



Open Library of Humanities



Part of the Ubiquity
Partner Network



Open Library of Humanities

Article

How to Cite: Ortolja-Baird, A, et al. 2019 Digital Humanities in the Memory Institution: The Challenges of Encoding Sir Hans Sloane's Early Modern Catalogues of His Collections. *Open Library of Humanities*, 5(1): 44, pp. 1–40. DOI: <https://doi.org/10.16995/olh.409>

Published: 26 June 2019

Peer Review:

This article has been peer reviewed through the double-blind process of *Open Library of Humanities*, which is a journal published by the Open Library of Humanities.

Copyright:

© 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>. Figures 1 & 3 excepted from the Creative Commons License and are public domain in most countries except the UK.

Open Access:

Open Library of Humanities is a peer-reviewed open access journal.

Digital Preservation:

The Open Library of Humanities and all its journals are digitally preserved in the CLOCKSS scholarly archive service.

ARTICLE

Digital Humanities in the Memory Institution: The Challenges of Encoding Sir Hans Sloane's Early Modern Catalogues of His Collections

Alexandra Ortolja-Baird¹, Victoria Pickering¹, Julianne Nyhan², Kim Sloan¹ and Martha Fleming¹

¹ British Museum, UK

² UCL, UK

Corresponding author: Julianne Nyhan (j.nyhan@ucl.ac.uk)

Catalogues are the core documents of museum structure and meaning. Yet no significant computational analysis has been made of how catalogues from the early modern period are constructed or of the way their structure and content relate to the world from which collections are assembled. The Leverhulme-funded 'Enlightenment Architectures: Sir Hans Sloane's catalogues of his collections' (2016–19), a collaboration between the British Museum and University College London, with contributing expertise from the British Library and the Natural History Museum, seeks to change this. The Enlightenment Architectures project is analysing Sloane's original manuscript catalogues of his collections to understand their highly complex information architecture and intellectual legacies. In this article we explore some of the challenges of seeking to integrate the methods of digital humanities with those of cataloguing, inventory, curatorial and historical studies and of bringing such interdisciplinary approaches to bear on early modern documentary sources. We do this through two case studies that highlight the approaches to encoding Sloane's catalogues in TEI that Enlightenment Architectures has employed and the major challenges that these have brought to the fore.

Introduction

What are the challenges of seeking to integrate the methods of digital humanities with those of cataloguing, inventory, curatorial and historical studies and of bringing such interdisciplinary approaches to bear on early modern documentary sources? What kinds of productive collisions and misalignments occur when the aforementioned fields' understandings of early modern documents and information meet the technical demands of present-day computational modelling? What barriers must be overcome in order to use the methods of the digital humanities to develop new understandings of the early modern period? We will explore these questions by drawing on our research from the Leverhulme-funded project 'Enlightenment Architectures: Sir Hans Sloane's catalogues of his collections' (2016–19), a collaboration between the British Museum and University College London, with contributing expertise from the British Library and the Natural History Museum. Enlightenment Architectures is blending research in the fields of cataloguing and inventory studies with curatorial studies and digital humanities to ask otherwise unanswerable questions about how Sir Hans Sloane's (1660–1753) manuscript catalogues of his collection were written, organised, annotated and used.

Sloane's manuscript catalogues are 'paper tools' that allowed him and his amanuenses to classify, cross-reference and document his collections and library. They were also instruments through which Enlightenment knowledge was produced and circulated, and reifications of how Sloane and others understood the world in the early modern period. The Enlightenment Architectures project is analysing Sloane's original manuscript catalogues of his collections to understand their highly complex information architecture and intellectual legacies. We place particular focus on the informational units of which the catalogues are composed and wish to understand the structural relations between these informational units. To support this, we are encoding, or marking up, Sloane's catalogues in line with the Guidelines of the Text Encoding Initiative (TEI). The TEI Guidelines set out XML-based encoding methods for making texts of the humanities, social sciences and linguistics machine readable. TEI is a de-facto standard (Jannidis, 2009: 258). It has been described as being among the

'most significant intellectual advances that have been made in [digital humanities] and [TEI] has influenced the markup community as a whole' (Hockey, 2004: 16).

In recent years, a number of projects with strong digital humanities elements have focused on catalogue descriptions of manuscripts, for example, FIHRIST (Union Catalogue of Manuscripts from the Islamicate World) (FIHRIST, n.d.). Evolving out of the pilot Islamic Manuscript Catalogue on-line (OCIMCO, n.d.), which built 'a sustainable data format using a tailored schema for the open source TEI/XML metadata standard and incorporating established library standards for description', FIHRIST is now a UK-wide union catalogue, whose schema is open for use by other TEI catalogues (FIHRIST, n.d.). Important work has also focused on harvesting bibliographical information about geographically dispersed manuscripts and federating this information in new catalogues and databases. Manuscriptorium, for instance, is a freely accessible digital library of manuscripts, old printed books and other documents. The catalogue assembles descriptive metadata about these works in XML and directs users to their respective complex digital documents (CDD) (Manuscriptorium, n.d.).

Though Enlightenment Architectures seeks to contribute to the conversations opened by such projects, it differs from them fundamentally. In this project we do not view the creation of a new digital representation of Sloane's catalogues as an end in itself; rather, our focus is on identifying and analysing the information architectures of Sloane's catalogues and his, and his amanuenses', cataloguing practices. We are modelling the catalogues and encoding them in TEI in order to study this. Our focus is therefore as much on the act of modelling as it is on the resulting computational model, and we view Sloane's catalogues as 'bifocal data'—a window frame that we must concurrently look 'at' and 'through' (Sperberg-McQueen, 2018). This requires us, as far as it is possible, to privilege a historically-accurate representation of the informational entities of Sloane's catalogues over achieving conformance with the views of information that are implicit in 21st-century encoding specifications like TEI. As a result, we believe that the difficulties that we are encountering are productive and meaningful and that our work is casting new light on epistemologies

of the digital in the context of the early modern while pointing to the particular demands that are made on digital methods and tools by early modern collections and situated, humanistic knowledge.

This article discusses the specific example of Sloane's early modern catalogues and the challenges that we have encountered when seeking to use TEI to encode this material. Our work nevertheless has resonance for the institutions, individuals and communities across the globe who manage, research, curate, archive and simply even browse the many and extensive digital heritage collections that are available online.¹ Poole (2016) has observed:

When, in 1977, an expert cataloguer looked at an object and made a few marks about it on a postcard-sized catalogue card, they would little have expected that one day the information they were creating would form the basis of a rich, complex and interwoven cultural experience on the World Wide Web. But fast-forward 35 years, and that is exactly what has happened.

The digitisation and publication of digital 'collections' (including, for example, digital images, 3D representations of objects, catalogue records and machine actionable metadata about objects) has been shown to benefit heritage institutions, researchers, educators, specialist communities like the media, the creative industries and the general public in many ways. Collections online can: support, *inter multa alia*, new opportunities for research and teaching on and with previously inaccessible collections (see e.g. Hughes, 2012: 5–7); offer heritage institutions new opportunities to work with communities, like community-based heritage groups (Roued-Cunliffe and Copeland, 2017); and support the creation of new knowledge by, for example, enabling previously dispersed collection records and surrogates to be interlinked and visualised (Dietrich and Pekal, 2012).

¹ Alone, the aggregator Europeana, for example, currently provides access to more than 50 million 'art works, artefacts, books, films and music from European museums, libraries, galleries and archives' (Europeana, n.d.)

Simple digitisation and publication of unstructured text is rarely sufficient for allowing online collections to be used and transformed in the ways suggested above (Stork et al., 2018). Rather, it is necessary to make machine readable information that a computer usually cannot decipher unaided. This can include information in and about a digitised object; for example, that the string of letters 'red' in a catalogue is actually a colour name or that a given catalogue was written by Hans Sloane. It is necessary to do this to support the sophisticated search, interlinking, remixing and other actions that collections online can ideally support. Languages like the XML-based TEI, which is the focus of this article, are thus crucially important pillars of digital collections because they can:

[make] it possible for people to embed additional knowledge in the text, including interpretative material. The purpose of text tagging is to facilitate retrieval and representation through applying what is essentially a controlled vocabulary of tags. A collection with an interpretative level of tagging is one where information is included in the tags that is otherwise not available in the text. (Ruecker, Radzikowska & Sinclair, 2016: 111)

Thus, a reader may ask 'why is TEI important? Why is it important to understand the benefits and complexities of applying TEI to early modern catalogues like that of Sloane?'. Our response is that TEI plays a crucial role in allowing sophisticated research questions to be asked of Sloane's catalogues and, in turn, it shapes the extent to which Sloane's catalogues can be intermeshed with the wider digital cultural heritage ecosystem that is discussed above.

Through the following case studies, we will discuss the approaches to knowledge representation that Enlightenment Architectures has employed and the major challenges we have encountered when seeking to apply TEI to early modern catalogues. In Case Study 1 we present examples of how we have customised and extended TEI so that it can better represent our historically-sensitive readings of Sloane's catalogues. In Case Study 2 we discuss the difficulties that we faced when

seeking to model object names and encode them in TEI. At stake, we argue, is not only how we can best use the methods of digital humanities to represent early modern catalogues but also the current limits of humanities and curatorial knowledge about such catalogues.

About Sloane

The globally significant collections of Sir Hans Sloane (1660–1753) were the foundations of three of the United Kingdom's national institutions: the British Museum, the Natural History Museum and the British Library. Sloane's collections of books, manuscripts, natural history, art, antiquities and ethnographic materials from around the world were a pivotal site of knowledge production and circulation during a period from the 1680s to 1750s and indeed in the British Museum after his death. Representing possibly the largest and most extensively documented of such collections, Sloane's handwritten catalogues are arguably among the first sustained attempts at collection management and information studies in the western world: as such, their intellectual legacies are unparalleled. As a royal physician, natural philosopher of apparently unlimited curiosity and both Secretary and President of the Royal Society, Sloane attempted to encompass the world and its knowledge through the creation of an encyclopaedic collection that would be left to the Nation upon his death in 1753.

Catalogues of collections were characteristic of early modern natural philosophy and went beyond a simple list or record of museum content (Findlen, 1996). In catalogues like Sloane's are the origins of modern methods for managing scholarly information (Blair, 2010). They can be conceptualised as 'human search engines' (Delbourgo, 2011). Sloane's catalogues are vital keys to unlocking not only his collection but also a greater understanding of the way knowledge was developed and produced. The differing priorities, rhetorics and documentation conventions of these catalogues provide rich information about the contribution of collecting to systems of 'rationality' that emerged in Sloane's era (Greenhill, 1992). Blakeway's research exemplifies the new knowledge that can be created through a close study of even a single subset of Sloane's catalogues (in this instance his library) (Blakeway, 2011). She

demonstrates how much work was involved in the act of cataloguing, by identifying through their handwriting the multiple authors of Sloane's library catalogues over time, the numbering and shelving systems adopted, and both contemporaneous and later chronology and re-ordering of the catalogue entries. Most recently, Kusakawa (2017) has examined inventory lists penned by William Courten which are held in Sloane's personal papers at the British Library and raises questions of how these items were re-listed and integrated into Sloane's catalogues once he acquired Courten's collection. Jones (1988) attempted to number and identify the present location of the original 54 Sloane manuscript catalogues and Caygill summarised their contents and potential for reconstructing 'the dazzling spectacle which amazed Sloane's visitors—to take a virtual tour of Sloane's museum, inspecting cabinets and opening drawers' (2012: 131). Yet, few of these publications have transcribed catalogue contents or analysed the evolving constituent parts of their structure and all emphasise the necessity of further research.

Inventory and cataloguing studies have been developed by historians of collecting for well over a century. Yet this research has often been carried out within a single discipline, such as art history (Keating and Markey, 2011). Sloane's 17th- and 18th-century manuscript catalogues present considerable research challenges to existing paradigms. No longer directly or consistently connected to his widely dispersed physical collections, they are far too extensive and complex to be studied without computational assistance. They also require broader disciplinary reach due to the encyclopaedic knowledge they represent (MacGregor, 1994). Sloane's catalogues and their complex, heavily annotated and indexical structures have consequently remained little understood and unanalysed; this project aims to change that.

Enlightenment Architectures focuses on five of Sloane's catalogues: two volumes of 'fossils', one volume of printed books and ephemera, one of 'miscellanies', and one of his collection of manuscripts.² All have been transcribed and we use this sub-set as a lens through which to best understand how collections and their documentation

² The titling of Sloane's manuscript catalogues is complex and current reference titles are not always those that were assigned by Sloane himself (see Caygill, 2012).

together formed a cornerstone of the laboratories of the emergent Enlightenment. To achieve this we are bringing cataloguing, inventory and curatorial studies into conversation with digital humanities and aim to devise and implement an interdisciplinary method bundle or bricolage that can create new knowledge about how Sloane's catalogues were written, organised, annotated and used. A cornerstone of our approach is the computational modelling of the catalogues, including making cataloguing-, inventory- and curatorially-informed readings of the catalogues machine readable, in line with the Guidelines of the TEI. We will now describe this process in greater detail.

Digital Humanities approaches

Data modelling is emblematic of computing. This is because:

models provide formalized perspectives on their subjects, expressed in a way that makes it possible to gather specific information about the subject. In short, the formalized model determines which aspects of the subject will be computable and in what form (Flanders and Jannidis, 2016: 229).

Modelling is accordingly a central activity of digital humanities and one of the main ways that it seeks to form and transform knowledge. In the main, it is 'modelling for' that is undertaken in digital humanities and this analytic approach aims to 'figure out how something works by taking it apart' (McCarty, 2008: 256; 2014: 26–29). Though such analytical work has a long history in the Humanities (Orlandi, 2002), the use of the computer as a partner in this process changes it substantially. When using a computer to model a catalogue such as that of Sloane, the model must be expressed within the constraints of computing technology: complete explicitness and consistency is required. In this way, computational modelling demands that humanities scholars identify and express interpretations of relevant textual features with an often-unprecedented degree of systematisation. Paradoxically, though, it has been argued that the greatest successes of modelling are to be found in its failures, or 'via negativa': '[modelling] gives us a tool for isolating that which will not compute and thus forces the epistemological question of how it is that we know what we really

know in the humanities' (McCarty, 2008: 256). This will be exemplified in case study 2 below with respect to the 'problem of the object' in Sloane's catalogues. Consequently:

models of whatever kind are far less important to the digital humanities than modelling. Modelling is crucial. If you only remember a single sentence from this brief essay, remember this one: the word 'computing' is a participle—a verbal adjective that turns things into algorithmic performances (McCarty, 2008: 254–5).

Thus, the ideal role of the computer and the purpose of computing in digital humanities is not to make research better, faster and/or cheaper. On the contrary, as a number of writers have argued, computing should be about making problems more difficult, more complex, more thrilling—computing is, or can be, 'a telescope for the mind' (Masterman, 1962).

Different approaches to the use of such a 'telescope' tend to be pursued in the institutions that specialise in cultural heritage and text-based humanities, or in the memory institution and the university. As a result, various studies (Eide, 2014; Ore and Eide, 2009) of how to bridge their modelling activities have been conducted:

Computer based modelling in cultural heritage has focused on database development, generalised as data standards and, since the 1990s, also formal ontologies. Modelling in digital humanities has had its core in textual scholarship, including close reading and text encoding of literary and historical sources as well as models of text corpora, usually relying on statistical methods (Ciula and Eide, 2014: 35).

As stated above, we are implementing the digital humanities modelling of Sloane's catalogues largely in line with TEI. This is an authoritative set of guidelines for making Humanities texts machine readable and is endorsed by agencies such as the NEH, AHRC and the EU's Expert Advisory Group for Language engineering (TEI Consortium, n.d.). Given that the primary location of the Enlightenment Architectures project (the British Museum) and that the locations of the objects that

are described in the catalogues are memory institutions, we carefully considered using a formal ontology such as CIDOC CRM rather than TEI as the basis of our work. CIDOC CRM is a conceptual model designed to provide 'definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation' (CIDOC, n.d.).

However, we concluded that TEI would be better for maintaining the integrity of Sloane's formally unstructured and continuous handwritten text, whilst exploring its information structures and its discursive development over time. This is because the aim of our modelling is to accurately represent the information architecture of Sloane's catalogues rather than to reconcile those information architectures with the concepts set out in a 21st-century encoding languages or ontologies. In CIDOC CRM, for example, '[t]he central idea is that the notion of historical context can be abstracted as things, people and ideas meeting in space-time' (Ore and Eide, 2009: 163). This is not a view of the catalogues that we were happy to commit to at the beginning of this work. Of course, implicit conceptual models underpin TEI too, but arguably not to the same extent and they are not articulated as such: 'The TEI guidelines are focused on how to annotate texts and do not prescribe any specific conceptual model' (Ore and Eide, 2009: 165). After careful consideration, we decided to use TEI as the master format for the project and integrate our annotations with that of an appropriate ontology at a later date, as other projects have done (e.g. Ciula, Spence & Vieira 2008). Nevertheless, the process of adapting and extending TEI to encode our material has presented significant challenges.

Though internationally recognised, TEI has been criticised from various angles. Earlier debates often centred on the theories of textuality that underpin it (e.g. DeRose et al., 1997; Renear, 1997; Renear, Mylonas & Durand 1993) and concerns of postmodern criticism, like performativity, that it poorly accommodates (Caton, 2000; McGann, 2007; 2004: 193–207). The appropriateness of embedded markup for cultural heritage texts has been questioned (see Schmidt, 2010). A recurrent point of concern is the complexity of TEI (e.g. Burghart and Rehbein, 2012; Dalmau and Hawkins, 2014; Dee, 2014), and the need for more user-friendly, TEI-compatible

tools. Communities like EpiDoc have developed specialist subsets of the overall guidelines for the encoding of epigraphic documents (EpiDoc, n.d.). In addition to the transcription and editorial treatment of texts, EpiDoc also addresses the history and materiality of the objects on which the texts appear (i.e., manuscripts, monuments, tablets) (Elliott et al., 2007). Work is also being undertaken on the visualisation of TEI (Del Turco et al., 2014) and on developing more user-friendly digital work environments for working with it (Dumont and Fechner, 2014).

Though TEI has maintained a distinction between 'document' and 'text', and shown an orientation towards the latter, the most recent iteration of the Guidelines facilitates manuscript description (Driscoll, 2006; The Text Encoding Initiative 2018) and a facsimile module has been added (Wittern, Ciula, & Tuohy, 2009). Nevertheless, the difficulties of using TEI to encode handwritten documents, including early modern manuscripts, is much discussed, as are special characters and abbreviations. The online edition of *The Book of Margery Kempe* has consequently:

established methods for embedding a custom font directly in a Web page—enabling Private Use Area character references to be used alongside their Unicode counterparts and resulting in the reliable display of special characters and abbreviations across Web browsers and operating systems (Fredell, Borchers IV, & Ilgen 2013).

The TEI additions needed to successfully encode Arabic manuscripts have resulted in proposed elements close to the 'semantic reality of the studied field (for example, a transliteration element and a copyist element which may be used similarly to the standard author element)' (Soualah and Hassoun, 2012: 9). Additionally, in addressing the difficulties of encoding the particular semantics of historical financial records, Tomasek and Bauman (2013) propose an encoding system not unlike the contextual markup of prosopographies or gazetteers using TEI P5.

A number of recent projects have also demonstrated how TEI can be extended to create rich digital scholarly editions. The Chymistry of Isaac Newton (The Chymistry

of Isaac Newton, n.d.) online edition of Newton's alchemical manuscripts encodes Newton's texts in TEI/XML so that they can be viewed in both diplomatic and normalised versions, and provides translations for Latin and other non-English texts as well as page images of the original manuscripts. The Newton Project (The Newton Project, n.d.) likewise provides similar usability, as well as encoding Newton's writings on the exact sciences with a combination of MathML and TEI-P5. A different approach is taken by The Map of Early Modern London (The Map of Early Modern London, n.d.) which comprises four distinct, interoperable projects (map, gazetteer, library and survey) whose databases share a common TEI tagset, thus enabling users to 'visualize, overlay, combine, and query the information in the MoEML databases' (Jenstad, 2018).

A number of projects have also sought to use TEI to interrogate historic or current catalogues. Adopted by libraries in particular, for example the Bodleian in Oxford, TEI has become a common framework for exploiting digitised catalogues. Another project, *The Digital Ark*, is:

a web-delivered virtual museum of collections of rarities and curiosities in England and Scotland from 1580 to 1700, comprising documentary and graphical representation of up to 10,000 specimens and artifacts collected in that period, some of them surviving in museums in England today (Nelson, 2016).

Similarly, the ASCH Project aimed to develop a metadata model to allow the contextualisation of different types of digitised resources (ASCH, n.d.). Using objects from the von Asch collection at the University of Göttingen, the project used TEI to encode the documents, such as letters and inventories, which referenced the objects, which were then linked to the metadata descriptions of the objects themselves.

However, Enlightenment Architectures fundamentally differs from these projects as it does not seek to address questions of provenance primarily, but rather of the organization of the information that is recorded in Sloane's catalogues. In the next sections we will discuss the difficulties that we have encountered when trying to

use TEI to encode the catalogues of Sir Hans Sloane, along with our current working solutions to such issues.

Case Study 1: applying and extending TEI

Upon his death in 1753, Sloane's library was estimated to contain some 50,000 volumes, over 400 of which were books and albums of prints and drawings and 2,666 were volumes of manuscripts—the rest were printed books (see Nickson, 1994). MS. 3972 C vol. VI is one of eight original volumes that contain the catalogue of Sloane's books and printed ephemera, now held at the British Library. Comprising 530 folio pages, in a variety of Sloane's and amanuenses' hands, it captures some of the richness of his library. It comprises catalogue entries for monographs, atlases and bound volumes of printed materials such as dissertations, treatises, proposals, letters, accounts and ephemera.

The majority of the catalogue's pages follow the entry layout in **Figure 1**. In the left-hand margin is the alphanumeric catalogue number, often crossed out, sometimes more than once, and replaced with a new number. To the right of this is the catalogue entry, which contains purely bibliographic detail: author, title

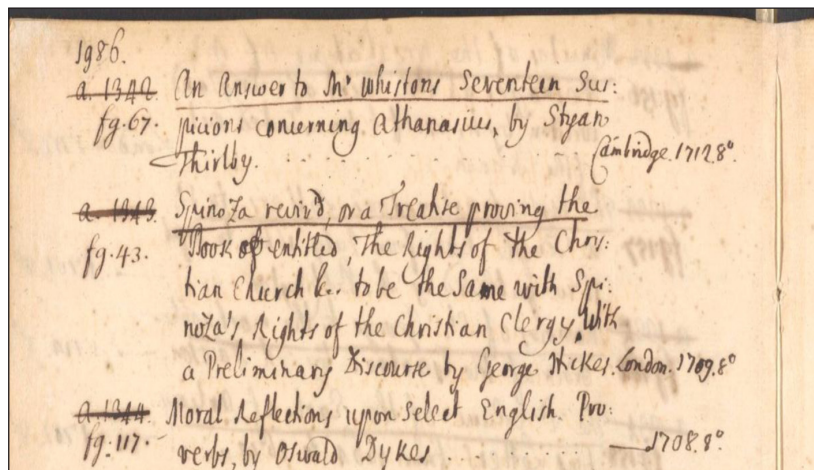


Figure 1: Extract from Sloane's manuscript catalogue of printed material, Sloane MS 3972C vol. VI, f. 8v, British Library (Public domain in most countries except the UK). These entries are in the hand of Johann Gaspar Scheuchzer, Sloane's amanuensis 1722–29. Figure 1 excerpted from the Creative Commons License and is public domain in most countries except the UK.

and sometimes edition or volume numbers. The right-hand margin holds further bibliographic detail of place and date of publication and the physical size of the text in abbreviation, such as folio (f°) or quarto (4°). Other important and common information includes underlining, pencil strikethroughs and the use of crosses and long dashes, as well as 'hands' other than that of Sloane. Sloane's catalogues of other kinds of collections are frequently broken down by object type; here with printed books, material is catalogued without the use of common structural elements like subheadings or page breaks to separate by form or genre. The exception is the final 37 folios of the catalogue, which contains the 'Min' entries documenting Sloane's 'books of miniature, painting, designs &c' (Sloan, 2012).

MS 3972 C vol. VI can be distinguished from those recording other parts of Sloane's collection by both the content and style of its catalogue entries. Unlike the 'fossils', 'miscellanies', antiquities or botanical specimens, the descriptions of which were related to observations and to early modern *qualia*, the objects represented in these bibliographic entries had standardised, printed characteristics. This difference between the unique objects in one set of catalogues and the relatively standardised objects that are printed books means that these catalogue entries only rarely include supplementary subjective descriptions of the text's content, condition, appearance or record of how it entered into Sloane's possession.

Early modern handwritten and printed library catalogues have been extensively studied by historians of science, the book and bibliography (see Walsby and Constantinidou, 2013). More importantly for this project, the general history of 17th- and 18th-century private library practices, book collections and the collecting habits of Sloane's contemporaries have been richly documented (Loveman, 2015; Edgington, 2016; Poole, 2015). TEI and other XML languages have been used extensively to digitally render historic book and library catalogues. Projects such as RICABIM: Repertorio di Inventari e Cataloghi di Biblioteche Medievali (RICABIM, n.d.) and Thecae (Thecae, 2018) have collated repertoires of catalogues, inventories and lists of books, incunabula and manuscripts in order to better understand the circulation and availability of these sources in the medieval and early modern periods. Moreover, in collaboration with Biblissima (Biblissima, n.d.), Thecae's searchable database of TEI

encoded inventories of medieval and modern books has been at the centre of the creation of new TEI standards for the encoding of ancient inventories and catalogues.

However, there is a gap in digital research on the materiality of catalogues, inventories and lists. To date, TEI has rarely been used to describe catalogues as objects in themselves, as opposed to vectors of bibliographic data. As TEI currently stands, it has an extensive capacity for the encoding of books, manuscripts and bibliographic detail more generally; however, more work is needed so that it can readily enable the encoding of the catalogue both as object and carrier of 'object detail'. The complex nature of Sloane's catalogues has meant that Enlightenment Architectures has faced productive conceptual and technical dilemmas in its markup activities. One of the most fundamental difficulties encountered was in the selection of appropriate TEI elements and the need to customise TEI in the absence of such elements: customisation has been crucial as it aids us in our bid to create an accurate representation of the catalogue and its descriptions. In customising some important features of markup for Sloane's catalogues, we anticipate that our research may help others working with other early modern printed and handwritten catalogues.

One such example is the `<ea:catent>` or 'catalogue entry' element, which has been created by Enlightenment Architectures in response to Sloane's catalogues. This tag serves to group all the information, be it descriptive, graphical or spatial, that corresponds to each catalogue number. This includes such varied information as the description of the object listed—its size, shape or condition; its provenance (person and place); the price paid for it; and many other details. Importantly, it can also contain the descriptions of multiple objects, all of which have been purposefully documented by Sloane and his amanuenses under one catalogue number. Although individual elements such as `<place>` (as in a geographic location) or `<ref>` (a reference to a location of any kind) are also tagged within the catalogue entry, we group these elements together within the `<ea:catent>` in order to convey the original cataloguer's choice to include this particular information when describing the object at hand. The information recorded (and not recorded) in the catalogue entry reflects how the individual perceived the object and the knowledge that they had about it.

Various possibilities exist for encoding this information in line with TEI. For example, we considered using a generic `<div>` element (a '(text division)[that] contains a subdivision of the front, body, or back of a text' (TEI Consortium, 2018b)) with an attribute to specify which kind of division is being referred to, for example, `<div type="catEnt">`. Yet we rejected this for a number of reasons. Firstly, the information that is supplied in the type attribute is crucial: first order information rather than qualifying information. Though the question about when to use attributes versus elements is one that is contested by XML experts (Cover, 2008) there is some consensus that, where possible, first order information should be recorded as an element (e.g. w3schools, n.d.). This verdict is also linked to the semantic limitations of XML, where relationships can be deduced from the nesting of elements but not from the order of attributes (Antoniou and van Harmelen, 2008: 32). So too, attributes can be more difficult to process than elements. As we do not see the encoding of the catalogues as an end in itself, but rather as something that can support the further interrogation of the catalogues, we concluded that it was therefore appropriate to devise a specialist element to encode this data.

Another option could have been to adapt elements currently found in the TEI header to apply to the content of the manuscript catalogue. For example, the TEI guidelines provide `<msContents>` ('describes the intellectual content of a manuscript or manuscript part, either as a series of paragraphs or as a series of structured manuscript items' (TEI Consortium, 2018d)) and `<msItem>` ('describes an individual work or item within the intellectual content of a manuscript or manuscript part' (TEI Consortium, 2018e)). However, not only are the decisions of the original cataloguer lost through the use of these `<ms>` elements, which do not strictly include the entire content of the catalogue entry, but crucially, the semantic import of an 'individual work or item' would not produce the necessary dataset for understanding the structure of the catalogue and its entries. The `<ea:catent>` element is thus a vital innovation for those who are not only seeking to extract data from catalogues but also attempting to understand the internal structure of the catalogue itself.

Related to `<ea:catent>` is the `<ea:catnum>` or 'catalogue number' element that we have also created. This element serves to identify each catalogue number

listed in the catalogues. The TEI guidelines suggest that <idno> ('supplies any form of identifier used to identify some object, such as a bibliographic item, a person, a title, an organization, etc. in a standardized way' (TEI Consortium, 2018c)) or <altIdentifier> ('contains an alternative or former structured identifier used for a manuscript, such as a former catalogue number' (TEI Consortium, 2018a)) would suffice for this purpose. **Figure 2** demonstrates how another catalogue could be embedded with the <idno> and <altIdentifier> tags.

The same logic could be applied to Sloane's catalogues. However, without any certainty as to whether the objects he listed have a new <idno> since being dispersed from his collection, the original catalogue number is the only identifier available, even though it is theoretically an <altIdentifier>. Curators at the British Museum who can match an object with a Sloane catalogue entry give it a Sloane registration number, which includes this catalogue number, which becomes its primary unique identifier. By contrast, Sloane's printed books at the British Library have been given a new shelfmark, which has resulted in the effective 'loss' of his collection 'in plain sight' within the library itself (Walker, 2016). Thus, in order to demonstrate the importance of these catalogue numbers as not just one identifier among several, but rather the only record which connects these objects to both the catalogue and also one another, the element <ea:catnum> has been created in order to underscore

28843. In Latin, on parchment: written in more than one hand of the 13th cent. in England: 7¼ x 5½ in., i + 55 leaves, in double columns: with a few coloured capitals.

'Hic incipit Bruitus Anglie,' the De origine et gestis Regum Angliae of Geoffrey of Monmouth (Galfridus Monumetensis): *beg.* 'Cum mecum multa & de multis.'

On fol. 54^v very faint is 'Iste liber est fratris guillelmi de buria de . . . Roberti ordinis fratrum Pred[icatorum],' 14th cent. (?): 'hanauilla' is written at the foot of the page (15th cent.). Bought from the rev. W. D. Macray on March 17, 1863, for £1 10s.

Now MS. Add. A. 61.

```

<msDesc>
<msIdentifier>
<settlement>Oxford</settlement>
<repository>Bodleian Library</repository>
<idno>MS. Add. A. 61</idno>
<altIdentifier type="SC">
<idno>28843</idno>
</altIdentifier>

```

Figure 2: Extract taken from The TEI Consortium Guidelines (2018: section 10.2, figure 10.1).

the significance of these catalogue numbers. This is particularly important as it is speculated that these numbers themselves can potentially aid our understanding of how Sloane acquired materials chronologically, and catalogued and grouped his items, and perhaps even how they were physically arranged in his house, both visually and for access and use, thereby allowing us to consider how Sloane and his contemporaries ordered and understood the world around them.

In other instances, we have made changes to TEI that have a more limited scope. Catalogue 3972C vol.VI, for instance, contains the element <textName>, which builds upon <msName> ('contains any form of unstructured alternative name used for a manuscript, such as an 'ocellus nominum', or nickname' (TEI Consortium, 2018f)) to indicate those printed texts listed in the catalogue under a title which cannot be found by the same name elsewhere. As with <ea:catnum>, in these instances the <textName> is the only identifier that exists. We therefore wish to be able to identify that these are published, if currently untraceable with the titles that Sloane gave them, even if we have no current record for them.³

The markup of MS 3972C vol. VI

The TEI markup of MS 3972C vol. VI captures the most important bibliographic content of the catalogue entries, as well as key physical and graphical elements (see **Figures 3** and **4**). **Table 1** shows the elements that are included in the markup (generic TEI structural elements such as <p> and <lb> are not listed here).

This markup effectively captures the core details of the catalogue in line with our historically-informed readings of them. At present, though, we are not encoding the particular language in which a title or other information is given;⁴

³ Some of these instances might well be published items which have been incorrectly recorded by the original cataloguer. The potential reasons as to why a text was recorded in an unstandardised way raise bibliographic questions regarding the editions, pirated copies, printed commentaries on texts, short titles and other sources which may have been available at Sloane's time and these original entries may contribute to our understanding of contemporary book cataloguing practices, the availability of texts during Sloane's lifetime and possibly even the book trade more generally.

⁴ While most of Sloane's catalogues are predominately written in English and Latin, the many languages found in 3972C vol. 6 reflect the general composition of his library, a quarter of which only, it is speculated, comprised English language books.

Table 1: Catalogue features of MS 3972C vol. VI and their XML markup.

Catalogue Feature	EA-defined elements
Catalogue number	<ea:catnum>
Text Name	<ea:textName>
Catalogue entry	<ea:catent>
Catalogue Feature	TEI-defined elements
Bibliographic reference	<bibl>
Title	<title>
Author	<author>
Publisher	<publisher>
Editor/Praeses/Other responsibility in text production	<respStmt> <resp> <name>
Publication place	<pubPlace>
Publication date	<date>
Book size	<dimensions type= ("folio" / "quarto" / "octavo" / "duodecimo")>
Volume	<biblScope= "vol">
Edition	<edition>
Graphical additions	<!--(comment) -->
Ibid/Ejusdem etc.	<ref>
Underline	<hi rend= "underline">
Strikethrough	<del rend= "striketrough">
Ticks	<add rend="pencil">✓</add>

neither time nor resources allow this. There is one exception to the decision to disregard foreign languages, which is the tagging of multi-language place names. The network of places from and through which Sloane's collection reached him stands to be a profitable line of enquiry and for this reason the names of publication places in different languages, such as London/Londra/Londres/Londinium, will be linked to one single georeferenced location. While this will not reflect the language composition of Sloane's library more broadly, it does ensure that this crucial information which links Sloane to the wider world around him is made identifiable and analysable.

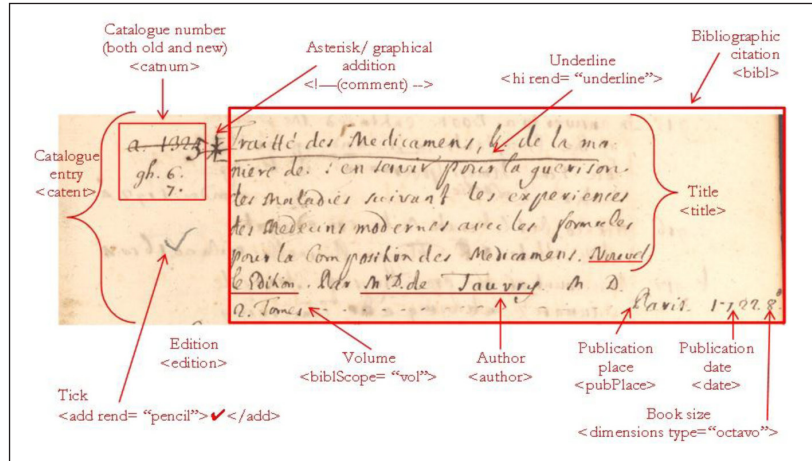


Figure 3: MS 3972C vol. VI, f.7. British Library (Public domain in most countries except the UK). An annotated extract from Sloane's catalogue of printed material showing composite parts of individual catalogue entries. For readability we have dropped the enlightenment namespace prefix. The markup of the transcription of this extract appears below in Figure 4. Figure 3 excerpted from the Creative Commons License and are public domain in most countries except the UK.

```

<div3>
<ea:catnum type="primary" place="margin">
  <del rend="strikethrough">a. 1325</del>. gh. 6. 7.</ea:catnum>
<ea:catent>
  <p>
    <bibl> <!--graphical addition -->* <title><hi rend="underline">
    Traitté des Medicamens, &amp;. de la ma</hi><lb/>
    niere de s en servir pour la guerison<lb/>
    des Maladies suivant les experiences<lb/>
    des Medecins modernes avec les formules<lb/>
    pour la Composition des Medicamens</title>. <edition>Nouvel:<lb/>
    le Edition</edition>. Par <author>M<hi rend="sup">r</hi> D. de
    Tavvry</author> M.D.<lb/>
    <biblScope type="vol">2. Tomes</biblScope>
    <pubPlace>Paris</pubPlace> <date>1722</date>.
    <dimensions type="octavo"> 8<hi rend="sup">o</hi>
  </dimensions>.</bibl></p>
</ea:catent>
<add rend="pencil">✓</add>
</div3>

```

Figure 4: The expanded TEI mark up of MS 3972C vol. VI, f.7.

This case study has raised fundamental questions about the conceptual limits of TEI and its capacity to support historically sensitive encoding. In spite of our experience of the limitations of TEI we have been able to propose workable solutions which will enable people to conduct a search of the catalogues that is not determined by contemporary concepts and categories.

Case study 2: The object problem

Sloane produced separate catalogues for different types of natural history, according to taxonomies significant at the time of each catalogue's production. Of the catalogues being studied by Enlightenment Architectures, 'Fossils 1' consists of 364 numbered folios containing 3,601 entries, which are divided across six sub-headings: 'Coralls', 'Serpents &c.', 'Echini', 'Crustacea', 'Starrfishes' and 'Humana' (Catalogue of 'Coralls, Sponges & some other submarines', n.d.). Sloane's catalogue of miscellanies comprises 4,551 object descriptions. They are grouped under the headings of 'Miscellanies', 'Antiquities', 'Bronzes', 'Impressions of seals', 'Pictures &c.', 'Mathematical instruments &c.', 'Agate handles' and 'Agate cups, bottles, spoons &c.' (Catalogue of 'Miscellaneous things', n.d.). The respective catalogues are composed of a series of chronologically consecutive numbered entries that describe an object and, where possible, Sloane includes other sorts of information including colour, size, material, provenance, names, uses and bibliographic references. Much like his other natural history catalogues, but unlike those of his library material, Sloane wrote the majority of these descriptions himself and was actively engaged in the process of recording and managing information about these natural and 'miscellaneous' objects.

The natural history catalogues of Sloane and his contemporaries have been described as interpretive 'repositories of multiple intersecting stories that textualized and contextualized each object' (Findlen, 1996: 36, note 61). The crux of these stories was the lengthy object descriptions given in each entry. They appear to have been part of a method of 'verbal description' (Wragge-Morley, 2010) that was used by Sloane and others to make the ever-expanding world knowable (not just by possessing the object itself, but by creating and retaining written information about its source and use). Early modern techniques for understanding the natural world included close observation of differences between specimens and the rendering of these differences in text form. Producing or reading these descriptions had a particular cognitive value that was central to understanding what was being described, especially for those without access to a collection. As Descartes enthused, the purpose of a worthy description is to cause a sensory impression and create correspondent images in the

imagination: something that can be effectively done by words alone (Wragge-Morley, 2010). This is implicit in the writings of Nehemiah Grew, a contemporary of Sloane (Wragge-Morley, 2010). Grew wanted his account of the Royal Society's Repository, *Musaeum Regalis Societatis*, to be comprehensive and offered the following reason for his detailed descriptions:

If any object against their length: perhaps they have not so well considered the necessity hereof, for the cleer and evident distinction of the several Kinds and Species, in so great a variety of Things known in the World [...] Besides, that in such Descriptions, many Particulars relating to the Nature and Use of Things, will occur to the Authors [sic] mind, which otherwise he would never have thought of. And may give occasion to his Readers, for the consideration of many more (Grew, 1681: Preface).

As discussed above, Sloane's catalogue entries consist of a catalogue number and an object description along with various annotations. TEI markup has been used by Enlightenment Architectures to encode a wide range of this information. For example, the standard element <name> is used to identify names and to distinguish them from additional information about a person. The element <addname> allows for references to nicknames and aliases, and is particularly useful for variant spellings of names. In addition, TEI offers various options for encoding the provenance of an object such as <placeName>, which identifies an absolute place name, and <geogName> for the identification of more specific geographical features. Similarly, the element <date> allows a date (in any form) to be tagged, which is useful for establishing the timelines of Sloane's catalogues. Additional information such as pencil location codes, monetary values, brackets, drawings and much later curators' comments can also be marked-up. For example, <add rend="pencil"> and <add rend="red"> encode 'additional' comments appearing in pencil and red ink. Capturing these (and what are thought to be location codes) in the margins of the catalogues is crucial to understanding Sloane's methods of arranging objects in his own home whether by theme, use, material or size, for example (see Caygill, 2012).

At an early stage of the project we identified the benefits that would result from encoding the objects that are described in Sloane's catalogues. The encoding of an object name in historical sources enables both humans and machines to identify and manipulate such names and, for example, to identify patterns in their descriptions and expand the potential for reuniting objects in the memory institution. Indeed, identifying a string of text that can function as a verbal signifier of an object is central to treating a historical document like a manuscript catalogue with 'curatorial sensibility' (Nelson, 2016). As we will show in this case study though, what has proved most difficult about attempting to encode object descriptions is not the application of TEI but a more fundamental issue—namely the problem of how to consistently and reliably identify references to individual objects and to identify the boundaries that exist between 'object names' and their qualifying descriptions. We asked ourselves, might it be better to side-step the issue of identifying the boundaries of the object and instead encode only the catalogue number, which keys the entry back to the physical object? Take, for example, 'Red corall growing on a rock wt. shells'(Catalogue of 'Coralls, Sponges, & some other submarines', n.d.: Entry no. 11, f. 2). In this example, the boundary between the object that is being described and additional descriptive information about that object is difficult to identify. Which, if any of the following suggestions specifies the object name?

corall

Red corall

Red corall growing on a rock

Red corall growing on a rock with shells

This problem has also been discussed by Nelson who has argued:

We must be able to determine and define the limits of what constitutes a mention of an object. This is crucial so that we retain all information that is immediately relevant to that object, but (ideally) no more than is pertinent. ... We must be able to define the relevant and relative contexts of each mention

of an object: ... To enable an articulation of an object's place in a hierarchy; ... To enable identification of relevant and related information; and ... To enable the articulation of events involving other entities (e.g., people, places, and other objects) (2016).

In his discussion of the difficulties of identifying the boundaries of object names and their descriptions, Nelson (2016) mentioned various document-dependent factors, including the fact that catalogue entries of the period often do not deal with proper nouns, 'the question of when an inventory becomes a catalogue' and the tendency to embed the object in continuous descriptive prose. However, we also wondered whether the problem of disambiguating the object from the surrounding text was fundamentally rooted in current forms of disciplinary knowledge rather than in understandings we were coming to through our document analysis. To explore this, we contacted five domain knowledge experts in separate but complimentary knowledge domains including natural history, collecting history, digital humanities and curatorial studies; all are areas pertinent to the EA project. We provided them with the following examples of object descriptions from Fossils 1 and 'Miscellaneous things':

1. Red corall growing on a rock with shells (Entry no. 11, f. 2).
2. The large claw of the triangular crab with tubercles from Jamaica (Entry no. 157, f. 240).
3. Two slates between which lives a shrimp (Entry no. 294, f. 253).
4. A piece of the keel of a ship eat by the worms ('Catalogue of "Miscellaneous things"', n.d.: Entry no 3).

We asked them to identify the 'object' in a longer description and to indicate which words should be marked-up as the object. These entries were chosen because they highlight the complexities of trying to disambiguate and encode an object in Sloane's catalogues.

Respondent 1, a digital humanities expert, saw the solution to tagging the object as relatively simple: 'each of these descriptions is grammatically a noun phrase, and as a general rule I think the head noun of the phrase is as good a candidate for identifying the object [that Hans Sloane] (or his assistants) saw as the object being described'. The head noun in each example then, is the one on which everything else in the noun phrase grammatically depends and this would change depending on how the description is phrased. In example 1, 'corall' is the object and in example 2, 'claw' is the object. In those instances where the head noun denotes a measure, as with example 4, 'the head noun and the prepositional phrase identifying the whole from which the part is taken is referred to as the object'. Even with this rationale it remains challenging to be consistent in the identification of the object name. With regard to example 4, we remain unsure whether it is the 'keel' or the 'piece' that should be tagged.

However, the historians of science, collecting and natural history whom we consulted reiterated an earlier argument, one of context. They consider grammar alone as insufficient for identifying an object. As Respondent 2, a historian of science, argued, 'if we want to be as "historically sensitive" as possible, then for Sloane, the *entire* entry is the object'. This means that all of the words that detail colour, material and size were chosen specifically by Sloane to describe the details of the object. Their order and place within the catalogue had meaning for Sloane. Respondent 3, a historian and philosopher of science, likewise argued that 'in all four cases it is the whole phrase that designates the object'. This is because the phrases are descriptive: they describe the objects but do not contain names in a technical sense that could be considered as labelling or indexing the object in question. What we find instead, are generic nouns like 'corall'.

But no analysis of what the object is can stop here, at the juncture between marking up one head noun or marking up the entire description. Indeed, what becomes clear is the difference in current interpretive analysis depends on domain knowledge but also on the potential end use of the data. Take for example the topic

of species and taxonomy. Respondent 3 argued that, if the object 'is the kind or species that the specimen is supposed to represent, [then the] "red coral" is the object, because corals were often distinguished by their colour'. Likewise, Respondent 4, a botanist, noted the importance of capturing 'red coral' for the purpose of indexing: 'an index entry might read "corall, red" because of the need to alphabetise the index entries'. Here the interpretive analysis takes into consideration the object as well as its features, such as colour, or its ecology.

If the information found in these catalogues is to form part of an institution's database, such as that of the British Museum, context is still important, especially in terms of what object categories might mean, both then and now. Respondent 5, a museum documentation expert, observed that 'Nuances are lost if only specific words can be tagged or highlighted and so there are "multiple key wording facilities" available that an institution like the British Museum "consider essential"'. This would mean that 'Red corall', 'rock' and 'shells' would all be tagged. In the case of example four, this becomes 'tricky', according to respondent 5, because it is the context of the description that alludes to why the object has been collected in the first place. While the British Museum database would include 'keel of the ship' as the object, it is the addition of 'eat by the worms' that makes the object interesting, both then and now—even if for different reasons. Moreover, as a result of such treatment, the object name loses its 18th-century meaning and context as it is divorced from the particulars of how Sloane described and organised the knowledge contained in his catalogues.

Overall then, while it would be possible to align our encoding of the object with a particular current disciplinary understanding of the data, each of the four views expressed bring complexities. With regard to the grammatical approach, for example, it can still get complex and such interpretive analysis is time consuming. Different researchers, whether they are historians, scientists or professionals working in galleries, libraries and museums, will always be interested in different aspects of the object being described. But as Respondent 2 pointed out, one cannot do everything in a finite life, let alone in a time- and resource-limited research project. Therefore, one suggestion is to prioritise the XML markup so as to focus on 'aspects

of a catalogue entry which would yield information that is difficult to gain otherwise at a discrete level' (Respondent 2). In other words, it is the information about issues like organisation (information about drawers and cabinets) that would yield the most conducive data for understanding the intellectual structures of these catalogues and their context. Tagging provenance for example, would allow for mapping or visualisation of source networks and object movements such as with Six Degrees of Francis Bacon (Six Degrees of Francis Bacon, n.d.) while encoding collection locations would eventually allow for the 'reconstruction' of what kind of objects were placed together in a drawer or cabinet.

At an earlier stage of this project we defined a new element `<ea:objectDescriptor>` for encoding descriptions of individual objects within entries. Yet the difficulties of attempting to apply this element, which are further validated by the varying responses that we received from the experts we surveyed, show that we are still left questioning what the object name is. While some would suggest that the object can be defined by one head noun, others find this insufficient. Either way, there is agreement that we must always consider context and be as historically sensitive as possible. In this way, our failure to model the object has opened McCarty's 'via negativa' (discussed above) for the project. Drucker has described how this approach can be understood as 'a rigorous approach to the study of knowledge through attention to ignorance, that pushes at assumptions to lay them bare' (2007). Thus, our failure to satisfactorily model the object has forced us to consider our implicit understandings of object descriptions and their boundaries and to interrogate the nature of the object with a rigour that is not necessary when undertaking more traditional scholarship. It has also helped us to refocus our research questions by recognising that our task does not depend on defining what the object is, but on understanding how Sloane's descriptions have been constructed. By doing this, we can begin to understand how Sloane and his contemporaries were managing, organizing and producing knowledge about the world around them.

In this case study we have reflected on our efforts to model and encode the object names that are recorded in Sloane's catalogues. While we initially understood the

challenge of encoding object names in Sloane's catalogues to be a question of how to use TEI appropriately, we came to realise that we faced something more fundamental, namely the challenge of computationally modelling highly interpretative historical information. We have argued that this issue raises important questions about the computational representation of early modern knowledge. At stake is not only how we can best use the methods of digital humanities to represent early modern catalogues such as that of Sloane but also the current limits of humanities and curatorial knowledge about early modern catalogues.

Conclusion

Collections documentation has often been described as making the difference between a museum and a junk shop. Catalogues are the core documents of museum structure and meaning, yet no significant computational analysis has been made to date of how catalogues from the early modern period are constructed or of how their structure and content relate either to the world from which collections are assembled or to the museums they form. Enlightenment Architectures is undertaking this task on some of the oldest, most detailed and most significant museum catalogues in the English-speaking world. The interdisciplinary approach that we are devising to pursue this in the context of Sloane has the potential to enable us to model the information structures of Sloane's catalogues computationally and interrogate them in ways that would otherwise be impossible.

Our research, and the new directions opened by it, will profit the communities of researchers, curators and information professionals who are addressed in this article. We expect that historians and curators will benefit from the ability to digitally search the catalogues in ways that would be impossible using paper catalogues alone, such as being able to search according to the colours, materials, weights and sizes that are mentioned in respective entries. It will also be possible for researchers to download the TEI-encoded versions of Sloane's catalogues and to extract, map and visualise information that is included in them. The TEI extensions and customizations that we have proposed can also be taken up by other projects working on early modern

archival materials. Finally, the more fundamental questions that we have raised about digital approaches to the modelling of early modern information may open new conversations between curators and digital humanists about current approaches to cataloguing and placing collections online.

It is widely understood that early modern catalogues such as those of Sloane are 'not simply lists that can be taken at face value' (Keating and Markey, 2011: 211). Rather, they are 'authored documents compiled under particular temporal, legal, political, and social constraints that affected their organization and the ways in which the objects they list were described' (Keating and Markey, 2011: 211) As such, they are sites where narratives of power and knowledge are made, unmade, silenced and sometimes imagined anew. This is no less true of the digital models of Sloane's catalogues on which the Enlightenment Architectures project is at work. The object problem that we have discussed above points to parallels between early modern and current classification systems. We still struggle, despite our "sophisticated" systems of classification, to determine what things are, where they belong and how to classify them. While we don't necessarily understand Sloane's cataloguing epistemology fully, we share his struggle regarding how best to use words to describe objects. The digital representations that Enlightenment Architectures is creating will contribute to what Poole described as the 'rich, complex and interwoven cultural experience on the World Wide Web' (2016). Therefore, it is crucial to interrogate the potential and limits of encoding languages like TEI for representing early modern catalogue materials, as we have done in this article.

Bowker has written of the totalizing imperatives of data-driven fields such as biodiversity and of how such efforts to homogenise and standardise data can make it incompatible with the user-generated datasets organised by 'local data cultures' (2000). We understand the work that we have undertaken in the Enlightenment Architectures project, and digitally-mediated research on the early modern period more widely, as indicative of a 'local data culture' and this article has demonstrated the nuance this research can contribute to a digital humanities that is sometimes

portrayed as totalising and elitist (e.g. Grusin, 2014; Pannacker, 2013). Of the black-boxing effect of technology more generally, Latour has written of:

the way scientific and technical work is made invisible by its own success. When a machine runs efficiently, when a matter of fact is settled, one need focus only on its inputs and outputs and not on its internal complexity. Thus, paradoxically, the more science and technology succeed, the more opaque and obscure they become (1999: 304).

In this article, we have pushed against the black box to reveal some of the ‘internal complexity’ of the Enlightenment Architectures project. We have shown the importance of attention to ‘internal complexity’ when thinking about the potential of interdisciplinary research across the digital humanities, history of knowledge, the library and the museum, especially in terms of the digital collections that such work can give rise to.

Ethics and consent

The aspects of this research that have involved human subjects have been carried out in accordance with the Declaration of Helsinki. Our work was given ethical clearance by the research director of the British Museum, in line with institutional norms. Informed consent was sought from participants and their identities have been anonymised.

Acknowledgements

The research that underpinned this article was funded by a Leverhulme Trust Research Project Grant (2016–19). We wish to thank most sincerely the individuals who agreed to be interviewed during the course of this research. We also wish to thank Alexandra Ault (British Library) and Mark Carine (Natural History Museum) for their assistance with, and comments about, various aspects of this article. Errors and omissions are, of course, ours.

Competing Interests

The authors have no competing interests to declare.

Sloane's Manuscript Catalogues

Catalogue of 'Coralls, Sponges, & some other submarines' (Fossils I), MS SLO, Paleontology Library, Natural History Museum

Catalogue of 'Miscellaneous things' (Miscellanies), Africa, Oceania and the Americas Library, British Museum

References

- Antoniou, G A** and **van Harmelen, F** 2008 *A Semantic Web Primer*. Cambridge, Mass: The MIT Press. *Biblissima* (n.d.) [online]. Available from: <https://projet.biblissima.fr/> (Last accessed 13 March 2019).
- Blair, A** 2010 *Too Much to Know: Managing Scholarly Information Before the Modern Age*. New Haven [Conn.]: Yale University Press.
- Blakeway, A** 2011 The Library Catalogues of Sir Hans Sloane: Their Authors, Organization, and Functions. *Electronic British Library Journal*. [online]. Available from: <http://www.bl.uk/eblj/2011articles/article16.html> (Last accessed 27 May 2019)
- Bowker, G C** 2000 Biodiversity Datadiversity. *Social Studies of Science*, 30(5): 643–683. DOI: <https://doi.org/10.1177/030631200030005001>
- Burghart, M** and **Rehbein, M** 2012 The Present and Future of the TEI Community for Manuscript Encoding. *Journal of the Text Encoding Initiative*. [Online] (Issue 2), [online]. Available from: <http://journals.openedition.org/jtei/372> (Last accessed 12 July 2018). DOI: <https://doi.org/10.4000/jtei.372>
- Caton, P** 2000 Markup's Current Imbalance. *Markup Lang*. [Online] 3(1): 1–13. (Last accessed 27 May 2019). DOI: <https://doi.org/10.1162/109966201753537123>
- Caygill, M** 2012 'Sloane's Catalogues and the Arrangement of his Collections', in Walker, A. & MacGregor, A. (eds.), *From Books to Bezoars: Sir Hans Sloan and His Collections*. London: The British Library Publishing Division. pp. 120–136
- CIDOC** (n.d.) *CIDOC Conceptual Reference Model (CRM)* [online]. Available from: <http://www.cidoc-crm.org/> (Last accessed 27 May 2019).
- Ciula, A, Spence, P** and **Vieira, J M** 2008 Expressing Complex Associations in Medieval Historical Documents: the Henry III Fine Rolls Project. *Literary and*

Linguistic Computing. [Online] 23(3): 311–325. (Last accessed 27 May 2019).
DOI: <https://doi.org/10.1093/lc/fqn018>

Ciula, A and Eide, Ø 2014 Reflections on Cultural Heritage and Digital Humanities: Modelling in Practice and Theory, in *DATECH '14 Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage*. New York, USA: ACM Press. pp. 35–41. [Online]. Available from: <http://dl.acm.org/citation.cfm?doid=2595188.2595207> (Last accessed 27 May 2019). DOI: <https://doi.org/10.1145/2595188.2595207>

Cover, R 2008 *SGML/XML: Using Elements and Attributes* [online]. Available from: <http://xml.coverpages.org/elementsAndAttrs.html> (Last accessed 17 July 2018). DOI: <https://doi.org/10.4000/jtei.1322>

Dalmau, M and Hawkins, K 2014 'Reports of My Death Are Greatly Exaggerated': Findings from the TEI in Libraries Survey. *Journal of the Text Encoding Initiative*. [Online] (Issue 8). Available from: <http://journals.openedition.org/jtei/1322> (Last accessed 12 July 2018). DOI: <https://doi.org/10.4000/jtei.1322>

Dee, S 2014 Learning the TEI in a Digital Environment. *Journal of the Text Encoding Initiative*. [Online] (Issue 7). Available from: <http://journals.openedition.org/jtei/968> (Last accessed 12 July 2018). DOI: <https://doi.org/10.4000/jtei.968>

Delbourgo, J 2011 Too Much to Know: Managing Scholarly Information before the Modern Age. *Times Higher Education (THE)*. 3 February. [online]. Available from: <https://www.timeshighereducation.com/books/too-much-to-know-managing-scholarly-information-before-the-modern-age/415024.article> (Last accessed 12 July 2018).

Del Turco, R R, Buomprisco, G, Di Pietro, C, Kenny, J, Masotti, R and Pugliese, J 2014 Edition Visualization Technology: A Simple Tool to Visualize TEI-based Digital Editions. *Journal of the Text Encoding Initiative*. (Issue 8). [Online]. Available from: <http://journals.openedition.org/jtei/1077> (Last accessed 12 July 2018). DOI: <https://doi.org/10.4000/jtei.1077>

DeRose, S J, Durand, D G, Mylonas, E and Renear, A 1997 What is Text, Really? *ACM SIGDOC Asterisk Journal of Computer Documentation*. [Online] 21(3): 1–24. (Last accessed 27 May 2019). DOI: <https://doi.org/10.1145/264842.264843>

- [ASCH]:** *Developing interoperable metadata standards for contextualising heterogeneous objects, exemplified by objects of the provenance von Asch* (n.d.) [online]. Available from: http://asch.wiki.gwdg.de/index.php/Main_Page (Last accessed 13 March 2019).
- Dietrich, D** and **Pekal, J** 2012 *Open data in cultural heritage institutions*. [online]. Available from: https://www.europeandataportal.eu/sites/default/files/2012_open_data_in_cultural_heritage_institutions.pdf (Last accessed 12 March 2019). DOI: <https://doi.org/10.16995/dm.12>
- Driscoll, MJ** 2006 P5-MS: A General Purpose Tagset for Manuscript Description. *Digital Medievalist*. [Online] 2(0). Available from: <http://journal.digitalmedievalist.org/articles/10.16995/dm.12/> (Last accessed 12 June 2018).
- Drucker, J** 2007 Philosophy and Digital Humanities: A review of Willard McCarty, *Humanities Computing* (London and New York: Palgrave, 2005). *Digital Humanities Quarterly*. [Online] 1(1). Available from: <http://digitalhumanities.org:8081/dhq/vol/1/1/000001/000001.html> (Last accessed 27 May 2019).
- Dumont, S** and **Fechner, M** 2014 Bridging the Gap: Greater Usability for TEI encoding. *Journal of the Text Encoding Initiative*. [Online] (Issue 8). Available from: <http://journals.openedition.org/jtei/1242> (Last accessed 12 July 2018). DOI: <https://doi.org/10.4000/jtei.1242>
- Edgington, J** 2016 Natural History Books in the Library of Dr Richard Richardson. *Archives of Natural History*. [Online] 43(1): 57–75. (Last accessed 27 May 2019). DOI: <https://doi.org/10.3366/anh.2016.0346>
- Eide, Ø** 2014 Ontologies, Data Modeling, and TEI. *Journal of the Text Encoding Initiative*. [Online] (Issue 8). Available from: <https://jtei.revues.org/1191> (Last accessed 25 August 2016). DOI: <https://doi.org/10.4000/jtei.1191>
- Elliott, T**, et al. 2007 *EpiDoc Guidelines: Ancient documents in TEI XML (Version 8)*. [online]. Available from: <http://www.stoa.org/epidoc/gl/latest/> (Last accessed 17 July 2018).
- Europeana** (n.d.) *Europeana Collections* [online]. Available from: <https://www.europeana.eu/portal/?locale=en> (Last accessed 12 March 2019).

- FIHRIST** (n.d.) *Fihrist: Union Catalogue of Manuscripts from the Islamicate World* [online]. Available from: <https://www.fihrist.org.uk> (Last accessed 6 December 2018).
- Findlen, P** 1996 *Possessing Nature: Museums, Collecting and Scientific Culture in Early Modern Italy*. Berkeley: University of California Press.
- Flanders, J** and **Jannidis, F** 2016 'Data Modeling', in Schreibman, S, Siemens, R & Unsworth, J (eds.), *A New Companion to Digital Humanities*. Chichester, West Sussex, UK: Wiley-Blackwell. pp. 229–237. DOI: <https://doi.org/10.1002/9781118680605.ch16>
- Fredell, J, Borchers, C, IV** and **Ilgen, T** 2013 TEI P5 and Special Characters Outside Unicode. *Journal of the Text Encoding Initiative*. [Online] (Issue 4). Available from: <http://journals.openedition.org/jtei/727> (Last accessed 12 July 2018). DOI: <https://doi.org/10.4000/jtei.727>
- Greenhill, E H** 1992 *Museums and the Shaping of Knowledge*. London; New York: Routledge.
- Grew, M** 1681 *Musaeum Regalis Societatis, or, A catalogue & description of the natural and artificial rarities belonging to the Royal Society and preserved at Gresham Colledge*. London.
- Grusin, R** 2014 The Dark Side of Digital Humanities: Dispatches from Two Recent MLA Conventions. *differences*. [Online] 25(1): 79–92. Available from: <https://read.dukeupress.edu/differences/article-abstract/25/1/79/60682/The-Dark-Side-of-Digital-Humanities-Dispatches>. (Last accessed 27 May 2019). DOI: <https://doi.org/10.1215/10407391-2420009>
- Hockey, S M** 2004 'The History of Humanities Computing', in Schreibman, S., Siemens, R. & Unsworth, J. (eds.), *A Companion to Digital Humanities (Blackwell Companions to Literature and Culture)*. Oxford: Blackwell Publishing Professional. [online]. Available from: <http://www.digitalhumanities.org/companion/> (Last accessed 27 May 2019).
- Hughes, L M** (ed.) 2012 *Evaluating and measuring the value, use and impact of Digital Collections*. London: Facet Publishing. DOI: <https://doi.org/10.29085/9781856049085>

- Jannidis, F** 2009 TEI in a crystal ball. *Literary and Linguistic Computing*. [Online] 24(3): 253–265. (Last accessed 27 May 2019). DOI: <https://doi.org/10.1093/lc/fqp015>
- Jenstad, J** 2018 About MoEML. The Map of Early Modern London, mapoflondon.uvic.ca/about.htm (Last accessed 27 May 2019).
- Jones, P M** 1988 A preliminary check-list of Sir Hans Sloane's catalogues. *British Library Journal* (14, 1 (Spring)), 38–51.
- Keating, J** and **Markey, L** 2011 Introduction: Captured objects: Inventories of early modern collections. *Journal of the History of Collections*. [Online] 23(2): 209–213. (Last accessed 27 May 2019). DOI: <https://doi.org/10.1093/jhc/fhq035>
- Kusukawa, S** 2017 William Courten's lists of 'Things Bought' from the late seventeenth century. *Journal of the History of Collections*. [Online] 29(1): 1–17. (Last accessed 27 May 2019).
- Latour, B** 1999 *Pandora's Hope: An Essay on the Reality of Science Studies*. Cambridge, Mass: Harvard University Press.
- Loveman, K** 2015 *Samuel Pepys and his Books: Reading, Newsgathering, and Sociability, 1660–1703*. Oxford: Oxford University Press Oxford. DOI: <https://doi.org/10.1093/acprof:oso/9780198732686.001.0001>
- Manuscriptorium** (n.d.) [online]. Available from: <http://www.manuscriptorium.com/en> (Last accessed 1 August 2018).
- MacGregor, A** (ed.) 1994 *Sir Hans Sloane: Collector, Scientist, Antiquary—Founding Father of the British Museum*. London: British Museum Press.
- Masterman, M** 1962 The Intellect's New Eye. *Times Literary Supplement*. 17 April.
- McCarty, W** 2008 What's Going On? *Literary and Linguistic Computing*, [Online] 23(3): 253–261. (Last accessed 27 May 2019). DOI: <https://doi.org/10.1093/lc/fqn014>
- McCarty, W** 2014 *Humanities Computing*. Basingstoke, Hampshire: AIAA.
- McGann, J** 2004 *Radiant Textuality: Literature after the World Wide Web*. 1st edition. New York; Basingstoke, Hampshire: Palgrave Macmillan. DOI: <https://doi.org/10.3366/ccs.2004.1.3.371>
- McGann, J** 2007 'Marking Texts of Many Dimensions', in Schreibman, S., Siemens, R. & Unsworth, J. (eds.), *A Companion to Digital Humanities*. [Online]. Wiley-

Blackwell. pp. 198–217. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470999875.ch16> (Last accessed 17 July 2018). DOI: <https://doi.org/10.1002/9780470999875.ch16>

Nelson, B 2016 Curating Object-Oriented Collections Using the TEI. *Journal of the Text Encoding Initiative*. (Issue 9). [Online]. Available from: <http://journals.openedition.org/jtei/1680> (Last accessed 12 June 2018). DOI: <https://doi.org/10.4000/jtei.1680>

Nickson, M 1994 'Books and Manuscripts', in MacGregor, A. (ed.) *Sir Hans Sloane: Collector, Scientist, Antiquary—Founding Father of the British Museum*. London: British Museum Press.

Nickson, M 1994 'Books and Manuscripts', in MacGregor, A. (ed.), *Sir Hans Sloane: Collector, Scientist, Antiquary—Founding Father of the British Museum*. London: British Museum Press. pp. 263–277 OCIMCO (n.d.) [online]. Available from: <https://www.bodleian.ox.ac.uk/weston/our-work/projects/ocimco> (Last accessed 13 March 2019).

Ore, C-E and Eide, Ø 2009 TEI and Cultural Heritage Ontologies: Exchange of Information? *Literary and Linguistic Computing*. [Online] 24(2): 161–172. (Last accessed 27 May 2019). DOI: <https://doi.org/10.1093/llc/fqp010>

Orlandi, T 2002 'Is Humanities Computing a Discipline?', in Braungart, G. et al. (eds.), *Jahrbuch für Computerphilologie*. Paderborn: Mentis. pp. 51–58. [online]. Available from: <http://computerphilologie.uni-muenchen.de/jg02/orlandi.html> (Last accessed 27 May 2019).

Pannacker, W 2013 Stop Calling It 'Digital Humanities'. *The Chronicle of Higher Education* [online]. Available from: <http://chronicle.com/article/Stop-Calling-It-Digital/137325/> (Last accessed 27 May 2019).

Poole, N 2016 *Create Once, Publish Everywhere (COPE)—Collections Trust*. [Online]. Available from: <https://collectionstrust.org.uk/resource/create-once-publish-everywhere-cope/> (Last accessed 12 March 2019).

Poole, W 2015 'Analysing a Private Library, with a Shelflist Attributable to John Hales of Eton, c.1624-41', in Jones, E. (ed.), *A Concise Companion to the Study of*

Manuscripts, Printed Books, and the Production of Early Modern Texts: A Festschrift for Gordon Campbell. Chichester, West Sussex, UK: Wiley-Blackwell. pp. 41–65.
DOI: <https://doi.org/10.1002/9781118635261.ch3>

Renear, A 1997 'Out of Praxis: Three (meta)Theories of Textuality', in Sutherland, K. (ed.), *Electronic Text: Investigations in Method and Theory*. Oxford: Oxford University Press. pp. 107–126. [online]. Available from: <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198236634.001.0001/acprof-9780198236634> (Last accessed 17 July 2018). DOI: <https://doi.org/10.1093/acprof:oso/9780198236634.003.0005>

Renear, A, Mylonas, E and Durand, D 1993 *Refining our Notion of what Text Really is: The Problem of Overlapping Hierarchies*. [Online]. Available from: <http://cds.library.brown.edu/resources/stg/monographs/ohco.html>.

Ricabim (n.d.) [online]. Available from: http://www.mirabileweb.it/ricerca_semplice.aspx. (Last accessed 13 March 2019).

Roued-Cunliffe, H and Copeland, A (eds.) 2017 *Participatory Heritage*. London: Facet Publishing. DOI: <https://doi.org/10.29085/9781783301256>

Ruecker, S, Radzikowska, M and Sinclair, S 2016 *Visual Interface Design for Digital Cultural Heritage: A Guide to Rich-Prospect Browsing*. Farnham, England; Burlington, VT: Routledge. DOI: <https://doi.org/10.4324/9781315547961>

Schmidt, D 2010 The Inadequacy of Embedded Markup for Cultural Heritage Texts, *Literary and Linguistic Computing*. [Online] 25(3): 337–56 <<https://doi.org/10.1093/lc/fqq007>> (Last accessed 27 May 2019). DOI: <https://doi.org/10.1093/lc/fqq007>

Six Degrees of Francis Bacon (n.d.) [online]. Available from: http://www.sixdegreesoffrancisbacon.com/?ids=10000473&min_confidence=60&type=network (Last accessed 26 July 2018).

Sloan, K 2012 'Sloane's 'Pictures and Drawings in Frames' and 'Books of Miniature & Painting, Designs, &c.', in Walker, A. & MacGregor, A. (eds.), *From Books to Bezoars: Sir Hans Sloan and His Collections*. London: The British Library Publishing Division. pp. 168–189.

- Soualah, M O and Hassoun, M** 2012 A TEI P5 Manuscript Description Adaptation for Cataloguing Digitized Arabic Manuscripts. *Journal of the Text Encoding Initiative*. [Online] (Issue 2). [online]. Available from: <http://journals.openedition.org/jtei/398> (Last accessed 12 July 2018). DOI: <https://doi.org/10.4000/jtei.398>
- Sperberg-McQueen, M** 2018 *Bifocal data: a problem and some solutions*. [online]. Available from: [http://blackmesatech.com/2018/02/London/#\(1\)](http://blackmesatech.com/2018/02/London/#(1)) (Last accessed 12 July 2018).
- Stork, L, Weber, A, Gassó Miracle, E, Verbeek, F, Plaat, A, Van Den Herik, J and Wolstencroft, K** 2018 Semantic Annotation of Natural History Collections. *Journal of Web Semantics*. [online]. Available from: <http://www.sciencedirect.com/science/article/pii/S1570826818300283> (Last accessed 12 March 2019). DOI: <https://doi.org/10.1016/j.websem.2018.06.002>
- TEI Consortium** 2018a *<altIdentifier>* [online]. Available from: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-altIdentifier.html> (Last accessed 26 July 2018).
- TEI Consortium** 2018b *<div>* [online]. Available from: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-div.html> (Last accessed 26 July 2018).
- TEI Consortium** 2018c *<idno>* [online]. Available from: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-idno.html> (Last accessed 26 July 2018).
- TEI Consortium** 2018d *<msContents>* [online]. Available from: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-msContents.html> (Last accessed 26 July 2018).
- TEI Consortium** 2018e *<msItem>* [online]. Available from: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-msItem.html> (Last accessed 26 July 2018).
- TEI Consortium** 2018f *<msName>* [online]. Available from: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-msName.html> (Last accessed 26 July 2018).
- TEI Consortium** (n.d.) *TEI: History* [online]. Available from: <http://www.tei-c.org/about/history/> (Last accessed 26 July 2018).
- Thecae** 2018 [online]. Available from: http://www.unicaen.fr/recherche/mrsh/document_numerique/projets/thecae (Last accessed 12 July 2018).

- The Chymistry of Isaac Newton** (n.d.) [online]. Available from: <http://webapp1.dlib.indiana.edu/newton/>; <http://www.newtonproject.ox.ac.uk>. (Last accessed 26 July 2018).
- The Map of Early Modern London** (n.d.) [online]. Available from: <http://mapoflondon.uvic.ca/>. (Last accessed 26 July 2018).
- The Newton Project** (n.d.) [online]. Available from: <http://www.newtonproject.ox.ac.uk/> (Last accessed 1 August 2018).
- The TEI Consortium** (2018) *Manuscript Description* [online]. Available from: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html#msdesc> (Last accessed 1 August 2018).
- Tomasek, K and Bauman, S** 2013 Encoding Financial Records for Historical Research. *Journal of the Text Encoding Initiative*. (Issue 6). [online]. Available from: <http://journals.openedition.org/jtei/895> (Last accessed 12 July 2018). DOI: <https://doi.org/10.4000/jtei.895>
- w3schools** (n.d.) *XML Elements vs. Attributes* [online]. Available from: https://www.w3schools.com/xml/xml_dtd_el_vs_attr.asp (Last accessed 17 July 2018).
- Walker, A** 2016 'Lost in Plain Sight: Rediscovering the Library of Sir Hans Sloane', Bruni, F. & Pettegree, A. (eds.), *Lost Books: Reconstructing the Print World of Pre-Industrial Europe*. Leiden: Brill. pp. 400–413. DOI: https://doi.org/10.1163/9789004311824_020
- Walsby, M and Constantinidou** (eds.) 2013 *Documenting the Early Modern Book World: Inventories and Catalogues in Manuscript and Print*. Leiden: Brill. [Online]. Available from: <https://brill.com/view/title/24308> (Last accessed 12 July 2018).
- Wittern, C, Ciula, A and Tuohy, C** 2009 The making of TEI P5. *Literary and Linguistic Computing*, [Online] 24(3): 281–296. (Last accessed 26 July 2018). DOI: <https://doi.org/10.1093/lc/fqp017>
- Wragge-Morley, A** 2010 The Work of Verbal Picturing for John Ray and Some of his Contemporaries. *Intellectual History Review*, [online] 20(1): 165–179. (Last accessed 26 July 2018). DOI: <https://doi.org/10.1080/17496971003638316>

How to cite this article: Ortolja-Baird, A, Pickering, V, Nyhan, J, Sloan, K and Fleming, M 2019 Digital Humanities in the Memory Institution: The Challenges of Encoding Sir Hans Sloane's Early Modern Catalogues of His Collections. *Open Library of Humanities*, 5(1): 44, pp. 1-40. DOI: <https://doi.org/10.16995/olh.409>

Published: 26 June 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Open Library of Humanities is a peer-reviewed open access journal published by Open Library of Humanities.

OPEN ACCESS The Open Access icon, which is a stylized 'A' inside a circle.