

Shifting perspectives and moving targets: from conceptual vistas to bits of data in the first year of the **MEDIATE** project

ALICIA C. MONTOYA (Radboud University)

The expression ‘Digitizing Enlightenment’, confidently used in the present collection’s title, might seem to presuppose some agreement not only on the existence of ‘the’ Enlightenment, but perhaps even on what, exactly, might constitute the digitizing of this ‘Enlightenment’.¹ Such basic definitional questions lie at the heart of a digital database project, **MEDIATE** (Measuring Enlightenment: Disseminating Ideas, Authors, and Texts in Europe), which started in September 2016 and will run until September 2021, at Radboud University in the Netherlands, with funding provided by the European Research Council.² The aim of the project is to study the circulation of books in eighteenth-century Europe, focusing on the movement of books and ideas traditionally associated with the ‘Enlightenment’, by building an open access database housing data from a corpus of catalogues of private libraries sold at auction in the Dutch Republic, France, the British Isles and Italy between 1665 and 1830. In the following pages, taking up Willard McCarty’s argument that ‘the point of all modelling exercises, as of scholarly research generally, is the process seen in and by means of a developing product, not the definitive achievement’,³ I

1. Substantial sections of the present chapter were published previously in ‘Middlebrow, religion, and the European Enlightenment: a new bibliometric project, **MEDIATE** (1665-1820)’, *French history and civilisation* 7 (2017), p.66-79, <https://h-france.net/rude/vol7/montoya7/> (last accessed 10 January 2020). The author gratefully acknowledges permission to reprint these sections here.
2. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme under grant agreement no.682022. See also the project website, www.mediate18.nl (last accessed 10 January 2020).
3. Willard McCarty, *Humanities computing* (London, 2014), p.22.

focus on how the definitional issues encountered in thinking through the technical and conceptual nuts and bolts of our first project year impacted notions of what that ‘Enlightenment’ we set out to study in September 2016 might, actually, turn out (not) to be.

The central hypothesis underlying the *MEDIATE* project is that, if we are to adequately understand the spread of ideas and books associated with the Enlightenment movement, and the processes of societal change supposedly engendered by these ideas and books,⁴ it is not enough to study only the ideas, books and authors that formulated them, or the traditional canon of Enlightenment *philosophes* and reformist thinkers. It is just as crucial to understand how these ideas and books were embedded in the cultural field at large. The project hence draws on the one hand on Pierre Bourdieu’s concept of the literary field, which productively foregrounds the relational nature of literary valuations and status, and the many connections between cultural and economic capital. On the other hand, the growing body of scholarship by literary historians on the concept of middlebrow, or forms of literature that situated themselves between elite forms and more popular ones, also provided important conceptual inspiration to the project (on which more below).⁵ *MEDIATE* postulates that the authors studied as participants in, or in relation to, the Enlightenment movement are significant not only as individuals, but, just as importantly, as part of a larger literary system of eighteenth-century authors, most of whom are unknown to us today. In order to understand the full cultural impact of any individual author, we need to view his or her texts as part of a complex set of relations between higher- and lower-prestige texts, geographic regions and languages, and between authors closer and farther away from centres of cultural authority. As Franco Moretti has argued, literary and intellectual historiography is commonly based on limited selections of material, or on a small corpus of well-known or canonized works, yet these represent, at best, no more than 1 per cent of the total. ‘A field this large’, he writes, ‘cannot be understood by stitching together separate

4. According to the famous thesis elaborated by Daniel Mornet in his *Origines intellectuelles de la Révolution française* (Paris, 1933) and taken up by Robert Darnton in *The Forbidden best-sellers of pre-Revolutionary France* (New York, 1996).
5. See, for a study that specifically links Bourdieu’s ideas to the concept of ‘middlebrow’, Caroline Pollentier, ‘Configuring middleness: Bourdieu, *l’Art moyen* and the broadbrow’, in *Middlebrow literary cultures: the battle of the brows, 1920-1960*, ed. Erica Brown and Mary Grover (London, 2012), p.37-51.

bits of knowledge about individual cases, because it isn't a sum of individual cases: it's a collective system, that should be grasped as such, as a whole.⁶ As he goes on to demonstrate, a literary system like this can be understood only by using digital tools allowing scholars to gain an overview of an enormous range of books, which would be simply impossible to count and analyze statistically by hand: a range that, in the case of the MEDIATE database, will ultimately document literally millions of individual copies of titles moving across Europe during the eighteenth century.

More specifically, by focusing on private libraries, MEDIATE aims to study the Enlightenment from a reception viewpoint, studying not only the circulation of books but also their potential readers. In its first phase, the project builds on the comprehensive corpus of Dutch auction catalogues put together by the late book historian Bert van Selm and his collaborators, starting in the 1980s, and now digitized by Brill as Book Sales Catalogues Online (BSCO). This corpus comprises digital images of microfiches, produced over several decades, of 4756 book sales catalogues printed in the Dutch Republic between 1599 and 1801 (including a very small number of doubles), harvested from over fifty libraries across Europe.⁷ Of these, 2536 are catalogues of private libraries, and about half of these are 'smaller' catalogues comprising fewer than 1000 items. In addition, there are 686 anonymous catalogues, some of which can through additional research be attributed to individual owners. The BSCO platform also provides metadata gathered by Van Selm and his collaborators on the catalogues (principally catalogue type, price, full text of title page and collation), collectors (name, profession, residence) and auctions (including place, date and name of auctioneer); the metadata that is currently inventoried – not all of it curated, as yet – is freely available on the Brill platform. Building on the BSCO corpus, MEDIATE seeks to substantially expand this corpus with private library catalogues from France, the British Isles and Italy, so as to create a union catalogue of private library catalogues, and a digital corpus that will eventually number 2000 catalogues of private libraries sold at

6. Franco Moretti, *Graphs, maps, trees: abstract models for literary history* (London, 2005), p.4.
7. Including collections in the Netherlands, Germany, Great Britain, France and Russia. For a full list of libraries, see Bert van Selm, J. A. Gruys and H. W. de Kooker's corpus Book Sales Catalogues Online, continued by Karel Bostoen, Otto Lankhorst, Alicia C. Montoya and Marieke van Delft, <http://primary-sources.brillonline.com/browse/book-sales-catalogues-online> (last accessed 10 January 2020).

auction between 1665 and 1830. The latter will be available through an open access database comprising also searchable full-text versions of the catalogues themselves. The MEDiate project will hence focus on four geographic-linguistic zones: the Dutch Republic, the British Isles, France and Italy. These regions were selected as being broadly representative of the movement of ‘Enlightenment’ ideas in eighteenth-century Europe, including both Catholic and Protestant contexts. In addition, they were selected because of their key role in European-wide networks of book production and distribution, and because these regions – with the possible exception of Italy – are central to most current accounts of the Enlightenment. For reasons of feasibility, MEDiate leaves out the German states for the time being, but might include them at a later stage.

Defining source material, moving targets: from one to two databases

The project’s primary source material, private library auction catalogues, contains information on books circulating in Europe in Latin, French, Dutch, German, English, Hebrew and several other languages. This material represents an extraordinarily rich primary source for research on the history of the book and libraries, as well as the history of ideas. The project posits that book ownership, regardless of whether books listed in private library catalogues were actually read, is in itself a significant indicator for the study of the spread and reception of ideas – even if catalogues do not necessarily reflect the full extent of an individual’s book ownership or reading choices made during a lifetime. Reported book ownership may in addition provide valuable indications about the intellectual aspirations of the collector, the association of specific social or professional groups with specific kinds of reading material, the relative prestige assigned to particular books as a form of cultural capital, and booksellers’ evaluation of books’ monetary worth.

Private library catalogues have long been central to research on book ownership, as well as broader questions regarding access to books, reading culture and the circulation of ideas in the eighteenth century. Yet, other than Daniel Mornet’s seminal article on ‘Les enseignements des bibliothèques privées (1750-1780)’,⁸ and a number

8. Daniel Mornet, ‘Les enseignements des bibliothèques privées (1750-1780)’, *Revue d’histoire littéraire de la France* 17 (1910), p.449-96.

of follow-up studies by other scholars,⁹ there have been few attempts to systematically analyze the contents of the enormous corpus of eighteenth-century private library auction catalogues still extant today,¹⁰ and to identify the mass of books actually consumed during this period – despite some attention paid by bibliographers and librarians to the auction catalogue genre in specific national contexts, such as the Ecole des chartes's electronic database *Esprit des livres: catalogues de vente de bibliothèques de l'époque moderne*.¹¹ There is clearly much pioneering work still to be done in this field, as our project team discovered during *MEDIATE*'s first project year.

Thus, work on the project started by looking more closely at our source material and defining its basic features in terms of possible data ontologies. *BSCO* distinguishes thirty-one distinct catalogue types, from the fairly straightforward 'auction catalogue' to the rarer 'satirical auction catalogue' or 'catalogue of a raffling of books' – yet even this typology is incomplete, as we discovered when faced with the first handwritten, domestic inventory of a library. It quickly emerged that practices of drawing up library inventories varied from country to country, for instance in relation to prevailing regimes of censorship. At the same time, sources the project team had initially thought to exclude from the corpus – most notably, manuscript inventories produced for non-commercial uses – later appeared important to include, especially in regions like France and Italy, if we were to properly address issues concerning the reach of books and ideas in the long eighteenth century. We decided therefore to adopt a catholic attitude, including in our

9. For example, for French books, S. A. Krijn, 'Franse lektuur in Nederland in het begin van de achttiende eeuw', *De Nieuwe Taalgids* 11 (1917), p.161-78, and Michel Marion, *Collections et collectionneurs de livres au XVIII^e siècle* (Paris, 1999). For an overview of work done on library auction catalogues, see *Les Ventes de livres et leurs catalogues, XVII^e-XX^e siècle*, ed. Annie Charon and Elisabeth Parinet (Paris, 2000), and Helwi Blom, Rindert Jagersma and Juliette Reboul, 'Printed private library catalogues as a source for the history of reading in seventeenth- to eighteenth-century Europe', in *The Edinburgh history of reading*, ed. Mary Hammond (Edinburgh, 2020), p.249-69.
10. Despite regularly repeated calls to do so, e.g. Dominique Bougé-Grandon, 'Vers la creation d'une base de données des catalogues de vente français?', in *Les Ventes de livres et leurs catalogues*, ed. A. Charon and E. Parinet, p.197-202; Didier Masseur, *Les Ennemis des philosophes: l'antiphilosophie au temps des Lumières* (Paris, 2000), p.6.
11. Ecole nationale des chartes (Annie Charon), *Esprit des livres: catalogues de vente de bibliothèques de l'époque moderne / Manuscrits mis en vente*, <http://elec.enc.sorbonne.fr/cataloguevente> (last accessed 10 January 2020).

database all catalogues of private libraries we encountered, even if MEDiate's primary focus remained on sourcing printed catalogues through existing printed bibliographies – Bléchet and Charon for France, Munby and Coral for the British Isles, Loh for Europe more broadly¹² – and through existing digital databases such as Gale Cengage's Eighteenth-Century Collections Online (ECCO).

We provisionally defined as a 'private library' any collection of at least ten books that could be linked in part or in whole to one or more named owners or collectors. No conceptual distinction was made at this stage, therefore, between a 'library' and a 'collection of books', or a coherently structured ensemble of books, presupposing some degree of intentionality on the owner's part, as opposed to a number of loose volumes in his or her possession. Similarly, we drew no conceptual distinction at this point between a book 'collector' and 'owner'. While noting that these concepts may require more critical unpacking at a later stage, in the first project year, the need to make practical choices in harvesting data encouraged us to cast the net as widely as possible. In terms of the metadata fields already used in the BSCO platform, MEDiate defined a private library catalogue as a printed or manuscript catalogue belonging to one of six major categories:¹³

- auction catalogue private library
- sales catalogue private library
- printed catalogue private library
- manuscript catalogue private library

12. Françoise Bléchet, *Les Ventes publiques de livres en France, 1630-1750, répertoire des catalogues conservés à la Bibliothèque nationale* (Oxford, 1991); Charon, *Esprit des livres*; A. N. L. Munby and Lenore Coral, *British book sale catalogues 1678-1800: a union list* (London, 1977); Gerhard Loh, *Die Europäischen Privatbibliotheken und Buchauktionen (1555-1732)*, 7 vols (Leipzig, 1997-2005).
13. In addition, all these types of library catalogue may be part of larger catalogues of household goods or other goods (prints, paintings, medical instruments, etc.) offered for sale or listed by the cataloguer. The only criterion for inclusion in the MEDiate database is that the catalogue list at least ten books by title. In practice, however, data harvesting focuses on catalogues devoted primarily to collections of books, and presented explicitly as library catalogues; the project team members will not actively be looking for collections of books that may have been included as part of larger inventories. Similarly, while making no fundamental distinction between printed and manuscript catalogues, data harvesting work will focus on printed catalogues for practical reasons of accessibility.

- probate inventory
- auction catalogue anonymous collection¹⁴

From the beginning, in defining data fields, we sought to reach a workable compromise between MEDiate's data needs on the one hand, and a desire to align our data with existing databases and datasets on the other. Thus, the first three categories of catalogues had been used already by Van Selm and his collaborators in their typology of thirty-one kinds of catalogue. As the MEDiate project moves forward, it will seek both to further consolidate existing data structures and categorization schemas, and to establish a workable schema crosswalk between ours and other projects' metadata – while yet remaining fully aware of the fact that 100 per cent equivalence is impossible to reach, and that some metadata fields will have to be discarded (as they already have been, in the case of BSCO's original data fields).

Furthermore, early discussions led the project team to decide to prioritize smaller and medium-sized library catalogues within the larger corpus of printed private (sales) catalogues, for both practical and conceptual reasons. Practically speaking, prioritizing smaller catalogues meant reducing the number of pages and items to be recorded in the database per catalogue, and subsequently maximizing the number of collectors who could be included in the corpus. At the same time, from a conceptual viewpoint, the choice to concentrate on smaller and medium-sized libraries would allow the project to counter to some extent the inherent bias produced by the survival rates of private library catalogues, which tend to privilege very large collections or collections that belonged to prominent scholars or political figures. Thus, book historians working on Dutch catalogues have estimated, on the basis of sales announcements in the periodical press, that the surviving auction catalogues represent only around 10 per cent of the total number of catalogue titles that were produced during that period, with wide variations between the cities where auctions were held and survival rates increasing in relation to the reputation of the collector.¹⁵

14. The MEDiate database also includes fields for 'auction catalogue multiple collections' and 'other', permitting us to include also borderline cases in the corpus of private libraries.
15. Hannie van Goinga, *Alom te bekomen: Veranderingen in de boekdistributie in de Republiek 1720-1800* (Amsterdam, 1999). For a breakdown per city and decade, see Rindert Jagersma, 'Dutch printed private library auction catalogues, 1599-1800: a bibliometric overview', in *Book trade catalogues in early modern Europe*, ed. Andrew Pettegree and Arthur der Weduwen (Leiden, forthcoming).

While recognizing that collectors who left library catalogues behind generally belonged to intellectual and social elites, MEDIATE's focus on smaller to mid-range collections was intended to help target what could provisionally be described as 'sub-elite', or perhaps even 'middle-class', collectors. Indeed, several studies suggest that smaller libraries are more likely to be 'choice libraries', reflecting the taste of the owner, as opposed to the well-known, large professional libraries of prominent public figures, and are more likely to belong to owners not part of the highest social or professional elites.¹⁶

A working definition of small and medium-sized libraries identified these as libraries reporting fewer than 1000 books, more or less – bearing in mind both that computing exact numbers of books, as opposed to auction lot items, entails complexities of its own,¹⁷ and that available metadata on some catalogues is limited, forcing us to estimate numbers of books based on number of catalogue pages.¹⁸ Thus Máire Kennedy reported a mean size of 500 to 1000 titles in eighteenth-century Irish private library auction catalogues.¹⁹ In his study of eighteenth-century Parisian libraries, rather than calculating mean library sizes, Michel Marion estimated that 60 per cent of the catalogues for the period 1740-1790 contained fewer than fifty pages, corresponding roughly to 1000 books or less.²⁰ In the Dutch BSCO corpus, similarly, 50 per cent of the catalogues contain fifty-two

16. Marion, *Collections et collectionneurs*; Alicia C. Montoya, 'French and English women writers in Dutch library (auction) catalogues, 1700-1800: some methodological considerations and preliminary results', in *'I have heard about you': foreign women's writing crossing the Dutch border, from Sappho to Selma Lagerlöf*, ed. Suzan van Dijk, Petra Broomans, Janet van der Meulen and Pim van Oostrum (Hilversum, 2004), p.182-216; Marie-Louise Coolahan and Mark Empey, "'There are numbers of very choice books": book ownership and the circulation of women's texts 1680-98', in *Women's writing, 1660-1830: feminisms and futures*, ed. Jennie Batchelor and Gillian Dow (London, 2016), p.139-57.
17. Some lots comprise multiple books; others do not include all volumes of a multivolume work; and in others, the bibliographic information provided is too sparse to determine exactly what is being offered for sale.
18. Since most bibliographies do not include numbers of books in their metadata, we made the initial estimate that a catalogue of fewer than fifty-three pages corresponded roughly to our 1000-book range, and consequently included all private catalogues of this size in our corpus – even if, as it emerged, our estimate turned out to be on the low side.
19. Máire Kennedy, *French books in eighteenth-century Ireland* (Oxford, 2001), p.130.
20. Marion, *Collections et collectionneurs*, p.80. However, do note that Marion also includes incomplete catalogues, or catalogues with missing pages, in his calculations.

pages or less. The decision to focus initial data harvesting work on library catalogues of 1000 books or less, or the lower 50 per cent of the extant corpus of catalogues in terms of size of collection, was therefore motivated by the statistical prevalence of this size of library in different geographic regions, together with the previously mentioned decision to privilege number of collectors over size of the libraries in our corpus.

This initial exploration of the MEDiate source material (i.e. private library auction catalogues and other types of library inventory) led to a basic database structure comprising four core classes: Collector, Catalogue, Lot and Item. Each of these core classes was associated with a set of properties that could further be defined using vocabularies drawn from the Europeana Data Model, Dublin Core, RDA and other widely used namespaces. However, once we moved into the realm of Item – the individual book or other item, listed in a catalogue lot – the properties began to multiply rapidly. Conversely, the richness of the metadata on catalogues already available through repositories such as BSCO, as well as the exponential growth of material that team members were able to harvest, added new layers of complexity to the database ontology. Indeed, such were the issues raised by the practical question of defining our source data with a sufficient degree of granularity that, a few months into the data harvesting phase of the project, the project team decided not to build a single database but two databases: one database for the books listed *within* the catalogues, the originally projected MEDiate database; and a second database, dubbed BIBLIO (Bibliography of Individually owned Book and Library Inventories Online), which, acting as a kind of union catalogue, would record metadata on all private library inventories we came across in the course of data harvesting. The additional database would thereby provide a repository for metadata on catalogues that had been sourced but not included in our final, more restricted MEDiate corpus of smaller and medium-sized libraries, but might nonetheless be useful to historians wanting to use them to address other kinds of research questions. Recognizing that some decisions regarding ontologies and what (not) to include in the database would also be dependent on future findings and the overall shape of our corpus, a second database would additionally permit us to go back and query the material anew later in the project, using new selection criteria. In short, this meant that the project output would now consist of two distinct but connected digital resources:

- BIBLIO database (Bibliography of Individually-owned Book and Library Inventories Online), containing bibliographic metadata for all extant private library (sales) catalogues and inventories in the Dutch Republic, France, Italy and the British Isles (including colonial possessions) for the period 1665-1830. BIBLIO's core classes, in order of importance, are Catalogue and Person; it contains no data in the Lot and Item classes;
- MEDIATE database (Measuring Enlightenment: Disseminating Ideas, Authors, and Texts in Europe), containing fully searchable transcriptions and book data extracted from a smaller corpus of around 2000 small to mid-range – fewer than 1000 books, approximately – private library (sales) catalogues for the same period. Besides data on books, this database contains more detailed biographical data on collectors. MEDIATE's core classes, in order of importance, are Lot, Item and Collector, with only basic data in the Catalogue class.

Defining the project's source data – private library auction catalogues and other kinds of library inventories – thus led us to far-ranging discussions on database ontologies, resulting in a first major overhaul of the project's projected output, and the move from a single to multiple databases. Three years into the MEDIATE project, its underlying data model remains a subject of ongoing, and sometimes heated, debate, as the project team moves forward in thinking about how to most accurately describe the object of our study.

Working transnationally: OCR and its discontents

Just as the practical questions raised by the initial work defining MEDIATE's source material underlay the move from a single database to two databases, the project's framing concepts, too, contributed to shifting our original perspective and rethinking ultimate project aims. MEDIATE's original framework privileged three aspects of eighteenth-century culture related to the Enlightenment movement: transnational relations; the heuristic concept of 'middlebrow culture'; and religious Enlightenment. Below, I address the first of these two briefly, having published elsewhere on the concept of religious Enlightenment.²¹

21. See among others Alicia C. Montoya, 'Madame Leprince de Beaumont et les "Lumières religieuses"', in *Marie Leprince de Beaumont: de l'éducation des filles à La Belle et la Bête*, ed. Jeanne Chiron and Catriona Seth (Paris, 2014), p.131-44.

A central framing concept was provided by the notion of transnational relations. Recent studies on literary history have argued for the importance of taking into account not only national but also transnational contexts, paying particular attention to the unequal power relations between various players in the literary field (i.e. to also interrogate centre–periphery relations in studying the spread of new books and ideas).²² Not only have Enlightenment studies as a whole undergone a global turn, with concepts such as the global or Atlantic eighteenth century now at the foreground of scholarship. In the field of book history, too, attention is increasingly being paid to interactions between different geographical regions.²³ This has led historians to question whether the diffusionist narrative of the Enlightenment as a movement spearheaded by an international group of intellectuals, including most notably the Parisian *philosophes*, truly reflects the historical reality of the Enlightenment as perceived on the ground.²⁴ Were the writings of this intellectual coterie really as influential, beyond elite circles, as they held themselves? How can we measure the impact of progressive intellectual debates beyond Paris, London and Berlin, and beyond court and elite circles, among more ‘ordinary’ readers?

MEDIATE’s original focus on transnational relations proved, from the outset, to be crucial in understanding our source material. Indeed, once the project team members started looking more closely at the corpus of private library auction catalogues, it became clear that viewing these within a narrowly nationalist framework, as much scholarship has in the past, would leave out much that gave them their unique interest. Even to the most casual of observers, it is immediately evident that many, if not most, of the catalogues list books in more than one language, and that they do not restrict themselves to books

22. For example, Pascale Casanova, *La République mondiale des lettres* (Paris, 1999).

23. See for example Jeffrey Freedman, *Books without borders in Enlightenment Europe: French cosmopolitanism and German literary markets* (Philadelphia, PA, 2012); *Books between Europe and the Americas: connections and communities, 1620–1860*, ed. Leslie Howsam and James Raven (London, 2011); Kennedy, *French books; Better in France? The circulation of ideas between Britain and the continent in the eighteenth century*, ed. Frédéric Ogée (Lewisburg, PA, 2005).

24. See among others Emma Rothschild’s proposal to study ‘the middling Enlightenment’. Emma Rothschild, ‘Inner shuddering in the French provinces’, paper presented at the Society for French Historical Studies 59th Annual Meeting, Cambridge, MA, 5 April 2012, <https://www.youtube.com/watch?v=ty340Q5GunI> (last accessed 10 January 2020).

printed within their own regional or national context. Thus an earlier study recorded a figure of 10-20 per cent French-language books in a corpus of a hundred Dutch auction catalogues from the period 1700-1750,²⁵ while a smaller sample of twenty-five catalogues from the years 1754-1802 produced a slightly different distribution: 15.4 per cent French-language titles, 2.9 per cent German books and 2.2 per cent in English. In a pilot study based on the *MEDIATE* corpus, comprising seventy-two Dutch catalogues sold between 1670 and 1750, Rindert Jagersma and I recorded 11.09 per cent books in French, 28.93 per cent books in Dutch and 59.98 per cent books in other languages, primarily Latin.²⁶ While Dutch catalogues probably present an extreme case of national and linguistic heterogeneity, the same picture of diversity holds also for catalogues printed in France, Italy and the British Isles. Readers, then as now, did not limit themselves to their own national contexts in defining their reading choices, even if literary historiography today, faithful to its nineteenth-century Romantic roots, continues to take national frameworks as its point of departure.

The catalogues thus presented us with immediate practical problems. Given their linguistic heterogeneity, we would need to develop optical character recognition (OCR) technology capable of capturing data in multiple languages, in many cases in different character sets. In the Dutch catalogues, it is quite common to find Dutch-language books listed in a Gothic character set, German-language books listed in a second, different Gothic font, Greek books listed in Greek characters, and French books listed in italic or Roman fonts – all on the same page. The decision to include Jewish catalogues and Jewish books in the corpus, hence also Hebrew character sets, added a new layer of technical problems that went well beyond UTF-8 encoding in our Excel files. These issues were further compounded by the poor image quality of the digital scans – in fact, reproductions of microfiches – and the often hasty typesetting of auction catalogues in general, which produced texts with inconsistently used fonts, fluctuating baselines and varying concentrations of ink.

Of course, *MEDIATE* is not the first project to run into the problem of early modern print material in different languages and

25. Krijn, 'Franse lektuur in Nederland', p.165.

26. Alicia C. Montoya and Rindert Jagersma, 'Livres français dans les catalogues de vente aux enchères des bibliothèques privées dans les Provinces-Unies (1670-1750)', in *Littérature, livres et librairie au XVII^e siècle: actes du colloque NASSCFL à Lyon* (Tubingen, forthcoming).

character sets, and the difficulties of using existing OCR technology to convert this material into mineable text files. Various initiatives have been undertaken to crack this problem, with varying success.²⁷ In the original grant proposal, I had naively announced that the MEDiate project would itself ‘develop innovative OCR technology and software enabling record linkage in the overall corpus of auction catalogues’. Given that auction catalogues are a semi-structured text form, we would additionally ‘develop a new text-mining algorithm’. However, it quickly became clear that developing an appropriate OCR pre-processing and post-correction workflow ourselves represented an enormous technical challenge. Not only would project team members need to put in a substantial amount of manual transcription work in order to create sufficient training material for our OCR software, but manual keying, even with the prospect of crowdsourcing, together with a test run, a workable transcription protocol and a dedicated group of a dozen undergraduate student assistants at our disposal, proved to be maddeningly slow, setting back the delivery date at which the first usable corpus of catalogue data would be ready.

Shortly after the decision to downscale the technical element, fortuitously for MEDiate, another solution to our data harvesting and transcription problem presented itself, through the intermediary of partners at the KB–Royal Library in The Hague, which hosts the Dutch Digital Library, or *Digitale Bibliotheek voor de Nederlandse Letteren* (DBNL). This is a Dutch–Flemish digital repository, launched in 1999, containing thousands of primary and secondary texts covering Dutch-language literature from the Middle Ages to the present, available both in text format and as digital scans. A large proportion of these texts date from the early modern period, and are similar in many ways to our own source material; the corpus even includes a few auction catalogues. Since 2003, the Royal Library has been outsourcing the digitization of these texts to a commercial

27. Thus the Early Modern OCR Project, led by Laura Mandell, and its various successors have for several years now been seeking to ‘make access to texts more transparent’, specifically ‘texts with fluctuating baselines, mixed fonts, and varied concentrations of ink (among many other variables)’. Although the eMOP homepage optimistically announced in 2012 that ‘by using innovative applications of OCR technology and crowd-sourced corrections, eMOP will solve this OCR problem’, to date early modern OCR continues to pose a challenge for many projects. See the eMOP project website <http://emop.tamu.edu> (last accessed 10 January 2020) and its successor project website *Reading the First Books*, <https://sites.utexas.edu/firstbooks> (last accessed 10 January 2020).

company, SPi Global, who has digitized more than 800,000 pages to date. In an ongoing, long-term partnership, SPi is now converting into searchable PDFs, and full text XML, a total of about 40,000 hardcopy books approximating over 10 million pages. Similar partnerships exist since 2000 between SPi and the US-based Text Creation Program (TCP), supporting the creation of full text electronic editions of thousands of works, including both the EBBO (Early British Books Online) and ECCO databases. To date, almost 1.5 million pages have been converted, with SPi performing full text conversion from TIFF inputs to SGML outputs via data entry and OCR, where possible.

Because the OCR software used in this process has by now been trained on thousands of pages of early modern material, this made it the best option for our own material. Work for MEDiate thus takes place through a combination of tailor-made OCR software, which combines multiple OCR engines, and manual post-correction. The initial output consisted of structured transcriptions in plain text format, CSV files of these transcriptions, and data ready to be loaded into an SQL database. As this volume goes to press, hundreds of outsourced, structured transcriptions are being loaded into our own MEDiate database. Further speeding up the work, we are automating the extraction of data such as dates and place names from these transcriptions. Outsourcing MEDiate's transcription work has allowed the project to readjust output targets so that, by project's end, transcriptions (of which 600 will be corrected and enriched) will be available of 2000 catalogues – as opposed to some 600, if we had done the manual keying ourselves with the help of our undergraduate student assistants.

Shifting perspectives: operationalizing 'Middlebrow'

A second framing concept in the original MEDiate project that pushed team members' work in specific directions was provided by the idea of 'middlebrow'. To better understand how texts functioned within the eighteenth-century literary field, the concept of middlebrow culture, first developed in English and American literary criticism,²⁸

28. For the concept of middlebrow as applied in nineteenth- and twentieth-century Anglophone literary history, see among others Brown and Grover, *Middlebrow literary cultures*, and Joan Shelley Rubin, *The Making of middlebrow culture* (Chapel Hill, NC, 1992).

provided a useful – if not altogether unproblematic – heuristic tool.²⁹ Backdating the historical emergence of the middlebrow publishing phenomenon, MEDiate postulates the existence of a corpus of texts in the eighteenth century that belong neither to the high Enlightenment studied by the history of ideas, nor to the popular, ‘low’ reading material studied by historians of the book, such as works of popular devotion and the chapbook literature of the *Bibliothèque bleue*. While the most radical philosophical writings, at one end, were actively consumed only by a small elite of readers, almanacs, chapbooks, catechisms and prayer books, at the other end, reached even peasant, non-literate audiences.³⁰ Both the high Enlightenment and, to a lesser extent, popular reading material, have been addressed by scholarship. What is missing, however, is a plausible account of how these two kinds of texts might be related to one another. In diffusing Enlightenment ideas among a broad reading audience, MEDiate hypothesizes, a crucial role was played by a middle tier of publications, composed of often religiously coloured, pedagogical writings targeting a non-learned, largely provincial and sub-elite audience that also included women and children. These texts combined ideas from the literature of the high Enlightenment with discursive structures and cultural traditions associated with more popular, often religious reading material, thereby ensuring their accessibility to a broad swathe of readers. Their mediating role was, therefore, a multidirectional one. Not only did they bring Enlightenment ideas to a non-elite audience; they also drew on non-elite, non-metropolitan cultural models and needs in reformulating the aims of the Enlightenment. This category of publications, hence, meets several of the criteria proposed by literary scholars to define the category of middlebrow as it arose in a later age and geographical context. These traits include commercial success; an overtly moralistic discourse; female authorial over-representation; accessibility to a non-elite reading audience; popularizing vocation and themes perceived as part of a middle-class

29. Space limitations do not allow me to address the issues involved in projecting a normative, twentieth-century category onto eighteenth-century material. For a more detailed discussion, I refer readers to my essay ‘Madame Leprince de Beaumont et la littérature “médiocre” (middlebrow)’, in *Une Educatrice des Lumières, Marie Leprince de Beaumont*, ed. Rotraud von Kulesa and Catriona Seth (Paris, 2018), p.205-26.

30. Roger Chartier, ‘Lectures et lecteurs “populaires” de la Renaissance à l’âge classique’, in *Histoire de la lecture dans le monde occidental*, ed. Guglielmo Cavallo and Roger Chartier (Paris, 1997), p.337-54.

ethos; and a supposedly ‘middling’ literary quality, in the eyes of influential literary and cultural gatekeepers. The values these texts reflected – including the usefulness of education, sentimental appeal, domestic ideals and economic self-reliance – helped shape the ethos of a specific social class, the bourgeoisie, which was coming into its own during the eighteenth century.

Indeed, a defining feature of more recent middlebrow literature that has been explored by literary scholars is its vexed links to class. As Erica Brown and Mary Grover write about twentieth-century middlebrow, the cultural authorities engaged in the 1920s in constructing the modernist aesthetic polemically and significantly opposed this new form of ‘high’ literature, produced by a self-conscious, self-designated intellectual elite, to “‘strap-hanging” typists commuting to and from the new suburban wastelands of interwar Britain’. Brown and Grover argue that the term ‘middlebrow’ has been subjected to a variety of rhetorical uses that need to be carefully examined, especially as these imply that middlebrow may in many cases be synonymous with middle-class. ‘The term “middlebrow” itself’, they note, ‘is the product of powerful anxieties about cultural authority and processes of cultural transmission. It is a nexus of prejudice towards the lower middle classes, the feminine and domestic, and towards narrative modes regarded as outdated.’³¹ Now in eighteenth-century Europe, cultural critics obviously did not use the term ‘middlebrow’, but they did insistently reference the middle-class values of authors they critiqued. Voltaire ironically referred to Marie Leprince de Beaumont, an author I have identified elsewhere with literary ‘middlebrow’,³² as a ‘shopkeeper’ or *magasinière*.³³ Another influential critic hypothesized that her titles had been inspired by ‘le génie de la Nation Angloise, singulièrement adonnée au commerce’.³⁴ Beaumont herself insisted that ‘si j’étais distributrice des marques d’honneur, je ne balancerai pas à accorder une statue au premier homme qui a eu le courage de s’élever au-dessus du préjugé ridicule, qui fait mépriser le commerce et l’agriculture.’³⁵ Leading

31. Brown and Grover, *Middlebrow literary cultures*, p.1.

32. ‘Madame Leprince de Beaumont et la littérature “médiocre” (middlebrow)’.

33. Voltaire, *Correspondence and related documents*, ed. Th. Besterman, in *Œuvres complètes de Voltaire*, vol.85-135 (Oxford, 1968-1977), vol.118, p.47 (D15202).

34. Joseph de La Porte, *Histoire littéraire des femmes françaises*, 5 vols (Paris, Lacombe, 1769), vol.4, p.364.

35. Marie Leprince de Beaumont, *Magasin des adolescentes, ou Dialogues entre une sage gouvernante et plusieurs de ses élèves de la première distinction*, 4 vols (London, n.n., 1760), vol.1, p.xxv.

Parisian critics spoke condescendingly of these authors' lack of literary style or, as they significantly put it, lack of 'style noble'.³⁶ Like a work that is in some ways similar, but much better remembered, Benjamin Franklin's *Poor Richard's almanac*, the corpus of middlebrow writings' 'real significance is not that it proclaims the virtues of capitalism, but rather that it linked high and popular culture [...] communicating the virtues of the common man to the elite and the benefits of scientific, literary, and historical knowledge to ordinary folk'.³⁷

The properties of middlebrow culture described above point to a complex series of interrelated questions that the MEDiate project sought to operationalize effectively. Foremost of these is what, exactly, is middlebrow: books, or people – and how do the two relate to one another? How can MEDiate effectively identify those books that could be described as belonging to the category of middlebrow? Is commercial success enough, and can we simply compile a list of most frequently cited titles in the catalogues to establish a thesaurus of 'middlebrow' titles? Or does commercial success need to be coupled with specific thematic elements present in those books? How does 'middlebrow' correlate with the kinds of books typically described in popularizing works on building an ideal library?³⁸ Here as elsewhere, it appears crucial to combine quantitative with qualitative approaches, broad statistical overviews with case studies, in order to reach satisfactory answers to these questions.

Conversely, considering middlebrow as cognate with 'middle-class' introduces a new set of questions to be resolved – apart from the difficulties, that is, of applying the category 'middle class' itself to a historical period in which modern class consciousness was arguably still in an embryonic state. Is it possible to operationalize social class in our corpus of collectors, and to establish representative sample populations? Given the thin spread of the MEDiate data over long periods of time and multiple countries, with sometimes no more than one catalogue per year per country, representativeness appears an illusory, or at best a very long-term, goal. A truly representative

36. Nadine Bérenguier, *Conduct books for girls in Enlightenment France* (London, 2011), p.144-45.

37. William Pencak, 'Poor Richard's almanac', in *A Companion to Benjamin Franklin*, ed. David Waldstreicher (London, 2011), p.275-89 (288).

38. Alicia C. Montoya, 'Private library auction catalogues and changing views on the ideal library, 1654-1832: from early modern self-fashioning to (imagined) reading communities', in *Book trade catalogues in early modern Europe*, ed. Andrew Pettegree and Arthur der Weduwen (forthcoming).

corpus of collectors, in terms of distribution across professions and income groups, might be envisaged (if at all) only many years, if not decades, after our own project's end, and will be dependent on collecting sufficient data on book collectors. We had run up against the fundamental problem, in much historical research dealing with early modern and Enlightenment source material, of incomplete datasets.

Indeed, it emerged quickly that the population of collectors represented by the auction catalogues was an exceptional one, by any measure, and would not yield sufficiently large numbers of 'middling-sort' individuals to warrant a productive use of the term 'middle class' as an analytic category. Based on a pilot sample of 254 Dutch collectors,³⁹ the population appears very similar, in terms of professional affiliation and social positioning, to the population of Parisian collectors studied by Michel Marion in a previous study (see Table 1: I retain Marion's original French-language terms in order to ensure maximum precision).

This is, by no measure, a population belonging to what could be termed the 'middle classes', but represents, rather, the members of intellectual and professional elites during the eighteenth century. The clear conclusion is that, if 'middlebrow' is to be retained as an analytic category, it will necessarily have to apply primarily to the content of books and the values they express, as opposed to a more positivistic characterization of the population of collectors.

By contrast, the non-representative nature of the population of library owners did mean that we also discovered, during the first project year, that there was a surprising amount of biographical data to be sourced about many of these individuals. For a spin-off case study on the reception of the books of the naturalist Maria Sibylla Merian,⁴⁰ student assistant Malou Brouwer collected basic biographical data on the collectors of Merian's books, and turned up much new information, for a large part through relatively straightforward Internet searches, to enrich our understanding of their collections. Perhaps the most

39. Drawn from Montoya, 'French and English women writers in Dutch library (auction) catalogues'. This sample is representative of the BSCO corpus as a whole, but not of the more restricted *MEDIATE* corpus, since it also includes larger libraries. The *MEDIATE* figures will therefore probably differ slightly from these.

40. Alicia C. Montoya and Rindert Jagersma, 'Marketing Maria Sibylla Merian, 1720-1800: book auctions, gender, and reading culture in the Dutch Republic', *Book history* 21 (2018), p.56-88.

Table 1: Occupational/social status of known library owners.
The figures add up to over 100 per cent because some collectors
fall into two categories.

Montoya 2004		Marion 1999	
law and government	27%	<i>parlement + offices + avocats</i>	36%
aristocracy	9%	<i>noblesse</i>	7%
military and navy	2%	<i>militaires</i>	4%
education and scholarship	10%	<i>académies</i>	7%
religion	11%	<i>clergé</i>	17%
medicine	7%	<i>médecins</i>	5%
industry and commerce	2%	<i>finances</i>	8%
the arts (including literature)	2%	[no equivalent category]	–
unknown	39%	<i>divers</i>	16%

important conclusion of this first case study was that, despite the presence most often of only a single owner's name on catalogue title pages, eighteenth-century book ownership needs to be conceived in broader terms, as a collective phenomenon, with access to books often passing through multiple family and professional networks. As James Raven notes, 'many apparently private libraries were designed for display and for use by friends and neighbours'.⁴¹ Similarly, the practice of opening scholars' libraries to interested students and visitors was widespread throughout Europe. Thus, while a catalogue may list a single collector's name on the title page, books could be part of a well-known scholar's collection or family library and hence circulate widely among professional colleagues, family members and beyond. Reading itself often took place in family and group settings, and was conceived of as a collective activity, as Abigail Williams has argued in her important recent study, *The Social life of books: reading together in the eighteenth-century home*.⁴² This collective dimension of book access and

41. James Raven, 'From promotion to prescription: arrangements for reading and eighteenth-century libraries', in *The Practice and representation of reading in England*, ed. James Raven, Helen Small and Naomi Tadmor (Cambridge, 1996), p.175-201 (176). On the collective or 'public' nature of many supposedly private libraries, see the essays collected in *Before the public library: reading, community, and identity in the Atlantic world, 1650-1850*, ed. Kyle B. Roberts and Mark Towsey (Leiden, 2017).

42. Abigail Williams, *The Social life of books: reading together in the eighteenth-century home* (New Haven, CT, 2017).

reading is particularly significant when studying the relationship and access to books of social groups under-represented in the surviving sources. For women, for example, these ‘borrowing and sharing networks’ could prove crucial instruments to secure books.⁴³ The pilot study on Merian collectors turned up scores of spouses, children and other family and household members who would have had access to the library of the collector named on the title page, underlining the importance of understanding owners not as individuals but as actors embedded in complex social networks. Moving from concepts of individual ownership to collective reading practices will, needless to say, entail a new rethink of the MEDiate data model, as well as the possible addition of new features to the future database interface to allow for fine-grained social network analysis.

Collaboration: linked open data and interoperability

From the outset, the MEDiate database planned to work towards collaboration through linked open data or even full-scale integration with a number of other digitally supported bibliometric projects with which contact had previously been established. These included most notably Simon Burrows’ Mapping Print, Charting Enlightenment (MPCE) database, a follow-up to the FBTEE or French Book Trade in Enlightenment Europe database, which privileges different types of book history sources from ours, including publishers’ archives and registries of *permissions simples*.⁴⁴ Such collaboration across databases is crucial in order to counter the biases inherent in printed auction catalogues as a source. More importantly, by creating an aggregation of eighteenth-century book history datasets, we aim to create tools allowing users to map the circulation of books among a large cross-section of readers in Europe, and to draw historically plausible conclusions about the movement of ideas during this period. In the original grant proposal, besides MPCE–FBTEE, I had suggested a number of other digital bibliometric projects for possible collaboration. During the first project year, MEDiate reached collaboration

43. Susan Staves, “‘Books without which I cannot write’: how did eighteenth-century women writers get the books they read?”, in *Women and material culture, 1660-1830*, ed. Jennie Batchelor and Cora Kaplan (London, 2007), p.192-211 (208).

44. Mapping Print, Charting Enlightenment, <http://fbtee.uws.edu.au/mpce> (last accessed 10 January 2020). See also the chapter by Simon Burrows in the present volume.

agreements with three other partner projects: the US-based Footprints: Jewish Books through Time and Place database;⁴⁵ the Short-Title Catalogue Netherlands (STCN),⁴⁶ based at the KB–Royal Library, with the curator for early printed editions (pre-1800) Marieke van Delft; and Andrew Pettegree’s Universal Short-Title Catalogue (USTC) project,⁴⁷ based at the University of St Andrews.

In all cases, the most immediate basis for collaboration with partner projects as presently being pursued is the exchange of datasets; initially, this was done at the artisanal level of exchanging Excel files, while establishing schema crosswalks between the various database schemas for future, more efficient forms of data exchange. This collaboration brought to the fore with ever greater clarity the fact that different projects are currently dealing with data and datasets that come in different shapes and sizes – and not only because the underlying source material is different. At present, there is no international agreement among librarians on minimum requirements in standardizing book descriptions in order to make data exchangeable. For example, in order to ensure interoperability further down the line, MEDiate decided to link all book titles identified in the auction catalogues to existing authority files, designated by Virtual International Authority File (VIAF) numbers for Works.⁴⁸ However, we discovered that in the case of Hebrew books, many VIAF numbers are non-existent. Current Hebrew cataloguing practice is to note VIAF numbers for People (now known as Personal Names), not Works. As Columbia University librarian Michelle Chesner explains,

VIAF has not been used much by Hebrew catalogers (at least in the U.S.) for Works in the same way that it is for People. The National Library of Israel made a conscious decision a couple of years ago to record People in VIAFs, and so many libraries have turned to do the same for people with Hebrew names because of that [...] For Footprints, we’ve been using the Library of Congress standardized

45. Footprints project, led by Marjorie Lehman, Michelle Chesner, Adam Shear and Joshua Teplitsky, <https://footprints.ccnmtl.columbia.edu> (last accessed 10 January 2020).
46. KB–Royal Library, Short-Title Catalogue Netherlands, <https://www.kb.nl/en/organisation/research-expertise/for-libraries/short-title-catalogue-netherlands-stcn> (last accessed 10 January 2020).
47. Universal Short-Title Catalogue project, <http://www.ustc.ac.uk> (last accessed 10 January 2020).
48. Thus, VIAF functions essentially as a search engine indexing existing library authority files worldwide: see <https://viaf.org> (last accessed 10 January 2020).

titles for literary works, and we have tended to go by their schema (with some small exceptions). [...] we see our primary ‘union catalog’ as the Bibliography of the Hebrew Book – and all records to Hebrew books in Footprints link to the BHB. Our ultimate goal is to work with the National Library of Israel (which hosts the BHB) to connect the two databases in a more concrete way.⁴⁹

Compounding the problem of sometimes non-existent VIAF numbers is the impossibility of following international library procedure that prescribes that, to identify a specific Work or Person, preference be given to the VIAF description provided by the national library of the Work or Person’s country of origin. Not only are countries of origin in some cases difficult to assign, given shifting political boundaries over the centuries, but not all present-day states have national libraries that produce authority files for ‘their’ authors – as we discovered was the case, for example, even for a luminary such as Jean-Jacques Rousseau, who to date lacks a Swiss authority file.

It is clear, then, that there is an urgent need to reach internationally agreed-on minimum standards, so as to ensure the sustainability of data currently being generated and housed in our various databases. Awareness of the issue is widespread, and several consortia are working towards data standardization: MEDIANE seeks to identify these and draw on their models as much as possible, from the critical, early conceptual stages of database design and development. Thus, our data model makes use among others of vocabularies created by the Europeana Data Model, established and co-financed by the European Union, which ‘is not built on any particular community standard but rather adopts an open, cross-domain Semantic Web-based framework that can accommodate the range and richness of particular community standards such as LIDO [LIDO] for museums, EAD1 for archives or METS2 for digital libraries’.⁵⁰ Other promising initiatives include the European-financed COST Action Reassembling the Republic of Letters, 1500-1800,⁵¹ led by Howard Hotson, which

49. Personal communication, 22 December 2017.

50. Europeana Foundation, ‘Europeana data model primer’, https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf (last accessed 10 January 2020).

51. Reassembling the Republic of Letters, 1500-1800, <http://www.republicofletters.net> (last accessed 10 January 2020). See also the chapter by Howard Hotson in the present volume.

co-ordinated discussion amongst librarians and archivists, IT experts and scholars on the steps required to establish common standards and digital infrastructure enabling the European-wide collection of data on the early modern Republic of Letters, and the PARTHENOS (Pooling Activities, Research and Tools for Heritage E-research Network Optimization and Synergies) consortium of national research organizations, cultural heritage institutions and existing research infrastructures, currently funded under Horizon 2020 by the European Commission.⁵² For MEDIATE, one partner that emerged as particularly important was the Consortium of European Research Libraries (CERL), established in 1992 and housing among others the CERL Thesaurus of imprint places, imprint names, personal names and corporate names for the early modern history of the book.⁵³ Like Europeana, with whom CERL works closely, CERL is also creating its own namespaces, specific to the kind of book history source material central to MEDIATE, and is developing capabilities to provide data hosting over the long term, after funding has ended for currently running digital book history projects.

Although the standardization challenges we face are daunting, as a first step towards addressing some of the issues of interoperability, MEDIATE aims to work towards aggregating the datasets at least of its partner projects, creating an ecosystem of book history databases that can productively work on common source material and research questions. To this end, as a spin-off of the MEDIATE project, there are also future plans being made to develop a new common user interface, dubbed E-ENABLE (Early-modern – Enlightenment Networks of Authors, Books, and Libraries in Europe), to allow researchers to query our own and partner bibliometric book history databases at the same time, so ensuring maximum analytic benefits to researchers making use of the accumulated datasets.

Conclusion: from concepts to modelling?

During its first year, the MEDIATE project encountered a number of issues and surprises as we moved from the project's initial conceptual stages to operationalizing some of its key concepts and theoretical

52. PARTHENOS project, <http://www.parthenos-project.eu> (last accessed 10 January 2020).

53. Consortium of European Research Libraries, <https://www.cerl.org> (last accessed 10 January 2020).

frameworks, and moving more generally from concepts to modelling. It is useful, in this context, to recall Willard McCarty's crucial distinction between models – and, by extension, the practice of digital modelling – and concepts. 'Unlike the latter', he explains, a model 'instantiates an attempt to capture the dynamic, experiential aspects of a phenomenon rather than to freeze it into an ahistorical abstraction'.⁵⁴ A database conceived in this way is not a repository, but a research instrument, with important conceptual work taking place in the process of developing database structures. The 'experiential' and 'dynamic' aspects of data modelling were certainly a defining feature of our work during MEDiate's first project year, and there were a number of major changes to the original project plan. These include the move from one to multiple databases; and the choice to replace OCR processes developed in-house by outsourced transcriptions. Most importantly perhaps, the data harvesting phase proved to be a full-blown project on its own, with the iterative, perfective process itself emerging as perhaps our most important research 'deliverable'. Learning how to describe our object of enquiry in terms understandable to a machine, we learned much about what it is, exactly, we are studying, with technical and conceptual issues inextricably linked, at the basic level of defining database ontologies. Thus in August 2017, in the course of our first discussions with the USTC team, project director Andrew Pettegree not-so-innocently asked us: 'But what do these auction catalogues actually *represent*?' Our answer today would be, I think: that's exactly what we're trying to find out, little by little, as we piece together the different elements making up this mutable object. Similarly, if asked what 'the Enlightenment' has turned out to be, viewed through the emerging digital lens provided by the MEDiate and BIBLIO databases, the answer is that it is too early to tell, but that the process of understanding some of the tools and human actors that historically enabled the circulation of its ideas – catalogues, books and collectors – has proved to be an exciting ride indeed.

54. McCarty, *Humanities computing*, p.23.