Lucia Vannini, MA student of Digital Humanities, UCL

Virtual reunification of papyrus fragments

Several ancient Latin and Greek papyrus fragments, although originally belonging to only one roll or codex, are currently separate and housed in different institutions. This paper, drawn from my MA dissertation, investigates how institutions that own the fragments might collaborate to produce an electronic edition of virtually reunited fragments.

Although there are many papyri with such an issue, there are no instances of online images of virtually reunified fragments, according to my survey; the few examples I found are in printed editions. It is also important to remember the case of the *Codex Sinaiticus*[1], whose digital surrogate represents the whole manuscript, currently scattered across four institutions (including Leipzig University Library); however, while the *Sinaiticus Project* website is exclusively dedicated to that artefact, I propose that an electronic edition should be designed with the possibility to be widened, to fit into a wider system, in order to allow more reunified papyri to be included in the future; in this way, each time a reunified papyrus is added to the database, the team in charge does not need starting over, and can work in an already established framework. A literature review shows a lack of studies on the topic of virtual reunification of papyri, unlike virtual reunification of other objects, like archaeological findings and manuscripts of one author[2].

Upon proposing a project of virtual reunification of papyri, I will follow a model that has been recently suggested by Ricardo Punzalan in his work *Understanding Virtual Reunification* (2014), the first systematic study on this topic. The other approaches only consider some aspects of virtual reunification (for example, only its goal, its final product, without verifying whether its completion is actually possible; or only the technical procedures and not the participants involved; or else the stakeholders involved are taken into account, but they are considered as isolated from each other)[3]. By contrast, Punzalan proposes a consolidated approach that integrates the existing models in a dynamic relationship: goals and procedures are viewed in the light of the constraints of the economical and technological resources, and stakeholders' interests are viewed as overlapping, so that virtual reunification is rightly intended as a collaborative process and the fruit of a compromise among the different participants' priorities[4].

---

[1] Cf. the *Sinaiticus Project* website, which contains images, text and a partial translation of the manuscript: British Library *et al*. 2009. A digital representation of the whole *Codex Sinaiticus*, along with a comment, is also available on the British Library website: British Library, n.d.

[2] An overview of the literature on virtual reunification is found in Punzalan 2014, pp. 295-300.

[3] Punzalan 2014, pp. 312-315.

[4] Punzalan 2014, pp. 315-317.

According to the approach illustrated, I propose that a workflow of a project of virtual reunification of papyrus fragments should include the following steps.

Each owning institution should encode the text and metadata of its own fragment, and produce a digital image. A different institution from those that own the papyri should be specialised in the curation of the process of virtual reunification and discuss with the owning institutions the permission of receiving copies of the texts and the images of their fragments, and the license to utilise the copies in order to realise the virtual reunification. Thus, the owning institutions are expected to send copies of their digital originals to another institution, which will be responsible for the process. This institution, which will involve expertise of several areas, will reassemble the copies in a new web-based database, preferably one equipped with tools that allow searching its material; the institution will make the new image available according to a license agreed with the owning libraries.

It is also possible that the institution that accomplishes the virtual reunification acts as a portal – it would only provide access to the resources on the various institutions' websites, without actually receiving any copy of the material. The *Europeana* project[5] is an example of this model: it exclusively shows thumbnails as a preview, contains the related metadata, and provides a link to the website of the owning institutions; an item can be actually viewed only from the database of its home institution. A portal is a useful framework to collect images of separate objects, when the goal is assembling of material that would be otherwise impossible to physically gather in only one place. But reunification of papyrus fragments also aims to achieve a visualisation of the whole[6] – texts and images of the scattered fragments will be pieced together to represent the complete extant artefact. Therefore, to this purpose, building a new website where a visualisation of the complete object may take place seems to be a more appropriate solution.

Regarding the text of the papyrus fragments, this should be encoded in XML, in particular according to a TEI standard such as EpiDoc[7], in order to be machine readable and thus to be able to be assembled. Furthermore, scholars would be hence provided with the ability to search, browse and find out information on the papyrus.

In order to represent every single characteristic of the fragments, additional features can be added to the TEI guidelines. It is true that the editors of some projects of virtual reunification, that is, the Rossetti and the Blake archives' editors, abandoned the idea of using TEI: they thought it not

---

[5] Europeana, n.d.

[6] "Visualizing the whole" is one of the possible goals of virtual reunification, as Punzalan highlights (Punzalan 2014, pp. 303-304).

[7] For the EpiDoc Guidelines, see Elliott et al. 2007-2014.

adequate to the representation of items of various nature (such as images along with texts)[8]; but it is also important to remember that the TEI standard can be adapted to the peculiarities of a project [9] – that is why the *Sinaiticus* editors decided to use it, even if with the addition of further features to fully describe the details of the item[10]. The TEI standard seems, therefore, very adaptable to the realisation of virtual reunification projects, both when they exclusively concern texts and when they include images as well, as in the project I wish to propose. In addition, the use of a diffuse method such TEI will grant that the project will be possibly integrable with others.

Large papyri can contain a remarkable amount of text, so that they cannot be encoded by a single person. The method followed in the Sinaiticus project[11] can employed in such cases: two electronic transcriptions may be made by two transcribers, who will work from the images to have a better view of damaged letters. Then a software solution, such as "Collate", will be utilised to provide an automatic comparison of the two transcriptions. The differences that will emerge will be checked on the images and, if necessary, on the original codex, to produce a definitive version of text. This final text will be then converted into XML, and eventually into HTML to be displayed on the website.

The owing institutions should agree a common digitisation strategy, as the institutions of the *Sinaiticus Projects* did[12]. The institution that will assemble their material may advise them about this issue. It is essential that the produced images are consistent, so that the final image of the whole papyrus can present a uniform appearance.

Obtaining images according to common practices is difficult because, as the scraps cannot be moved from their institutions, they have to be photographed at their venues, that is, in different environments and with different equipment[13]. Anyway, it is fundamental that at least the following key points are established in a common imaging practice of papyrus fragments: the writing on the papyrus has to be readable, to enable scholars to examine the text; the material of the support has to be reproduced with its natural appearance, to allow the understanding of the physical features of the book[14]; the whole text-bearing support has to be visible, including the margins and the *verso*, even if

---

[8]   Browner et al. 2000, p. 156; Eaves et al. 2015, in part. *Technical Summary*, available from: http://www.blakearchive.org/blake/public/about/tech/index.html.

[9] Browner et al. 2000, p. 156.

[10]   British Library *et al.* 2009 http://www.codexsinaiticus.org/en/project/transcription_download.aspx. One more example of a customisation of TEI according to the needs of a project, is the *Jane Austen's Fiction Manuscripts Digital Edition*; its editors modified the TEI encoding to accommodate numerous interlinear additions inserted in the manuscripts: see University of Oxford - KCL 2015, at the page http://www.janeausten.ac.uk/edition/ technical.html, concerning the technical information on the project.

[11] http://www.codexsinaiticus.org/en/project/transcription.aspx.

[12] http://www.codexsinaiticus.org/en/project/digitisation.aspx.

[13] Cf. http://www.codexsinaiticus.org/en/project/digitisation.aspx, in part. the introduction.

[14]   These two criteria are established as the digitisation standards of the *Sinaiticus Project*: see http://www.codexsinaiticus.org/en/project/digitisation.aspx.

this is blank or preserves a work not related to the *recto*; the natural appearance of the papyrus may not be reproduced in case of seriously damaged fragments, which necessitate an intensive image editing to become readable (for example, the carbonised scrolls from Herculaneum) – in this event, providing a readable image is preferable to the faithful reproduction of the actual current aspect of the papyrus.

In order to achieve these goals, each owing institution should follow agreed recommendations regarding the equipment (e.g., camera, lens, image editing software, lighting) and the process (e.g., equipment setup, colour profiling), upon digitising their fragments[15]. A common representation of the colour of the papyrus should be decided, in order to have a uniform image of the complete papyrus; a valid option can be a compromise colour that allows both readability and a relatively faithful representation of the true colour of the support[16].

Following the EpiDoc guidelines, the code of the electronic edition of the recomposed fragments may contain the following elements pertaining to the reunification.

Firstly, the TEI Header should present a "file description" element with information about the different stakeholders involved in the projects, such as sponsors, funding bodies, people with responsibility in the transcription, proofreading and conversion to XML. There, we may also find information about the publication, including the last update and the agreed licence of use; and, wrapped in a "source description" element, details of the owning institutions and the features of the papyrus considered as a whole: a unique identification number, total dimensions, the layout of a complete page, and all its contents.

The second part of the file should present a "text" element with the transcription. In the *Sinaiticus* transcription code, the text of each page is preceded by a "page break" element whose attributes report, among other information, the indication of the library where that page is housed. But in most papyri the break in the support does not coincide with a unit such as a page, or a column: in most papyri the break is casual, the fragmentation is accidental, because of the unfavourable material conditions in which papyri are found, that is, below ground, among ruins and interspersed with other material. Then, to pinpoint the beginning of a fragment housed in one institution the <milestone> empty element can be used, as it is specialised to describe the text-bearing support, especially when there is a change in it, rather than to describe the content structure of a text[17]. It is in fact defined as a marker of "any kind of section of a text", provided that "the change is not represented by a structural element"[18]. The <milestone> element may be followed by a @unit

---

[15] Cf. http://www.codexsinaiticus.org/en/project/digitisation.aspx.

[16] Cf. http://www.codexsinaiticus.org/en/project/digitisation.aspx, in part. the *Background* section.

[17] Cf. http://www.stoa.org/epidoc/gl/latest/trans-nonstructural.html.

[18] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-milestone.html.

attribute with a "fragment" value; a further attribute, @n, may detail the name of the fragment, according to the place where it is housed.

In conclusion, the visualisation of the fragments joined together would be of greater help to the researchers than their seeing the fragments in two separate images, as happens now, with the advantages that digital images offer in comparison to printed images, and it would also allow a simultaneous comparison between images and transcription. The presence of encoded transcriptions and metadata will facilitate a systematic analysis, thanks to a text that is searchable, of long-term duration and suitable for interchange of information.

Bibliographical references

British Library *et al*. 2009 = British Library, Leipzig University Library, Monastery of St Catherine at Mount Sinai, National Library of Russia at St Petersburg, 2009, *The Codex Sinaiticus Project*, available from: codexsinaiticus.org [accessed 19 January 2016].

British Library, n.d. = British Library, n.d., *Turning the pages. Codex Sinaiticus*, available from: http://www.bl.uk/turning-the-pages/?id=b00f9a37-422c-4542-bfbd-b97bf3ce7d50&type=book [accessed 19 January 2016].

Eaves et al. 2015 = M. Eaves, R. N. Essick, J. Viscomi (eds.), 2015, *The William Blake Archive*, available from blakearchive.org [accessed 19 January 2016].

Elliott *et al*. 2006-2016 = T. Elliott, G. Bodard, H. Cayless *et al*., 2006-2013, *EpiDoc: Epigraphic Documents in TEI XML. Online material*, available from: epidoc.sf.net [accessed 19 January 2016].

Europeana = Europeana foundation, n.d., *Europeana*, available from: europeana.eu [accessed 19 January 2016].

Punzalan 2014 = R. L. Punzalan, *Understanding Virtual Reunification*, Library Quarterly: Information, Community, Policy 84.3, 2014, pp. 294-323, available from: http://rpunzalan.com/wp/wp-content/uploads/2014/10/Virtual-Reunification-2014.pdf [accessed 19 January 2016].

TEI = Text Encoding Initiative (TEI), n.d., *TEI: Text Encoding Initiative*, available from: tei-c.org [accessed 19 January 2016].

University of Oxford - KCL 2015 = University of Oxford, King's College London, 2015, Jane Austen's Fiction Manuscripts Digital Edition, available from: janeausten.ac.uk [accessed 19 January 2016].