9. A Catalogue of Digital Editions

Greta Franzini, Melissa Terras and Simon Mahony

Introduction

Since the earliest days of hypertext, textual scholars have produced, discussed and theorised upon critical digital editions of manuscripts, in order to investigate how digital technologies can provide another means to present and enable the interpretative study of text. This work has generally been done by looking at particular case studies or examples of critical digital editions, and, as a result, there is no overarching understanding of how digital technologies have been employed across the full range of textual interpretations. This chapter will describe the creation of a catalogue of digital editions that could collect information about extant digital editions and, in so doing, contribute to research in related disciplines. The resulting catalogue will provide a means of answering, in the form of a quantitative survey, the following research questions: What makes a good digital edition? What features do digital editions share? What is the state of the art in the field of digital editions? Why are there so few electronic editions of ancient texts, and so many of texts from other periods? By collecting data regarding existing digital editions, and corresponding directly with the projects in question, we provide a unique record of extant digital critical editions of text across a range of subject areas, and show how this collaboratively edited catalogue can benefit the Digital Humanities community.[1]

---

1   For example, the *Digital Classicist* (http://www.digitalclassicist.org), *Digital Medievalist* (http://www.digitalmedievalist.org) and *Digital Byzantinist* (http://www.digitalbyzantinist.org) communities.

Digital editions and cataloguing

There is no universally accepted definition of *digital scholarly edition*.[2] Scholars continuously experiment with old and new tools in order to achieve the optimal digital experience of a manuscript and although there are online guidelines on how to produce scholarly editions,[3] the resulting projects often differ greatly. The term *edition* is generally used to describe the result of an interpretative study of a text. No matter how malleable, diverse and dynamic an edition is, it must be original or, in other words, must add new knowledge. Work that does not produce new knowledge is considered to be a mere reproduction of the primary source. Digital editions move beyond the mere translation into the digital. A digital facsimile is a good example of duplication inasmuch as it is a high-quality, faithful[4] photographic reproduction of the primary source, which can be used as an alternative consultation medium, thus avoiding repeated handling of the original. As a replica, this type of publication does not bear any new information and cannot, therefore, be considered an edition. Our area of interest is in the interpretative, digital publications of texts that allow new understanding of the original source material to be generated.

Unlike the past, where scholarly merit derived from expert and monumental pieces of work, (digital) editions today are constantly assuming different shapes; whether standalone projects or pieces of a larger whole,[5] digital editions are reassessing the notions of engagement and completeness. The latter often depends on the former, in that today's editions seek to embrace crowds—from both a reception and production standpoint—whose goal is to socialise, to exchange

---

2   Much literature exists on the topic; see e.g. Kenneth Price, 'Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?', *Digital Humanities Quarterly*, 3.3 (2009), http://www.digitalhumanities.org/dhq/vol/3/3/000053/000053. html; Hans Walter Gabler, 'Theorizing the Digital Scholarly Edition', *Literature Compass*, 7 (2010), 43–56, http://dx.doi.org/10.1111/j.1741-4113.2009.00675.x; Mats Dahlström, 'How Reproductive is a Scholarly Edition?', *Literary and Linguistic Computing*, 19 (2004), 17–33, http://dx.doi.org/10.1093/llc/19.1.17. See also Patrick Sahle's chapter in the present volume.

3   For example: *MLA Guidelines for Editors of Scholarly Editions* (2011), http://www.mla. org/cse_guidelines

4   As close as possible to the original.

5   Such as, digital libraries or archives showcasing various subprojects, items and collections.

views, to produce community knowledge and to help users read,[6] thus advancing research. This push for advancement is not only informed by our immersion in the rapid technological evolution but it is also dictated by people's need to assert their presence in an increasingly competitive and interdisciplinary field.

How might we understand the remit of digital editions, given this pace of technological change? Patrick Sahle's *Catalog of Digital Scholarly Editions*[7] presents a taxonomy which identifies when a digital edition is scholarly, providing various indicators to help understand the outputs of digital textual projects:[8]

> **S—Scholarly**: An edition must be critical, must have critical components. A pure facsimile is not an edition, a digital library is not an edition.[9]
>
> **D—Digital**: A digital edition cannot be converted to a printed edition without substantial loss of content or functionality. Vice versa: a retro-digitised printed edition is not a scholarly digital edition (but it may evolve into a scholarly digital edition through new content or functionalities).
>
> **E—Edition**: An edition must represent its material (usually as transcribed/ edited text)—a catalogue, an index, a descriptive database is not an edition.
>
> **Complete/Prototype**: An SDE (Scholarly Digital Edition) is a publication of the material in question; an SDE project is not the same as an SDE, that means an SDE is more than a plan or a prototype.

In Sahle's model, a *scholarly* digital edition is a *critical* digital edition, understood as an analytical and accurate contextual study offering hypotheses and new insights into the source text under examination,

---

6   Peter Robinson, 'The One Text and the Many Texts', *Literary and Linguistic Computing*, 15 (2000), 5–14 (p. 13), http://dx.doi.org/10.1093/llc/15.1.5

7   As the name suggests, the *Catalog* lists only *scholarly editions*. Personal correspondence (14/06/2012) with Patrick Sahle revealed that the *Catalog* began in 2006 and saw only fifty new entries in the four years 2008 to 2012. The *Catalog* is available at http://www.uni-koeln.de/~ahz26/vlet/vlet-about.html

8   What follows is a summary of Patrick Sahle's analysis in the *About* page of his *A Catalog of Digital Scholarly Editions* website http://www.digitale-edition.de/vlet-about.html

9   Scholars in the field use the term *digital library* to describe a collection of electronic texts and/or visual materials, which typically does not add new knowledge to the primary source. It can be considered a digital exhibit. Digital libraries should not be confused with *variorum editions*, which are collections of variants of the same copy-text with appended commentary.

as elaborated by Peter Robinson.[10] Sahle's definitions help us define in more precise terms the intention and scope of the edition. To express it in Espen Ore's words:

> Any scholarly edition is better than none even if it is not a critical edition, and […] editions that may not be critical digital editions do indeed have value and represent a kind of edition which are in fact the basis for critical text editions.[11]

These words bring to the fore the present and more social nature of digital editions, whereby multiple people can contribute—some more, some less—to a single edition.[12] Sahle's work also provides the starting point for analysing all digital editions, and offers an overarching catalogue that allows us to understand this field.

The need for a catalogue of digital critical editions is dictated by the absence of up-to-date, analogous resources. Caroline Macé called for a repository of digital editions at the *IV Incontro di Filologia Digitale*[13] in September 2012. Her rationale was that such a catalogue would provide a means of discovery, linking and advertisement of digital texts that may otherwise go unnoticed. We believe that an up-to-date, online catalogue of digital editions would provide an accessible, unique record of manuscripts that have had digital editions created based on them; allow an understanding of the digital editions created which are allied to a range of distinct historical periods; and providing a data bank of features, tools, licences, funding bodies and locations. This will give an insight into past, present and future digital edition projects, providing

---

10   Peter Robinson, 'What is a Critical Digital Edition?', *Variants: The Journal of the European Society for Textual Scholarship*, 1 (2002), 43–62. *Contextual* here is to be understood as a comprehensive study of the history, materiality and reception of the primary source under investigation.

11   Espen Ore, 'Monkey Business — or What is an Edition?', *Literary and Linguistic Computing*, 19 (2004), 35–44 (p. 35), http://dx.doi.org/10.1093/llc/19.1.35

12   Or, as Siemens et al. write: '[…] the 'social' edition is process-driven, privileging interpretative changes based on the input of many readers; text is fluid, agency is collective, and many readers/editors, rather than a single editor, shape what is important and, thus, broaden the editorial lens as well as the breadth, depth, and scope of any edition produced in this way'. Ray Siemens et al., 'Toward Modeling the Social Edition: An Approach to Understanding the Electronic Scholarly Edition in the Context of New and Emerging Social Media', *Literary and Linguistic Computing*, 27 (2012), 445–61 (p. 453), http://dx.doi.org/10.1093/llc/fqs013

13   The fourth meeting on Digital Philology was held in Verona, Italy (2012), http://www.filologiadigitale.it

the means to identify and view trends or patterns across the corpus (for example what time periods are covered most, which features are most prominent, or which institutions produce the largest number of digital editions), areas for improvement of errors, as well as projects which are no longer maintained or even available. This will inform future development of digital editions (from both technical and subject area perspectives), establish a hub around which collaborators can engage in community discussions, and become the source of updated information as it becomes available.

Fragmentary lists of digital editions projects already exist, but these do not record project features or provide an easy means of browsing, viewing and downloading the data, and often maintain links to projects which are no longer available.[14] Minor catalogues are curated by Paolo Monella, Cinzia Pusceddu, Aurélien Berra, the *Monastic Manuscript Project* and the wikis of Hunter College, the Digital Classicist and the *Associazione per l'Informatica Umanistica e la Cultura Digitale*.[15] With the exception of Sahle's *Catalog of Digital Scholarly Editions*, which currently records 386 projects, the catalogue emanating from our research is the most recent, and is certainly the most detailed in circulation, providing an overview of features and approaches, as well as details of the projects themselves. While Sahle's *Catalog* aims to record extant scholarly editions, our project brings together both scholarly and

---

14 See, for example, Ian Lancashire's *The Humanities Computing Yearbook 1989–1990* (Oxford: Oxford University Press, 1991). More recent resources include *arts-humanities.net* (http://www.arts-humanities.net), which also lists AHRC-funded projects (http://www.arts-humanities.net/ahrc_projects, last accessed February 2013), the *Zotero Digital Humanities Group* (http://www.zotero.org/groups/digital_humanities), *Romantic Circles Electronic Editions* (http://www.rc.umd.edu/editions) and *Rotunda Publications* (http://rotunda.upress.virginia.edu).

15 Monella's catalogue was formerly available at https://docs.google.com/document/d/1rmCkvtVJmLcJrJsUOXs90dSEcgs7MOOJdLdEb7, section 2.2 [last accessed February 2013]; Pusceddu's at http://www.digitalvariants.org/e-philology; Berra's at http://philologia.hypotheses.org/corpus; *Monastic Manuscript Project* list at http://earlymedievalmonasticism.org/listoflinks.html#Digital; Hunter College's at https://www.zotero.org/groups/hunter_college_engl_390.81/items/collectionKey/34ST6AVS; *Digital Classicist* lists, http://wiki.digitalclassicist.org/Greek_and_Latin_texts_in_digital_form and http://wiki.digitalclassicist.org/Digital_Critical_Editions_of_Texts_in_Greek_and_Latin; *Associazione per l'Informatica Umanistica e la Cultura Digitale* wiki, http://www.digitalclassicist.org/wip. Another notable catalogue is UCLA's *Catalogue of Digitized Medieval Manuscripts*, which, however, records some 3126 fully digitised manuscripts as opposed to digital editions http://manuscripts.cmrs.ucla.edu

non-scholarly resources, though it distinguishes between the two. Where Sahle gives useful summaries of each project, our catalogue also provides information on different categories and lists of technical and scholarly features. To date, we have examined and categorised 187 out of the total 325 editions harvested thus far.[16] The reason for not yet fully cataloguing all the digital editions collected is practical: listing editions is a relatively effortless process, but examining a project, on the other hand, is a slower and more labour-intensive activity, with much time spent looking for information (whether available via the project website or gathered directly from the creators). The number of projects suitable for further study will grow as our catalogue expands, and we aim fully to examine and categorise all projects listed as this research progresses.

Methodology

The editions present in the catalogue come from numerous sources, and their selection follows basic criteria: the electronic texts, whether available online or on CD-ROM, can be ongoing or complete projects,[17] born-digital editions[18] as well as electronic reproductions of print volumes. They were gathered from the previously mentioned catalogues, from lists such as *Projects using the TEI*,[19] RSS feeds,[20] publications (articles, reviews and books), Google Scholar alerts, Twitter, word of mouth, web browsing and chaining.[21]

---

16   The number is subject to change as the project progresses. The collection progress described in this chapter ran from August 2012 until December 2013.
Examined means that the authors have looked at the editions in great detail. In this sense, the catalogue will never be *complete*. New editions are systematically added to a queue waiting to be analysed in the same detail. The editions in Sahle's *Catalog* will also be included in our database, with the exception of those listed by him but appear no longer to exist, for example *Con2: An Edition of the Anglo-Saxon Chronicles, 924–983* (formerly available at http://www.slu.edu/departments/english/chron/index.html).

17   Still active on the web.

18   *Born-digital edition* refers to text born digital and edited for a digital publication.

19   Projects using TEI, http://www.tei-c.org/Activities/Projects/index.xml

20   Such as *The Ancient World Online* (http://ancientworldonline.blogspot.co.uk); *arts-humanities.net* (http://www.arts-humanities.net) and *Digital Classicist seminars* (http://www.digitalclassicist.org/wip).

21   For chaining, see David Ellis, 'Modeling the Information-seeking Patterns of Academic Researchers: A Grounded Theory Approach', *The Library Quarterly*, 63 (1993), 469–86 (p. 482), http://www.jstor.org/stable/4308867

In line with Klaus Krippendorf's content analysis method, the data was carefully collected and assessed both quantitatively and qualitatively, in order to make reliable inferences from which further research can stem and develop.[22] The content analysis was carried out along two parallel tracks: a data gathering approach, whereby each project team was contacted with a short questionnaire aimed at gaining a deeper understanding of both the production and user needs of the edition (see section 3.1.); and an observational examination of extant knowledge about the electronic edition through the analysis of the project website and its related publications. To date (March 2014), our catalogue showcases 187 digital editions, collected and examined over a period of sixteen months. As previously mentioned, the *Catalogue of Digital Editions* makes a distinction between scholarly and non-scholarly digital editions, and replicas of existing print volumes.[23] Of course, there are many more editions left to include and, indeed, many more to come.[24] Launched in May 2013,[25] the website showcases visualisations of the catalogue data, providing contextual information, as well as encouraging the community to contribute information, suggestions for improvement or feedback.

Data gathering

Creators of the digital editions collected were contacted directly between August 2012 and March 2013 and asked to provide information about their projects in structured categories. The questionnaire was disseminated and completed through email due to the scale of the

---

22 Klaus Krippendorf, *Content Analysis: An Introduction to Its Methodology*, 2nd ed. (Thousand Oaks: Sage, 2004). Statistically, the higher the number of editions, the more accurate and revealing the results (as they are less subject to sampling risk). Similarly, the more meticulous the qualitative analysis (the number of features under investigation), the clearer the implications.

23 Once again, it is important to stress that while making this distinction, this catalogue lists all electronic texts, regardless of their academic purpose.

24 To the authors' knowledge, there is no definitive or rough estimation of the total number of electronic editions active today on the web. However, based on the lists mentioned earlier, it seems fair to suggest a combined total of some 500 editions (excluding forthcoming and inactive projects archived in the *Wayback Machine* http://archive.org/web/web.php).

25 *A Catalogue of Digital Editions*, https://github.com/gfranzini/digEds_cat. The catalogue was initially set up as a *Google Sheet* (spreadsheet) viewable with any online reader but has now moved to GitHub.

research, and aimed to discover what the project goals and achievements were, what type of user enquiries or requests had been received, if the project had gathered any statistics regarding use of the resource, if the project understood who their main audience was, the project budget and team size needed to create the resource,[26] what lessons the team felt they had learnt from undertaking the project and sustainability issues in making the resource available over the longer term. Of the seventy-eight people contacted,[27] thirty-seven replied (some in full, some partly). Of the remaining forty-one, six emails bounced back due to expired email addresses and thirty-five have yet to, and may never, reply. Information gathered from this correspondence is stored in a separate file which is used as a log, an address book and as a reference tool to track changes and developments over time.

Content analysis of available information

Once a project has been identified for inclusion in the catalogue, the edition is subject to in-depth analysis, depending on the information that can be found in the available project data. The *Catalogue of Digital Editions* contains a variety of information: the *Catalogue*, or editions examined so far; *Institution coordinates* (a list of all institutions encountered thus far and their geographical coordinates, for reference and spatial/visual analysis); *Funding body coordinates*; *Repository coordinates* (the source document's current location or home); and *Place of origin coordinates* (the source document's presumed or known hometown or country and where it is now, if different). In the future, we will also note *Linked open data* when projects share source data. Additional features are grouped into *subject areas* encompassing ontological, technological and philological aspects.

The categories are the result of a comparative study of a number of editions whose aim was to identify commonalities between editions and

---

26   Given the potentially sensitive nature of the question, interviewees were given the option to not respond.

27   The number of emails sent out to project managers is smaller than that of editions as, more often than not, project managers are either co-investigators or work on multiple projects at the same time (in these instances, the questionnaire email addresses all relevant projects).

draw out the more useful and desirable features. None of the categories represent in any way compulsory features and while one might be more important than another to certain user groups, the catalogue does not make use of any weighting system. As is the nature of content analysis, categories are rather clear-cut and some projects may not entirely subscribe to their specificity. For example, ongoing projects will figure as incomplete or the digital text may be part replica and part born-digital. In such cases, the comments field allows details about classification choices. The edition illustrated in Figure 9.1, for example, is classified as scholarly but not digital for this version of the *Carmina* does not in any way enhance the printed text. It is also not complete, and so does not satisfy the aforementioned requirements to be considered critical.

	C	D	E	F	G	H	I
1	Edition_ID	URL	Scholarly	Digital	Edition	Prototype	Language
83	Claudii Claudiani Carmina Latina	http://www.curcu	1	1	1	0	LAT

Fig. 9.1 Example application of Sahle's rules to the *Claudii Claudiani Carmina Latina* project.

There is a full and detailed table available online that lists and describes all the fields in the *Catalogue of Digital Editions* and their respective numerical scoring, indicating the breadth of useful information that can be captured about digital editions to facilitate their analysis.[28] Whenever project websites are unclear or we cannot explicitly know the information, this is noted. Examples of the information recorded include: title and web address; the historical period the text belongs to (ancient, medieval or modern, with clarification of those terms); the project scope and perceived audience; does it qualify as a scholarly digital edition according to Sahle's classification?; does it include textual criticism and/or any *apparatus criticus*?; is the content encoded in TEI-compliant XML?; are there digital facsimiles of the primary sources?; what platform is used to host the project?; details of any analytic tools provided; ease of access; is there a CD-ROM edition?; details about the project itself; translations and

---

28 Table describing all fields and numerical scoring used in our catalogue, https://github.com/gfranzini/digEds_cat/wiki/Contribute

languages; any other useful and desirable information such as licences, open or proprietary, Open Source/Open Access.

Integration and visualisation

Once recorded, the information in the catalogue can be integrated and visualised. At present, the catalogue contains fifty-four columns and an ever-increasing number of rows, showcasing a large set of data, which can only be satisfactorily viewed and understood through visualisations and detailed queries; the data will only become more complex as more examples of editions are added.

Google Fusion Tables was used to extract data from the *Google Sheet* and visualise it as a map.[29] Maps can be used as data filters, displaying only certain sets of information, helping us contextualise data and better understand distribution and relationships. In our case, a map can help institutions survey their place in the world.[30]
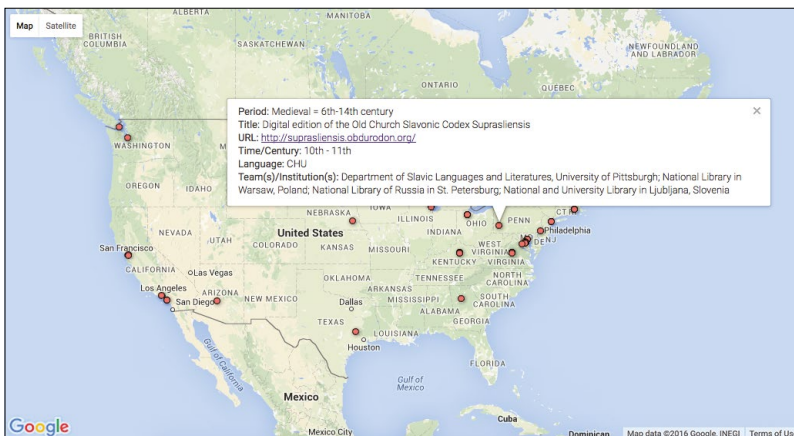

Fig. 9.2a Screenshot of a section of the map visualisation (March 2014). Location markers identify projects and pop-up windows provide information to the user.

---

29 See https://sites.google.com/site/digitaleds/mapping-editions
30 Martin Jessop, 'The Inhibition of Geographical Information in Digital Humanities Scholarship', *Literary and Linguistic Computing*, 23 (2008), 39–50 (p. 39), http://dx.doi.org/10.1093/llc/fqm041
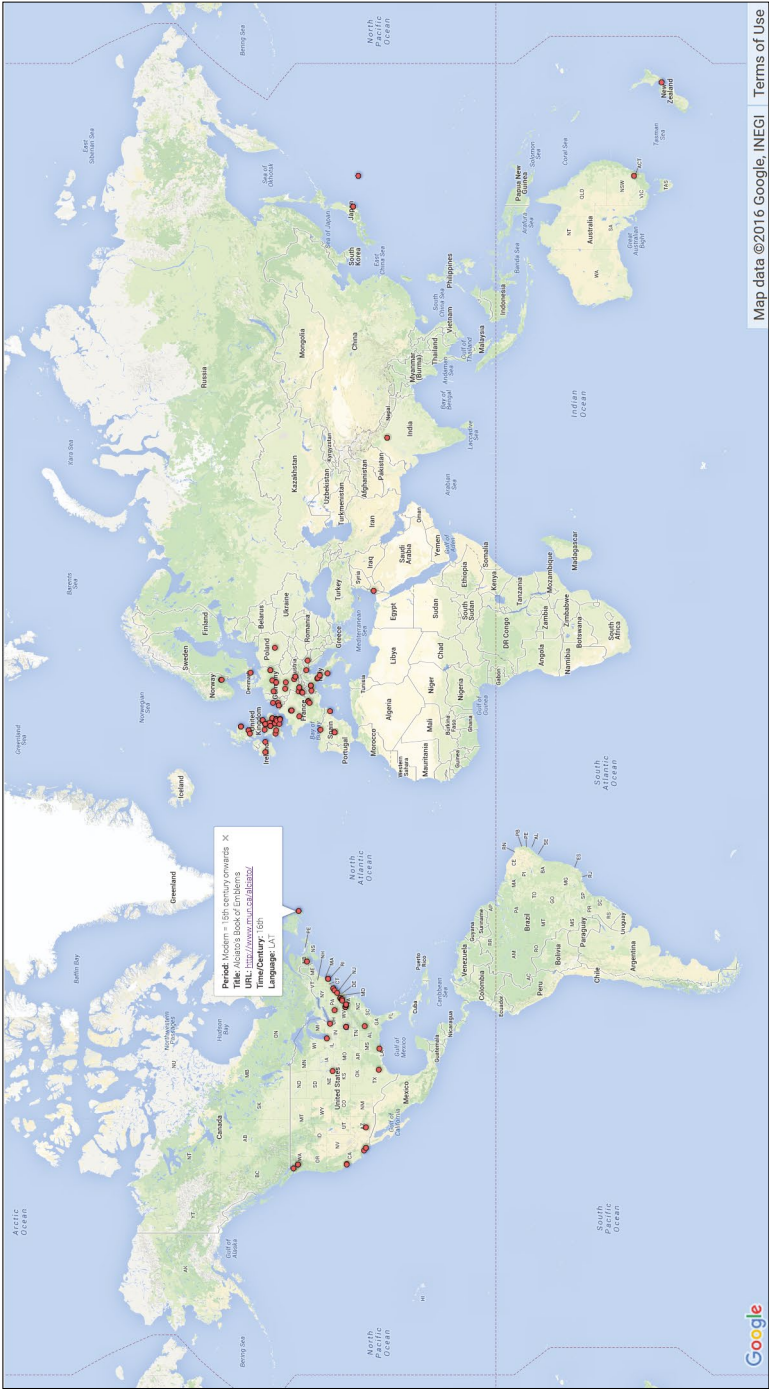
Fig. 9.2b Screenshot of the map visualisation of the editions present in the catalogue to this day (March 2014).

As our dataset grows, we will be able to generate maps of manuscript place of origin and current repository,[31] which may tell us more about the travels of a particular document and the culture and cultural factors which surround manuscript collecting. In Figure 9.2a, institutions are highlighted using location marks which, upon hovering, open a pop-up window with information about project title, author and date of project.

Figure 9.2b is a snapshot of the state of the catalogue in March 2014: the reader will notice a shortage of, for example, Asian and Arabic editions as we work through those in the catalogue. Nevertheless, digital editions appear to be a Western phenomenon, led by the United States and the United Kingdom, two of the wealthiest and most influential countries in the world, both economically and politically.[32] A recent study by the Oxford Internet Institute (OII) reveals a correlation between wealth and data openness; data openness, OII explains, is partly dependent on Internet penetration.[33] Countries affected by limited access to the Internet such as those in Asia, Africa and Latin America appear at the bottom of the OII charts. While there might not be an obvious correlation between production of digital editions, openness and wealth, our initial results seem to point towards this.[34] Dissemination is part of the European Digital Agenda, which provides a funding infrastructure to promote growth in the educational, cultural and commercial sectors, helping Europe build a competitive research and innovation profile.[35] While North American investment in the information and communications technology (ICT)[36] infrastructure is

---

31   The Schoenberg Database of Manuscripts (SDBM) is conducting such a study, http://dla.library.upenn.edu/dla/schoenberg/index.html; another example is that of the *Digitized Medieval Manuscripts* (*DMMMaps*) project, which is producing a map of current manuscript repositories in an attempt to link libraries and documents across the world, http://digitizedmedievalmanuscripts.org

32   65% of the projects are Anglo-American. That is 123 out of 187 editions.

33   Emily Badger, 'Why the Wealthiest Countries are also the Most Open with their Data', *The Washington Post* (14 March 2014), http://www.washingtonpost.com/blogs/wonkblog/wp/2014/03/14/why-the-wealthiest-countries-are-also-the-most-open-with-their-data

34   It should also be noted that the present catalogue is not a comprehensive survey and might, therefore, be overlooking editions written in alphabets and languages beyond the authors' reach (e.g. Cyrillic or Chinese).

35   *Digital Agenda for Europe*, http://ec.europa.eu/digital-agenda

36   The above website states: 'Currently, EU investment in ICT research is still less than half US levels', http://ec.europa.eu/digital-agenda/en/our-goals/pillar-v-research-and-innovation

much higher than that of Europe, both continents are at the forefront of the educational sector. Finally, we must remember that major associations and portals in the Digital Humanities are based in the US and UK.[37] We will be better able to define this relationship as the catalogue grows, and the data can be merged with other public tables to allow further analysis of trends emerging from the data collection.[38]

Results from the catalogue

Identifying and listing the features editions provide in a detailed and methodical fashion can help us refine our thoughts about digital editions. Projects are highly affected by the size of the corpus they select, the financial backing they can rely on and the timeframe within which they develop. We were able to gather much more information about the production than the usage of editions: many projects do not keep track of user statistics and, therefore, cannot provide information about how the resource is used.[39] Most editions address an intended scholarly audience but many projects do not provide basic editorial or technical information. There appears to be a tendency to leave out important information about the production process (imaging, for instance) and to place too much trust in the audience's knowledge of the field by way of assumption. Regardless of the intended audience, more could—and should—be done in terms of clarity and transparency, from both a content and contextual standpoint. However, as John Lavagnino writes:

> When we create editions, we are thinking about readers in two disciplines: readers who are editors, and readers who are not editors. […] Making editions that work for both editors and for the popular audience will always be tricky, and moving into the digital world does not really make it much easier. […] The most obvious problems that the popular audience has with editions stem from the apparatus, and such

---

37 Among many others, the Association for Computers and the Humanities (ACH); Alliance of Digital Humanities Organizations (ADHO); The European Association for Digital Humanities (formerly ALLC); arts-humanities.net; DHCommons; Digital Humanities Now; Humanities, Arts, Science, and Technology Advanced Collaboratory (HASTAC) and The Humanities and Technology Camp (THATCamp).

38 Some *Google Fusion Tables* users choose to make their data publicly available so that other users can merge multiple datasets to create custom and diverse visualisations.

39 One example is the digital edition of the Old Church Slavonic Codex Suprasliensis, http://suprasliensis.obdurodon.org

> problems frequently have the undesirable effect of leading readers to ignore the apparatus or consider it too hard to use.[40]

It follows that editors should either strive for utmost clarity (with the end result that experts in the field might find some of the information redundant) or create layered content. The often-complex critical notes, for instance, could be organised in such a way that would allow users to filter by level of detail.

Projects urging the digital reunification of fragments or manuscript leaves housed in different locations are often internally fragmented themselves, having split the project management between different institutions. For instance, the computer and web development section of the British Library's Codex Sinaiticus project is managed by the University of Leipzig, which does not share user information with the British Library and the curators.[41]

The budget structures of projects are diverse: some small projects (e.g. *Phineas Fletcher's Sylva Poetica*) are the result of lengthy and free labour, whereby costs are either defrayed by the authors themselves or the work is carried out in their spare time; some editions hide behind pay walls or registrations forms (the fourteen *Rotunda* editions by the University of Virginia Press, for example), or can be fully accessed only by borrowing or purchasing the CD-ROM.

Technical trends emerge: the pie chart below (Figure 9.3) represents the 187 editions covered so far by the catalogue. The chart showcases the technologies used to encode the source texts and how many projects in the catalogue employ those technologies. Each slice is broken down even further to show, for example, how many ancient texts have been encoded using TEI as opposed to custom XML. Of the electronic editions collected and examined thus far only sixty-nine (37%) follow TEI encoding standards; fourteen (7%) use a bespoke set of XML tags to suit the features of the source text; thirty-seven (20%) employ other technologies, namely HTML, Cascading Style Sheets (CSS) and Plain Text; and the remaining sixty-seven projects (36% of the entire catalogue) do not clearly state if and how encoding was carried out, so we have

---

40 John Lavagnino, 'Access', *Literary and Linguistic Computing*, 24 (2009), 63–76 (pp. 65–66), http://dx.doi.org/10.1093/llc/fqn038

41 As orally reported to Greta Franzini by one of the British Library curators in charge of the project in May 2012.

labelled those 'unclear'.[42] Two reasons were given why projects prefer to devise custom XML rather than adopt the recommended TEI tag-set. The first is purely practical: TEI has to be learnt and some projects feel they do not have the time to develop and apply this skill effectively. The second reason is editorial in that custom XML can be designed to better fit the nature and features of the source text. As for HTML, CSS and plain text editions, these technologies are typical of older projects, such as *D. Iunii Iuvenalis Saturae* or *Supliciae Conquestio*.[43]
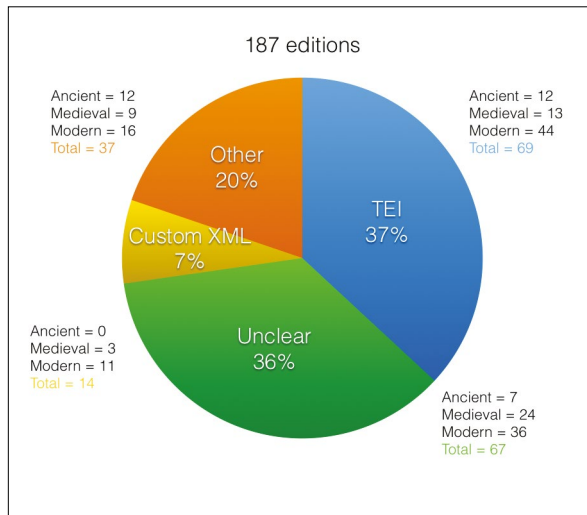

Fig. 9.3 Use of XML-TEI vs. other technologies in the digital editions featured in the catalogue.

These results are rather surprising considering TEI's promotional strategy.[44] The number of projects adopting TEI guidelines is gradually increasing, perhaps reflecting TEI's systematic growth and improvement.

---

42 These figures are based on the study of information explicitly available on the project website. It is likely that any implied or non-explicit information will fall under the 'unclear' slice of the pie chart (see Figure 9.3).

43 The former is available at http://www.curculio.org/Juvenal, the latter at http://www.curculio.org/Sulpiciae

44 The chief dissemination avenues being *The Journal of the Text Encoding Initiative* (http://journal.tei-c.org/journal), workshops, conference and seminar presentations given not only by members of the TEI community but also by project investigators who are adopting the standard, as well as publications and word of mouth. More recently (March 2014), the TEI advertised a Social Media Coordinator position as it seeks to improve outreach.

Nevertheless, the number of TEI editions is still low, suggesting some resistance to implementing this specific set of tags and structures. Today, at a time when scholars are gradually recognising the advantages of community-driven projects, standards need to be adopted if we are to push digital editions in a social direction or integrate their resources. Without guidelines such as the TEI, exchange and repurposing of data will not be possible and electronic editions will be used as standalone objects with their own set of characteristics, objectives and requirements.

The vast majority of texts encoded in TEI are written in English (and older variants thereof, i.e. Old and Middle English); the second most represented language is Latin, followed by French (including Old and Middle), Old Norse, German (including Middle High German), Welsh, Spanish, Ancient Greek, Italian, Old Irish and Hebrew.

Figure 9.4 below shows the languages covered by the catalogue and the predominance of English and Latin primary sources (which may change as we come to catalogue further Asian editions).
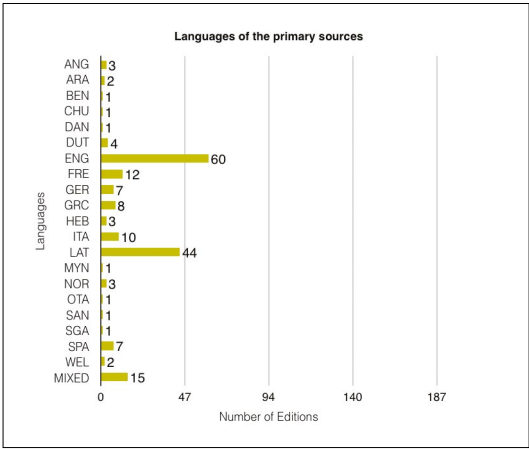

Fig. 9.4 Languages of the primary sources present in the catalogue. Some projects contain multiple texts in different languages. For the purpose of this calculation and illustration, these projects are categorised under *Mixed*. *Mixed* covers projects focussing on multiple texts and combinations of Ancient Greek, Aramaic, English, French, German, Hebrew, Latin, Middle English, Polish, Sanskrit and Vietnamese.

As the catalogue grows and more editions are added, it will be interesting to discover whether there is a relationship between a particular language and the use of XML. Are some languages easier to encode and therefore more likely to be digitally rendered using XML? As it stands,

the predominance of English and Latin could be indicative of three things: apart from being the most widely known languages out of those featured in the catalogue, English and Latin texts are more likely to be encoded in XML because both are Latin-script languages;[45] XML is not yet easily implementable for non-Latin scripts (e.g. Cyrillic or Arabic); or the XML (and TEI) penetration is higher in academic environment where English and Latin are studied.

Most scholars working in the field of digital editions will be aware of Creative Commons licences[46] and how important it is to make work available under these conditions in order to promote research and further knowledge. Creative Commons licences appear to be becoming increasingly popular, and yet out of the 187 editions examined thus far, only thirty-two are available under a Creative Commons licence. The content of the remaining editions is either proprietary or available under different licences.

Other issues emerge from the survey, such as broken links, unavailable projects or expired email addresses, due to poor maintenance (for instance, the *purchase* button of the *Domesday Explorer* CD-ROM edition returns a 404 error[47]). Only a few people have provided information about project costs, or will allow this information to be shared.[48] Open projects are not always transparent, and funding appears to be a taboo topic.[49]

Digital editions of ancient vs. modern texts

Our catalogue of extant digital editions has led us to recognise a numerical disparity between the electronic reproduction of medieval and modern documents and manuscripts predating the fifth century AD.

---

45 Latin-script languages present fewer challenges when it comes to digitisation and encoding because computer software has been trained to recognise these characters.

46 For more information about these licences, see http://creativecommons.org

47 *Domesday Explorer*, http://www.domesdaybook.net

48 However, it does not seem unreasonable to expect publicly funded projects, such as those supported by JISC, to be open to such questions.

49 When asked about funding, one of the investigators of the *Orlando Furioso Hypertext Project*, http://stel.ub.edu/orlando, funded by the *Ministerio de Educación y Ciencia* and the *Istituto Italiano di Cultura di Barcelona*, replied: '*Su questa questione, capisce che non voglia dare forse i dati concreti*' [With regard to this matter, I am sure you can appreciate why I prefer not to disclose concrete numbers] (10 March 2013).
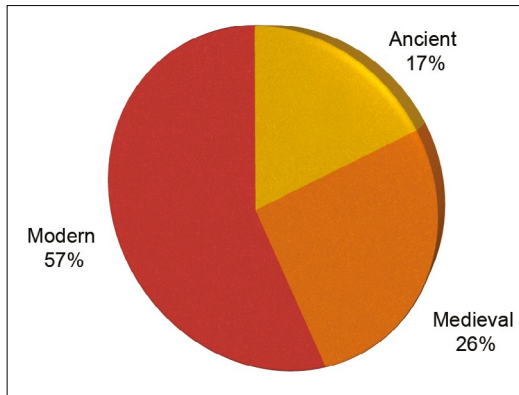
Fig. 9.5 Out of the 187 projects fully analysed, over half are electronic
editions of modern texts.

Possible reasons for this could be that the number of modern manuscripts is larger than that of ancient manuscripts; that eighteenth- to twentieth-century manuscripts might be of more interest as they more closely relate to present times; that the absence of suitable, user-friendly tools supporting ancient scripts stops people from creating these resources. There are perhaps issues of funding: are modern topics better supported by funding councils? Or is the perceived degree of the usefulness of providing editions of texts less in cases where a relatively small proportion of society will have the skills to read them? Mats Dahlström, who also noticed a shortage of scholarly editions of classical works, offers further reasons:

> […] shortages of time, resources, and competence. […] Another reason for this conspicuous lack is the varying degrees of meritocratic prestige of print and digital media. Further reasons include: copyright restrictions; presumed illucrativity and consequential difficulty in finding financial support; authenticity, security, and long-time preservation uncertainties; as of yet severely primitive software for storing, presenting, encoding, and displaying the kind of complexity inherent in classical works. Finally, both the construction and the usage of existing digital SE:s [scholarly editions] need probably be thoroughly evaluated.[50]

---

50   Mats Dahlström, 'Digital Incunabules: Versionality and Versatility in Digital Scholarly Editions', in *ICCC/IFIP Third Conference on Electronic Publishing 2000, Kaliningrad State University, Kaliningrad/Svetlogorsk, Russia, 17th–19th August 2000* (Washington: ICCC Press, 2000).

Moreover, as John Lavagnino noted twenty years ago:

> In the last few decades, many textual scholars have come to believe that classical texts and modern texts have very different kinds of textual problems and constitute different kinds of literary works. Texts from classical antiquity have great textual problems: any manuscript that has survived to our day of such texts is the product of a long sequence of copyings and recopyings, so that it's likely to be full of errors in transmission that need to be corrected. These are errors on such a scale that the works are often simply unreadable without editorial correction. But for modern texts, the body of surviving evidence is very different. For texts circulated since the invention of movable type, and particularly for texts written since the seventeenth and eighteenth centuries, the problem of mistransmission is less and less imposing. The texts have been copied only a few times prior to the creation of our sources, rather than many times, and we often have many more of the sources, sometimes going back to the author's own drafts. Error will always be present, and is still sometimes a great problem, but it ceases to be the central problem.
> What we have for many modern works is not a shortage of reliable information, but an excess: often there is far too much textual information to include in any printed edition. We find, for example, cases in which a writer made extensive revisions over a span of many years, so that there may be a number of versions that were all produced by the same person and that all have good claims to our attention; but which one should be the text that a scholarly edition prints?[51]

The difference in numbers between classical and modern digital editions is a complex issue which deserves further attention.

A good digital edition? Some recommendations

The catalogue reveals just how different editions can be, despite them all sharing a core objective to disseminate and advance new knowledge about the text by means of the digital medium.[52] Structures and outputs are dictated by numerous variables, including readership, usage and resources. In this inevitably dynamic, ever-changing reality, there is perhaps no real necessity to set an axiom or formulate a strict definition

---

51   John Lavagnino, 'Reading, Scholarship and Hypertext Editions', *TEXT*, 8 (1995), 109–24 (p. 111), http://www.stg.brown.edu/resources/stg/monographs/rshe.html

52   *Digital medium* understood as an aid to, not a replacement of, the print publication, where available.

of digital scholarly edition. How does defining an edition affect that edition and its audience? Should we not be more concerned about generating usable and useful content?[53]

From creating the catalogue and analysing various digital editions in detail, we suggest that in order to be comprehensive and widely used, a digital edition should include descriptive information about the purpose of the edition, the manuscripts' history, production, significance and use, high definition images of the manuscript, optimised for the web, with the possibility of downloading or purchasing image files for personal or educational use,[54] documentation about the photographic process and technical metadata about the capture equipment and a transcription of the text and marginalia, including non-standard textual features (abbreviations, punctuation marks etc.) to make the manuscript more accessible to a non-expert audience. If the nature of the project is more scholarly, then

> The transcribed text must attain the usual levels of critical accuracy, meaning that the edition needs to follow diplomatic standards and be the product of expert work. The modern reader must have confidence in the edited text.[55]

Transcriptions should also conform to XML standards for searchability, worldwide integration, interchange and repurposing of data. 'A well-made electronic scholarly edition will be built on encoding of great complexity and richness'.[56] The digital edition should include searchable text and images[57] made possible with the use of appropriate and meaningful metadata, a critical apparatus, indices and word lists

---

53   Melissa Terras, 'Should We Just Send a Copy? Digitisation, Usefulness and Users', *Art Libraries Journal*, 35.1 (2010), 22–27, http://discovery.ucl.ac.uk/171096/1/Terras_Sendacopy.pdf

54   Greta Franzini purchased images of the primary source that her edition is based on for €100 (September 2011). The purchase came with an agreement whereby these images could only be published online if appropriately credited.

55   Jonas Carlquist, 'Medieval Manuscripts, Hypertext and Reading: Visions of Digital Editions', *Literary and Linguistic Computing*, 19 (2004), 105–18 (p.115), http://dx.doi.org/10.1093/llc/19.1.105

56   Peter Robinson, 'Where We Are with Electronic Scholarly Editions, and Where We Want to Be', *Jahrbuch für Computerphilologie*, 5 (2003), 125–46.

57   Susan Hockey, *Electronic Texts in the Humanities* (Oxford: Oxford University Press, 2010), pp. 141–42.

to facilitate filtering and more advanced searches, downloadable data (XML files), as well as print-friendly outputs or versions,[58] links to external resources, such as word or abbreviation dictionaries, clarifications of palaeographical terms and biographical information about the people mentioned within the text, different views of the text 'XML for analysis, XHTML for consultation on the screen and PDF for printing out as a reading edition',[59] a translation if necessary (whether internal to the website or a link to an existing, third party translation) for a wider appreciation of the text, space for users to comment, suggest improvements or corrections and discuss the material. The web platform or content management system (CMS) on which the edition runs should conform to W3C (World Wide Web Consortium) standards;[60] provide project documentation to allow the user to appreciate the edition's limitations or customisations and thus better utilise the resource; provide metadata (METS: Metadata Encoding and Transmission Standard) for all types of files (transcriptions, images etc.) in order to identify easily information and edition components as well as to understand better their mutual relationships,[61] text-image linking, as well as hyperlinking (desirable, not essential) and variant readings. If, however, the project is not a *variorum* edition, variant readings should at least be mentioned or referred to. If variants are used, these should be indicated as such by the authors.[62] Finally, the project should clearly state the type of licence the work is released under, not only as a means of stressing ownership but, more importantly, of telling users the extent to which they are allowed to repurpose content. Findings from this analysis will go onto inform the production Greta Franzini's diplomatic edition of St. Augustine's *De Civitate Dei*.

---

58 Carlquist, 'Medieval Manuscripts, Hypertext and Reading'.

59 Edward Vanhoutte, 'Every Reader his own Bibliographer—An Absurdity?', in *Text Editing, Print and the Digital World*, ed. Marilyn Deegan and Kathryn Sutherland (Farnham: Ashgate, 2009), pp. 99–110 (p. 109).

60 The *World Wide Web Consortium*, http://www.w3.org

61 The *Metadata Encoding and Transmission Standard*, http://www.loc.gov/standards/mets

62 Robinson, 'Where We Are with Electronic Scholarly Editions'.

Conclusion

The ultimate aim of our catalogue of digital editions is to record extant digital editions of texts and their features and functionality, helping us to establish an overview of past and current practice in the creation of digital editions and draw up a digital edition best practice profile. In building the catalogue this way, data is centralised and systematically organised into a unique bank which could be used in other studies and is useful to the community. While initially curated by Greta Franzini, a larger group of administrators are now carrying out regular updates, ensuring accurate data, providing support and broader outreach while integrating and populating the database on an ongoing basis. The catalogue itself raises issues about the nature of digital editions and the relationship of digital editions to source texts, institutions, and funding structures, while encouraging us to pause and establish the best way in which to support the user experience when engaging with such digital content. Trends emerge regarding the types of manuscripts which are supported and explored in digital editions, and we can begin to understand the larger issues that direct the work of the Digital Humanities community when they undertake a scholarly digital edition project. It is only by thoroughly cataloguing and analysing the hundreds of digital editions that now exist that we can understand and question the scope of the field, spot technical and procedural trends, and make recommendations as to how best to build digital editions that will provide the information required by users and expected by the Digital Humanities community.