

Identifying Best Practice in Recording Copy-Specific Elements in Special Collections Cataloging

Alex Kither & Leo Appleton

To cite this article: Alex Kither & Leo Appleton (2024) Identifying Best Practice in Recording Copy-Specific Elements in Special Collections Cataloging, *Cataloging & Classification Quarterly*, 62:1, 7-55, DOI: [10.1080/01639374.2023.2290725](https://doi.org/10.1080/01639374.2023.2290725)

To link to this article: <https://doi.org/10.1080/01639374.2023.2290725>



© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC



Published online: 21 Feb 2024.



Submit your article to this journal [↗](#)



Article views: 196



View related articles [↗](#)



View Crossmark data [↗](#)

Identifying Best Practice in Recording Copy-Specific Elements in Special Collections Cataloging

Alex Kither^a and Leo Appleton^b

^aSOAS University of London, London, UK; ^bInformation School, University of Sheffield, Sheffield, UK

ABSTRACT

This paper considers various approaches to the practice of copy-specific descriptive cataloging in rare books and special collections and seeks to identify how practice can be improved to ensure greater discoverability. Through applying a survey method, it identifies some of the core requirements of library professionals and researchers working with these materials and proposes solutions toward the standardization of this practice.

ARTICLE HISTORY

Received August 2023
Revised November 2023
Accepted November 2023

KEYWORDS

Special collections;
provenance; controlled
vocabularies; copy-
specific cataloging;
descriptive cataloging

Introduction

This paper seeks to identify how information professionals and researchers discover, create, manage, and access copy-specific information in special collections environments through examining their own practice. This includes the practical limitations of cataloging standards and metadata encoding systems for copy-specific cataloging, and to develop and evaluate theoretical solutions to these problems. Within the scope of this study, “cataloging” is broadly defined as the process of library metadata creation, typically recorded as discrete records via a library catalog. Specifically, this study concerns “descriptive cataloging”, being the description of a specific instance of a resource, as opposed to “subject cataloging”, which primarily records the intellectual contents of the resource.¹ “Copy-specific cataloging” is a more granular level of descriptive cataloging, concerned with capturing unique aspects of an individual copy of a resource such as material evidence of use or ownership. The “copy-specific” level can also be understood as item-level description; that is, the description of a “single physical object.”² This level of cataloging is not commonly practiced with modern research collections and is often only required to record aspects of resources which have a financial, historical, or cultural value aside from their intellectual contents (i.e., special collections). The purpose

CONTACT Alex Kither  ak155@soas.ac.uk  SOAS University of London, London, UK.

© 2024 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

of special collections cataloging differs from that of modern collections and its rules and practices must be considered separately from the general requirements and standards for bibliographic description. Special collections require additional details of description in order to identify the specific edition, impression, or issue of an item and to provide a detailed and accurate description of the resource as a material object or artifact.

Literature review

The practice of recording copy-specific, material aspects of books in library catalogs has been motivated by the development of book studies over the last three decades. Rose³ considers book studies as an interdisciplinary field – encompassing textual bibliography, intellectual history, print history, literary criticism – which focuses on the book as a material object, rather than simply considering the text as the primary subject of study. Rose’s concept of book studies has more recently been promoted by Stead⁴ as one that is not restricted by disciplinary rigidity or conventions, but that facilitates a collaborative approach to the study of books across various places, periods, and contexts. This positions book studies within a larger material turn in the humanities that considers the physical attributes of texts as providing archaeological evidence to aid the study of ideas across disciplines.

The practice of studying the materiality of books as evidence of use and ownership was pioneered by academics such as Jardine and Grafton⁵ whose analysis of marginal annotations made by Gabriel Harvey in his books deduced information about his reading and study habits. This method of “marginalia studies” – that is, studying the markings made in individual copies of books – was continued by renaissance scholars such as Sherman in his study of John Dee’s books⁶ and eventually formalized in his seminal work, *Used books*,⁷ which presents marginalia studies as a unified and important sub-discipline in its own right.

This field of study approaches books in a new way; one which requires library and information professionals to reconsider how they acquire, organize, and describe their collections. To this point, Sherman indicates the role of information professionals in aiding this emergent form of research, “while some libraries have traditionally kept index card files of former owners, better catalogues of marginalia are badly needed to help future scholars locate the past readers and readings that interest them, and researchers will need to collaborate with cataloguers.”⁸ Sherman emphasizes not only the need for developed cataloging resources to aid book studies but also encourages collaborative efforts between academics and information professionals. Information professionals have remained aware of the need to create records of the document as an historical and cultural

artifact, a practice which Smiraglia considers as “narrating the evolution of the work across time.”⁹

Parallel to the study of marginalia is the study of provenance in books, which similarly focuses on the study of individual copies of books, and the material evidence they carry. Just as marginalia studies use copy-specific material evidence to investigate the use of the book, provenance research seeks to use the same evidence to enquire about the historical ownership of the book. As Pearson observed: “There is a steadily growing literature on the ownership and use of books, embracing works on particular private libraries, studies of marginalia, and the new academic vogue for the history of reading.”¹⁰

Copy-specific cataloging can also be adopted as a means of widening access to special collections through catalog discoverability. Most research libraries hold their own, often unique, special collections which represent a valuable resource for potential research and engagement projects. These collections can provide a basis for the development of innovative programs and services that enable research libraries to engage a diverse audience in new ways, and create tangible value for the parent institution.¹¹ The pedagogical value of recording the unquities of special collections has also been illustrated in projects that have used catalog data to increase access to special collections as tools for research, teaching, and learning.¹² Capturing unique, copy-specific data enriches catalogs, but this endeavor is in vain if the user cannot find what the librarian has recorded. Just as these collections represent opportunities, they also pose challenges to the profession and to the capabilities of library management systems.

As user search behaviors move away from specific library catalogs and toward an open web environment, catalogers have to rely less on their institutional discovery platform and more on the descriptive capacity of the metadata itself. Schaffner argues just this point, and surmises that static, non-user informed metadata creation methods have created a “gap between the expectations of users and historical descriptive practices in archives and special collections.”¹³ Overcoming this gap requires special collections practitioners to identify user needs and reflect those needs in how items are described, and how metadata is standardized, structured, and presented in our systems to facilitate effective discovery, access, and use.

Descriptive Cataloging of Rare Materials (Books) (DCRM(B)) is a popularly adopted cataloging standard for rare books. Despite its precedence, DCRM(B) offers only a single entry for recording “provenance” information (7B19.2.), which states: “Make a local note to describe details of an item’s provenance, if considered important.”¹⁴ The importance of creating discoverable metadata that describes the unique material evidence of ownership and usage in books is no longer a question of “if”. This opinion

is reflected in the competencies for special collections cataloging by the IFLA Rare Books and Special Collections Section, which stress the importance of recording, “context, use and provenance of materials.”¹⁵ However, in the case of recording copy-specific material evidence in books, metadata standards are insufficiently vague and inconsistent. More recently, DCRM RDA edition (DCRMR) has elaborated on guidance both for recording physical provenance evidence (9.41.33) and recording the custodial history of an item (9.42). This resource offers guidelines for constructing both brief and detailed descriptions, prioritizing the inclusion of named agents, places, dates, and the location of the evidence in the item. DCRMR provides more detailed guidance on this area of descriptive cataloging than its predecessor, and encourages users to identify key data elements from provenance evidence, and apply controlled vocabularies, where possible. While DCRMR works to fill a gap in the existing guidance, an agreed international framework for consistently and accurately recording copy-specific information has not been sufficiently developed, and in this absence, local decisions have been made as to how to interpret guidance and represent this evidence.¹⁶ These local cataloging decisions have created a practice of inconsistency across special collections.¹⁷ Consequently, the metadata we have is sparse, inconsistent, and undiscoverable.

If, as Green and García-Monge hope, new technologies will facilitate rare book catalogs, “to integrate data from the collections in our care into the semantic universe,”¹⁸ we will need to fundamentally change how we record metadata. As Maxwell and Attig¹⁹ note in their discussion paper on the subject on RDA implementation in DCRM, there may be a need for total structural change in our metadata encoding standards, and to abandon the traditional bibliographic record altogether. Tennant²⁰ famously argued that “MARC must die,” citing its archaic underlying principles as being incompatible with modern data requirements. Yet, the MARC format remains deeply ingrained in most library systems. MARC forces us to consider our collections as discrete records, rather than a network of interrelated nodes, which consequently focuses on the textual content of fields and not the relationships between elements.²¹ A recent OCLC report²² acknowledged that dominant data encoding standards, such as MARC, are poorly adapted for special collections. Alternatively, the report suggests linked data could be used to better represent special collections, especially for the inclusion and increased visibility of item-specific information and the ability to gain a greater understanding of inter-related networks within collections.

Efforts have already been made to go beyond the limitations of the traditional MARC environment to record material evidence metadata for special collections. The Consortium of European Research Libraries (CERL) has developed a suite of bespoke databases, thesauri, and supporting digital

resources dedicated to the identification of material evidence in fifteenth-century printed books. These resources complement and enrich information already held in library catalogs.²³ Even further removed from MARC, the Sampo Semantic Web portals have presented metadata, drawn from historical collections, in a linked data environment to enable complex discovery and access to information.²⁴ Projects such as these illustrate the possibilities available to recording material evidence of book ownership in an alternative data environment.

The existing literature reveals the current state of special collections discoverability is poor, with many collections being effectively “hidden”. Yet, growing areas of academic research require additional, detailed metadata to highlight unique aspects of special collections. Underpinning this lack of visibility in OPACs is a fundamental problem with how special collections records are structured. Though these limitations have been identified, a uniform best practice has not been established, and consequently collections metadata suffers due to inconsistencies and lack of necessary detail. This study uses the following primary research questions (RQs) in an attempt to meet the predefined aims and objectives: RQ1: How best can metadata be structured and created to record copy-specific information in special collections catalogs? RQ2: What are the functional metadata requirements of stakeholder groups working with these collections? RQ3: How best can a data ontology be devised to meet core requirements for copy-specific cataloging?

Materials & methods

This research project aims to better understand the current limitations within an area of practice, and evaluate potential alternatives based on the subjective experiences and reflections of professionals. This approach requires the collection of qualitative data from a group of engaged participants to gain an understanding of their professional experiences, opinions, and practical requirements. To this end, a phenomenological method was initially posited as a means to gather qualitative data of how individuals experience metadata in their professional lives. However, this approach relied solely on the experience of the participants and did not provide scope for consideration of how these experiences could be synthesized to inform a broader understanding of best practice in the sector.

Specifically, the method draws upon the action research model employed by Elliot.²⁵ The initial literature review enabled the formation of the initial concept, before conducting data “reconnaissance” through a survey designed to identify the requirements of information professionals and researchers. Based on the qualitative data from this first stage, an ontology was designed to respond to the functional requirements of participants as articulated

in their survey responses and to provide a guide for best practice. The next stage was to critically reflect on the perceived successes and failures of this implementation by asking a sample of participants to engage with this ontology through the process of an analytic discussion.

Participants were primarily sought through an open call circulated to eight JISC mailing lists generally used for discussion purposes by practitioners and users of library services, primarily in the UK. The survey was open for two weeks and attracted fifteen participants. The study applied noninvasive research methods, including an online survey and analytic interviews, to generate qualitative data. Ethics approval was obtained from the University of Sheffield Information School ethics review panel, approval number 048242.

In order to conduct this research, it was essential to have access to specific heritage collections so that participants could have access to consistent examples. To this effect the research team used the John Collins Collection, which is part of the printed heritage materials at the British Library. British Library collection items and metadata have therefore been used to inform areas of this research project.

Results

Stage one: Reconnaissance: Survey analysis

The first stage of research sought to identify the functional requirements for a prospective metadata schema. This was achieved through the circulation of an online survey designed to investigate how participants interact with special collections metadata and identify requirements for best practice metadata creation and access. The survey which was used during the first stage of the research made use of open-ended questions asking “How” or “Why” rather than eliciting closed, binary responses. Participants were able to respond to open-ended questions using a free-text box and provide as much detail as they deemed appropriate, allowing participants the option to decline to respond, or respond unconventionally, to elicit detailed and representative responses.²⁶

Overview of respondents

Of the fifteen survey respondents, 66.7% professionally self-identified as librarians and 13.3% identified as academics or researchers. One respondent was an “archivist”, one an “antiquarian cataloguer”, and one a “rare book curator”. Of the respondents, 73.3% established that they create catalog records in a professional capacity, while 40% use them and 33.3% manage them. Most respondents selected multiple options, suggesting they work with catalog records in a variety of ways. Eighty percent also stated

that they interact with catalog records on a daily basis, while the remaining 20% do so on a weekly basis.

Descriptive exercise evaluations

The first section of the survey consisted of seven practical descriptive exercises where respondents were asked to describe image(s) of examples of copy-specific information in books. The examples provided for these exercises were taken from the Collins collection at the British Library, a collection especially dense in unusual copy-specific elements. Each example was selected to test participants' ability to describe various types of copy-specific material evidence, including textual and visual material, hand-written and printed, by multiple persons, with varying amounts of information presented.

Exercise 1 analysis

Exercise 1 presented a blue ink library stamp from the Fürstlich Fürstenbergische Hofbibliothek Donaueschingen (see Appendix 1A). The significance of this evidence is reflected in the descriptions provided by respondents. Most descriptions focused on the stamp itself and providing the name of the library that it is indicative of. Providing this named information is a priority to respondents when producing a description. Many respondents also highlighted the location of the stamp in their description; being on the title page. A minority of respondents described the appearance of the stamp in any detail. One respondent described the stamp as an "oval-shaped armorial stamp", and two others also indicated that the stamp was "armorial", with a fourth respondent using the term "heraldic". Only two respondents noted the color of the ink, however they did not use the same color – one was blue, the other purple (Figure 1).

The sample in exercise 1 also included a 5-digit accession number (37943) written in pencil, above the imprint. Two respondents identified the 5-digit number in their descriptions, though one identified it as a "catalogue number" while the other did not attempt to interpret the purpose of the number at all. A third respondent also identified what they interpreted as being a "catalogue number" but did not transcribe the 5 digits. The other 12 respondents all omitted the number from their description, and another respondent went as far as to explicitly note that they would *not* include an accession number in any description.

Exercise 2 analysis

The second exercise presented an image containing two distinct elements: a printed bookplate, and an inscription (see Appendix 1B), which together contained five key pieces of information: (1) The name of an individual

Frequency	Word
13	hofbibliothek
11	page
11	stamp
10	title
9	donaueschingen
6	furstenberg
5	imprint
5	ink
5	ownership

Figure 1. Exercise 1 common word list.

agent (Gillie Potter – as printed on the bookplates and written in the inscription); (2) The name of a **place** (Hogsnorton); (3) The physical **location** in the resource (front pastedown, endpaper/flyleaf); (4) The **type** of element (bookplate, inscription); (5) Transcription.

Only one respondent described the illustrated contents of the bookplate in any detail while others only specified the bookplate was “pictorial” or “illustrated”. One respondent noted the period of the bookplate as “twentieth century”. The hand-written inscription on the flyleaf proved difficult to decipher for multiple respondents. Most respondents produced transcriptions of all text, and each adopted different methods of recording the elements that they could not fully read. One respondent used an ellipsis to indicate illegible text. Another described the section as “[Unreadable]”, and another used the abbreviation “illegal.” [sic] to indicate illegibility. A fourth respondent wrote, “[S_?]” for the illegible section. The hand-written section is written in abbreviated Latin, and reads, “Gillie Potter e Coll. Vigorn. in Univ. Oxon.” identifying the owner as Gillie Potter (as is confirmed by the bookplate) who studied at Worcester College, Oxford. Three respondents included this interpretation of the abbreviated Latin inscription in their description, suggesting that the role of the special collections cataloger can go beyond description, and allow space of basic interpretation.

One respondent also identified Gillie Potter's true identity (being Hugh William Peel) and the fact that "Hogsnorton" is a fictional place (Figure 2).

Exercise 3 analysis

Three distinct elements were identified in the third exercise: the price ("£7.50") written in pencil in the top-right hand corner, the gift inscription in pen and the printed cutting which has been pasted beneath (see Appendix 1C). All respondents acknowledged and described the inscription and the cutting, with varying degrees of detail. However, six respondents did not acknowledge the price at all. Of those who did, most provided the amount written, however some did not and a few also explicitly described it as a "bookseller's price". However, the optional inclusion of the price raises the question of what is deemed worthy of recording, just as was the case with the library accession number in the first exercise (Figure 3).

The inscription itself was identified as the primary copy-specific element. All but one respondent provided a transcription for this inscription. By contrast, none provided a full transcription of the printed clipping. While some only described this element as an "inscription", or "manuscript inscription", other respondents referred to it as an "ownership inscription", "autograph inscription", or a "dedication note". The function and type of

Frequency	Word
19	potter
17	gillie
13	front
12	bookplate
12	coll univ oxon
12	hogsnorton
9	vigorn
8	inscription

Figure 2. Exercise 2 common word list.

Frequency	Word
25	pagett
17	mother
15	amy
14	inscription
13	loving
12	catherine
11	death
9	price
8	newspaper

Figure 3. Exercise 3 common word list.

copy-specific element takes precedent before other details, such as physical location or materials used. Additional elements such as a place or time have not been prioritized in this instance. Only one respondent attempted to approximately date the interventions “[ca. 1950?]”, and one other noted the absence of any indication of year. One description of the clipping includes the place, Handsworth Wood, where the subject died.

Four of the respondents recorded a connection between the inscription and the newspaper cutting. These asserted that Catherine Pagett (as named in the printed death notice which has been pasted into the book) was the mother of Amy G. Pagett, the subject of the gift inscription. This connection is an assumption, albeit a logical one, and again questions whether the role of the cataloger is to be purely descriptive, or to also provide interpretation. Those respondents who made this connection used words such as “presumed” to illustrate this information as unconfirmed. One respondent provided a note stating, “I would check before making this assumption.”

Exercise 4 analysis

Participants were presented with an inscription framed within a visual representation of a tombstone (see Appendix 1D). All but one respondent made mention of the visual nature of this element, variably describing it as a “tombstone” (6 respondents), “headstone” (3), “gravestone” (2), “memorial stone” (2), or “tomb” (1). One respondent referred to the element as, “Autograph bookplate drawn in the style of a gravestone.” A bookplate is typically a printed plate, pasted into an item to indicate ownership, which is not the same as the example used in this exercise (Figure 4).

Of the textual contents, five respondents provided full transcription, while others provided abridged transcriptions (making use of ellipses). Across the responses, three key elements were recorded: (1) Name of a primary **agent** (Reverend George Lawson), along with birth and death dates; (2) Name of a secondary **agent** (Jane Macallum Lawson); (3)

Frequency	Word
18	lawson
13	george
9	reverend
7	jane
6	macallum
6	husband
6	tombstone
5	inscription
5	ink
5	drawing

Figure 4. Exercise 4 common word list.

Relation between agents (spousal). Identifying individual named agents was prioritized by respondents; however, the relation between named agents was also commonly recorded. Nine respondents identified a spousal relationship. This presents a similar outcome as observed in the responses to exercise 3.

Exercise 5 analysis

Exercises 5 to 7 included multiple images of copy-specific elements which respondents were expected to describe in a single entry (see Appendix 1E). These longer, more complicated exercises make full transcription difficult and encourage respondents to rely on alternative descriptive practices to create succinct but detailed descriptions (Figure 5).

Responses to exercise 5 varied significantly differently to exercise 5. One respondent created three distinct descriptions for each page. Several others, perhaps overwhelmed with a vast amount of copy-specific evidence illustrated, instead provided much more concise descriptions, such as, “multiple autograph inscriptions” or just, “manuscript annotations”. Most attempted to collate all the information they saw across all three images into a single description. These descriptions sought meaning between disparate pieces of evidence. Multiple individual names and dates are inscribed in the images, but most respondents chose to group individuals together as the “Hedges and Andrews families” or “the family of Hannah Hedges”. “Family” is a commonly occurring word across all responses, despite it not featuring explicitly in the contents of the examples. The cataloger’s ability to interpret the relations of all these individuals as being of a single family has been to the benefit of accurately and efficiently describing the copy-specific elements.

Multiple dates are listed alongside these names, indicating years of significance (deaths, marriages, births, etc.). Only two respondents provided each date individually. One respondent alternatively provided a date range (1820–1846) and another simply grouped them together as being of the “nineteenth century”. A few others only provided the date “1843” which is found on the first inscription page by Hannah Hedges, and five respondents provided no dates.

Only two respondents acknowledged the price (“£5”) in their description, suggesting the inferiority of this evidence in contrast to the other information present in this example.

The term “inscription” is being overwhelmingly used to describe the copy-specific element. After the names of individuals who are described in the text, the term “inscription” is the most common term. This illustrates that “inscription” may be the preferred terminology to describe additional textual markings as opposed to terms such as “annotation”,

Frequency	Word
22	hannah
15	hedges
13	andrews
10	inscription
8	family
8	mary
7	births
6	october

Figure 5. Exercise 5 common word list.

“note”, “writing” or “manuscript”, as have been evidenced elsewhere in minority.

Exercise 6 analysis

Exercise 6 provided four images of evidence found in a book, including visual and textual material (see Appendix 1F). The visual elements have been described as “drawings” (6 respondents), “doodles” (4), “graffiti” (1) and “scribbles” (1). Seven respondents attributed the additions to a “child”, “girl” or “schoolgirl”, despite this not being made explicitly clear in the images provided. No respondents attempt to describe what the drawings depicted.

Some respondents attempted full transcriptions of textual elements and one respondent noted, “I would describe what appears on each page.” Alternatively, many of the respondents kept their answers concise, providing a general overview of what they saw. Information such as physical location in the resource, place, or time have been of lesser importance in this example. Materials were also not deemed particularly important, with only two respondents noting that the drawing and writing were composed in “biro” (Figure 6).

Frequency	Word
12	robert
11	joy
10	annotation
8	hand
6	eliza
6	drawings
5	throughout
4	doodles
4	child's

Figure 6. Exercise 6 common word list.

Whereas the textual element of the exercise 5 was largely described as “inscriptions”, here the most commonly applied term for textual matter is “annotations”. This can be explained due to the nature of the textual elements, which are often written in the printed text of the book, as opposed to on flyleaves, and are directly engaging with the printed text (paratext). Therefore, the authorial function of these para-textual elements determines how they are perceived by catalogers and, consequently, the terminology used to describe them. Any classification or categorization of these terms needs to consider the subtle variations in the functionality of markings.

Exercise 7 analysis

The final exercise was chosen to be the most difficult to describe (see Appendix 1G). Multiple distinct types of material evidence are present, and the amount of information is too complex and vast for full transcription. The intention of this example was to force participants to provide a general description that would be able to capture what they perceived

to be the most important aspects of the copy-specific elements in the book. The following list illustrates some of the main areas that the respondents recorded in their descriptions: “postcard”, “annotations”, “bookplate”, “hand-drawn map”, “underlinings” (Figure 7).

Some of these elements are self-contained. The bookplate provides a regular piece of copy-specific information which catalogers are familiar with describing. Thirteen respondents acknowledge the bookplate explicitly, and they all mention it as the first thing in their description. The bookplate reads, “From the library of Paul Kenneth Baillie Reynolds”; therefore, providing the name of a former owner. This personal information has been prioritized in almost all descriptions, and illustrates the importance of identifying provenance in books. One respondent also described the style of the bookplate as having an “arts and craft border”.

Other elements, however, are far more irregular, and contain various pieces of information within them. The postcard – which thirteen

Frequency	Word
25	reynolds
15	postcard
13	paul
12	annotation
11	kenneth
11	inserted
10	bookplate
10	baillie
9	front
9	throughout
8	map

Figure 7. Exercise 7 common word list.

respondents directly recorded – is an example. Firstly, there is the printed contents of the postcard on the recto, which suggests it was originally acquired at “Hotel Beauséjour – Hyères les Palmiers” (this was recorded by four respondents). On the verso, there is an inscribed message (no respondents transcribed this, but one did summarize the contents and another noted “In a more detailed description I would read the postcard and describe the type of handwritten annotations”), the names of the recipient and sender of the postcard (8), the date (2), the postage stamps and library stamp (neither recorded by any respondents).

“Underlining” was also specified by four respondents as a separate intervention. Two of the respondents used “ephemera” to refer to all inserted materials. Another commonly used word across the descriptions was “extensive(ly)” – used to indicate a large quantity of annotation, without providing itemized detail. Both “extensive” and “ephemera” are used as terms which can reduce the size of the description, but also provide readers with a useful “signpost” as to what they can expect from the item.

The materials used to create copy-specific interventions were described by four respondents, who identified “pen”, “pencil”, and “ink” to describe how annotations and other additions were created. Physical location in the book was also noted to some extent by nine respondents, who made references to “front pastedown”, “front covers”, and “flyleaves”. Two respondents also provided specific pages for certain items. The first image illustrated four distinct entities: the bookplate, a hand-drawn map, a price (“£2.50”) and a shelfmark (“Collins.And.74”). Only three respondents recorded the price, and none recorded the shelfmark.

General observations on descriptive exercises

Some fundamentally different approaches to descriptive practice can be observed across the responses to the above exercises. The most notable difference is in the depth of description. Several respondents aimed to maximize detail in their description, including full transcriptions. In a response to exercise 5, one respondent included, “[transcribe all]” at the end of their description to suggest that in a real-life cataloging situation they would provide a full transcription of what appeared on the page. In contrast, two of the respondents chose to consistently apply the most minimal of descriptions, only providing an indicator of what the reader might find in the book, such as: “Manuscript annotations: genealogical notes”, “Gift inscription with note on donations”, or “Bookplate, ownership inscription”. In the final exercise, the longest description consists of eighty-nine words, while the shortest is only five.

Both these approaches seem to have their own respective merits and limitations. A lengthy description provides the user with more information about the source they are seeking to consult. However, the time required to produce such a description may be considered more than the description is worth. Brief descriptions also allow for greater consistency in terms, which can be observed in the examples listed above. A notable exclusion from shorter descriptions is personal names. The common word analysis generally indicates that the inclusion of names are prioritized by most catalogers, however this information is regularly omitted in shorter entries, with one respondent not using personal names in any of their answers.

The value of copy-specific cataloging

Following the descriptive exercises, the survey continued with a series of questions on the value of copy-specific cataloging, seeking to understand what catalogers believed was worth recording, and why. These questions were optional and some of them provided “tick box” answers, though most still provided free-text options to ensure a depth of qualitative data could be collected.

Respondents were initially asked how they would define copy-specific cataloging. Many focused on the “uniqueness” of the item and provided some examples. Some descriptions were extremely broad, e.g., “anything that has happened to the book after printing, including the binding”, “everything to do with the history of the object”.

A majority of respondents believed the recording of copy-specific elements in books was either “Extremely important” (53.3%) or “Important” (33.3%). Two respondents provided their own answer, one stating that the importance depends on the type of book, and the other stating that it depends on the type of library. This latter answer raised practical considerations regarding the purpose of the library collection, the interests of its users, and the availability of time and human resources for cataloging.

Respondents cited that special collections users required adequate copy-specific cataloging in order to efficiently conduct research. One respondent noted, “Researchers are increasingly interested in the historical usage and transmission of books through society” and this opinion was corroborated by another respondent who wrote, “if the cataloguer does not record them, researchers have no way of finding them other than looking at a *lot* of books”. A further respondent also shared this view, and stated, “I need this information to know better the book’s history”. The need to meet user’s requirements was the most common reason provided for ensuring good copy-specific cataloging practices. Most respondents shared the opinion that copy-specific cataloging helps in telling the “history”/“story” of collections. Some participants also noted the

importance of recording copy-specific aspects of collection items as a way of promoting the collection itself.

The next questions in this section sought to establish some of the practical considerations that should be taken when recording information. Participants were asked what they expected to be recorded in a catalog record. The most popular elements chosen were: inscribed names, bookplates/book labels (100% of respondents believed these should be recorded), library stamps, additional inserted materials (93.3%), marginalia and notes (86.7%), inscribed initials, doodles (80%). Elements of lesser importance included bookstore labels, waste fragments, prices and booksellers' notes. When considering the specific attributes of these individual types of copy-specific material evidence, names and dates were perceived to be the most important information to capture (93.3%) followed by the physical location of the copy-specific element in the book (80%) and numbers (shelfmarks, lot numbers, etc.) (73.3%). Properties such as color, size, and materials feature lower in priority.

Participants were also asked what *should not* be included. Suggestions included accession numbers, lengthy transcriptions, and details of annotations. One respondent also noted that the extent of the description is dependent on the value or importance of the book itself: "It depends on the nature of the evidence and age of the book. The older the book the more that should be recorded".

Of the respondents, 35.7% believed that more than 30 minutes should be spent recording copy-specific information for a single item, while 28.6% believed 10–20 minutes was appropriate, with the same number stating that 5–10 minutes was appropriate. Only one participant suggested that 1–5 minutes should be spent.

Many respondents also believed that catalogers should conduct their own research to enhance records – from "brief googling" to "as much as necessary" – and provide transcriptions, where practical. Most respondents promoted an, "as much as possible" attitude to copy-specific cataloging and emphasized that this additional time is well spent to provide historical context to the item. However, two respondents disagreed with this position. One stated that the record should "only point the user to the evidence" as opposed to providing any interpretation or context. Another respondent also argued that a cataloger's time should not be spent researching: "I would record what I see, and leave the additional research to ... researchers! Cataloguers are pathfinders, not research staff!" Ten participants responded positively to the idea that external researchers could be expected to contribute to catalog records, thus saving the time of the cataloger, though some respondents held reservations of the practicality of this strategy.

Metadata requirements for copy-specific cataloging

The final section of the survey primarily focused on the technical parameters in encoding copy-specific information. As most of the respondents were self-identifying library professionals, 100% of them were familiar with MARC as a metadata standard, and most (86.7%) with AACR2. Fewer, though still a majority, were familiar with RDA and DCRM, though a minority were familiar with the concept of linked data.

When asked if they believed there was adequate existing guidance for recording copy-specific information, most responses were unsure. Many stated that standards, such as DCRM and MARC 21, provided some useful instruction for recording copy-specific information, though there were gaps. Two respondents claimed that there was a lack of standardization in terminology used, though a third respondent claimed that existing controlled vocabularies provided useful guidance but are not always properly applied in UK libraries. Multiple respondents also raised the point that, unlike bibliographic cataloging, copy-specific cataloging guidance is often more specific to institutions. Many also acknowledged that there is a great deal of room for interpretation in prescribed standards, often meaning that consistency is lost between institutions.

When asked if MARC provided an adequate framework for recording copy-specific information, respondents offered a similar variety of opinions. Six of the respondents believed MARC has presented serious limitations to copy-specific cataloging. Two respondents specifically highlighted MARC's inability to "link" as a serious limitation to the system. One respondent wrote, "you can't hyperlink to names in a note or inscription in the way you would if it was a webpage." Other critics argued that MARC lacked sufficient fields to capture the granularity of detail required for copy-cataloging. Five respondents, however, believed MARC was adequate, with one respondent arguing that the structure of the data was secondary to the data itself: "The content is more important of the form which contains it." Multiple respondents identified the problem with "consistency of terminology", and the lack of an agreed vocabulary as well as the inability to include authorized entries within a MARC holdings record.

Discoverability

Of the respondents, 46.7% believed that copy-specific information is "hard to find" across most online library catalogs, with 40% believing it to be "discoverable with the right knowledge" and only 13.3% believe it is "easily discoverable". This corroborates the data presented in the RLUK *Hidden Collections* report.²⁷ When asked how to improve discoverability, most respondents believed that systems needed to change to better suit complex search behaviors. One respondent specifically suggested that special

collections professionals should do more to influence the development of library management systems and cataloging software:

“[Greater discoverability requires] bringing together relevant people from Special Collections to influence systems development. In many libraries the needs of SC are seen as a specialist; it is difficult to make changes to a standard library discovery.”

Respondents offered a variety of specific technical features which they believed would improve systems to allow greater discoverability of copy-specific elements in special collections, including specific provenance-related search facets, advanced field searching, standardized terminologies, full text retrieval from notes fields, the creation of more detailed metadata, linking data, and data visualizations to enhance search behaviors. Other non-system specific suggestions focused on the importance of providing adequate instruction to users so that they understand how to navigate library catalogs efficiently.

As a final practical exercise, participants were asked to test the discoverability of an existing record with significant copy-specific information in the *Explore the British Library* OPAC. The item was a copy of Brontë poetry which was owned and annotated by Emily Brontë, and respondents were tasked with discovering it. Of the fifteen respondents, only five were able to successfully identify the shelfmark of the item, while two others found shelfmarks for the wrong holding of the item (one without the annotations), and the rest were completely unable to find the item. One respondent (who, despite finding the item, admitted it was difficult) also provided their method of discovery:

“I restricted by date to Emily’s lifetime and searched for ‘Bronte’ Anywhere and Material Type ‘Books’ which gave me 22 results and this was the only one called ‘Poems.’”

This final exercise reinforces the inadequacies of the discoverability of copy-specific records in most library catalogs. Even within a sample group largely consisting of information professionals, only a minority could locate this record. The record in question was created adhering to current cataloging practices, and included authorized added entries for “Emily Brontë” as former owner, as well as detailed 5XX copy-specific notes detailing the provenance of the copy. Despite these efforts, the evidence suggests that the record remains largely hidden.

Stage 2: Action: Data classification and modelling

The survey responses highlighted two primary obstacles to best practice in copy-specific cataloging: (1) a lack of standardization and consistency

in terminology and methodology; (2) a lack of meaningful structure of data encoding in existing cataloging systems. To address these issues, a data ontology was created (see [Appendix 2](#)), derived from the comments and suggestions offered by participants in their responses. The ontology consisted of: (1) Core requirements of copy-specific cataloging; (2) List of fourteen main data entities; (3) Sample controlled vocabulary; Data modeling in MARC and RDF triples. The ontology provides a set of core elements and attributes, in the form of a data dictionary, to provide a logical plan establishing the relationships between metadata elements and the semantics of these elements and relationships.

Stage 3: Evaluation

The data ontology was evaluated through a discussion with two special collections practitioners with knowledge and experience of copy-specific cataloging for rare materials. Both participants were also respondents to the initial survey. This discussion was informed by a set of guiding questions, but respondents were also given the opportunity to discuss other areas which they found to be of interest.

Core requirements

Both participants shared thoughts on what they believed to be the core requirements of copy-specific cataloging in special collections. The first participant believed the primary requirement of any ontology was to identify the “significant elements” of an item. When pressed on what made an element “significant”, they prioritized elements which illustrated the history of the object, and specifically noted that they believed recording provenance was the most important aspect of copy-specific cataloging.

The second participant was more concerned with the “searchability” of records as a core requirement: “If you’re doing loads of detailed work, [...], but you can’t actually search in that way, then I just think, ‘what’s the point?’” Any data ontology for copy-specific cataloging had to prioritize access to information, not just instructing what information to capture in a record. The second participant also highlighted that the inclusion of a controlled vocabulary was a core requirement for any instructional document, as a means to overcome inconsistency.

Controlled vocabularies

Both participants reacted positively to the inclusion of a controlled vocabulary in the ontology and believed that a controlled vocabulary was necessary to ensure consistency and clarity. The consistent application of controlled vocabularies in metadata facilitates the categorization,

indexing, and retrieval of information.²⁸ To this end, the second participant raised their concerns that inconsistency of terminology would negatively impact discoverability of records, “[The user] may miss things because they are searching for terms that may not be the ones that people have indexed it as.”

When asked if they already use any kind of controlled vocabulary in their own practice, the first respondent said that they do, and the second said that they do not currently, but they did in their previous institution. Both participants highlighted that their use of terminology largely depended on institutional practice. The first respondent explained that they use an in-house vocabulary, which is based around the specific requirements of their collections. Similarly, the second respondent also used an in-house controlled vocabulary at a previous institution.

When asked if they believed the construction of a cross-institutional controlled vocabulary would be beneficial, both agreed it would, and suggested that ideally a controlled vocabulary be followed as closely as any other prescribed cataloging rules. Despite this, both had reservations about the practicality of a universal vocabulary being implemented, used, and maintained. The first respondent noted that the successful adoption of a universal controlled vocabulary required proper consultation across all institutions. They reflected on the failure of the RBMS controlled vocabulary to be properly adopted across many UK research libraries: “I think we [UK research libraries] are very reliant on standards that are set in the US. Maybe that means that we should get more involved in setting those standards.” The RBMS Controlled Vocabularies Editorial Group do periodically accept proposals for new terms to be added to the vocabulary, so perhaps this offers an opportunity for greater involvement from UK special collections professionals.

When asked to comment on the contents of the controlled vocabulary in the ontology, both participants responded positively. The second participant suggested that a controlled vocabulary could never be too detailed and argued that areas such as binding description could go into a level of extreme granularity. The first participant responded positively to the terminology for annotations in the controlled vocabulary, and suggested, “That is what the researcher was most interested in.” When considering the limitation of existing vocabularies for annotations, they cited a specific experience with a researcher: “he was looking for more specific evidence, which just something labeled as annotations doesn’t provide because annotations come in so many different ways.” Generally, both respondents believed the sample vocabulary was an important part of the ontology, but that it could be improved and extended.

Data modelling and structuring data

The ontology also illustrated potential modeling solutions for copy-specific cataloging. It provided a pictorial representation of a bookplate, followed by three methods of description: a rich-text note, a structured MARC record, and RDF triples.

When asked if they believed the rich-text description was adequate, the first respondent answered, “No – but I think it is the most common.” The second participant had a more positive response to the note, and stated, “I do like a chunk of text when I'm reading something. It really annoys me if there were just headings and there's no kind of meaty detail.” Both respondents agreed that the description needed to provide adequate detail, though the first respondent believed this could be achieved in a better, more structured way.

Both responded most positively to the second representation. The first respondent noted that this was the same approach they took in their institution. The first respondent cited a recent instance where they were attempting to draw copy-specific metadata from across collections via the Jisc Library Hub union catalog and identified that using the 655 MARC fields was the easiest way to gather this information. The second participant also raised some concern about the time resource needed to construct a MARC record in the way illustrated in the second representation and suggested that creating a single note can often be a better use of limited resources.

The final representation was articulated in RDF triples and introduced concepts of linked open data. When asked if they believed MARC was an adequate format to record copy-specific information, there was a difference of opinion. The second respondent believed MARC was adequate, and that the limitations of the format existed more within the search capabilities of user-facing library management systems. However, the first respondent believed firmly that MARC was inadequate to this end. They elaborated on this point:

“I think the issue comes down to making the links between items, which in the RDF is a more concrete link between the evidence and the agent, the date range and the physical location. At the moment, in MARC these are all in different fields and the more evidence there is in a book, the more difficult it is to actually interpret.”

Both respondents suggested that “linking” data was beneficial and crucial for recording copy-specific information in a meaningful way. Creating links between records across different datasets was perceived to be a good thing, and respondents saw this as a potential way to enrich records. Both respondents articulated an awareness that special collections cataloging practice would need to change fundamentally to cope with the expectations of current researchers. They both envisioned this change taking the form

of a decentralization of information, whereby individual institutions fed their data into larger shared systems.

Being able to query this information was also raised by the second respondent as a crucial consideration when modeling data. They raised a concern, similar to Taylor's fear of "mega-choice", where too much information and too many links could become unnavigable for the user. To this end, ensuring access and searchability was a priority: "Catalogues can't do complex searches. So why are you complex cataloguing in the first place?"

Discussion

In an OCLC report on special collections discoverability, "note-ification" was identified as a serious limitation to current descriptive data models and their lack of navigable structure.²⁹ This largely reflected the common practice in MARC cataloging of inputting most copy-specific metadata as a single string into 5XX fields which, due to a lack of structure and full-text search capabilities, could not be queried by users. This lack of structure inhibits discoverability of the unique, copy-specific aspects of collection items. The "hidden" nature of special collections metadata was also identified by survey participants. Inconsistency in structure and terminology were identified as key contributors to this challenge.

Over half of survey respondents acknowledged the limitations of MARC as an encoding standard, with a majority responding favorably to the prospect of greater structure to copy-specific information. The benefits of adopting alternative models for metadata description have been promoted by practitioners, such as Harpring, who claim these models allow institutions to, "share our knowledge and expertise with the international art and cultural heritage communities in their broadest sense."³⁰ Despite this, respondents were cautious of moving away from MARC. Despite existing non-MARC bibliographic databases being better suited to recording copy-specific information, special collections librarians often cannot justify duplicating cataloging efforts. As one discussion participant said, "we don't have time to enter data in 2 places, because the 2 systems are structured very differently."

Identifier-based linked data offers a more structured approach to metadata, which allows "inclusion and increased visibility of item-specific information as well as variances and the surfacing of resource uniqueness."³¹ Alternative ontologies for rare materials metadata creation are already being developed by the Linked Data for Production (LD4P) cooperative, whose Rare Materials Affinity Group have produced an Art and Rare Materials Ontology designed to be expressed in BIBFRAME. Many of the challenges identified in this paper could be overcome through the enhancement and conversion of MARC data to linked data. However, most of the survey participants were not familiar with the possibilities of linked data for special

collections cataloging, and in analytic interviews, respondents acknowledged the value of “linking” records, but admitted that they could not visualize how this could be implemented. Practical concerns were articulated as being the greatest barriers to implementation of any new practices in cataloging. These considerations concur with observations made by those who argue that, while the ambitions of linked data may be promising, implementation requires resources that are currently unavailable.³²

In the ontology, the second data model was most popular with respondents, on the basis that it represented a compromise between traditional MARC cataloging and structured data. Consequently, this study identified that structuring MARC fields – using added entries and 561 sub-fields – offers a manageable step toward best practice in capturing copy-specific information in special collections.

Consistency was a priority among the functional metadata requirements of most participants in both stages of the research process. This consistency was achieved, in part, through the development of standardized terminology and data types, represented in the creation of a data dictionary and controlled vocabulary. The variation in description types across the descriptive exercises illustrated the inconsistencies of local practice, as was suggested by Green and García-Monge.³³ Through the analytic interviews, these differences were observed in case studies, where individual institutions relied on in-house documentation which served the unique needs of that collection. These “silos” of information proved problematic for researchers who wished to conduct work which surveyed items from various collections. Therefore, the development of a suite of controlled vocabularies that could be applied to all special collections was one that respondents felt was certainly necessary. While existing resources such as the CERL or RBMS thesauri were discussed, overall respondents made limited use of these resources due to the limited ability of individual institutions to influence their development.

However, it is also important to note that the priority requirements of catalogers varied between institutions. Through both stages of the research process, respondents provided detailed insights into their own professional experiences, and the individual needs of the institutions they represent. A discussion participant noted that their institution prioritized markings of provenance as the most important requirement, and consequently all practice and documentation supported that goal. In contrast, a respondent to the survey made multiple references to the importance of detailed binding descriptions in their practice – something a third respondent saw as being of lesser importance due to their complexity.

It is evident that, while an RDF-based model for special collections cataloging may provide an ideal format for detail capture and discoverability, its implementation cannot be realized without a fundamental change

to library systems and knowledge. The resources required for this change are substantial and will not be achievable across all collections. Therefore, we must continue to think in MARC for current best practice. However, there is the possibility to anticipate an eventual change in how we model our data. As Welsh³⁴ noted, some national libraries such as the British Library or Bibliothèque nationale de France are already converting their MARC records into RDF datasets and making them available via repositories for others to query and use. Anticipating a workflow to convert MARC records to RDF or XML *en masse* could ensure that copy-specific data are recorded in such a way that, when individual institutions are able to start making the changes from MARC to a new encoding system, their metadata can be easily migrated. In the short term, this also allows catalogers to develop a best practice model which encourages structured and standardized data input, without superseding MARC cataloging.

Though respondents to this study were skeptical of the idea of adopting wholly new data models for descriptive cataloging, there was a consensus that systems would need to change to reflect the broader information landscape and new research behaviors. Similar hesitation was articulated during the scoping phase for developing the Getty vocabularies (and particularly the Art & Architecture Thesaurus) in LOD. Despite this, when implemented, Getty observed how a standardized and consistent data model, shared through linked data, allowed their resources to be exploited in new ways for retrieval and discovery purposes.³⁵ Favorable attitudes toward resources such as Material Evidence in Incunabula (MEI) and similar copy-specific focused projects suggest there is an appetite among special collections practitioners to adopt new ontologies to overcome the challenges of hidden collections. Yet, limited resources and expertise present barriers to successful implementation. It is these practical limitations that must be addressed first. As Byrne and Goddard observed, “[linked data] has a long way to go before it is seen as a standard foundation for library data [...] the majority of issues are non-technical in nature. The technology is ready; it is now a matter of getting libraries and librarians ready as well.”³⁶

Notes

1. Fortis Lazarinis, *Cataloguing and Classification: An Introduction to AACR2, RDA, DDC, LCC, LCSH and MARC 21 Standards* (Waltham, MA: Elsevier, 2014), 12–13.
2. Pat Riva, Patrick Le Boeuf, and Maja Žumer, *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information* (Den Haag: IFLA, 2017).
3. Jonathan Rose, “The Horizon of a New Discipline: Inventing Book Studies,” *Publishing Research Quarterly* 19, no. 1 (2003): 11–19, <https://doi.org/10.1007/s12109-003-0019-1>.

4. Elaine Stead, *Reading Books and Prints as Cultural Objects* (Cham: Palgrave Macmillan, 2018), <https://doi.org/10.1007/978-3-319-53832-7>.
5. Lisa Jardine and Anthony Grafton, "Studied for Action': How Gabriel Harvey Read His Livy," *Past and Present* 129 (1990): 30–78.
6. William H. Sherman, *John Dee: The Politics of Reading and Writing in the English Renaissance* (Amherst: University of Massachusetts Press, 1995).
7. William H. Sherman, *Used Books: Marking Readers in Renaissance England* (Philadelphia: University of Pennsylvania Press, 2008).
8. Sherman, *Used Books*, 186.
9. Richard P. Smiraglia, "Rethinking What We Catalog: Documents as Cultural Artifacts," *Cataloging & Classification Quarterly* 45, no. 3 (2008): 25–37, https://doi.org/10.1300/J104v45n03_04.
10. David Pearson, *Provenance Research in Book History: A Handbook* (London: British Library & Oak Knoll Press, 1998), xi.
11. Christina Kamposiori and Susan Crossley, *Evidencing the Impact and Value of Special Collections: RLUK Report* (London: Research Libraries UK, 2019).
12. Kirsten Jones, "The Pedagogical Value of Provenance Research in Rare Book and Cultural Heritage Collections: Rediscovering the First Collection at the Frick Fine Arts Library" (2020).
13. Jennifer Schaffner, "The Metadata IS the Interface: Better Description for Better Discovery of Archives and Special Collections," in John Michalko, ed., *Making Archival and Special Collections More Accessible* (Dublin: OCLC Research Report, 2015).
14. Bibliographic Standards Committee, Rare Books and Manuscripts Section, Association of College and Research Libraries, *DCRM(B): Descriptive Cataloging of Rare Materials (Books)*, third printing with corrections, November 2011 (Washington, DC: Cataloging Distribution Service Library of Congress, 2007). <https://rbms.info/files/dcrm/dcrmb/DCRMB3.pdf>.
15. IFLA Rare Books and Special Collections (RBSC) Section, "Competency Guidelines for Rare Books and Special Collections Professionals" (Den Haag: IFLA, 2020), <https://repository.ifla.org/handle/123456789/1382>.
16. Isabel García-Monge and Daryl Green, "Introduction," *Cataloging & Classification Quarterly* 54, no. 5–6 (2016): 277–281, <https://doi.org/10.1080/01639374.2016.1209799>.
17. Sarah A. Buchanan, "Special Collections and Archives Code (SPAC): Preserving Provenance in a Bibliographic Catalog," *Cataloging & Classification Quarterly* 49, no. 5 (2011): 349–360, <https://doi.org/10.1080/01639374.2011.580446>.
18. Green and García-Monge, "Introduction," 280.
19. John Attig and Robert Maxwell, "Reconsidering DCRM in the Light of RDA," RBMS Bibliographic Standards Discussion Paper (2010).
20. Roy Tennant, "MARC Must Die," *Library Journal* 127, no. 17 (2002): 26.
21. Peter Murray, "MARC isn't Dead, but it is a Dead End," Disruptive Library Technology Jester (blog), October 29, 2010. <https://dltj.org/article/marc-as-dead-end/>.
22. OCLC Research Archives and Special Collections Linked Data Review Group, *Archives and Special Collections Linked Data: Navigating between Notes and Nodes* (Dublin: OCLC Research, 2020). <https://doi.org/10.25333/4gtz-zd88>.
23. Cristina Dondi, "CERL's Work and Vision for Provenance Research I: CERL Thesaurus, Material Evidence in Incunabula, and the 15cBOOKTRADE Project," *Bibliofilia: rivista di storia del libro e di bibliografia* 117, no. 3 (2015): 317–320.
24. Eero Hyvönen et al., "BiographySampo – Publishing and Enriching Biographies on the Semantic Web for Digital Humanities Research," *Lecture Notes in Computer Science* 11503 (2019): 574–589. https://doi.org/10.1007/978-3-030-21348-0_37; Mikko

- Koho et al., "Linked Death—Representing, Publishing, and Using Second World War Death Records as Linked Open Data," in Eva Blomqvist et al., eds., *The Semantic Web: ESWC 2017 Satellite Events. Lecture Notes in Computer Science 10577* (Cham: Springer, 2017). https://doi.org/10.1007/978-3-319-70407-4_45.
25. John Elliott, *Action Research for Educational Change* (Buckingham: Open University Press, 1991), 71.
 26. Jezz Fox, Craig Murray, and Anna Warm, "Conducting Research Using Web-Based Questionnaires: Practical, Methodological and Ethical Considerations," *International Journal of Social Research Methodology* 6, no. 2 (2003): 167–180, <https://doi.org/10.1080/13645570210142883>.
 27. Mike Mertens and Dunja García-Ontiveros, *Hidden Collections: Report of the Findings of the RLUK Retrospective Cataloguing Survey in Association with the London Library: RLUK Report* (London: Research Libraries UK, 2014).
 28. OCLC, *Archives and Special Collections Linked Data*.
 29. Patricia Harpring, *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works*, updated edition (Los Angeles: Getty Research Institute, 2013).
 30. Harpring, *Introduction to Controlled Vocabularies*, x.
 31. OCLC, *Archives and Special Collections Linked Data*, 8.
 32. Melanie Tam, "Improving Access and 'Unhiding' the Special Collections," *The Serials Librarian* 73, no. 2 (2017): 179–185, <https://doi.org/10.1080/0361526X.2017.1329178>.
 33. Green and García-Monge, *Introduction*.
 34. Anne Welsh, "The Rare Books Catalog and the Scholarly Database," *Cataloging & Classification Quarterly* 54, no. 5–6 (2016): 317–337, DOI: 10.1080/01639374.2016.1188433.
 35. Jennifer Cobb, "The Journey to Linked Open Data: The Getty Vocabularies," *Journal of Library Metadata* 15, no. 3–4 (2015): 142–156, <https://doi.10.1080/19386389.2015.1103081>.
 36. Gillian Byrne and Lisa Goddard, "The Strongest Link: Libraries and Linked Data," *D-Lib Magazine* 16, no. 11/12 (November/December 2010).

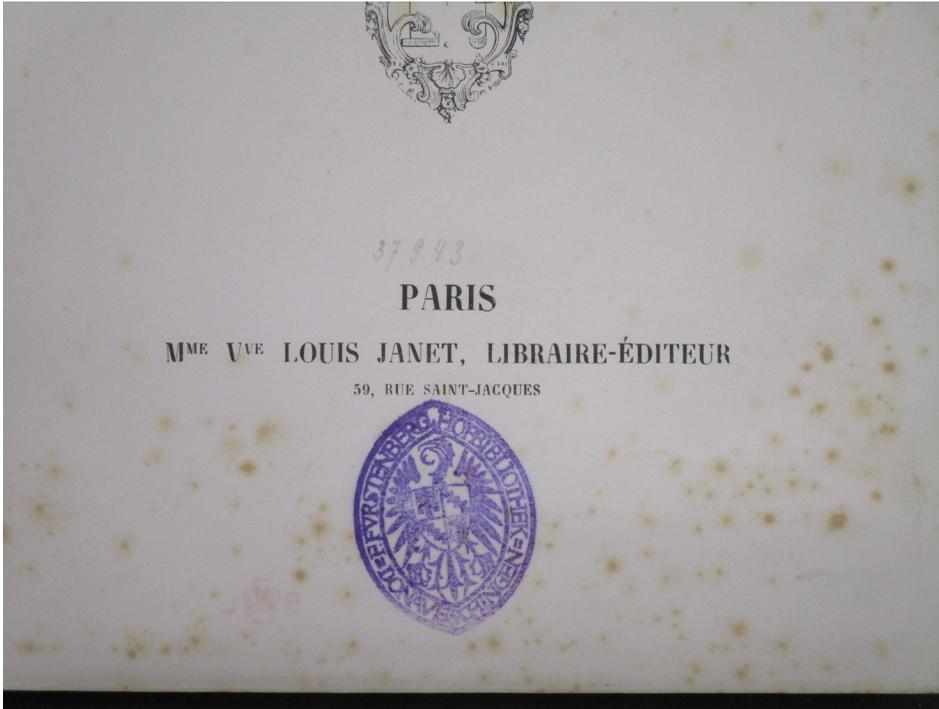
Disclosure Statement

No potential conflict of interest was reported by the author(s).

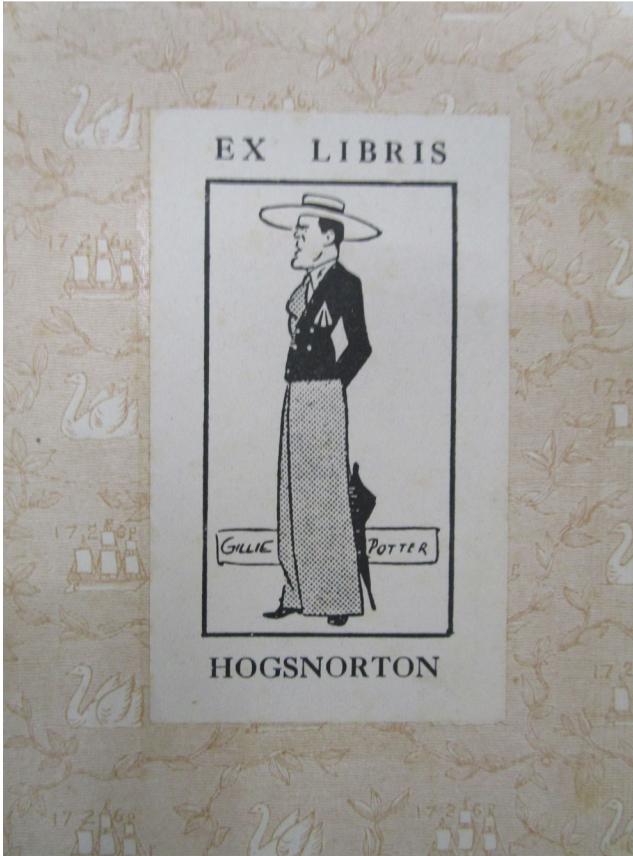
Appendix 1. Exercise images

This appendix includes the images that were provided to participants during the seven descriptive exercises that made up the core of this research.

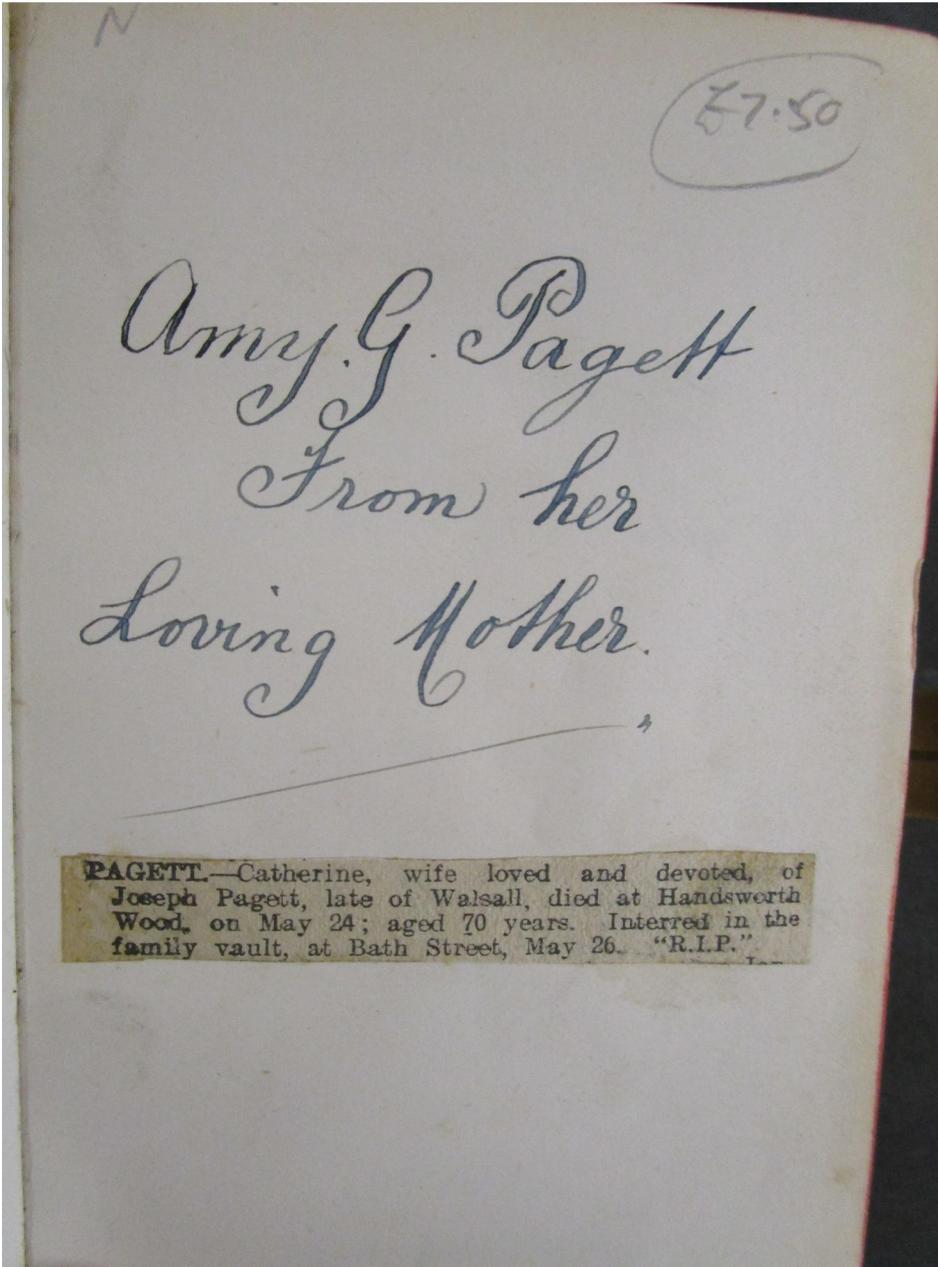
1A. Exercise one (BL Collins.Add.186)



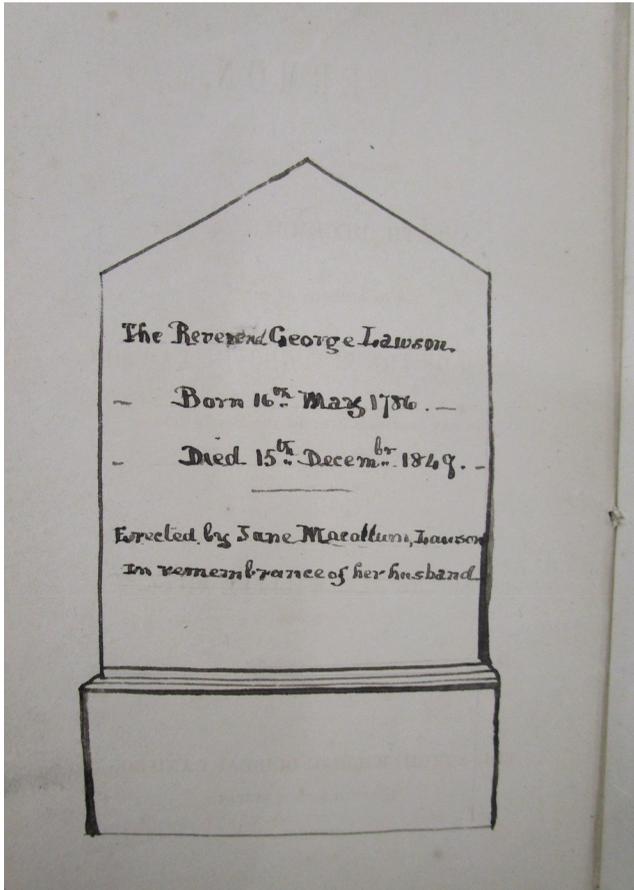
1B. Exercise two (BL Collins.Add.71)



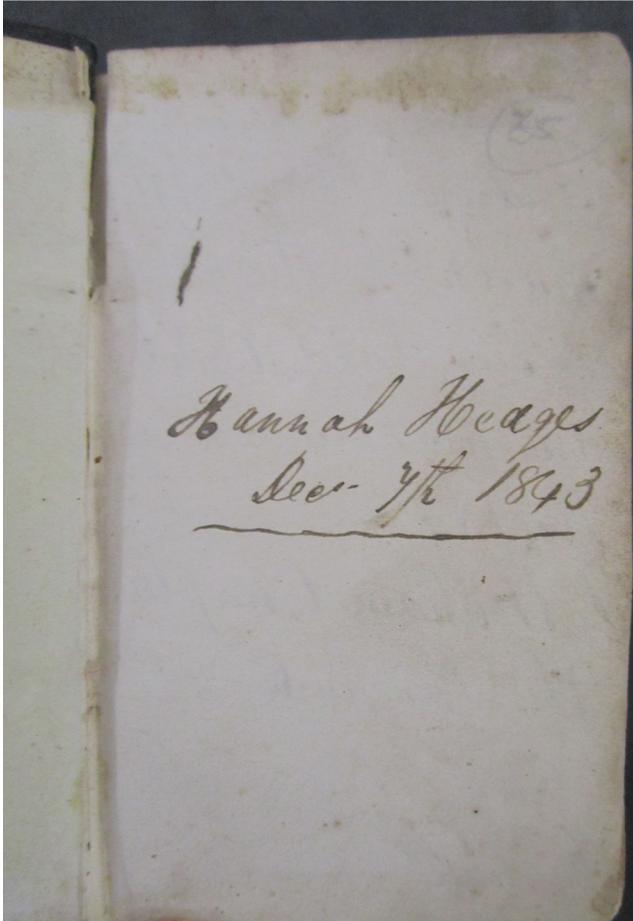
1C. Exercise three (BL Collins.Add.199)



1D. Exercise four (BL Collins. Add.84)



1E. Exercise five (BL Collins.Add.67)

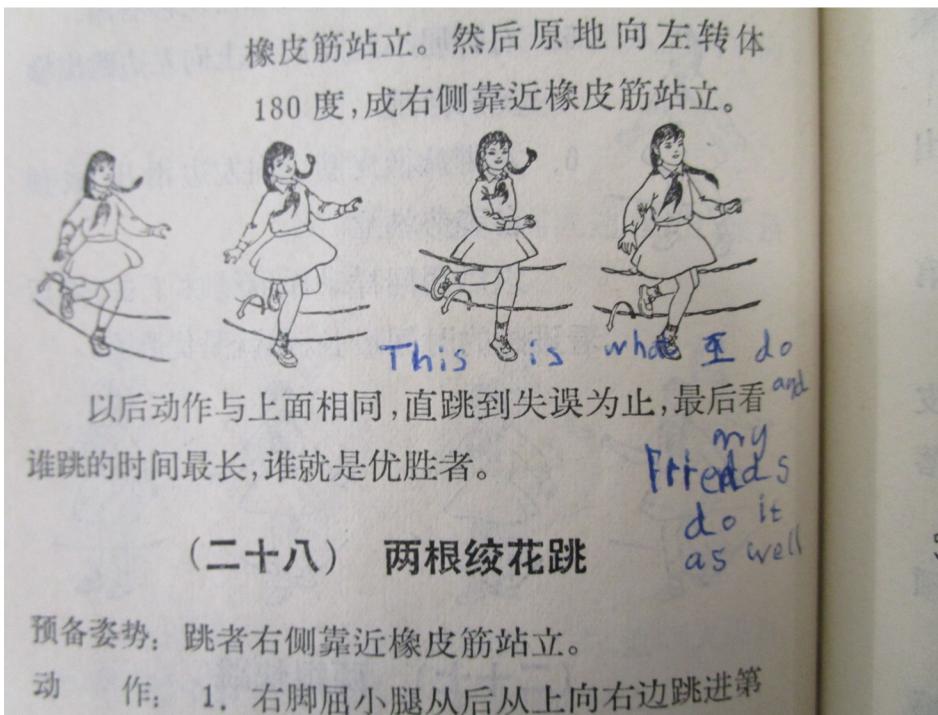
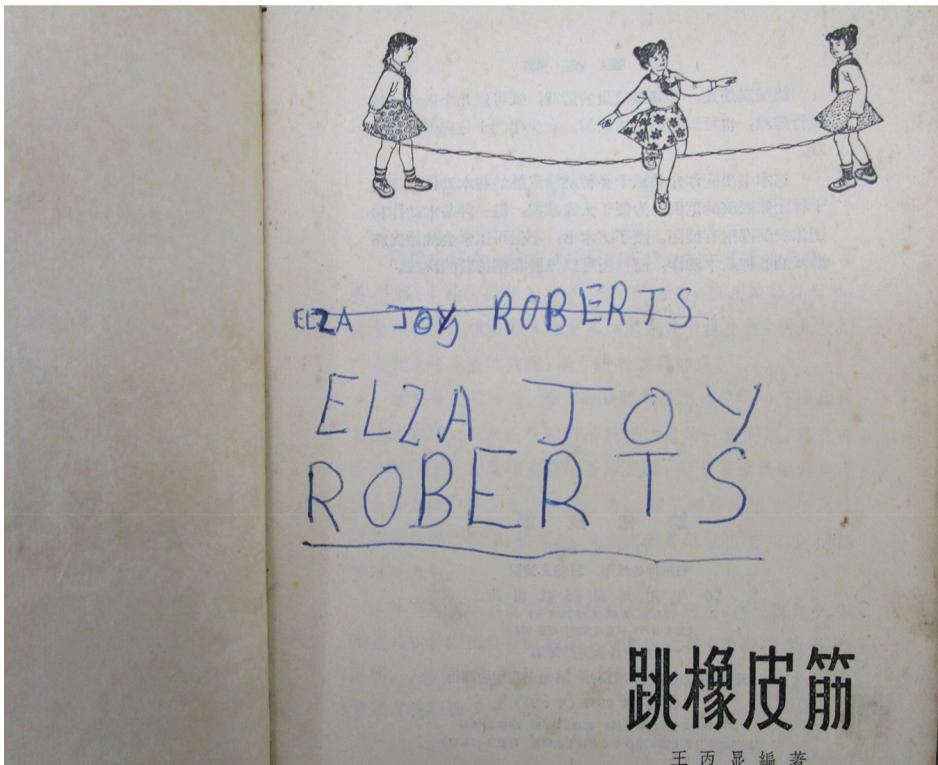


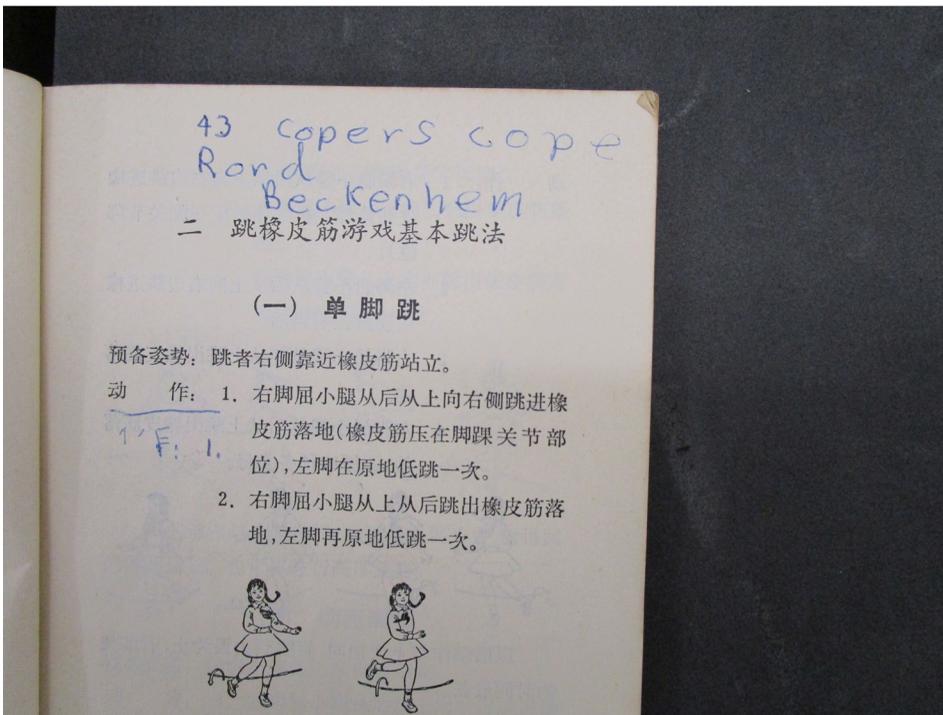
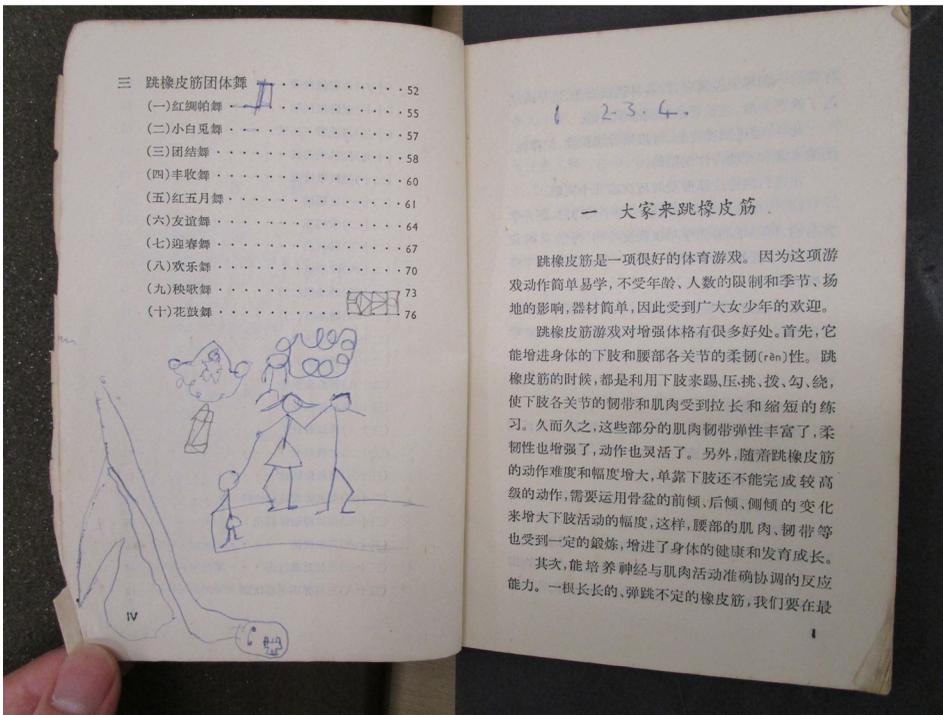
Hannah Mary
Andrews. Died 11th
March, in the year
of our lord 1846
and was buried on
the 18 of March at
Tottenham Chapel
St. Pancras

THE B
COMMON
AND ADMINI
THE SAC
AT
OTHER RITES ANI
THE CI
ACCORDING TO THE USE
OF ENGLAND A
TOGETHE
THE PSALTER OR
POINTED AS THEY ARE
CHUR
AND THE FORM AND MANN
AND CONSE
BISHOPS PRIESTS
OXFO
PRINTED AT THE UN
SOLD BY E. GARDNER AND SON,
PATERNOSTER R
Nonp. 32's.
M.DCCC.X

<p> ven race 67 ve render 75 tho to my 77 r-falling 81 O Lord 86 and 88 d, shall 89 of d 107 y Lord 110 , who 112 ft my 121 who 123 ly Lord 126 tetest 139 y cries 141 rful 142 God 145 govern'd 2 r, since 27 loud 45 us off 74 th 93 t all 100 y soul 102 114 let 117 God 125 sons 126 less 127 ed 137 t 138 ght 29 113 148 </p>	<p> Francis John Andrews Born October 9th in the year of our lord 1823. </p> <hr/> <p> Hannah Hedges Born October 4th in the year of our lord 1820. </p> <hr/> <p> F. J. Andrews & Hannah Hedges married June the 2nd 1844. </p> <hr/> <p> Hannah Mary Andrews Born May 1st 1845 in the year of our lord 1845. </p> <hr/> <p> Francis John Andrews Born the 30th October in the year of our lord 1846 christened at newbury Church Oakes </p>
---	--

1F. Exercise Six (BL Collins.Add.143)





橡皮筋站立。然后原地向左转体
180度,成右侧靠近橡皮筋站立。



以后动作与上面相同,直跳到失误为止,最后看
谁跳的时间最长,谁就是优胜者。

(二十八) 两根绞花跳

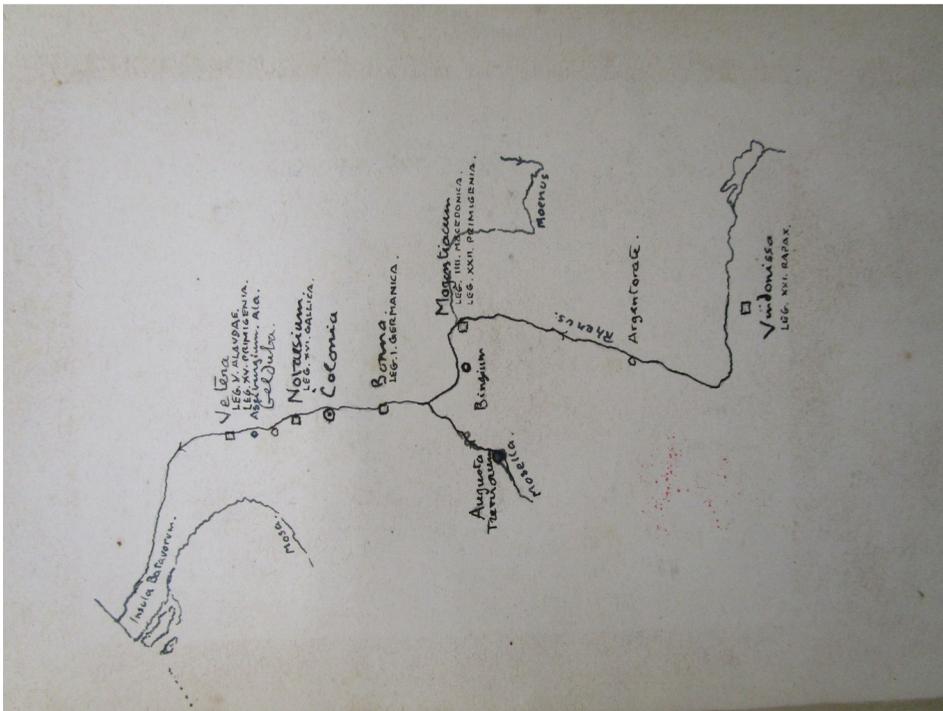
预备姿势: 跳者右侧靠近橡皮筋站立。

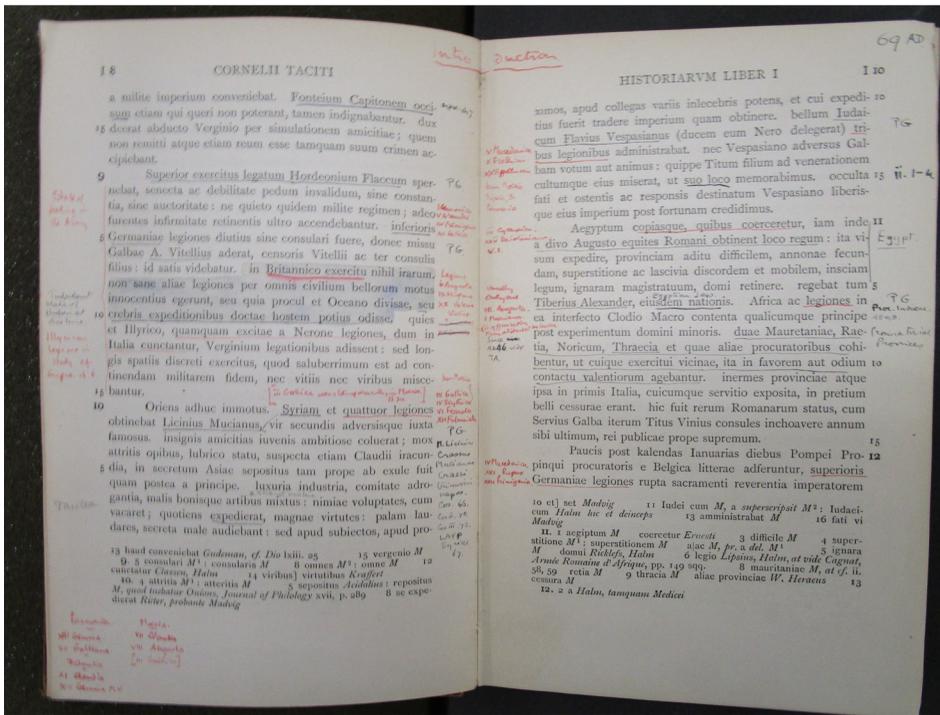
动作: 1. 右脚屈小腿从后从上向右边跳进第

This is what I do

*my friends
do it
as well*

1G. Exercise seven (BL Collins.Add.74)





HOTEL BEAUSÉJOUR - HYÈRES-les-PALMIERS

at Hotel Beau-séjour Hyères
 Var. France 31.6.34
 We are glad & interested to hear
 of your appointment & send
 congratulations. Wonder if it
 is likely to bring you to London
 sometime this season. Anyhow
 if you could let pay us a few
 days visit, we would be pleased
 & especially if ^{my} the two ladies
 tickets for Danlough Wickham
 & Mullingham would interest you.
 We have had rather wet
 weather here but have seen
 not to do something every day.
 My couch is much better, tho
 not so long I have a better
 prospect of it 4/4

H. Reynolds



P. B. Reynolds Esq

University College

Aberystwith

Angleterre



Radio-cavare A. Egerer Frères, 5, rue Théard, Paris - Reg. int.

Appendix 2. Ontology

Standardising and Structuring Approaches to Copy-Specific Cataloguing of Special Collections

Contents

Core Requirements	1
Data Entities	3
Controlled Vocabularies	5
Data Modelling	6
Pictorial Representation	7
Rich Text Description	7
MARC Representation	7
Alternative Representation	8

Core Requirements

This document has been designed to respond to the feedback received from the initial survey on the subject of copy-specific cataloguing practices in special collections. The aim of this document is **not** to provide a comprehensive guide to copy-specific cataloguing, nor to propose a wholly new approach to descriptive practice, but to devise possible, data-focused solutions to overcome some of the challenges raised by respondents to the survey.

In preparation of our scheduled interview, I would like you to review some of the concepts presented in this document and consider their respective merits and limitations in theory and practice. The interview will mostly discuss the ideas in this document, and consider how they could (or already do) play a role in your own cataloguing practice.

This document takes the form of a data ontology designed to tackle some of the core challenges of copy-specific descriptive practice. This ontology seeks to work towards a framework of standardising and structuring descriptive practices. Without an ontology designed specifically for copy-specific description in rare materials, our metadata will encounter two main problems:

1. Data encoding will be inconsistent - this will diminish data interoperability.
2. Data encoding will be high-level - this will diminish discovery of granular content.

It is therefore the intention of this ontology to provide a conceptual framework to ensure consistency, interoperability, discoverability and detail when capturing metadata of copy-specific elements in special collections materials.

Data Entities

An important first step towards a copy-specific data ontology is to identify the key data entities present across most copy-specific descriptions. These entities will provide the foundation for the latter stages of data modelling in this document.

Based on the most commonly identified entities across the responses to the descriptive exercises in the initial survey, I have identified 14 core entities which could form the basis of a data ontology for copy-specific metadata:

Type

Definition: The nature of the copy-specific element.

Comment: Recommended best practice is to use a controlled vocabulary.

Example: bookplate

Sub-Type

Definition: An additional class to specify the precise nature of a

Comment: Recommended best practice is to use a controlled vocabulary.

Example: armorial bookplate

Function

Definition: The purpose or significance of a copy-specific element.

Comment: Recommended best practice is to use an authorised vocabulary.

Example: ownership, bibliographic.

Agent

Person

Definition: An individual entity.

Comment: Recommended best practice is to source from an authorised record.

Example: Wilde, Oscar, 1854-1900.

Collective Agent

Definition: A gathering or organisation of persons bearing a particular name and capable of acting as a unit

Comment: Recommended best practice is to source from an authorised record.

Example: Pre-Raphaelite Brotherhood.

Relation

Definition: The relationship between other entities. .

Comment: Applied to define complex entity-relationships within the structure of the copy-specific data record.

Example: mother, recipient, bookseller, friend.

Date

Definition: A temporal extent having a beginning, an end and a duration

Comment: Date may be used to express temporal information at any level of granularity.

Example: 1892, 19th Century, 5th December 2002.

Place

Definition: A given extent of space.

Comment: This refers to a cultural construction of a place, such as a human identification of a geographical location.

Example: Germany, London (England), The Himalayas.

Physical Location

Definition: The location of the copy-specific element within the related resource.

Comment: Recommended best practice is to use an authorised vocabulary.

Example: front pastedown, spine.

Resource

Definition: A related resource.

Comment: Recommended best practice to provide a UID to link to a catalogue record.

Example: UID for a catalogue entry of a book.

Material

Definition: The physical material(s) that the copy-specific element consists of.

Comment: Recommended best practice is to use an authorised vocabulary.

Example:

Technique

Definition: The method used to create the copy-specific element.

Comment: Recommended best practice is to use an authorised vocabulary.

Example: copper engraving, woodcut, yotsume.

Size

Definition: Physical dimensions of the copy-specific element.

Comment: Recorded regularly in centimetres.

Example: 15 x 25 cm

Additional Note

Definition: A string of data which cannot be encoded elsewhere within the ontology.

Comment: The granularity of our data cannot continue indefinitely. Yet, the irregularity of copy-specific cataloguing may require the capture of some data which cannot otherwise be encoded, but may be useful to record. That is the purpose of this entity.

Controlled Vocabularies

Most of the data represented in the entities above will be textual string data, designed to be human readable. When encoding this kind of descriptive data, there is the inevitability that different cataloguers will use different terms to describe entities and their attributes. This was something that was clearly observed in the responses to the survey questions. A solution to this problem is the creation of a controlled vocabulary, to ensure consistency and accuracy of descriptive terminology.

The RBMS Bibliographic Standards Committee is responsible for a suite of controlled vocabularies designed to standardise use of terms in rare materials cataloguing. These vocabularies are commonly used by rare materials practitioners in most anglophone institutions. Among these vocabularies is the [Provenance Evidence Terms](#), which contains the class, [Physical Amendments](#). It is this area that most closely corresponds to the requirements of copy-specific cataloguing as outlined in this project. Below I have provided a revised sample from that vocabulary, which incorporates some of the findings from the survey. These terms are designed to define 'type'/subtype' entities specifically:

. Labels

- .. Bookplates
 - ... Armorial (UF Heraldic)
 - ... Pictorial
- .. Book Labels
- .. Stickers
 - ... Tickets

. Stamps

- .. Ink stamp
 - ... Armorial
 - ... Textual
- .. Perforated stamp

. Inserted ephemera

- .. Printed clippings
 - ... Newspaper clippings
 - ... Book clippings
- .. Manuscript Notes
 - ... Textual
 - ... Pictorial
- .. Object

. Enclosures

- .. Bindings
- .. Wrappers

. Markings

- .. Annotations
 - ... Marginalia
 - ... Underscoring
 - ... Highlighting
 - ... Strikethrough
 - ... Bracketing
- .. Inscriptions
 - ... Indexes
- .. Numbers
 - ... Accession numbers
 - ... Auction lot numbers
 - ... Codes
 - ... Shelfmarks
 - ... Prices
- .. Dates
- .. Pictorial markings
 - ... Drawings (UF Doodles)
 - ... Borders
 - ... Decorations
 - ... Diagrams
 - ... Symbols
 - Manicules
 - Arrows
 - Stars (UF: Asterisk)
 - Crosses
- .. Pen-tests (UF Scribbles)

Additionally, it is recommended that authorised terms are used to describe other aspects of the copy-specific element, such as **physical location**. For example, across the responses to the descriptive exercises, the terms 'flyleaf' and 'endpaper' were used interchangeably by respondents. A controlled vocabulary for physical location can standardise this area of descriptive practice too. For example:

- . **Endpapers (gathering term)** : *double leaves bound into a book at the front and rear*
 - .. **pastedown** : *portion of the endpaper that is glued to the inner boards of a book*
 - ... **front pastedown**
 - ... **back pastedown**
 - .. **flyleaf** : *the free inner paper of the endleaves (UF: front paper)*
 - .. **front flyleaf**
 - .. **back flyleaf**

Data Modelling

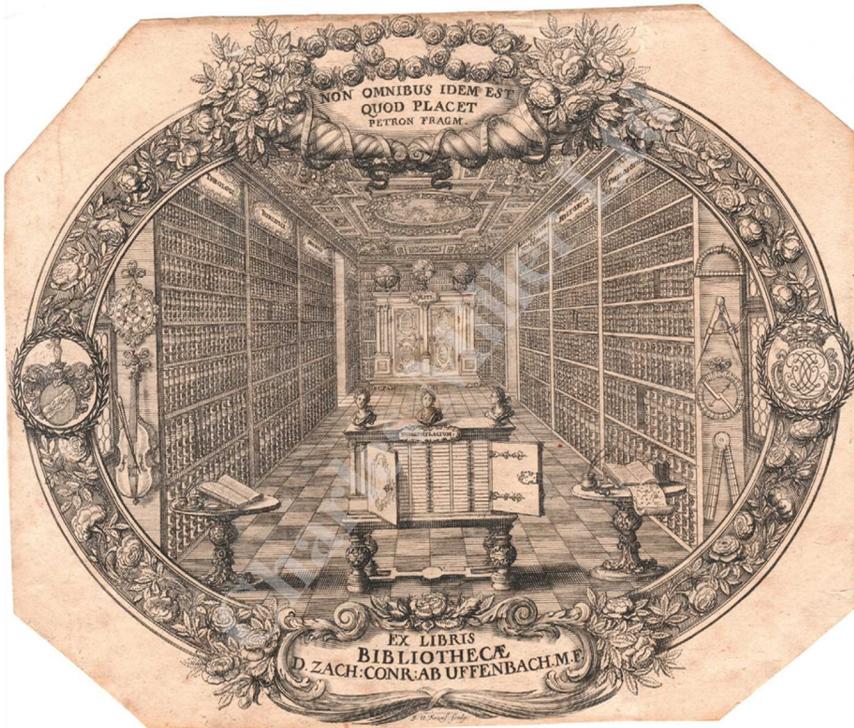
Standardising data encoding through the application of defined entities and controlled vocabularies works to overcome the challenge of interoperability. However the second challenge, of moving away from high-level encoding to more granular data, requires an assessment of how data is structured within records.

Respondents provided their descriptions to the seven exercises in the survey as free-form, rich text strings. This reflects the common practice of devising 5XX 'notes' in MARC cataloguing, which are generally unstructured. The purpose of these notes fields are used to capture bibliographic information, with 561 (Ownership and Custodial History), 562 (Copy and Version Identification Note) and 563 (Binding Information) being specifically designed for capturing copy-specific information.

The usefulness of metadata is largely determined by how it can be interpreted by a system and, consequently, how that data can be grouped, organised and queried. Unstructured metadata does not facilitate dynamic interaction and discovery with data. MARC allows data to be structured using sub-fields. Other systems typically represent data as 'triples', nested within a hierarchy.

To consider the merits of these different data modelling options, I have selected an example - a bookplate (see below) - which will be described using two distinct data structures: first, a MARC representation, and secondly, an alternative RDF schema.

Pictorial Representation



Rich Text Description

An early [18th Century] [German] [bookplate] of a gentleman's cabinet by [J.U. Kruf] for [Conrad Zacharias von Uffenbach], with finely worked [copper engraving] depicting a library with instruments and globes, etc., [12.5 x 15 cm].

MARC Representation

In MARC, this data can be structured using specific sub-fields. So, instead of writing an unstructured, rich text note reading “An early 18th Century German bookplate of a gentleman's cabinet by J.U. Kruf for Conrad Zacharias von Uffenbach”, one could write:

655_7 \$a Bookplates (Provenance) \$z Germany \$y18th century. \$2 rbprov

700_1 \$a Uffenbach, Zacharias Konrad von \$d 1683-1734 \$e former owner

This description manages to convey a lot of information, in a structured format. There are certain areas of information that are not provided here, such as a visual description or name of the engraver of the plate, however these could be incorporated into a 561 note as ‘leftover’ unstructured rich text:

561_0 \$a Includes an early 18th Century German bookplate of a gentleman's cabinet by J.U. Kruf for Conrad Zacharias von Uffenbach.

However, this non-hierarchical representation of metadata presents limitations when attempting to represent some of the more complex examples of copy-specific material evidence. Many respondents to the survey noted that the importance of recording copy-specific elements in books was that these elements told the “history of the book as object” and often provided a social historical context to the ownership, use, sale and custodianship of books. These histories are often complex, involving multiple agents, at multiple times, who may have certain relationships that give context to the history of the object. This kind of data may be difficult to model within a MARC environment.

With relationships, for example, a former owner of a book can be identified in a 7XX field, using the subfield \$e to indicate “former owner”. However, this function only serves to relate the agent to the resource, and cannot additionally relate one agent to another. This inability to link elements freely becomes a constraint when posed with examples such as gift inscriptions, where the recipient and receiver are related to each other, just as they are to the resource itself.

Alternative Representation

This second data model moves away from MARC in an attempt to provide some of the semantic information that can be found in copy-specific material evidence. The data is modelled using RDF triple statements, where individual entities are related to one another hierarchically:

Copy-Specific Element

Is a **Type**: label

Is **Sub-type**: bookplate

Is **Sub-type**: pictorial

Looks like **Note**: “Gentleman’s cabinet depicting a library with instruments and globes, etc. and owner’s armorial left and right and inscribed in cartouche Ex. libris bibliotecæ”

Has a **Function**: ownership

Is related to **Person**: [Uffenbach, Zacharias Konrad von](#), 1683-1734

Is associated with **Person**: [Uffenbach, Zacharias Konrad von](#), 1683-1734

Is of **Time**: 18th Century

Is of **Place**: Germany

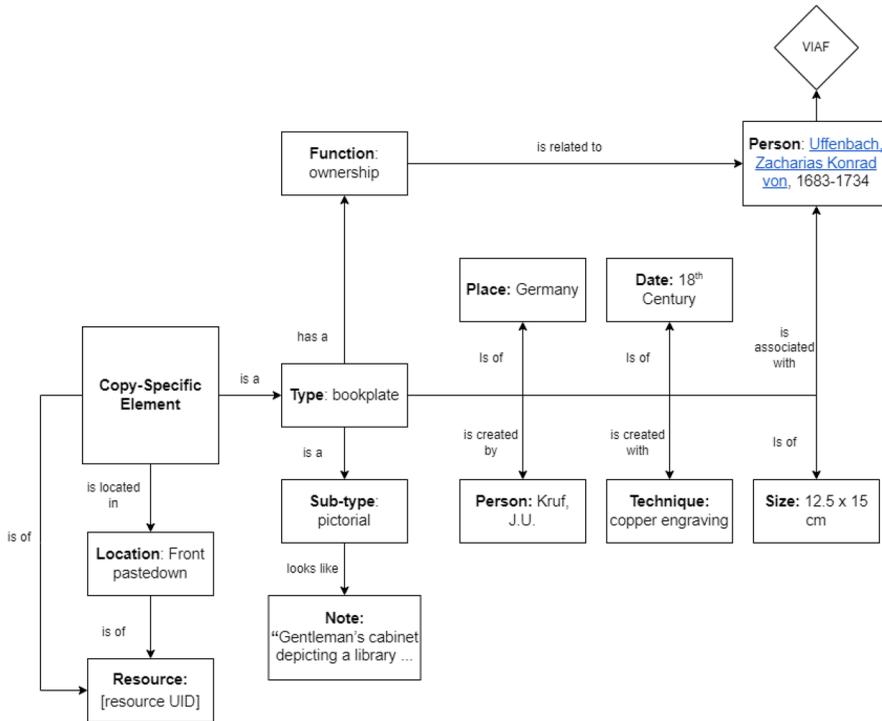
Is of **Size**: 12.5 x 15cm

Is created by **Person:** Kruf, J.U.

Is created with **Technique:** copper engraving

Is found at **Location:** Front pastedown

In **Resource:** [resource UID]



With this structure, more properties can be captured, but also the hierarchical nature of the ‘nested’, multi-level data is designed to allow relations between entities to be created within the description. This increases the granularity of information which can be captured and maximises the potential for advanced catalogue queries.

There is also the possibility of linking data in this record to additional data found in external sources. For example, I have linked the individual agent element for Zacharias Konrad von Uffenbach to the [VIAF record](#), which contains additional biographical information about the person, such as life dates, sex, profession, nationality and other relations to works or agents. This external information can then be linked into any prospective queries to the library catalogue. Many of these data entities - such as names, dates, places - could be linked to a wider web of interrelated datasets, thus providing more ‘meaning’ to records. Linked open data (LOD) environments such as [Sinopia](#) illustrate how this type of schema could be practically implemented.