VisColl. A new model for the description of the collation of manuscripts and related tools

When books are digitized, typically, photographs are taken for each folio (recto and verso), and, increasingly, also of the different views of the object (left and right cover, edges, and spine). Information on the order of items in these sequences of images is preserved as metadata and used for navigation in online viewers of digitized book materials and digital editions. Facing pages are also often shown together to represent the opening of the page in the original object. Sometimes, as for example in the digital edition of the Codex Sinaiticus,[1] metadata and navigation allow to browse the images by quire, each group of images belonging to pages from the same quire having been mapped to a virtual gathering; no information, however, is recorded on which folios are physically conjoined, and then nested, to form the gathering structure. As it has been pointed out by Dot Porter (Schoenberg Institute for Manuscript Studies - SIMS, University of Pennsylvania, Philadelphia) in a recent talk at the Parker Library (Cambridge, 26th March 2018),[2] the set of typical visualizations for digitized library materials (single-page view, double-page view, scroll view, and thumbnail view) are not an invention to deal with digital images of books, but they find their legacy in non-digital representations and reformatting of documents. The single-page and the double-page view have their counterparts in physical facsimiles of manuscripts and image reproductions of books; the scroll view was the typical mode of browsing books that had been microfilmed; whilst the thumbnail view reminds us of microfiche images. Contemporary viewing of digital images of book materials is therefore not novel, however, these are not the only modes of accessing and navigating digitized books through digital means, and new ones that take into consideration the physical make out of the original objects are possible, and should be welcome.

Books, in codex format, i.e. 'collections of sheets of any material, folded double and fastened together at the spine, and usually protected by covers',[3] are formed of pages connected at the spine and not of floating pages in sequences, as portrayed in most digital projects. At the core of the technology of the codex lies the **gathering** (or quire), a 'group of folded or single leaves which can be used either singly or with other gatherings to create a textblock'.[4] The gathering is in fact the ultimate working unit of the codex.[5] The gathering structure of books in codex format has been recognized as an important aspect to be recorded and described in catalogues of both manuscripts and early printed books, and this is traditionally accomplished through *collation formulas*. These are however not standardized, and come in a series of different flavours, depending of the tradition and the scholar.

VisColl, conceived by Dot Porter in the early 2010s, tackles these issues. The original idea aimed at devising a new visualization tool to show the structure of the original codex, based on TEI[6] manuscript descriptions[7]

---

[1] The British Library, Leipzig University Library, National Library of Russia and St. Catherine's Monastery (2018). *Codex Sinaiticus. Experience the Oldest Bible*. The Codex Sinaiticus Project http://www.codexsinaiticus.org/en/ (accessed June 2018). NB. All websites links checked in June 2018.
[2] Porter, D. (2018). Using VisColl to Visualize Parker on the Web: Reports on an experiment *Dot Porter Digital* http://www.dotporterdigital.org/using-viscoll-to-visualize-parker-on-the-web-reports-on-an-experiment/.
[3] Roberts, C. H. and Skeat, T. C. (1983). *The Birth of the Codex*. London; New York: Published for the British Academy by the Oxford University Press, p.1.
[4] Ligatus (2015). Gatherings *Language of Bindings*. London: University of the Arts London http://w3id.org/lob/concept/2286.
[5] Andrist, P., Canart, P. and Maniaci, M. (2013). *La syntaxe du codex: essai de codicologie structurale*. (Bibliologia 34). Turnhout: Brepols, p. 50.
[6] TEI (2018). *Text Encoding Initiative P5 (v. 3.3.0): Guidelines for Electronic Text Encoding and Interchange.* P5 revised and re-edited edition. Oxford; Providence (RI); Charlottesville (VA); Nancy (KY): Text Encoding Initiative Consortium http://www.tei-c.org/Guidelines/P5.
[7] TEI (2018). 10. Manuscript Description. *Text Encoding Initiative P5 (v. 3.3.0): Guidelines for Electronic Text Encoding and Interchange*. P5 revised and re-edited edition. Oxford; Providence (RI); Charlottesville (VA); Nancy (KY): Text Encoding Initiative Consortium http://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html.

where one typically finds both collation formulas or other descriptions of the gathering assembly, and digital images of each page. A prototype project[8] was set in collaboration with Doug Emery (SIMS), Dennis Mullen (SIMS), and Alberto Campagnolo.[9] The project took collation formulas from the TEI descriptions of manuscripts in the Digital Walters collection[10]; a parser was programmed to read collation formulas and transform them into XML[11] files according to an ad hoc schema devised for the project; from these XML files, a series of XLST[12] scripts would then generate SVG[13] collation diagrams and webpages visualising the conjoined pages and their position within the gathering structure. This approach, proved difficult to scale up because, as mentioned, collation formulas are not standardized, and a new parser would need to be written for each style and adaptation of the formulas. Furthermore, manuscript quires can have complex structures, with sub-quires and haphazardly inserted leaves that are difficult to describe in collation formulas. To avoid these problems, in a second phase of the project, a collation modeller[14] was programmed as a web service that through a series of web forms would guide the user in building collation models according to the project's XML schema. The XML files thus generated could then be transformed with the usual pipeline of XSLT scripts.

The project's schema, however, could not model sub-quires, stubs, indicate the attachment method for each folio, signal the writing orientation (useful to visualize right-to-left manuscripts), and indicate uncertainties. A new schema needed to be devised to accommodate these complex modelling characteristics, VisColl 2.0,[15] turning the project more into a model rather than just a way to visualize collation information. The new model is based on the single folio as its building block (as opposed to the bifolio), and this allows to model very complex structures with sub-quires and single leaves. The attachment methods can also now be described, and uncertainties can be flagged at any step of the model. The new model has already been implemented into other project such as the Digital Tools for Manuscript Studies,[16] EVT 2,[17] and Bibliotheca Philadelphiensis.[18] New features will include modelling and visualizing flaps and fold-outs, synoptic charts, page-level semantic tagging, and the use of the automated collation diagrams as data validation during modelling to secure better data being recorded.

Alberto Campagnolo

Washington, DC, June 2018

---

[8] Schoenberg Institute for Manuscript Studies (2013). SIMS Collation Project http://dorpdev.library.upenn.edu/collation/.

[9] At the time working on his PhD at the Ligatus research Centre, University of the Arts London. Campagnolo, A. (2015). Transforming structured descriptions to visual representations. An automated visualization of historical bookbinding structures London: University of the Arts London PhD Thesis http://ualresearchonline.arts.ac.uk/8749/.

[10] The Walters Art Museum (2018). The Digital Walters http://www.thedigitalwalters.org/.

[11] Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., Yergau, F. and Cowan, J. (2006). Extensible Markup Language (XML) Version 1.1 (Second edition) https://www.w3.org/TR/xml11.

[12] Clark, J. (1999). XSL Transformations (XSLT) Version 1.0 W3C http://www.w3.org/TR/xslt.

[13] Dahlström, E., Dengler, P., Grasso, A., Lilley, C., McCormack, C., Schepers, D., Watt, J., Ferraiolo, J. and Fujisawa, J. (2011). Scalable Vector Graphics (SVG) 1.1 (Second Edition) W3C http://www.w3.org/TR/SVG.

[14] SIMS (2018). Collation Modeler. https://protected-island-3361.herokuapp.com. See also Emery, D. (2018). Collation Modeling. https://github.com/demery/collation-modeling.

[15] Porter, D. and Campagnolo, A. (2018). *VisColl: Modeling and Visualizing Physical Manuscript Collation*. XSLT https://github.com/leoba/VisColl. For an overview on the history of the project see Porter, D., Campagnolo, A. and Connelly, E. (2017). VisColl: A New Collation Toll for Manuscript Studies. In Busch, H., Fischer, F. and Sahle, P. (eds), *Kodikologie & Paläographie Im Digitalen Zeitalter 4*. Norderstedt: Books on Demand GmbH, pp. 81–100.

[16] University of Toronto Libraries (2018). *VisualCollation: Visual Collation Builder*. JavaScript University of Toronto Libraries https://github.com/utlib/VisualCollation.

[17] University of Pisa (2018). Edition Visualization Technology http://evt.labcd.unipi.it/

[18] Bibliotheca Philadelphiensis (2018). Bibliotheca Philadelphiensis http://bibliophilly.pacscl.org/.