

RECORDING PROVENANCE in TEI-ENCODED MANUSCRIPT CATALOGUES for MORE EFFECTIVE DATA SHARING

Toby Burrows (Oxford e-Research Centre, University of Oxford) August 2020

The effective reuse of information from catalogues of medieval and Renaissance manuscripts depends, ultimately, on the approach taken by the manuscript cataloguer towards recording provenance. There are two possible approaches:

- Recording the physical evidence found in the manuscript itself, usually in the form of a series of notes or narrative statements about the manuscript's history;
- Assembling a structured list of successive stages in the ownership of the manuscript, usually in chronological order, together with information about the evidence for each stage.

An example of the first approach, taken from Rodney Thomson's catalogue for Merton College,¹ can be found in the Bodleian Library's online catalogue, *Medieval Manuscripts in Oxford Libraries*:

https://medieval.bodleian.ox.ac.uk/catalog/manuscript_10100

The TEI-XML encoded version can be found on the Bodleian Library's GitHub site:

https://github.com/bodleian/medieval-mss/blob/master/collections/Merton/Merton_College_MS_1.xml

This approach is often found in older printed manuscript catalogues, as well as in MARC-based library catalogue records, where the information is recorded in otherwise unstructured "Note" fields.²

The American Library Association's guidelines for manuscript cataloguing embody this approach:

"Make notes on the history of ownership of an item. Note all marks of ownership such as bookplates, ex libris and ex dono inscriptions, armorial bindings, etc., as well as other physical or textual evidence that contribute to establishing provenance. Provide references to sales catalogs, inventories, or other documentation as available."³

The second approach may be found in more recent printed catalogues as well as in newly created or revised online descriptions. It can be seen in action in Peter Kidd's catalogue of

¹ Thomson, R. M., *A descriptive catalogue of the medieval manuscripts of Merton College, Oxford* (Cambridge: D. S. Brewer, 2009).

² See: <https://www.loc.gov/marc/bibliographic/bd561.html> and http://documents.cerl.org/provenance/CERL_VII_art_04.pdf

³ Pass, Gregory A., *Descriptive Cataloging of Ancient, Medieval, Renaissance, and Early Modern Manuscripts* (Chicago: Bibliographic Standards Committee, Rare Books and Manuscripts Section, Association of College and Research Libraries, American Libraries Association, 2003), pp. 66-68
http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/booksanddigitalresources/digital/AMRE_MM_full.pdf

manuscripts in the Queen’s College Oxford⁴ and Richard Gameson’s catalogue for Trinity College Oxford.⁵ An example from the Queen’s College catalogue is given in Appendix 1.

Peter Kidd describes his approach as follows:⁶

“Provenance is here taken to mean every definable stage in the history of the manuscript from its creation to the present day, whether this can be reconstructed from internal or external evidence. Each stage is presented in a chronological numbered list, though the relative order of some items may be uncertain.”

There are two key goals in this approach:

- (1) assemble (as far as possible) the chronological chain of ownership; and,
- (2) record the evidence for each step in the chain.

This is similar to the practice used by art museums and galleries for recording the chain of ownership for paintings and other works of art. A typical example, from the National Gallery of Art, can be found here: <https://www.nga.gov/collection/art-object-page.19.html> The Art Tracks project at the Carnegie Museum of Art has developed a Digital Provenance Standard with the aim of providing a machine-readable structure for recording the provenance of art works.⁷ This defines three levels of provenance information:⁸

- Level 1: Basic Provenance – a “provenance text”; similar to printed manuscript catalogue entries like Thomson’s;
- Level 2: Provenance with Entities – adding mappings for parties, locations, and URLs; similar to TEI encoding of manuscript catalogue entries like Thomson’s; and,
- Level 3: Event-Based Provenance – documenting the individual transfer events; similar to manuscript catalogue entries like Kidd’s.

Encoding with the TEI Guidelines

Both approaches to recording manuscript provenance can be encoded using the TEI Guidelines, but the evidence-focused, narrative, or notes-based approach is less satisfactory as a basis for extracting structured data for reuse in other settings. The [Mapping Manuscript Migrations](#) (MMM) project managed to extract more than 20,000 provenance statements from *Medieval Manuscripts in Oxford Libraries*. These were then converted to RDF triples and mapped to the MMM Data Model, which is based on a combination of the FRBR_{oo} and CIDOC-CRM ontologies with some additional entity classes and properties.⁹ But only the TEI <origin> element could be mapped to a specific event type in the CIDOC-CRM ontology:

⁴ Kidd, Peter, *A descriptive catalogue of the medieval manuscripts of the Queen’s College, Oxford* (Oxford: The Oxford Bibliographical Society, 2016).

⁵ Gameson, Richard, *The Medieval Manuscripts of Trinity College, Oxford: a Descriptive Catalogue* (Oxford: The Oxford Bibliographical Society, 2018).

⁶ Kidd, p. 40.

⁷ Art Tracks, Carnegie Museum of Art, *The CMOA Digital Provenance Standard*. Draft version 0.2, Oct. 14, 2016. <http://www.museumprovenance.org/reference/standard/>

⁸ http://www.museumprovenance.org/pdfs/provenance_model_diagram_v.0.2.pdf

⁹ Burrows, Toby, Athanasios Velios, Matthew Holford, David Lewis, Andrew Morrison, and Kevin Page, ‘Transforming TEI Manuscript Descriptions into RDF Graphs,’ *Scholarly Digital Editions, Graph Data-Models and Semantic Web Technologies (GraphSDE) 2019 proceedings*, forthcoming 2020

E12_Production. All other events had to be mapped to a generic MMM-specific entity “Manuscript related activity”, which was created specifically for this purpose.

The second approach, using a structured list of events, has the advantage of being much more suitable for extracting and re-using the data in platforms like [Mapping Manuscript Migrations](#) than a narrative approach to provenance is. Assuming that a manuscript’s provenance is recorded in a form similar to that used by Kidd and Gameson, how should a structured list of this kind be encoded using the *TEI Guidelines*?

<provenance>

<provenance> can be used for all the steps in the ownership chain, including origin and acquisition by the current owner. There are two possible encoding structures:

- (1) A separate <provenance> element for each step. Each <provenance> element should preferably contain one event with persons, organisations, places, and dates, as well as the evidence for that event.
- (2) A single <provenance> element containing a <listEvent> with multiple <event> elements for each provenance event. Persons, organisations, places, and dates should be encoded within each <event> element, together with the evidence for that event.

Alternatively, <origin> and <acquisition> can still be used for information about the production of a manuscript and for the acquisition of a manuscript by the current owner, respectively. All other steps in the ownership chain would then be encoded using separate <provenance> elements or an event list, as above.

Attributes

Attach these to <provenance> or <event> elements, depending on which structure is preferred.

Provenance – types

[@type](#)

Use a specified vocabulary for different types of provenance events. A detailed vocabulary of provenance event types is the “Acquisition Method Vocabulary” of the Art Tracks project, which also includes origin event types.¹⁰

Use with either <provenance> or <event>

Provenance – dates

[@when](#), [@notBefore](#), [@notAfter](#), [@from](#), [@to](#) [@calendar](#), [@period](#)

Use with either <provenance> or <event>

Alternatively, use a <date> element with appropriate attributes

Provenance –places

[@where](#)

Use with either <provenance> or <event>

Alternatively, use a <placeName> element with appropriate attributes

¹⁰ http://www.museumprovenance.org/reference/acquisition_methods/

Provenance – <persName> and <orgName>

@role

Use a consistent vocabulary for roles of persons and organisations in the production of a manuscript and its ownership history. Library of Congress relator codes should be preferred.¹¹

The most relevant codes are:

auc	Auctioneer	dpt	Depositor
bsl	Bookseller	fmo	Former owner
cli	Client	own	Current owner
col	Collector	sll	seller
dnr	Donor		

<origin> and <acquisition> - if these are used, follow a similar structure:

<origin>

Encode persons and organisations mentioned

Encode places mentioned

Add date or date range to <origin> or <origDate>

<acquisition>

Encode persons and organisations mentioned

Encode places mentioned

Encode dates and/or add date range to <acquisition>

Ensure the name of the current owner is present and encoded, together with the current collection and location if relevant

General principles

For each provenance event, record and encode as many as are known (or can be deduced) of the following elements:

- Person
- Organisation
- Place
- Date
- Type of provenance event

Include the evidence for a specific event in the <provenance> or <event> element for that event.

Pressmarks mentioned as part of the evidence for a provenance event should be encoded as <q> with attribute @type = “pressmark”.

If there is evidence which cannot be related to a person, organisation, place, or date of an event, record it as a separate <provenance> element, even when an event list is being used.

¹¹ <http://www.loc.gov/marc//relators/relaterm.html>

Appendix 1: Chronological provenance list for a manuscript

The Queen's College, Oxford, MS. 305

PROVENANCE

1. Written in France in the third quarter of the 15th century, probably in or near Avignon or Carpentras in the 1460s (see Decoration).
2. Unidentified early 17th-century owner; perhaps rebound for him: inscribed 'Ihon ... [deleted and uncertain up to this point, only the following numerals are clear] 1616' (fol. lii').
3. Inscribed 'Isaac Crommeling. 1653. lxvii' (fol. ii'), presumably the Dutch historian Isaac Crommelin (1598-1676).
4. Peter Causton (1656-1700), London merchant, author of the poem *Tunbrigialia* (1686, etc.) and others: inscribed 'Quanta, eheu, miserum caligo oppresserat orbem! | ... | Somnia, quae sapiens usque legenda neget. | [signed] P. C.'
5. The Queen's College: given by Causton in 1697; inscribed 'Donum Petri Causton. Merc: Lond:' and below this, in paler ink, 'Qui et Tunbrigialia cecinit quorum exemplar deauratae | affixum tabellae huic collegio dono dedit – A^o 1697' (fol. iii'), thus too late to be included in Bernard, *CMA*; inscribed with a note in pencil by Samuel Meyrick (1783-1848) (on whom see *ODNB*) (fol. iii'); College shelfmarks: 'C. 9' in ink (fol. iii'); 'S. 1' and '305' in pencil (fol. iii'; cf. spine), and in ink (fol. iv'); and inscribed in pencil 'MS Queens 305 (in M^r Coxe's Catalogue)' (fol. ii^v, the former pastedown).

Kidd, Peter, *A descriptive catalogue of the medieval manuscripts of the Queen's College, Oxford* (Oxford: The Oxford Bibliographical Society, 2016), p. 141.