

University of Pennsylvania

From the SelectedWorks of Zachary Lesser

October 2008

Early Modern Digital Scholarship & DEEP: Database of Early English Playbooks

Contact
Author

Start Your Own
SelectedWorks

Notify Me
of New Work



Available at: <http://works.bepress.com/zacharylesser/17>



Early Modern Digital Scholarship and *DEEP: Database of Early English Playbooks*

Alan B. Farmer*

The Ohio State University

Zachary Lesser**

University of Pennsylvania

Abstract

This paper discusses recent trends in digital resources for early modern literary studies, as well as the implications of these resources for research and scholarship. In addition to comparing the use by scholars of print reference works and online databases, the essay analyzes the recent shift from ‘first-generation’ digital resources, such as the *English Short Title Catalogue* (ESTC) and *Early English Books Online* (EEBO), to newer ‘second-generation’ resources like *DEEP: Database of Early English Playbooks*. Rather than strive for comprehensive coverage of early modern print culture, as ESTC and EEBO do, these ‘second-generation’ sites typically aim for in-depth coverage of a particular kind of text or document. *DEEP*, for example, is a searchable database of all extant plays printed in England to 1660, while the *English Broadside Ballad Archive* focuses on sixteenth- and seventeenth-century ballads. This shift in emphasis – from comprehensiveness to specialized subject matter – has resulted in, and been driven by, changes in thinking about the fundamental architecture of the databases, their searchability, and their analytical and editorial principles, all of which have significant ramifications for the type of research they enable.

It has become a truism that early modern literary studies has been revolutionized by the advent of online databases such as the *English Short Title Catalogue* (ESTC) and *Early English Books Online* (EEBO). As George Justice pointed out in an earlier volume of *Literature Compass*, for example, the appearance of these digital resources

has meant that the critical lip service given to expansion of the ‘canon’ has become a practical reality for scholars, who can survey titles, authors, dates, imprints, editions, and other information from the comfort of their desks. (1)

Research that previously could be done only by consulting reference works like A. W. Pollard and G. R. Redgrave’s *Short-Title Catalogue*, typically in the non-circulating reference sections of university libraries, could now be

accomplished by connecting to ESTC from any computer with online access. Similarly, the availability of digital images on EEBO radically curtailed that time-honored scholarly ritual of scrolling through microfilm reels and, perhaps even more significantly, meant that visiting rare book libraries was no longer the only, or even the most efficient, way to conduct certain kinds of research. For scholars who simply needed to track down a quotation in a 1629 sermon, or who wanted to read a series of sermons from that year, EEBO offered unparalleled speed and access, at least for those lucky enough to work at institutions that subscribe to it.

We are now more than a decade into the 'digital age' of early modern studies, and while the effects of this transformation continue to reverberate through the field, a shift in the nature of online resources now also seems to be taking place, one that has not received the same kind of attention. In this essay, we will outline what we see as the crucial differences between what might be called 'first-generation' databases like ESTC and EEBO and the newer, 'second-generation' databases. We focus specifically on an online resource that we have created, *DEEP: Database of Early English Playbooks*, which is a database of all plays that were printed in England from the early sixteenth century through 1660. It offers a wealth of information about these plays and playbooks, such as the information on their title pages, the presence of paratextual matter (like dedications and addresses to readers), various advertising features (such as performance blurbs or indications of genre), and bibliographic and theatrical backgrounds. Although we primarily discuss *DEEP* as an exemplar of this second generation, a number of other databases have recently come online that seem to have been created with similar underlying principles and goals (we point to some of these resources at the end of the essay). Our aim is not to undertake a full review of ESTC or EEBO, but rather to highlight several important differences between them and second-generation resources. The most important distinction is that whereas ESTC and EEBO aim for comprehensive treatment of early modern print culture, *DEEP* and similar resources aim for in-depth coverage of one particular kind of text or document. This key distinction, moreover, entails different approaches to the basic architecture of the databases, their searchability, and their analytical and editorial principles, all of which have significant ramifications for the kinds of research that these databases enable.

Before we examine these shifts in detail, we want to address two persistent critiques of first-generation resources that have sometimes, but we think wrongly, seemed to indict the digital nature of these resources as such. Reviews of ESTC and EEBO have tended to criticize them on two grounds, mainly centering on (1) the accuracy of their records; and (2) the practice of consulting digital reproductions rather than original documents. As prominent bibliographers such as Peter W. M. Blayney, William Proctor Williams and William Baker, David McKitterick, and James E. May have all pointed out, ESTC still does not include all the records in the revised *Short-Title*

Catalogue or the revised *Wing Short-Title Catalogue*, and those records it does contain are too often incomplete and incorrect.¹ What we might call the ‘facsimile’ critique has addressed a slightly different issue, that of consulting reproductions instead of original documents. Microfilm and digital images are prey to any number of errors in reproduction, both intended and unintended, such as omitted leaves, pages in the wrong order, and ‘opaquing’ (to eliminate show-through). Beyond mistakes like these, there is the more theoretical problem of consulting reproductions when making historical arguments. As G. Thomas Tanselle has observed,

[a]ny reproduction, whether clear or indistinct, must be suspect simply because it is not the ultimate source: documentary texts, like all other artifacts, must be examined first-hand if one is serious about approaching them as historical evidence[.]

which, presumably, is how most historicist literary critics do wish to approach these texts (‘Reproductions’ 34).

Whatever the validity of these critiques, however, we would argue that they could apply equally to print resources and so should not be understood to imply any essential limitations of the digital medium for bibliographic research. The accuracy critique could of course be made of any reference work that had not been created with sufficient scholarly rigor. And neither are Tanselle’s caveats exclusive to digital resources: they apply to any attempted reproduction or transcription, including those landmarks of bibliography like the *Short-Title Catalogue* or W. W. Greg’s *Bibliography of English Printed Drama*. Once we put aside issues of accuracy, there is no *theoretical* difference between using a digital resource like EEBO or *DEEP* to research a Shakespearean playbook and using a print reference like Greg’s *Bibliography* – or, indeed, an Arden or Cambridge edition of a play. Likewise, for other kinds of historical research, and especially for detailed bibliographic work, one would always need (as Tanselle argues) to consult the original documents rather than any of these ‘reproductions’, whether digital or printed. Both the accuracy and the facsimile critiques of online resources, therefore, prove too much: they apply to *any* reference work that attempts to reproduce, transcribe, or describe aspects of physical documents.

Two great advantages of resources like ESTC and EEBO, however, do arise from their digital form – their comprehensiveness and their searchability – and it is here that the differences between first- and second-generation databases become most evident. ESTC advertises itself as listing ‘nearly 470,000 entries’ for items printed from 1473 to 1800, which are held in the collections of over 2,000 libraries. EEBO, meanwhile, says that it ‘contains digital facsimile page images’ of over 100,000 titles from 1473 to 1700. And unlike printed bibliographies, which cannot be easily changed as new editions or variants are located or as scholars discover new bibliographical evidence, ESTC and EEBO can be continuously updated to reflect these developments.² Furthermore, since neither resource is limited by the codex

structure, users can search these comprehensive listings in ways that would be all but impossible in a printed book, for example by performing keyword and subject searches.

What most interests us here is a paradox that underlies these digital resources: the benefits of comprehensiveness and searchability can actually work against each other. Comprehensive sites have been designed with the goal of allowing searches that are relevant to *all* of their records (such as searches by author, title, subject, or keyword), but not specialized searches that may pertain to only a small subset of their records – such as whether a particular theater is advertised on a playbook's title page. As a result, precisely because of their comprehensiveness, ESTC and EEBO cannot enable the types of searches that many scholars of printed drama would like to perform. And as scholars have grown more familiar with both the benefits and the limitations of ESTC and EEBO, second-generation databases that offer detailed search capability for a more narrow range of texts – what we might call 'analytic databases' – have arisen precisely to address this need. We can think of such databases as the digital analogues to Greg's specialized *Bibliography of English Printed Drama* as compared to the comprehensive *Short-Title Catalogue*.

In creating *DEEP*, we wanted to exploit the benefits of searchability that a digital resource offers while also tailoring it specifically to printed plays. To do so, we had to pre-analyze and extract bibliographic and theatrical information from a variety of sources and categorize that information with regard to the particular questions that scholars of plays tend to ask: this is what we mean by an *analytic* database. In a very real sense, it must be acknowledged, all databases are analytic. Creating any database requires organizing data into categories, which are neither self-evident nor inherent in the data itself; rather, as Jerome McGann writes, this process 'represents an initial critical analysis of the content materials' (1588).³ Even the most comprehensive of databases such as ESTC and EEBO encode this initial critical analysis, but in their attempt to account for *every* printed book of the period, these sites tend to downplay this analytical process because there is a limited number of questions that scholars might ask of every early modern printed book. As an analytic database focused on a narrower set of records, *DEEP* allows us to foreground this critical activity, picking up significant debates and problems in the field and organizing the data so as to enable scholars to investigate them.

As we have found, one of the most important steps in analyzing and organizing data to improve searchability is deciding how to structure information in the database itself – that is, deciding on its architecture – and even seemingly small decisions about how to transcribe an individual word or spell a person's name can have significant consequences for users. To explore some of these decisions and their effects, we might consider the example of Thomas Goffe's play *The Raging Turk, or Bajazet the Second*, first printed in 1631 as a quarto, and then printed again in the 1656 octavo collection

of Goffe's *Three Excellent Tragedies*. Searching for 'Raging Turk' in ESTC returns *Three Excellent Tragedies* but not the 1631 quarto, and if we search for that first edition by its STC number (11980), we can see why: the title of this quarto is listed as 'The Raging Turke', and searching for 'Turk' without the final 'e' does not yield a match. But the story is even more complicated than it initially appears. ESTC transcribes the full title of the 1631 quarto as:

Title: The raging Turke, or, Baiazet the Second. A tragedie vvritten by Thomas Goffe, Master of Arts, and student of Christ-Church in Oxford, and acted by the students of the same house.

Publisher/year: London: printed by August. Mathevves, for Richard Meighen, 1631.

The title page for the 1631 quarto, as transcribed in Greg's *Bibliography*, however, reads somewhat differently (Fig. 1).

THE | RAGING | TVRKE, | OR, | BIAIAZET | THE SECOND. | A Tragedie
 vvritten by THOMAS | GOFFE, Master of Arts, and Student of | Christ-Church in
 Oxford, and Acted by the | Students of the fame houfe. ||
Monstra fato, scelera moribus imputes
Det ille veniam facile cui venia est opus.
 || [ornament] | LONDON: | Printed by AVGVST. MATHEVVES, for | RICHARD
 MEIGHEN. | 1631. [variant: . . . RANING . . .]

Fig. 1. Entry for *The Raging Turk* in Greg's *Bibliography* (2: 597). Reproduced by permission of The Bibliographical Society.

Comparing the two transcriptions, we see that ESTC reproduces some, though not all, of the original spelling and letter forms: it uses the double *v* for *w* in 'written' and 'Mathewes', for example, but silently changes 'TVRKE' to 'Turke' in the title and 'AVGVST' to 'August' in the imprint.⁴

These discrepancies point to the problems with searchability that arise when issues connected to old-spelling word forms are not adequately addressed. The usefulness of ESTC and EEBO as scholarly resources suffers greatly because users are often required to know beforehand the original, idiosyncratic spellings on title pages (not to mention idiosyncratic spellings of ESTC's own creation) in order to find the records for particular editions. Hence the inability to find the first quarto of *The Raging Turk* by searching either for the modern-spelling 'Raging Turk' or for the old-spelling 'Raging Tvrke'. As this example begins to illustrate, while historicist scholars will generally prefer title-page transcriptions in the original spelling, they will find modern-spelling forms of book titles, authors, printers, booksellers, and so on, to be far more effective for searching.

ESTC has tried to ameliorate this problem by establishing a field called 'Variant Titles', which contains alternate spellings of some book titles. The

record for the 1631 quarto of *The Raging Turk* lists the alternate title of ‘Raging Turke, or, Bajazet the Second’ – thereby ensuring that searches for either ‘Baiazet’ or ‘Bajazet’ will turn up this edition – but, surprisingly, it does not include either the fully modernized title ‘The Raging Turk’ or the old-spelling ‘The Raging Tvrke’. While the idea of including variant titles could help offset the problem of forcing users to search for old-spelling word forms, the effectiveness of the field is limited when book titles are not fully modernized. Partially or incompletely modernized titles do not greatly improve the usability of the site.

Searches for particular stationers in ESTC do not include variant forms, and so the difficulties here are even more stark. Trying to research a particular stationer using ESTC is thus almost impossible because the site has not been created with this type of search in mind. For instance, finding all the books printed or published by Augustine Mathewes, the printer of *The Raging Turk*, requires searching in the ‘Publisher’ field for at least *forty-one possible forms of his name*: six variant spellings of his first name (‘Augustine’, ‘Avgvstine’, ‘August’, ‘Avgvst’, ‘Aug’, and ‘Avg.’); six of his last name (‘Mathewes’, ‘Mathevves’, ‘Matthewes’, ‘Matthevves’, ‘Math.’, and ‘Matth.’); his initials (‘A. M.’, which returns 269 records, if the search is restricted to the exact phrase ‘A. M.’, not all of which are actually for Augustine Mathewes); and several forms of his name in which ESTC has inserted brackets (‘A[ugustine] M[athewes]’, ‘A[ugustine] M[atthewes]’, ‘Aug[ustine] Mat[hewes]’, and ‘Aug. Mat[hewes]’). ‘Avgvstine Mathevves’ yields no records, even though this is the spelling of his name on the title page of Thomas Tomkis’s *Lingua* (1632; STC 24108); the same is true of ‘Avgvst. Mathevves’, the spelling of his name on the title page of Goffe’s *The Raging Turk*. ‘Avg. Mathevves’ does return two records, but not the one for Richard Brome’s *The Northern Lass* (1632; STC 3819) because ESTC has changed the ‘Avg.’ in its title-page imprint to ‘Aug.’.

EEBO fares somewhat better. Using its ‘Variant spellings’ checkbox, a search for ‘Raging Turk’ or ‘Raging Tvrke’ does return both the 1631 quarto and the 1656 collection. With stationers, however, EEBO runs into similar problems as ESTC. Searching for ‘Mathewes’ in the Imprint field on the Advanced Search page does not return the 1631 edition (because EEBO’s record mistranscribes the name as ‘Mathevves’, adding an extra *t*), nor does it return a host of other books printed by Augustine Mathewes, because (just as with ESTC) the numerous variant spellings of his name are not unified under a single search term.

There is one last problem worth addressing: the title page for the 1631 quarto of *The Raging Turk* exists in two states, one with the correct spelling of ‘RAGING TVRKE’ (STC 11980), and one with the incorrect ‘RAN-ING TVRKE’ (STC 11981). Neither ESTC nor EEBO contains a record for STC 11981, and although these sites might eventually remedy this particular lack, they handle variants inconsistently. When they work as intended, ESTC and EEBO contain multiple records for variant issues and states of a single

edition, general notes about the existence of the other variant(s), and a short statement describing the differences between them. There are two issues, for instance, of the ninth edition of *The Spanish Tragedy* (1623): as both ESTC and EEBO explain, one issue lists Thomas Langley as bookseller (STC 15093a) while the other lists John Grismand (STC 15093). Not all variants, though, are described with this level of consistency. The first edition of Chapman, Jonson, and Marston's *Eastward Ho* (1605) exists in two states; both ESTC and EEBO include a note in their records for STC 4971 explaining that this is a later state of the first edition, but the corresponding records for STC 4970 do not mention that it is the earlier of two variant states.

Even when the variants are fully recorded, however, the architectures of ESTC and of EEBO are at odds with their search functions, as the underlying structure of these databases often results in multiple variants of the same edition being returned in a single search. For example, ESTC lists twelve records, and EEBO eleven, for the ten editions of *The Spanish Tragedy* printed before the Restoration; ESTC lists nine records, and EEBO six, for the single edition of the Shakespeare second folio (1632); and both sites list two records for the single quarto edition of Shakespeare's *Troilus and Cressida* (1609). In both ESTC and EEBO, each variant is treated as its own independent record, and there is no hierarchical organization relating records to each other nor any links between variant records. As a result, a user searching for *The Spanish Tragedy* is easily misled into believing there were twelve pre-Restoration editions instead of ten, a consequential error for scholars interested in the relative popularity of different plays, or in the theatrical revivals of Kyd's tragedy, or in a host of other issues. And in general, the absence of a hierarchical organization among variant records makes it more difficult for scholars to pursue any number of lines of research on these sites. The relationships among variants of a given edition *can* be readily perceived in the printed *STC* and Greg's *Bibliography* – which do clearly distinguish among editions, issues, and states – but of course those references cannot enable the kind of searchability that digital resources allow.

The massive comprehensiveness of first-generation resources like ESTC and EEBO has great benefits, of course, but their size also makes it more difficult to include fine-grained or genre-specific searches, or to undertake the kind of labor necessary to bring together, for search purposes, a host of variable spellings of a given name, or to link all the disparate records that derive from a given edition – not to mention to ensure the accuracy of all transcriptions.⁵ Second-generation resources, in contrast, tend to have a narrower focus that, while losing some of the benefits of a wide scope, allows them to alleviate these drawbacks.

In creating *DEEP*, we have tried not only to correct many of the errors in ESTC and EEBO but also to structure the site in a way that is convenient for searching modernized and old-spelling forms and that creates a hierarchy of different types of records for variants of a single edition. We

have benefited, of course, from years of experience with ESTC and EEBO, which has allowed us to hone our understanding of the ways that searchability is affected by such issues as old-spelling or variants. Unlike ESTC, for example, *DEEP* has been designed so that its searches return only one record per edition; if an edition has any variants, the relationships among these are described in a brief note that includes links to the other records. In this way, *DEEP* provides an accurate count of the number of editions matching the user's search, while still allowing the user easily to see all the variants within those editions. These variants are based on Greg's *Bibliography* and the revised *Short-Title Catalogue*: when either of these bibliographies contains multiple entries for a playbook edition, *DEEP* follows suit. *DEEP*'s categorization of variants, however, also tries to build on these two standard reference works; based on the principles laid out by Tanselle, *DEEP* systematically distinguishes between variant states within an edition and variant issues of an edition. As a result, users can not only grasp specific differences among variants but also understand, to use Tanselle's terms, whether these differences are the result of 'correcting errors' on a specific sheet, or whether the variants represent two or more 'consciously planned publishing unit[s]' ('Bibliographical' 65–6).⁶

To make searches more practical, we use drop-down menus with regularized and modernized forms of the information that appears on title pages. Selecting the 'Stationer' search generates a drop-down menu of all the stationers involved in producing playbooks through 1660. Choosing 'Mathewes, Augustine' from this menu returns all of the playbook editions that he printed (and sometimes also published), regardless of the precise spelling of his name on the title page or even whether it appears at all.⁷ In the results listing, *DEEP* provides an old-spelling transcription of the title page of each record, broken down into several categories:

Title: THE RAGING TVRKE, OR, BIAZET THE SECOND. A Tragedie

Author: written by THOMAS GOFFE, Master of Arts, and Student of Christ-Church in *Oxford*,

Performance: and Acted by the Students of the same house.

Latin Motto: [in single column] *Monstra fato, scelera moribus imputes | Det ille veniam faciliè cui venia est opus.*

Imprint: LONDON: Printed by AVGVST. MATHEWES, for RICHARD MEIGHEN. 1631.

Users thus get the best of both worlds, benefiting from the ease and effectiveness of modern-spelling searching without forsaking the historicist desire for old-spelling results. While *DEEP* does not reproduce early modern letter forms (the long-*s*, ligatures, the double-*v* for *w*, swash letters, or black-letter fonts), it does retain the original use of capitalization, italics, small capitals, and punctuation. (Some of these decisions, such as the omission of black letter, are due to the limitations of Web browser display, while others, such as the transcription of double-*v* as *w* or the use of short- for long-*s*, result

from our fidelity to old spelling but not to the particular forms of the pieces of type that were set.⁸) *DEEP* also includes fully modernized title-page transcriptions for each record, which are not displayed but can be searched by users in a separate field. We believe that allowing users to search in modern and regularized terms while still seeing original spelling in results should be a crucial aspect of second-generation database design going forward.

The most distinctive aspect of second-generation databases like *DEEP* is the way they have been tailored to address the kinds of inquiries that scholars tend to ask in a given field of research. With its exclusive focus on playbooks, *DEEP* offers numerous search options important to theater and book historians. Users can search for whether a specific theater is named on the title page – ‘*diuerse times acted at the Globe*’ (the 1622 *Othello*) or ‘*sundry times priuately acted in the Blacke Friers*’ (Ben Jonson’s 1602 *Poetaster*) – or whether any theater or no theater appears. (In the pre-Restoration period, a total of twelve different commercial theaters were so named, from the Globe and Blackfriars to the New Theater in St Werburgh Street, Dublin, but interestingly enough, not the Theater or the Rose, which have played such a large role in theater history.) We have also analyzed these theater attributions to enable users to search for whether a court performance is mentioned or whether any indoor theater or any outdoor theater appears, a division that has loomed large in studies of early modern drama since Alfred Harbage’s *Shakespeare and the Rival Traditions* (1952). As with stationer names, these searches use drop-down menus to obviate any problems with variable spelling or naming. The playhouse in Drury Lane, for instance, was sometimes referred to as the Cockpit, sometimes as the Phoenix, and sometimes simply by its street location, but *DEEP* unifies these records as ‘Phoenix/Cockpit’ in the search menu. Similarly, playing companies often went by varying names (the Admiral’s Men were also Nottingham’s Men), and more than one company could share a name over time (the Queen’s Men were Queen Elizabeth’s, Queen Anne’s, and Queen Henrietta Maria’s Men). *DEEP* provides a standardized search for whether a specific playing company is named on the title page, or whether any or no company appears.

Creating these searches involves analytical and critical work that goes beyond simply modernizing and regularizing information found on a title page and that touches on some of the most important issues in book-historical research. For example, what exactly does it mean to say that Shakespeare is ‘the author’ of a particular set of printed plays? While such a question might sound pedantic, it and others like it have been the subject of extended inquiry over the past two decades in ‘new textualist’ work on printed drama. Like ESTC and EEBO, of course, *DEEP* allows users to search for plays that we now believe Shakespeare wrote – what *DEEP* calls the ‘modern attribution’ – but *DEEP* goes beyond those sites in also enabling a search for the ‘playbook attribution’ of authorship: on which playbooks does Shakespeare’s name appear as author? The two categories are not, of course, identical: the 1605 edition of *The London Prodigal* ‘By William Shakespeare’ and

the 1619 edition of the first part of *Sir John Oldcastle*, 'Written by William Shakespeare', will appear in the results list for the playbook attribution but not for the modern attribution. As numerous scholars have shown, play-book attributions of authorship have a great deal to tell us about the construction of dramatic authorship and the transformation of drama from low entertainment to literature. Neither EEBO nor ESTC allows this kind of specialized search, however, because their very comprehensiveness means that they are far less likely than a narrowly focused site like *DEEP* to take the cues for their organization of data from ongoing scholarly debates.

DEEP's analysis and categorization of information goes beyond the title page to include other aspects of the physical book that have been and are sure to remain important to scholars of early modern drama. Many critics have been interested in what the paratexts of playbooks can tell us about their intended audience, the readings of the plays imagined by their publishers, or the political and social coteries in which they were embedded. *DEEP* can search for the presence or absence of paratextual material in general, or for the presence of specific kinds of paratexts, such as dedications, commendatory verses, addresses to the reader, 'arguments' of the play, character lists (*dramatis personae*), and lists of actors who performed the play.

One of the most important ways that *DEEP* enables new kinds of research that are not possible with ESTC or EEBO, and that are difficult and time-consuming using printed resources, is the ability to perform combinatory searches. Because *DEEP's* search terms are tailored to printed drama, combining them with AND or OR operators creates a powerful research tool. *DEEP* can quickly answer questions such as: Which plays from the professional theater that were first performed before 1600 were still being printed after 1630? The results might tell us something about dramatic canon-formation in the early modern book trade, as old plays from an earlier era of theater became 'classics' that were still relevant much later (Farmer and Lesser, 'Canons'). Some of the results might be expected: Marlowe's *Doctor Faustus*, Kyd's *Spanish Tragedy*, and several of Shakespeare's plays. But this search might lead us to wonder why *Fair Em*, *The Miller's Daughter* was reprinted in 1631, forty years after its first edition, or why Marlowe's *Jew of Malta* was first printed in 1633, more than four decades after its stage debut (Lesser ch. 3).

To take another example: Which editions of plays from the professional theater contain a title-page author attribution and include a list of the actors who performed the play on stage? Since critics have often argued that professional drama is transformed into suitably literary matter only as it is distanced from its theatrical origins – both by authors, paradigmatically Ben Jonson, and by stationers – and since including an actor list is surely one way that printed plays could be *connected to*, rather than distanced from, the theater, one might be surprised to learn that the results of such a search are dominated by none other than Ben Jonson. The first plays to include an actor list are all contained within the Jonson folio of 1616, which should

remind us that Jonson's collection contains a wealth of explicitly theatrical paratexts; as we have argued elsewhere, in many ways the dramatic author was created *within the commercial theater*, not apart from it (Farmer and Lesser, 'Vile Arts'; Bentley 10).

One might also approach the intersection of theatricality and authorship by examining differences between boy- and adult-company plays. Were plays performed by boy companies more or less likely to advertise their authors' names on their title pages? DEEP's 'Play Type' field sorts all the records by their auspices – boy company, adult company, closet, private, university, and so forth, as well as broadly by whether they were professional or nonprofessional – and using this term in conjunction with other categories allows researchers to fine-tune their investigations in numerous ways. What kinds of paratexts tended to appear in books of closet drama, for instance, as compared to university or professional drama? Which types of playbook were more likely to include Latin on their title pages?

Another combinatory search might investigate generic labeling. Which plays now considered to be histories of some kind (by the *Annals of English Drama*) were originally labeled as tragedies on the title pages of their first editions? Of the 59 plays that *Annals* identifies as histories (including those in collections), 15 were initially advertised as tragedies, including three of Shakespeare's plays, as well as *Lochrine*, which was attributed to 'W.S.' on its title page. Two other histories refer to the 'tragical' ends of characters in the plays. Combinatory searches such as this one can illuminate changing perceptions of dramatic genre, and taking the additional step of restricting such searches to particular date ranges can further reveal patterns in the rise and fall of particular genres over time.

These are only a few examples of the kinds of searches that can be performed with DEEP, which currently includes over twenty search terms, many of them specific to the scholarly issues that arise in relation to printed plays. It is this type of fine-grained search that distinguishes second-generation digital resources both from their first-generation counterparts and from printed reference works. In much of the existing criticism on digital resources for early modern studies, digital scholarship has been represented as a break with the past – sometimes utopian, sometimes dystopian – but we see as much continuity as disjuncture. DEEP would certainly not have been possible without essential printed reference works like Greg's *Bibliography* and the *Short-Title Catalogue*. The crucial difference between digital and printed databases, as we see it, lies in the ability to allow users easily to search across categories of data in various combinations – or, in the idiom of database mavens, to 'slice and dice' the data in multiple ways, thereby revealing patterns that might have otherwise remained invisible. The rise of second-generation digital resources now allows us to see patterns that could not be easily perceived using either print or earlier digital resources.

Other early modern digital resources seem to be arising out of a similar desire for more narrowly focused, analytic databases. The *English Broadside*

Ballad Archive at the University of California–Santa Barbara provides facsimiles and transcriptions of early ballads printed in English, mainly from the seventeenth-century, as well as audio recordings of some of them. It allows users to search by categories such as title, first line, refrain, tune, author, publisher, and by Samuel Pepys's own categorizations of the ballads in his extensive collection. *LEME: Lexicons of Early Modern English* 'searches and displays word-entries from monolingual English dictionaries, bilingual lexicons, technical vocabularies, and other encyclopedic-lexical works, 1480–1702' (some of its more advanced search functions are disabled without a subscription). The *Bath Chronicle Georgian Newspaper Project* covers the years 1770 to 1800, allowing users to search the newspaper by month, year, keyword, and subject categories.

Other sites cover non-printed materials. *The Proceedings of the Old Bailey* are now online for the years 1674 to 1913 and are searchable by keyword, by the name, gender, occupation/status, and age of defendant and victim, by place, by type of crime, by verdict and punishment, and by dates; the site also provides advice on searching for specific demographic groups and social issues, including 'Black Communities of London', 'Gypsies and Travellers', 'Homosexuality', 'The Irish in London', 'Jewish Communities', and 'Huguenot London'. *REED – Patrons and Performers* organizes some of the records covered by the Records of Early English Drama project; the site offers searches by patron, event type, troupe, county, performer type (including acrobats, minstrels, players, and performing animals), and other categories, as well as providing interactive maps.

Precisely because of their narrow focus on particular kinds of texts and records, sites like these open up new kinds of digital research that take us beyond the advances of the 1990s, when ESTC and EEBO went online. At the same time, by enabling and encouraging specialized searches relevant to their particular fields, these resources also create new demands for assessing how particular and perhaps idiosyncratic fields of evidence fit into the larger structures of early modern culture. And as more and more of these specialized sites come online, we believe that theorizing this relationship between the general and comprehensive, on the one hand, and the particular and intensive, on the other hand, will be central to the future of early modern scholarship in the digital age.

Short Biographies

Alan B. Farmer is an Assistant Professor of English at The Ohio State University whose research focuses on Renaissance drama, the history of the book, and news and politics in 1620s and 1630s England. His current book project is entitled *Playbooks, Newsbooks, and the Politics of the Thirty Years' War in England*. He has co-edited with Adam Zucker, *Localizing Caroline Drama: Politics and Economics of the Early Modern English Stage, 1625–1642*, and published essays on Ben Jonson, Shakespeare, and foreign books in early modern England.

He is co-creator with Zachary Lesser of *DEEP: Database of Early English Playbooks*, and they have also co-written several articles on the publication of playbooks in Renaissance England. They are currently working on a book entitled *Print, Plays, and Popularity in Shakespeare's England*. He earned a Ph.D. in English Literature from Columbia University, a B.A. in English Literature from the University of Pennsylvania, and a B.S. in Economics from the Wharton School of Business at the University of Pennsylvania.

Zachary Lesser is an Assistant Professor of English at the University of Pennsylvania. He is the author of *Renaissance Drama and the Politics of Publication: Readings in the English Book Trade* (Cambridge University Press, 2004), winner of the 2006 Elizabeth Dietz Prize as the best book of the year in early modern studies. He has published essays on Shakespeare, early modern drama, and the history of the book in *Shakespeare Quarterly*, *ELH*, *ELR*, and elsewhere, and has edited three plays for the Barnes & Noble Shakespeare and Thomas Dekker's *The Noble Spanish Soldier* for the Globe Quartos. He is co-creator with Alan B. Farmer of *DEEP: Database of Early English Playbooks*, and they have also co-written several articles on the publication of playbooks in Renaissance England. They are currently working on a book entitled *Print, Plays, and Popularity in Shakespeare's England*. He received his Ph.D. in English Literature from Columbia University and his B.A. in Renaissance Studies and Religious Studies from Brown University.

Notes

* Correspondence address: Alan B. Farmer, 421 Denney Hall, Ohio State University English Department, 164 W. 17th Ave., Columbus, OH 43210, USA. Email: farmer.109@osu.edu

** Correspondence address: Zachary Lesser, Department of English, University of Pennsylvania, 3340 Walnut St, Philadelphia, PA 19104, USA. Email: zlesser@english.upenn.edu

¹ As Stephen Tabor, the curator of the Huntington Library collection and himself a contributor to ESTC, has noted, ESTC was not initially designed to give complete and accurate transcriptions of title pages: 'Until recently, the eighteenth-century records used liberal elisions in the title and imprint transcriptions, with the latter omitting most addresses and truncating long lists of publishers' names' (370). The principles used for pre-1700 books were not much better, for 'the editors adopted a full-title policy for most records' (371; emphasis added), though apparently not all, and there appears to have been a deliberate policy to truncate the imprints and titles of early sixteenth-century books (Blayney 72). As a result, the accuracy of all of these records is cast into doubt, since users can never be sure whether the title-page information in any given record is complete or truncated – even for those records that are, in fact, complete.

² The ability of sites like EEBO and ESTC to be continuously updated, however, can paradoxically become a drawback. These sites are always 'in process', and scholars can therefore never be sure about either the completeness or the accuracy of any particular section of the site, a problem that is compounded if the creators of the electronic database do not proceed with the same dedicated rigor that went into the printed *Short-Title Catalogue* (and its revision) or Greg's *Bibliography*.

³ McGann's essay responds to Folsom, 'Database'; see also Folsom's 'Reply'.

⁴ Note also that ESTC omits the Latin motto from the title page; for scholars interested in the kind of readership imagined for this book, such information is potentially vital. *DEEP* allows users to search by the presence or absence of Latin on the title page.

⁵ Tabor estimates that 'ESTC has a minimum of 80,000 records containing at least one error in description', and because accurate title-page transcriptions 'tend to overshoot the ability of [ESTC] cataloguers', he concludes that 'a rigorous programme' of correcting these errors 'is clearly not feasible' (372, 375).

⁶ The lines between an edition, an issue, and a state, while conceptually clear, can be fuzzy in practice, and different scholars will reach different conclusions about the relationships of variants in copies of a particular text. The descriptions of variants in *DEEP* represent our own critical analysis of the nature of variants in those playbooks that have multiple Greg and STC entries. Users may of course disagree with our judgments, but the crucial point is to alert users of *DEEP* and other bibliographic databases not only to the existence of but also to the relationships among these variants.

⁷ When printers have been attributed by the *Short-Title Catalogue* or recent bibliographical research, we have followed these attributions.

⁸ All of the principles underlying our bibliographic descriptions and our collection and organization of data can be found on the site, in the section 'How to Use DEEP'.

Works Cited

- Annals of English Drama, 975–1700*. Eds. Alfred Harbage, Samuel Schoenbaum, and Sylvia Stoler Wagonheim. 3rd ed. London: Routledge, 1989.
- Bath Chronicle Georgian Newspaper Project*. 28 September 2008 <<http://www.bathnes.gov.uk/BathNES/leisureandculture/recordsarchives/georgian/default.htm>>.
- Bentley, G. E. *The Profession of Player in Shakespeare's Time, 1590–1642*. Princeton, NJ: Princeton UP, 1984.
- Blayney, Peter W. M. 'Correspondence'. *The Library*, 7th ser. 1 (2000): 72–4.
- DEEP: Database of Early English Playbooks*. 28 September 2008 <<http://deep.sas.upenn.edu>>.
- Early English Books Online*. 28 September 2008 <<http://eebo.chadwyck.com/home>>.
- English Broadside Ballad Archive*. 28 September 2008 <http://emc.english.ucsb.edu/ballad_project/>.
- English Short Title Catalogue*. 28 September 2008 <<http://estc.bl.uk>>.
- Farmer, Alan B. and Zachary Lesser. 'Vile Arts: The Marketing of English Printed Drama, 1512–1660'. *Research Opportunities in Renaissance Drama* 39 (2000): 77–165.
- . 'Canons and Classics: Publishing Drama in Caroline England'. *Localizing Caroline Drama: Politics and Economics of the Early Modern English Stage, 1625–1642*. Eds. Adam Zucker and Alan B. Farmer. New York, NY: Palgrave, 2006. 17–41.
- Folsom, Ed. 'Database as Genre: The Epic Transformation of Archives'. *PMLA* 122 (2007): 1571–79.
- . 'Reply'. *PMLA* 122 (2007): 1608–12.
- Greg, W. W. *A Bibliography of English Printed Drama to the Restoration*. 4 vols. London: Bibliographical Society, 1939–1959.
- Harbage, Alfred. *Shakespeare and the Rival Traditions*. New York, NY: Macmillan, 1952.
- Justice, George. 'The ESTC and Eighteenth-Century Literary Studies'. *Literature Compass* 1 (2004): 1–3, doi:10.1111/j.1741-4113.2004.00002.x.
- LEME: Lexicons of Early Modern English*. 28 September 2008 <<http://leme.library.utoronto.ca/>>.
- Lesser, Zachary. *Renaissance Drama and the Politics of Publication: Readings in the English Book Trade*. Cambridge: Cambridge UP, 2004.
- McGann, Jerome. 'Database, Interface, and Archival Fever'. *PMLA* 122 (2007): 1588–92.
- McKitterick, David. '“Not in STC”: Opportunities and Challenges in the ESTC'. *The Library*, 7th ser. 6 (2005): 178–94.
- May, James E. 'Who Will Edit the ESTC? (and Have You Checked OCLC Lately?)'. *Analytical and Enumerative Bibliography* 12 (2001): 288–304.
- Pollard, A. W. and G. R. Redgrave, eds. *A Short-Title Catalogue of Books Printed in England, Scotland, & Ireland and of English Books Printed Abroad, 1475–1640*. Rev. W. A. Jackson, F. S. Ferguson, and Katharine F. Pantzer. 3 vols. 2nd ed. London: Bibliographical Society, 1976–1991.
- The Proceedings of the Old Bailey, London's Central Criminal Court, 1674 to 1913*. 28 September 2008 <<http://www.oldbaileyonline.org/>>.

- REED – *Patrons and Performers*. 28 September 2008 <<http://link.library.utoronto.ca/reed/index.cfm>>.
- Tabor, Stephen. 'ESTC and the Bibliographical Community'. *The Library*, 7th ser. 8 (2007): 367–86.
- Tanselle, G. Thomas. 'The Bibliographical Concepts of *Issue* and *State*'. *Papers of the Bibliographical Society of America* 69 (1975): 17–66.
- . 'Reproductions and Scholarship'. *Studies in Bibliography* 42 (1989): 25–54.
- Williams, William Proctor, and William Baker. 'Caveat Lector. English Books 1475–1700 and the Electronic Age'. *Analytical & Enumerative Bibliography* 12 (2001): 1–29.