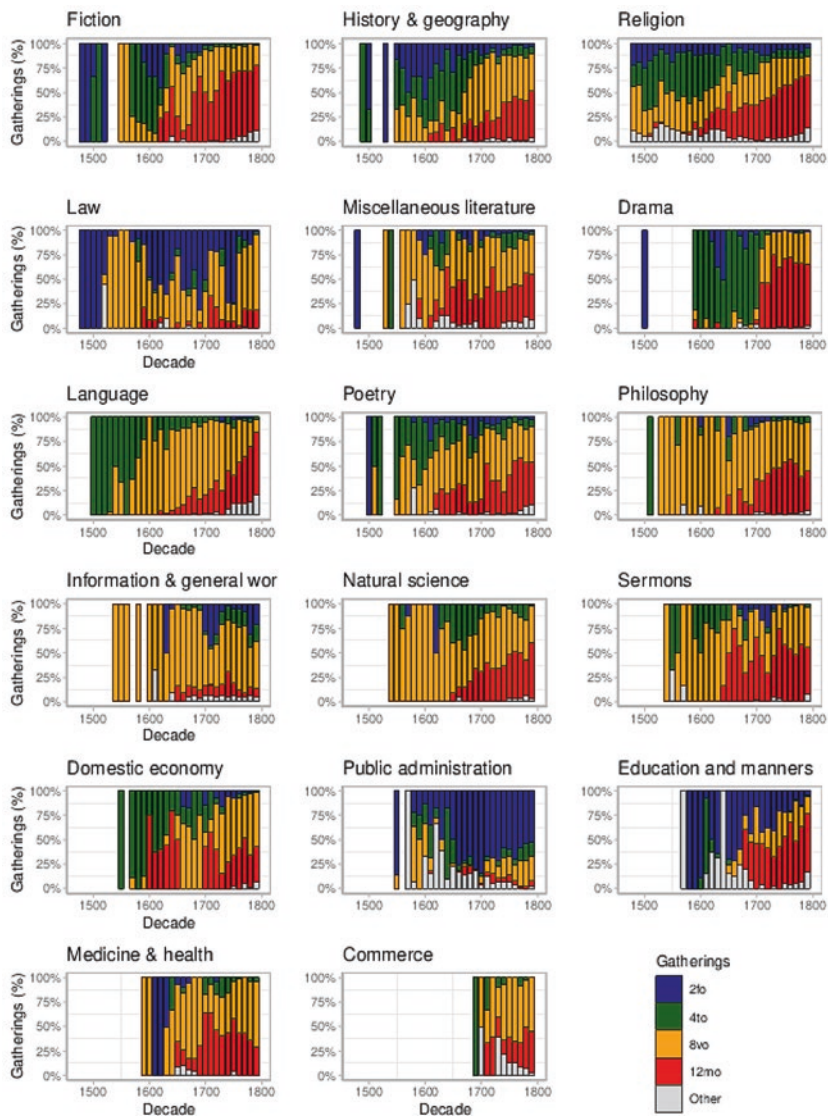
Fig. 3.19   Dominant book formats for the most frequent subject-topics
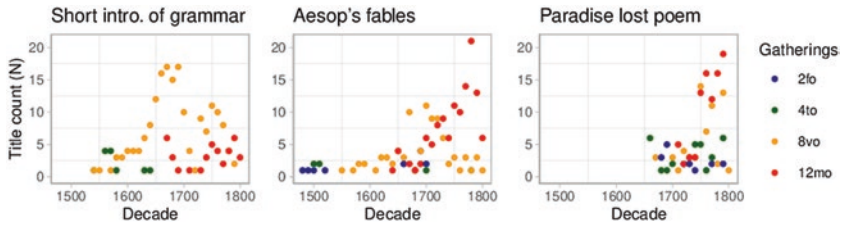
Fig. 3.20 The distribution of common book formats for selected canon-works over time

literary and philosophical works. Legal and public administration books, on the other hand, were slower to change, remaining in folio for much longer. Interestingly, we have also noticed a regional variation with respect to preferred formats, with duodecimo being the most popular format in the New World.

When looking at the changing materiality of the canon with respect to some of the most popular works published between 1500 and 1800, we notice that early publishing mixed folio, quarto, and octavo formats. This trend continues, in some cases, until the end of the eighteenth century (see, for instance, *Paradise Lost*, in Fig. 3.20), but with time the octavo and duodecimo formats become dominant (see *Aesop's Fables* and *Short Introduction to Grammar* in Fig. 3.20). The choice of format, therefore, is tied to complex relations among economic viability, perceived importance, and pragmatics. A folio edition of Milton, for example, was a worthy and desirable endeavor in the late eighteenth century, while a grammar was much less likely to be imbued with the same subjective value and was, therefore, more convenient to own in a smaller format. Thus, this type of analysis allows us not only to touch upon the realities of printing practices but also to gain insight into the changing preferences of the public with respect to materiality and early modern reading habits.[89] At the same time, when looking at the overall paper consumption for different formats of books included in the canon, we can see that octavo and duodecimo

---

[89] See, for example, Reinhard Wittmann, "Was There a Reading Revolution at the End of the Eighteenth Century?" in *A History of Reading in the West*, ed. Guglielmo Cavallo and Roger Chartier (Amherst, MA: University of Massachusetts Press, 1999), 284–312.
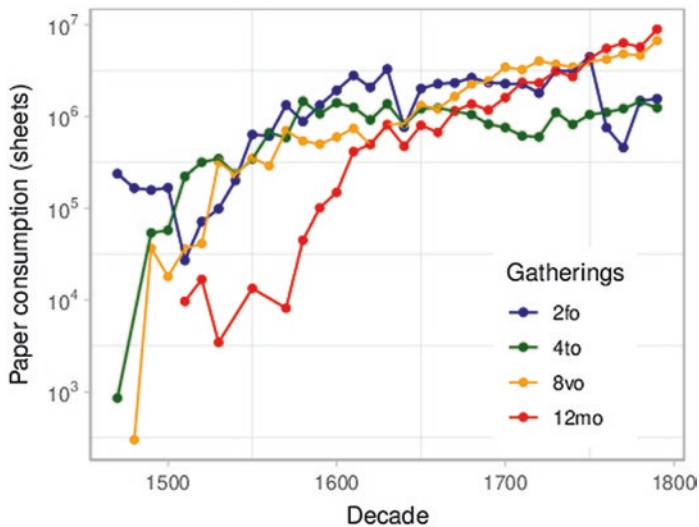
Fig. 3.21 Estimated paper consumption for different formats over time for books included in the canon

formats started to dominate the book printing market only in the second half of the eighteenth century (Fig. 3.21).

CONCLUSION

The goal of this study has been to extract from the ESTC a data-driven canon which could be used to demonstrate that quantitative investigations of this type are valuable for historical research. While quantitative analyses of the history of the book trade exist, there has been no attempt so far to engage with the complex process of canon formation at such a large scale.[90]

To this effect, we have constructed a method for extracting a list of "canonical" works from the ESTC based on three publication features: count, frequency, and longevity. We have thus generated a data-driven list of canonical works that considers subject-topics, top-works, authors and publishers, publication place, and materiality from a historical perspective.

[90] It should be noted that this is an ongoing process; we continue working on further harmonizing the data.

While we believe this quantitative approach is in itself a methodological contribution worth reporting, it also allows us to make a number of historical claims worth studying further.

At the same time, it is important to recognize the limitations of this type of analysis. Data reliability, representativeness, and completeness will improve over time, and this will influence all quantitative estimates derived from the data. Algorithmic questions, such as the exact definition of the canon index or genre classification, and choices made in model parameters, such as investigated time window, will also affect this analysis. From our perspective, however, making such interpretations explicit allows one to evaluate these choices and propose alternative solutions. The reproducibility of the analysis will then allow us (and others) to explore how sensitive the qualitative conclusions are to different analytical choices. Here, however, we have primarily focused on broad historical patterns and trends that are expected to remain stable to variations in data and algorithmic details.

When examining the early modern English canon from this data-driven perspective, it becomes obvious that an epistemological shift takes place during the late seventeenth century-early eighteenth century, when religious works lose their dominant position within the canon and are increasingly replaced by literary works (Figs. 3.4, 3.5, 3.7, and 3.8). Although literature in all its forms was historically an important part of the canon, changes in its production and consumption allowed for its growth in the eighteenth century.[91]

Additionally, this analysis allows us to highlight the essential role played by the publisher in the process of canon formation, besides that of the literary critic. In particular, the role of the elite, London-based publishers (Fig. 3.11) and that of the arguably more dubious printers operating outside of London (Table 3.5, Figs. 3.17 and 3.18) become evident. Overall, we can now visualize the lengthy and arduous process of a work becoming canonical. While the total number of publications grew exponentially during the latter half of the eighteenth century, the distribution of canonical works remained relatively stable in comparison (Fig. 3.1), which indicates that the works that are most often reprinted over long stretches of time are comparatively few. This is perhaps a finding worth reflecting on further:

---

[91] Other historical moments, Donaldson v. Becket in particular, are also visible in our data (Figs. 3.17 and 3.18). However, as shown in the canon comparisons, the impact of such moments on the ESTC from a data-driven canon perspective should not be overestimated.

this description of large-scale cultural production and competition within the literary market, which directs the canonization process, may allow scholars of the period to extrapolate further and use these statistics to develop prediction models for an author's or a work's likelihood of becoming canonical.

The broader claim of this chapter is that the development of the print market as a cultural producer has driven the changes we are able to witness in the ESTC when studied in a data-driven manner. This builds on previous work by Bourdieu, Anderson, and Habermas, who tied print capitalism to historical, cultural, political, and social changes.[92] Our contribution is to apply quantitative methods to demonstrate the accuracy of these qualitatively-grounded studies in a manner which has not been attempted before.

As this analysis demonstrates, what Wendell V. Harris wrote more than thirty years ago is truer today than ever: "The 'canon question' … proves much more complex than contemporary ideological criticism admits."[93] While large-scale cultural production is certainly a key factor in the canon-making process, were it to be taken as *the only* factor, we would dismiss numerous individual voices and would offer yet another version of the revisionist approach to canon formation. There is no such thing as an "absolute" canon, only different takes on it. While it is inevitable that different works matter in different ways, our main concern in this study is with the impact of print culture on canon formation. By considering these recorded works and their historical availability over extended periods of time, we hope to offer a more nuanced understanding not only of the history of the book trade but also of the cultural context from which it emerged.

[92] According to Bourdieu, there are two important factors that make a difference in the print market: the restricted production of literature for like-minded audiences, and the large-scale cultural production. We have limited our study to the latter category because of gaps in the information related to particular works. For example, in Shakespeare's case, we cannot judge whether his late eighteenth-century success is due to earlier restricted production. We may only note that Shakespeare enters large scale cultural production after a gap in publishing his works in the seventeenth century, as seen in Fig. 3.8.

[93] Harris, "Canonicity," 115.

Bibliography

Altieri, Charles. 1983. An Idea and Ideal of a Literary Canon. *Critical Inquiry* 10 (1): 37–60. https://www.jstor.org/stable/1343405.

Anderson, Benedict. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. New York: Verso.

Barnard, John, D.F. McKenzie, and Maureen Bell, eds. 2002. *The Cambridge History of the Book in Britain. Volume IV: 1557–1695*. Cambridge: Cambridge University Press.

Bell, Maureen, and John Barnard. 1992. Provisional Count of STC Titles, 1475–1640. *Publishing History* 31: 47–66.

Bloom, Harold. 1994. *The Western Canon: The Books and School of the Ages*. Harcourt: Brace.

Bonnell, Thomas F. 2008. *The Most Disreputable Trade: Publishing the Classics of English Poetry 1765–1810*. Oxford: Oxford University Press.

Bourdieu, Pierre. 1983. The Field of Cultural Production, or the Economic World Reversed. *Poetics* 12 (4–5): 331–356. https://doi.org/10.1016/0304-422X(83)90012-8.

———. 1985. The Market of Symbolic Goods. *Poetics* 14 (1–2): 13–44. https://doi.org/10.1016/0304-422X(85)90003-8.

Bourdieu, Pierre, and Jean-Claude Passeron. 1977. *Reproduction in Education, Society and Culture*. Trans. Richard Nice. London: Sage.

Bradford, Richard. 2015. *Is Shakespeare Any Good? And Other Questions on How to Evaluate Literature*. Malden: Wiley-Blackwell.

Bruni, Flavia, and Andrew Pettegree, eds. 2016. *Lost Books: Reconstructing the Print World of Pre-Industrial Europe*. Leiden: Brill.

Carroll, Noel. 1984. Hume's Standard of Taste. *The Journal of Aesthetics and Art Criticism* 43 (2): 181–194. https://doi.org/10.2307/429992.

Collins, Irene. 1998. The Rev. Henry Tilney, Rector of Woodston. *Persuasions* 20: 154–164.

Collinson, Patrick, Arnold Hunt, and Alexandra Walsham. Religious Publishing in England 1557–1640. In *The Cambridge History of the Book in Britain. Volume IV*, ed. Barnard, McKenzie, and Bell, 29–66.

Curtius, Ernst Robert. 2013. *European Literature and the Latin Middle Ages*. Princeton: Princeton University Press.

Depledge, Emma. 2018. *Shakespeare's Rise to Cultural Prominence: Politics, Print and Alteration, 1642–1700*. Cambridge: Cambridge University Press.

Disraeli, Isaac. 1823. *Curiosities of Literature*. Vol. 4. 7th ed. London: John Murray.

Eagleton, Terry. 1985. The Subject of Literature. *Cultural Critique* 2: 95–104. https://doi.org/10.2307/1354202.

Eisenstein, Elizabeth. 1979. *The Printing Press as an Agent of Change: Communications and Cultural Transformations in Early Modern Europe*. Cambridge: Cambridge University Press.

Fairer, David. 2000. Historical Criticism and the English Canon: A Spenserian Dispute in the 1750s. *Eighteenth-Century Life* 24 (2): 43–64. https://doi.org/10.1215/00982601-24-2-43.

Feather, John. 1995. *Publishing, Piracy and Politics: An Historical Study of Copyright in Britain*. London: MansFigell.

Febvre, Lucien, and Henri-Jean Martin. 1976. *The Coming of the Book: The Impact of Printing, 1450–1800*. New York: Verso.

Fielder, Leslie A., and Houston A. Baker. 1979. *English Literature: Opening Up the Canon*. Baltimore: Johns Hopkins University Press.

Fowler, Alastair. 1979. Genre and the Literary Canon. *New Literary History* 11 (1): 97–119. https://doi.org/10.2307/468873.

Gants, David L. 2002. A Quantitative Analysis of the London Book Trade, 1614–1618. *Studies in Bibliography* 55 (1): 185–213. https://doi.org/10.1353/sib.2005.0004.

Gorak, Jan, ed. 2001. *Canon vs. Culture: Reflections on the Current Debate*. New York: Garland.

———. 2013. *The Making of the Modern Canon: Genesis and Crisis of a Literary Idea*. London: Bloomsbury.

Green, Ian, and Kate Peters. Religious Publishing in England c. 1640–1695. In *The Cambridge History of the Book in Britain. Volume IV*, ed. Barnard, McKenzie, and Bell, 67–94.

Green, James. The British Book in North America. In *The Cambridge History of the Book in Britain. Volume V*, ed. Suarez and Turner, 544–559.

Guillory, John. 1993. *Cultural Capital: The Problem of Literary Canon Formation*. Chicago: University of Chicago Press.

Habermas, Jürgen. 2003. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Cambridge: Polity Press.

Halasz, Alexandra. 1997. *The Marketplace of Print: Pamphlets and the Public Sphere in Early Modern England*. Cambridge: Cambridge University Press.

Harris, Wendell V. 1991. Canonicity. *PMLA* 106 (1): 110–121. https://doi.org/10.2307/462827.

Hill, Mark J. 2016. Invisible Interpretations: Reflections on the Digital Humanities and Intellectual History. *Global Intellectual History* 1 (2): 130–150. https://doi.org/10.1080/23801883.2017.1304162.

Hill, Mark J., Ville Vaara, Tanja Säily, Leo Lahti, and Mikko Tolonen. 2019. Reconstructing Intellectual Networks: From the ESTC's Bibliographic Metadata to Historical Material. In *Proceedings of the Digital Humanities in the Nordic Countries 4th Conference: Copenhagen, Denmark, March 5–8, 2019*, ed. Constanza Navarretta, Manex Agirrezabal, and Bente Maegaard, 201–219. Aachen: CEUR-WS.org.

Hoxby, Blair. 2015. *What Was Tragedy? Theory and the Early Modern Canon*. Oxford: Oxford University Press.

Hume, David. 1987. Of the Standard of Taste. In *Essays Moral, Political, Literary*, ed. Eugene F. Miller, 226–250. Indianapolis: Liberty Fund.

Karian, Stephen. 2011. The Limitations and Possibilities of the ESTC. In *The Age of Johnson: A Scholarly Annual*, ed. Jack Lynch, vol. 21, 283–297. New York: AMS Press.

Kramnick, Jonathan Brody. 1999. *Making the English Canon: Print-Capitalism and the Cultural Past, 1700–1770*. Cambridge: Cambridge University Press.

Lahti, Leo, Niko Ilomäki, and Mikko Tolonen. 2015. A Quantitative Study of History in the English Short-Title Catalogue (ESTC), 1470–1800. *LIBER Quarterly* 25 (2): 87–116. https://doi.org/10.18352/lq.10112.

Lahti, Leo, Jani Marjanen, Hege Roivainen, and Mikko Tolonen. 2018. Bibliographic Data Science and the History of the Book (c. 1500–1800). *Cataloging & Classification Quarterly* 57 (1): 5–23. https://doi.org/10.1080/01639374.2018.1543747.

Lahti, Leo, Ville Vaara, Jani Marjanen, and Mikko Tolonen. 2019. Best Practices in Bibliographic Data Science. In *Proceedings of the Research Data in the Humanities 2019 Conference: Data, Methods, and Tools*, ed. Jarmo Harri Jantunen, Sisko Bruni, Kunnas Niina, Santeri Palviainen, and Katja Västi, 57–65. Oulu: Faculty of Humanities Linnanmaa.

Levin, Harry. 1981. Core, Canon, Curriculum. *College English* 43 (4): 352–362. https://doi.org/10.2307/377120.

Lopez, Jeremy. 2014. *Constructing the Canon of Early Modern Drama*. Cambridge: Cambridge University Press.

Low, Gail, and Marion Wynne-Davies, eds. 2006. *A Black British Canon?* Hampshire: Palgrave Macmillan.

Mace, Nancy A. 2006. The History of the Grammar Patent from 1620 to 1800 and the Forms of Lily's Latin Grammar. *The Papers of the Bibliographical Society of America* 100 (2): 177–225. https://www.jstor.org/stable/24293669.

Mäkelä, Eetu, Mikko Tolonen, Jani Marjanen, Antti Kanner, Ville Vaara, and Leo Lahti. Exploring the Material Development of Newspapers. *TwinTalks at DHN 2018 - Understanding Collaboration in Digital Humanities*. https://cst.dk/DHN2019Pro/TwinTalksWorkshopProceedings.pdf

Mukherjee, Ankhi. 2013. *What Is a Classic? Postcolonial Rewriting and Invention of the Canon*. Stanford: Stanford University Press.

Munch, Thomas. 2019. *Conflict and Enlightenment: Print and Political Culture in Europe, 1635–1795*. Cambridge: Cambridge University Press.

O'Brien, Karen. 2001. The History Market in Eighteenth-Century England. In *Books and Their Readers in Eighteenth-Century England: New Essays*, ed. Isabel Rivers, 105–134. London: Continuum.

Patey, Douglas Lane. 1988. The Eighteenth Century Invents the Canon. *Modern Language Studies* 18 (1): 17–37. https://doi.org/10.2307/3194698.

Pettegree, Andrew, ed. 2017. *Broadsheets: Single-Sheet Publishing in the First Age of Print*. Leiden: Brill.

Proudfoot, Richard. 2001. *Shakespeare: Text, Stage & Canon*. London: Arden Shakespeare.

Raven, James. 2007. *The Business of Books: Booksellers and the English Book Trade 1450–1850*. New Haven: Yale University Press.

———. 2014. *Publishing Business in Eighteenth-Century England*. Woodbridge: Boydell Press.

Robinson, Lillian. 1997. *The Canon's Mouth: Dispatches from the Culture Wars*. Bloomington: Indiana University Press.

Robinson, Peter. 2008. Introduction to *Textual Scholarship and the Canon*, ed. Hans Gabler, Peter Robinson Walter, and Paulius Subačius, 1–8. Leiden: Brill.

Ross, Trevor Thornton. 1998. *The Making of the English Literary Canon: From the Middle Ages to the Late Eighteenth Century*. Montreal: McGill-Queens University Press.

Rounce, Adam. 2015. Authorship in the Eighteenth Century. *Oxford Handbooks Online*. https://doi.org/10.1093/oxfordhb/9780199935338.013.38.

Rousseau, Jean Jacques. *Politics and the Arts: Letter to M. D'Alembert on the Theatre*. Trans. Allan Bloom. Ithaca: Cornell University Press, 1968.

Salzman, Paul, ed. 2010. *Expanding the Canon of Early Modern Women's Writing*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Sher, Richard. 2007. *The Enlightenment and the Book. Scottish Authors and Their Publishers in Eighteenth-Century Britain, Ireland, and America*. Chicago: University of Chicago Press.

Skinner, Quentin. 2002. *Visions of Politics. Volume I: Regarding Method*. Cambridge: Cambridge University Press.

Smith, R. 2008. Lily, William (1468?–1522/3), Grammarian and Schoolmaster. *Oxford Dictionary of National Biography*. https://doi.org/10.1093/ref:odnb/16665.

Spencer, Jane. 2005. *Literary Relations: Kinship and the Canon 1660–1830*. Oxford: Oxford University Press.

Stallybrass, Peter. 2004. The Library and Material Texts. *PMLA* 119 (5): 1347–1352. https://doi.org/10.1632/003081204X17914.

Steinberg, Sigfrid H. 1955. *Five Hundred Years of Printing*. London: Penguin.

Stewart, Mary Beth. 1938. William Lily's Contribution to Classical Study. *The Classical Journal* 33 (4): 217–225. https://www.jstor.org/stable/3291195.

Suarez, Michael F. 2015. Book History from Descriptive Bibliographies. In *The Cambridge Companion to the History of the Book*, ed. Leslie Howsam, 199–219. Cambridge: Cambridge University Press.

Suarez, Michael F. Towards a Bibliometric Analysis of the Surviving Record, 1701–1800. In Suarez and Turner, *The Cambridge History of the Book in Britain. Volume V*, 37–65.

Suarez, Michael F., and M.L. Turner. 2009. *The Cambridge History of the Book in Britain. Volume V: 1695–1830*. Cambridge: Cambridge University Press.

Terry, Richard. 1997. Literature, Aesthetics, and Canonicity in the Eighteenth Century. *Eighteenth-Century Life* 21 (1): 80–101. https://www.muse.jhu.edu/article/10407.

The Works of John Locke, Esq.; in Three volumes. 1714. London: Printed for John Churchill and Sam. Manship.

Thomson, Alex. 2014. What Is Literature? In *The Edinburgh Introduction to Studying English Literature*, ed. Dermot Cavanagh, Alan Gillis, Michelle Keown, James Loxley, and Randall Stevenson, 3–16. Edinburgh: Edinburgh University Press.

Tolonen, Mikko, Leo Lahti, Hege Roivainen, and Jani Marjanen. 2019. A Quantitative Approach to Book-Printing in Sweden and Finland, 1640–1828. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52 (1): 57–78. https://doi.org/10.1080/01615440.2018.1526657.

Underwood, William, David Bamman, and Sabrina Lee. 2018. The Transformation of Gender in English-Language Fiction. *Journal of Cultural Analytics*: 1–25. https://doi.org/10.22148/16.019.

van Oostrom, Frits. 2007. *A Key to Dutch History: The Cultural Canon of the Netherlands.* Amsterdam: Amsterdam University Press.

Weedon, Alex. 2009. The Uses of Quantification. In *A Companion to the History of the Book*, ed. Simon Eliot and Jonathan Rose, 33–49. Malden: Wiley Blackwell.

Werner, Sarah. 2019. *Studying Early Printed Books, 1450–1800: A Practical Guide.* Malden: Wiley Blackwell.

Wittmann, Reinhard. 1999. Was There a Reading Revolution at the End of the Eighteenth Century? In *A History of Reading in the West*, ed. Guglielmo Cavallo and Roger Chartier, 284–312. Amherst: University of Massachusetts Press.