Title	Encoding Caxtons in XML : a preliminary note on the Caxton and beyond project
Sub Title	キャクストン版のオンラインデータベース化に向けて
Author	徳永, 聡子(Tokunaga, Satoko)
	加藤, 誉子(Kato, Takako)
Publisher	慶應義塾大学日吉紀要刊行委員会
Publication year	2017
Jtitle	慶應義塾大学日吉紀要. 英語英米文学 (The Hiyoshi review of English studies). No.69 (2017. 3) ,p.79- 90
JaLC DOI	
Abstract	
Notes	
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN10030060-20170331-0079

Encoding Caxtons in XML:
A Preliminary Note on the Caxton and Beyond Project[1]

Satoko Tokunaga and Takako Kato

Incunabula were hand-crafted objects and post-printing every copy had its own individual history. Hence, even when studying copies of the same edition, every book is different: editions changed as they moved through the press, individual rubricators, illuminators, binders and owners rendered each copy unique. The development of the history of book as a discipline has seen increasing scholarly attention paid to copy specific information as a means of gaining new insights into the production and reception of individual books. Online catalogues such as the Incunabula Short Title Catalogue (ISTC) and the English Short Title Catalogue (ESTC) have allowed modern researchers easily to track the locations of verified copies.[2] Furthermore, recent years have seen the publication of new incunabula catalogues, which provide extensive descriptions of individual copies, alongside conventional bibliographical descriptions of editions. Key amongst these have been the *BMC XI* and the Bodleian Library's incunabula catalogue.[3] *BMC XI*, based on an extensive survey of the copy specific details of English incunabula housed in the British Library, is a monumental overview of early English printing and its reception clearly demonstrates the potential and significance of this approach to study for enhancing our

understanding of early English print culture. Since the beginning of the twenty-first century there has been significant progress too in the online cataloguing of incunabula, with projects placing the copy-specific at the heart of their work. Cambridge University Library's Incunabula Cataloguing Project, for example, launched in 2009 and completed in 2014, revised and updated the printed incunabula catalogue edited by J.C.T. Oates in 1954. The Cambridge project described every copy, with special emphasis on copy-specific features including anomalies and imperfections, binding, decoration, illumination, annotations, marks of ownership and provenance.[4] Likewise, the Glasgow Incunabula Project at the University of Glasgow, which unified and standardized records for incunabula scattered throughout the Library's online catalogue, focused much of its attention on the characteristics of individual copies.[5]

There can be no doubt that the pool of copy-specific data about early printed books has been significantly enriched, both in print and online, over the past decade. There remains, however, a fundamental step yet to be taken: the cross-examination of records of multiple copies across multiple editions. Traditionally, the primary function of rare book catalogues has been to help users identify a particular edition of a particular work, and then to locate copies of that edition on library shelves. Theoretically, in a digital realm the abundant data available in these various resources should allow the user to conduct all sorts of exciting combined searches, both within a single database and across multiple resources, with the click of a few buttons. In reality, however, this is not as simple as it seems. The vital copy-specific data are recorded differently in different repositories, can be dispersed in different formats in different places, and described in different ways. Although catalogues provide indexes to specific focused areas, which allow a single collection to be studied in great depth, undertaking a comprehensive

survey of copy specific records still requires the researcher to resort to conventional methods. To do cross-edition comparison, or comparison across multiple institutions, still requires the manual accumulation and compilation of data from various repositories.

What if we could bring all of the available copy-specific data together into a single resource, and then supplement and build that resource with the addition of new copy specific records? One of the greatest tools of the digital age is the ability of algorithms to sort and search large blocks of data; the computer's ability to search across multiple strata of information, using complex and layered searches, is unparalleled. Drawing together copy-specific data and making it searchable through a single interface would potentially open up a whole new set of research possibilities, addressing questions almost unanswerable prior to the digital revolution spanning different editions, printers or texts, and multiple copies of a single edition. Such an interface would allow us to interrogate the data with different combinations of keywords, and analyze them from a variety of different perspectives, focusing on, for example, combinations of typography, layout, format, language and choice of texts, use of woodcuts, provenances, annotations, hand-written facsimiles, rubrication, etc.

It is well recognized that this approach is highly desirable. It has, for example, been adopted by Material Evidence in Incunabula (MEI), hosted by the Consortium of European Research Libraries (CERL). MEI is a database specifically designed to record and "combine searches of bibliographical records (extracted from ISTC) with copy specific records".[6] Its main focus is the provenance of individual copies, which it aims to trace chronologically. MEI has implemented a sophisticated search engine; its advanced search allows the user to interrogate the data by combining search options which include provenance, the current location, bibliographical and

physical aspects.[7] The research value of search engines like MEI is immense.

It was in this fertile research area that the authors of this paper embarked on "Caxton and Beyond: Copy Specific Features of English Incunabula". Our research objective is to shed new light on the late fifteenth-century book culture of England during a critical transitional period between manuscript and print.[8] Our initial focus is on books printed by William Caxton, England's first printer, but the ultimate aim is to expand the analysis to encompass a wider, more holistic view of early English book culture. The project focuses on creating a tool to record and analyse rubrication within English incunabula, broadly conceived of as encompassing handwritten initials, paragraph marks, underlines, and marginalia added after printing by professionals. The aim is to amass evidence of this sort of copy-specific detail, and then to consider to what extent these manuscript additions were systematized in the contemporary printing house and literary culture.[9] Furthermore, the evidence can be used to examine the continuity and links between manuscript culture and print culture: how are copy-specific features of incunabula related to those found in contemporary English manuscripts?

The comparison of printed books with manuscripts raises some interesting methodological questions. The traditional approach characterized the former as "exist[ing] in multiple copies, and [able to] be described adequately by well-established and formalized bibliographic conventions", whereas the latter were viewed as "unique objects, often of great cultural or political value".[10] As a consequence of this perceived difference, different methodological approaches were adopted for the two classes of object. The movement in book history with which this article opened, however, which recognizes each incunable copy as a "unique object", bearing material evidence of cultural value, necessitates these methodological approaches to

be drawn together. The authors therefore opted to record edition specific information following the well-established and formalized bibliographical conventions, but to model the methods of recording copy specific information on the conventions of manuscript description.

The differences between edition and copy specific features are fundamentally connected to the theoretical implications of how we understand and study the book as a material object. *Functional Requirements for Bibliographical Records* (FRBR) addresses this question from a broader perspective: there are different "entities" to be described in bibliographical records.[11] Some points of interest will be connected to the work contained, such as "Chaucer's *Canterbury Tales*"; some will be connected to the manifestation and/or expression of the work; for example, the edition specific features of "the first edition of Chaucer's *Canterbury Tales* printed by Caxton". Others are connected to the item, a single book on the shelf, and its copy specific features such as "provenance of a copy of the first edition of Chaucer's *Canterbury Tales* printed by Caxton now in the British Library".

How we treat the relationship between these different entities depends on the aims which underlie the descriptive endeavour. Basic requirements set by FRBR for national and international bibliographic records mainly focus on the identification of the work and its manifestation and/or expression.[12] As it is often the case that union catalogues of incunabula utilize the relational database structures and standards developed for the description of these features, the resulting datasets are not well suited to the storage and manipulation of copy specific information. This sort of information is considered as a secondary function, whereas the primary function is the identification of work, manifestation and/or expression. This database structure, and the standards which underpin it, require data fields to be strictly pre-defined, and hence it lacks the flexibility necessary for

recording edition and copy specific information as separate groups of data.

The format selected by the authors for the Caxton and Beyond Project, XML, Extensive Markup Language, is far more flexible. XML allows users for the description of incunabula in natural language—in readable, manageable and understandable sentences. XML is also extendable: it is not constrained by pre-defined fields. Hence, if a researcher encounters a feature which is worthy of recording, and worth making searchable, but which has not previously appeared in the XML structure, it is simple to create a new set of tags without changing the document structure of already completed descriptions. For these reasons, the authors have adopted the well-established Manuscript Description module, set by the Text Encoding Initiative (TEI), which they note is "potentially useful for any kind of inscribed artifact".[13] Incunabula, however, are not manuscripts: manuscripts are unique objects, and hence the item and manifestation and/or expression fall into one, an incunabulum forms part of an edition, and so the FRBR distinction between item and manifestation and/or expression is needed.

For the Caxton and Beyond project, therefore, the authors have customized the Manuscript Description module substantially. The primary objective of the interface is to integrate data and to allow users to conduct complicated and sophisticated searches across work specific, edition specific, and copy specific features; an example might be: "Which copies of which editions contain the style 'A' of rubrication, who owned these copies in the sixteenth century, and how many of those copies also contain hand-written facsimiles from the nineteenth century?" To this end, the information needs to be described at three clearly-defined levels: work, edition, and copy. So, the authors are using three sets of TEI-compliant XML files. The first contains work specific information, names of authors and works, and a list of editions with basic publication details, such as printer, publication

place and date. Editions in this list will be linked to edition specific descriptions, using ISTC as the spine. The second provides publication details, as well as more detailed edition specific information, such as format, the ideal collation, textual contents, printing type, page layout (numbers of columns and lines), printed headings, signature and folio numbers, size of untrimmed page, use of woodcuts and the printer's device, and space for rubrications with guide letters. This second level also contains a list of existing copies, which functions as the gateway to the third-level descriptions. The third level contains the copy-specifics of the individual copies: copy-specific collation, details of the paper stock, the size of leaves, condition of the copy, rubrication and decoration, additions to the copy, including marginalia and leaf facsimiles added in later centuries, and binding. Other accompanying details will also be included in level three, details of provenance and former owners, with the historical prices paid for an individual copy if known. All three levels of description will be cross-searchable, and will also include reference to essential secondary resources.

Copy-specific information is by its nature complicated to describe in words, and the provision of full-colour digital images linked to individual records would greatly enhance users' understanding of the books. With technological advances and a shift in the economical focus of digital culture, it is now possible to produce and store significant bodies of digital data at relatively low cost. In the early days of digitization it was expensive and time-consuming to capture images of historical materials: it required a specialist camera, a custom-made cradle, and a professional digitization team. Now many libraries have their own in-house digitization departments, and are actively engaged in digitizing their own material and making the digital facsimiles available online. Significantly, many larger institutions in the UK and the USA have changed their library policies to allow users to

take their own photographs, with their own digital cameras. Moreover, various institutions have ventured into new digital enterprises which have at their heart the free use of digital images.[14] The opening of access to resources which previously would have required a physical visit to view them has advanced scholarship significantly. The publication since the 1970s of a large number of black and white facsimiles, for example, has inspired the concentrated study of medieval manuscripts,[15] whilst resources such as Early English Books Online (EEBO) and Eighteen Century Collections Online (ECCO) have revolutionized the textual study of pre-1801 English books.[16] Digital resources such as EEBO and ECCO have made a major contribution to our understanding of textual transmission; it is hoped that the Caxton and Beyond resource will similarly illuminate the history of individual copies.[17]

The Project already has a considerable body of data to call upon; many individual incunabula have already been recorded in some detail, notably in the *BMC XI* and other incunabula catalogues. Capturing and encoding this pre-existing printed scholarship will create a useful preliminary dataset for the Caxton and Beyond project, just as Goff's catalogue served as a basis for the ISTC. The authors' project will supplement these existing printed descriptions with digital images, whilst simultaneously creating fresh descriptions from scratch. The accumulation and integration of old and new data is essential in realizing the project's objectives, and here collaboration will be key. Rather than repeating or competing with existing scholarship, the objective of the interface is to facilitate further collaboration of scholarship and integration of data, irrespective of the format of that data. Some will come from printed resources, but much will come from digital catalogues where the collaboration of the institutions hosting those catalogues will be essential. Furthermore, that collaboration will need to be

two-way; our data needs to be compatible with existing online catalogues so that information can be shared, exchanged, and integrated effectively. Creating links between our resource and others *is* useful, but such links do not optimize the full potential of the digital environment. The automatic retrieval of data from different resources is on the research agenda of several recent digital humanities projects, including the ERC-funded "The Digital Resource for Palaeography" and the JISC-funded "Manuscripts Online: Written Culture from 1000 to 1500".[18] More recently, the International Image Interoperability Framework (IIIF) is radically changing the display and delivery of image-based resources on the web. Previously, institutions worked independently through locally built applications, but IIIF endeavours to draw together an international community of leading research libraries and repositories to work collaboratively "to develop an interoperable technology and community framework for image delivery".[19]

Thus, in the present climate, where an international standard for the sustainable storage, display and sharing of digital images is on the cusp of being realized, the merits of sharing and integrating scholarship in digital form should be self-evident. Though our current corpus is focused, the structure of the interface will be flexible and expandable, and able to flex to accommodate an ever-broadening body of material, and hence to answer more and more complex and varied research questions. The Caxton and Beyond Project responds directly to the current dynamic trends in digital humanities; the resulting interface has the potential to help rewrite the established history of English written and print culture from the fifteenth century and beyond.

Notes

1) At its inception, this project was awarded a Katharine F Pantzer Jr Research Scholarship by The Bibliographical Society (UK), and a Research Grant for Global Initiative Research Projects made available by Keio University (Japan), for which the authors are grateful. The last date of access to all of the electronic references listed in this paper was 25 January 2017. The authors would like to thank Ed Potten for his support in editing this paper for publication.

2) ISTC (http://www.bl.uk/catalogues/istc/) records all known editions and issues of fifteenth-century books. A new version has been published by Consortium of European Research Libraries: The International Database of 15th-Century European Printing (https://data.cerl.org/istc/_search). For the ESTC see (http://estc.bl.uk).

3) *Catalogue of Books Printed in the XVth Century now in the British Museum [British Library]*, Part XI, ed. by Lotte Hellinga ('t Goy-Houten, 2007); *A Catalogue of Books Printed in the Fifteenth Century now in the Bodleian Library, Oxford*, ed. by Alan Coates, Kristian Jensen, Cristina Dondi, Bettina Wagner, and Helen Dixon, with the assistance of Carolinne White and Elizabeth Mathew; blockbooks, woodcuts and metalcut single sheets by Nigel F. Palmer; an inventory of Hebrew incunabula by Silke Schaeper, 6 vols (Oxford: Oxford University Press, 2005), which is now available online (http://incunables.bodleian.ox.ac.uk).

4) See "About the Project", Cambridge University Library Incunabula Project Blog: (https://inc.blog.lib.cam.ac.uk/?page_id=11).

5) See "Project Introduction" (http://www.gla.ac.uk/services/incunabula/projectintroduction/).

6) The Consortium of European Research Libraries, Material Evidence in Incunabula (http://incunabula.cerl.org/cgi-bin/search.pl).

7) Recently, MEI has been integrated into The 15th Century Booktrade Project: (http://15cbooktrade.ox.ac.uk/project/). According to its online description, it is to use the material evidence from thousands of surviving books, as well as unique documentary evidence, to address five fundamental questions relating to the introduction of printing in the West: (1) Distribution, use, and reading practices; (2) The books' contemporary market value; (3) Transmission and dissemination of

texts in the 15th-century; (4) The circulation and reuse of illustrations; (5) Visualization of the circulation of books and of the texts they contain.

8) The continuity between manuscript and print culture has been recognized as an important issue in the history of the book. One of the earliest advocates for such continuity was Curt F. Bühler, *The Fifteenth-Century Book: The Scribes, the Printers, the Decorators* (Philadelphia: University of Pennsylvania Press, 1960). For more recent studies, see, for example, Margaret Smith, "The Design Relationship between the Manuscript and the Incunable", in *A Millennium of the Book: Production, Design and Illustration in Manuscript and Print 900—1900*, ed. by Robin Myers and Michael Harris (Winchester and New Castle: St Paul's Bibliographies and Oak Knoll, 1994), pp. 23–43.

9) The importance of research of rubrication in incunabula has been advocated by scholars such as William Blades, Margaret Smith, A. S. G. Edwards and Lotte Hellinga. See: William Blades, *The Life and Typography of William Caxton, England's First Printer: With Evidence of his Typographical Connection with Colard Mansion, the Printer at Bruges*, 2 vols (London: J. Lilly, 1861–63); Margaret Smith, "Patterns of Incomplete Rubrication in Incunables and What They Suggest about Working Methods", in *Medieval Book Production: Assessing the Evidence*, ed. L. L. Brownrigg (Los Altos Hills: Anderson-Lovelace, 1990), pp. 133–46; *ibid*., "Red as a Textual Element during the Transmission from Manuscript to Print", in *Textual Cultures, Cultural Texts*, ed. Orietta Da Rold and Elaine Treharne, a special issue of *Essays and Studies* (Cambridge: Brewer, 2010), pp. 187–200; A. S. G. Edwards, "Decorated Caxtons", in *Incunabula: Studies in Fifteenth-Century Printed Books Presented to Lotte Hellinga*, ed. Martin Davies (London: British Library, 1999), pp. 493–506; *BMC XI*, pp. 346–47; and most recently, Satoko Tokunaga, "Rubrication in Caxton's Early English Books, c. 1476–78", *Incunabula on the Move: The Production, Circulation and Collection of Early Printed Books*, ed. Ed Potten and by Satoko Tokunaga (Transactions of the Bibliographical Society, 12.1 (2012): 58–78.

10) Digital.Humanties@Oxford, *Manuscript Description* (http://tei.oucs.ox.ac.uk/ Talks/2009-03-vilnius/talk-02_msDesc.xml).

11) The International Federation of Library Associations and Institutions, *Functional

Requirements for Bibliographic Records: Final Report (1997; amended and corrected through 2009; (http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf).

12) *Ibid.*, pp. 98–99.

13) "Manuscript Description" in *P5: Guidelines for Electronic Text Encoding and Interchange*, version 2.1.0, 2012 (http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html).

14) See for example, JRUL's "About Digitisation Activities and Services": (http://www.library.manchester.ac.uk/using-the-library/staff/teaching/services/digitisation-services/about/) and the Cambridge University Digital Library (http://cudl.lib.cam.ac.uk/).

15) Derek Pearsall, "Introduction" to *New Directions in Late Medieval Manuscript Studies*, ed. by Derek Pearsall (Woodbridge: York Medieval Press, 2000), pp. xi–xv (p. xi).

16) See http://www.eebo.chadwyck.com and http://gale.cengage.co.uk/product-highlights/history/eighteenth-century-collections-online.aspx.

17) Moreover, Text Creation Partnership (TCP) provides searchable standardized SGML/XML-encoded electronic texts transcribed and marked up from pages of EEBO and ECCO, a great of which has been made freely available to the public. See the Text Creation Partnership (http://www.textcreationpartnership.org/).

18) See: The Digital Resource for Palaeography (DigiPal) (http://www.digipal.eu/) and Manuscripts Online: Written Culture from 1000 to 1500 (https://www.manuscriptsonline.org). The recent project Catalogue of Digital Editions (https://dig-ed-cat.eos.arz.oeaw.ac.at) "seeks to examine best practice in the field of digital editions by collating relevant evidence in a detailed catalogue of extant digital projects".

19) International Image Interoperability Framework (http://iiif.io).