Old Books, New Books, and Digital Publishing

Elena Pierazzo, Peter Anthony Stokes

CHAPTER TWENTY-TWO

Old Books, New Books, and Digital Publishing

ELENA PIERAZZO (UNIVERSITÉ DE TOURS) AND PETER STOKES (ÉCOLE PRATIQUE DES HAUTES ÉTUDES – UNIVERSITÉ PARIS SCIENCES ET LETTRES)

Digital publishing is a big topic that means different things to different people, and the Digital Humanities has an important theoretical and critical contribution to offer. An important part of this involves reflecting on how the digitization of old books and documents is done and what is changing in the way we read and use these digitized books and editions; and, on the other hand, on the way we digitize scholarly practices and methods. In a previous contribution (Blanke et al. 2014) we concentrated our attention on the aspects of modelization, standardization, and data infrastructure. While some of these aspects are still very central to our discourse (and in fact we will return to them shortly), other aspects have emerged as also being fundamental, such as the sharing and publication of research data as scholarship as well as or instead of polished websites, articles, or monographs, and all the questions that this entails. Digital publications are models, insofar as they are (more or less) conscious selections and representations of certain elements. The digitized object may be presented on a screen as text (a sequence of characters) or as images of the physical object, but it may also comprise data communicated through application programming interfaces (APIs), in mashups and metadata. It follows that the digital object both embeds and is the result of scholarly practice, and therefore that academic output now extends to computer software (code) and even platforms for presentation, archiving and "informal" publications such as blogs and social networks. We therefore argue for a continuity between the digitization of a medieval manuscript, the sharing of the metadata, and the publication of gray literature, namely the so-called "mesotext" (Boot 2009). However, this position raises further questions about academic practices, such as how to publish, and the status of these activities and their outputs. Practices and methods that come from other disciplines are being used in new contexts; the speed of scholarship is also changing, not to mention the context in which it is practiced. Nevertheless, in some circles the publication of an article in an online-only journal is still frowned upon, let alone the recognition of data or code as valid (Digital) Humanities scholarship deserving of academic recognition. The use of these new practices, including new publication formats, should not compromise the quality of discussion, but these transformations are profound and they *can* lead to such problems. The question for us is therefore how to understand and use these changes in a positive and productive way, in which scholarly rigor and the advancement of knowledge still remain at the center of our work. The Digital Humanities have been defined as "project-based scholarship" (Budrick et al. 2012), a

Pierazzo, Elena, and Peter A. Stokes. 2022. "Old Books, New Books and Digital Publishing." In The Bloomsbury Handbook to the Digital Humanities, edited by James O'Sullivan, 233–44. London: Bloomsbury Academic. DOI: 10.5040/9781350232143.ch-22.

definition that indicates a change in the heuristics of the research itself, as well as the implication that such research can be sustained only for the duration of a (funded) project. We still lack the distance to judge this evolution in full, but we can and must continue to survey its impact and the relevance of these changes. A full analysis is of course impossible in a short contribution like this, but in the discussion that follows we will attempt to raise some of the key points and provide some indication of where the issues lie.

DIGITAL PUBLISHING OF OBJECTS FROM THE PAST

Modeling objects, texts, documents, and books

The digitization and publishing of historical books may seem simple: one only needs an appropriate camera and setup, one takes better-than-life photographs of the pages, and presto, the work is done. However, the reality is of course very much more complex than this. Among other things, digitizing requires first establishing what we want to digitize (the book? the text?), then deciding for whom we are digitizing it (scholars? of which disciplines?), how this digitization will be published and accessed (on an existing platform? from a home computer or mobile device?), and for how long this digitization will be needed (for a quick one-off reading? the foreseeable future?). All these questions and others demand answers, particularly when the object is fragile or restricted and may not be available for digitization again.

In the past two decades, solid guidelines and recommendations have been developed for the creation of digital surrogates able to serve different purposes and to last for as long as one can reasonably expect. Metadata standards and protocols now allow for much wider and easier sharing of digitized images than was imaginable only a few years ago: stable and open standards help to ensure longevity and interoperability of images, while protocols for sharing and archiving are helping findability and interoperability. Despite this progress, obstacles still remain, and technical improvements cannot address the need for scholarly reflection around the digitization of historical objects. In particular, the large-scale and easy availability of digital surrogates has opened the door to theoretical discussions around the nature of these representations with respect to the original, as well as about the different uses and cognitive experiences that they provide. For instance, recent publications have brought attention to the need of critically reflecting on the modeling of the objects that we digitize (Pierazzo 2015; Eggert 2019; Flanders and Jannidis 2019). Following McCarty, modeling in this sense can be defined as "the heuristic process of constructing and manipulating models," where a "model" is "take[n] to be either a representation of something for purposes of study, or a design for realizing something new" (2005, 24). McCarty has added that models "are by nature simplified and therefore fictional or idealized representations" of real objects (24). The process of modeling is also an epistemological process: when modeling an object, through this activity we come to know this object, and this activity. Modeling is therefore interpretative, and it is required both to better know the object we model, and to build something new (the object we publish, for instance). By asking which are the main features of books, texts, and works that we need to preserve for their digital publication we investigate the object and we anticipate what the user will do with the digitized resource. In other words, our interpretation of the physical object is reflected in the digital object we publish and will influence how the resulting publication can be used.

In the domain of texts and books, the scholarly community has produced a large number of models, the two best-known being perhaps the TEI and FRBR.[1] The former is focused primarily on the text (hence *Text* Encoding Initiative) and the latter on the relationship between published books as objects in a collection and the content of these books, namely the works produced by authors. As is the case for all models, the two mentioned here each have strengths and shortcomings, and their widespread use should not be confused with their fitness for the purpose for which they have been used. Nevertheless, their usefulness has been clearly demonstrated in practice. A model, even when incomplete and partial, is also a cognitive tool and the iterative nature of the modeling activity is a powerful heuristic device, as demonstrated by the rich literature surrounding these (and other) models (McCarty 2005; Flanders and Jannidis 2019).

Conceptual models should ideally be independent of the technologies used to implement them, but in practice this is rarely the case. The critical apparatus, which can be considered a model for representing textual transmission, has been heavily shaped by the constraints of the printed book: the scarcity of space and the dimension of the page have led to the invention of a system of conventions that are economic and synthetic but hard to grasp for the uninitiated. The now familiar, if not standard, format for the publication of the digital scholarly editions sees the edited text on one side of the screen and the facsimile image on the other. This may seem "natural" but is a result of technological factors, namely the rectangular dimension and "landscape" orientation of the standard computer screen, and this dependency is demonstrated by how poorly it works on mobile devices, for instance. Another example is the method of photographing old books which usually relies on cradles able to flatten pages even for the tightest bindings. This leads to a digitization which proceeds page by page, rather than by two-page opening which is in fact how the original object is seen and used. It is important here to note, then, how many of the features of print and digital publishing which we consider "natural" or which represent scholarly pillars are indeed the results of compromises due to technological constraints.

Multifold Editions

While digital editions are slowly making their way into the academic community, print editions are still the most likely outlet for textual scholarship. This is partly because publishing houses provide not only professional support in the production of the edition but also because a printed edition is stable and more easily accepted as a scholarly output. These considerations have led several digital editorial projects to provide multifold embodiments of their work, that is, publishing the "same" work in different formats such as print and perhaps several different digital forms. This tendency has now become almost a given for any form of new book publication, with publishing houses offering the same "book" as printed objects, print-on-screen (normally PDF) and ePub; digital scholarly editions tend to add the source files as well, often in the form of TEI XML as mentioned above. In addition to these more or less "official" publications, one often finds pre-prints deposited in various open archives, or earlier versions of editions that have been later updated. In fact, one of the most striking features of digital publications is that they can be easily and seamlessly updated at any moment, and while some editions provide access to previous versions for the sake of transparency and citability, most of them do not, leaving this task to tools like the Way Back Machine offered by the Internet Archive.[2] This proliferation of formats does not apply only to digital editions of old books, but is found in all sorts of digital publication. In fact it is complicated

and magnified by the fact that all these embodiments can potentially be multiplied over and over, since in practice people can often easily download and republish the content (legally or not), but also thanks to the fact that search engines often provide snippets of content as results of users' queries, decontextualizing and recontextualizing the content each time, a point we will come back to shortly.

Editions as Data, Metadata, and APIs

As suggested here, a digital edition is more than just a representation of a textual object from the past that is shaped by the scholarship of its editor. While printed editions are most easily treated one at a time, and at close inspection, the digital environment allows (and almost invites) a wider set of usages. In a seminal contribution of 2005, Franco Moretti has introduced the concept of "distant reading," namely the possibility of mining large quantities of texts in order to study phenomena that escape close investigation. Beside being texts to be read, digital editions can also be used as data to be investigated and queried, potentially alongside other data from elsewhere. This activity goes well beyond "counting words," a simplistic label that has sometimes been used to define these computational approaches, extending instead to disciplines such as stylometry and authorship attribution. The call to publish the "raw" data that underlies digital editions (for instance the XML source) alongside the finalized edited versions is more current than ever, in a moment where the sharing and publishing of the data and metadata of one's resource are becoming increasingly normal and even required. Even scholars that were afraid to let people look into their encoded texts a few years ago are now publishing their pre-prints and preliminary ideas in institutional or private repositories (see infra) or on blogs, a practice that shows how much our culture has changed in a very short span of time. The availability of source files (XML or otherwise), as well as that of metadata (Dublin Core, METS or other formats) implies also that expectations about the deliverables of editions have changed as well.[3] The publication of FAIR principles and their recommendations concerning metadata (Wilkinson et al. 2016) is changing radically the face of digital publishing of cultural heritage texts, where the availability of standardized data and metadata is in many ways more important than the publication of the edited text within a self-contained, dedicated website.[4]

This principle of "editions as data" can be taken even further through application programming interfaces (APIs). The concept here is to publish the content in a format that is adapted not for people, but rather for other software and computers. It is then up to others to produce the necessary tools for interacting with the content, be that an interface for (human) reading, or any number of other possibilities such as distant reading, stylometry, mashups, and so on. Such an approach is very flexible and potentially very powerful, but it is also a very different and sometimes unsettling form of publication. Principles that seem essential to the very foundation of scholarly publishing are quickly undermined, citation being just one example of this. This question is helped in part by initiatives such as the Canonical Text Service (CTS), which defines a minimal protocol for requesting and therefore referring to specific parts of texts (including sections, chapters, words, and so on).[5] In this way, the text can be published not (or not only) as a complete entity, but (also) as fragments, which can then be republished in new interfaces, combined with other texts from elsewhere, or presented not as text at all but as graphs, charts, or other analyses. The principle of APIs is of course not limited to CTS but is very widespread in the informatics community and is becoming increasingly so also for the Digital Humanities. In particular, another important

standard for publication is the International Image Interoperability Framework, or IIIF.[6] As the name suggests, IIIF is primarily focused on the image rather than the text, and so is often (but by no means only) used as a means of publishing images of books and documents. Repositories can publish their content via the IIIF standard, and this content can then be collected automatically by software and treated in some way, for instance shown to the user. A number of different viewers already exist for IIIF content, meaning that one can very easily work with images of books in different environments with different software.

Publishing Datasets and Code

An increasingly important area of change in publishing for the Digital Humanities is the publication not only of final articles and monographs but also the raw data, code, images, and other research outputs that make up the whole. This is important for several reasons: on the one hand, the ideal at least of research is that work should be as transparent and reproducible as possible, so that others can understand exactly what was done, verify the methods and data, and so on. Indeed, this question of the scientific method applied to at least some parts of the humanities has been discussed for well over a century, with palaeography and philology being two examples among many (Derolez 2003, 1–10; Canart 2006; Stokes 2015). Although many aspects of humanities research cannot be quantified and are not meaningfully reproducible, nevertheless it is an ideal that is often sought where possible, and this implies that conclusions based on digital analyses should be supported by publishing not only the article or monograph but also all of the code, data, and the precise steps and parameters used to produce the results. This is already a requirement in at least some branches of the "hard" sciences and is being discussed more and more in computer science as well as in the Digital Humanities.[7]

The line between code and article is also blurred in the principle of "literate programming," which was first proposed by Donald Knuth (1984) but which did not find widespread use until the last decade or so. The principle here is to reverse the normal paradigm in programming: rather than centering the code and adding occasional comments to the users to explain the content, the principle of literate programming is instead to focus on the explanatory text and then embed the code into this. This means that the human thought process takes priority over the machine code structure, and Knuth argued that this in turn would help support intelligibility and transparency in coding. Probably the most common form of this at present, at least for Digital Humanities and data science, is Jupyter Notebooks which allows one to include both formatted text and lines of code, mixing the two in a way that allows one to produce fully functional software embedded in the context of a written discussion.[8] This hybrid form of publication is still not used for software of any complexity but is extremely widespread for discussions particularly of data analysis as well as other uses in Digital Humanities and elsewhere, so much so that a search of GitHub at the time of writing returns close to nine million examples of Jupyter Notebooks.[9]

Despite this very widespread use of notebooks, blogs, and other less formal publication formats, scholars can still be reluctant to publish their raw research data, for various reasons including that they may be concerned that others will use the data to publish before the original authors, or even that the data is "too messy" for publication. Indeed, many questions arise here, one of which is what should be published, and indeed what even should be considered as "research data," particularly for the humanities. Few researchers in the humanities would consider publishing their

private notes taken during a visit to the library, for instance, and in general this seems reasonable as publication is traditionally reserved for the final analysis rather than working notes. However, as we have seen, the role of publishing and indeed the notion of finality in research is changing, such that the question of what should be published becomes less and less clear.

SCHOLARLY INFRASTRUCTURE FOR DIGITAL PUBLICATIONS

The changing nature of scholarly publication discussed above raises a number of practical and theoretical questions about the nature and format of publication, and the impact of this on scholarship and vice versa. Furthermore, publication is at the core of academic practice and evaluation, and so these changes necessarily have an impact on scholarship and how it is carried out. Digital publishing of scholarly outputs, whether digital editions of books of the past or new contributions, is still struggling to find recognition in the appropriate academic venues, particularly when it comes to early career scholars. Digital publications can be tainted with being self- or, worse, vanity publications, since in many cases they do not go through the filter and quality checks that characterize print. To publish, it is enough to have an Internet connection and an account in some publishing service or have some server space. From here, it is easy to generalize, thinking that all digital publications are just the result of a self-evaluation, or of our inner circles. Even project websites hosted by universities do not escape this logic, in the sense that while securing some funding requires going through sometimes extremely strict processes, the publication of the website is in a sense left to the care of the project teams and their internal editorial workflow with little or no external verification.

The challenges of establishing scholarly infrastructures for digital scholarly output go well beyond the creation of an equivalent to peer review, since academic trust and quality require a series of facilities like the possibility of citing texts which do not disappear in the meantime, of verifying who did what and what one can do with the outputs that are offered online, all points that will be discussed here.

Existing Infrastructures

One of the many challenges in digital publication is the degree to which standard frameworks can be used for publication. In some ways this "prêt-à-porter" model would be ideal, since it would enable scholars and indeed publishers to publish their material with a relatively modest investment in terms of funding, programming expertise, and so on.[10] On the other hand, practice has shown that this is very difficult to achieve, since editions and indeed digital publications in general are very different, with different needs, requirements, and data models as discussed above, and this means that most editions take a "haute couture" approach of custom-made software. This has implications for cost, as discussed above, and sustainability, since as a general principle that the more a publication is based on ad hoc data formats and software, the more difficult it is to sustain in the longer term. It is also becoming increasingly clear that the existing body of highly specialized and idiosyncratic publications of the last twenty years is no longer sustainable, and publishers (including academic centers) are increasingly being forced to make difficult decisions around closing down publications even when in some cases they are still important and in regular use (Smithies et al. 2019). Once again, the problem is somewhat easier to manage if one considers only the data, but even here significant difficulties arise, for instance around who should bear the responsibility of maintaining this content.

Many attempts have been made to balance this tension between "prêt-à-porter" and "haute couture," and in practice one can now go a relatively long way towards standard frameworks, with sophisticated systems that do most of the work with standard templates and can then be customized for the final details (examples of which include the TEI Publisher, the Edition Visualisation Technology, and TEI plugins for Omeka and Omeka S, among others) (Pierazzo 2019). Another approach is to focus on the "publication as data." This approach is more limited but is very much more manageable in practice. Here, rather than attempting to preserve the full interface and "experience" of the publication, the decision is instead to focus on preserving the data and perhaps the accompanying code, but not in a way that users can interact with directly. This is becoming increasingly easy to achieve, given the number of infrastructures for scientific data which are being used as archival and publication formats for the humanities. In a European context, the best-known example is probably Zenodo, which is supported by CERN, the OpenAire initiative and the European Commission's research program. Described as a "catch-all repository for EC funded research,"[11] Zenodo is specifically aimed to encourage "Open Science" by providing near-unlimited storage for researchers regardless of their institution, country, or source of funding. Researchers are explicitly encouraged to submit "data, software and other artifacts in support of publications," but also other content such as "the materials associated with the conferences, projects or the institutions themselves, all of which are necessary to understand the scholarly process." Upon submission, a dataset is automatically given a stable identifier (a DOI), and updates to the dataset are automatically versioned and given new identifiers in order to ensure long-term citability and stability. Although "Science" is here used in the European sense of scholarly research of any type (including the humanities), the reality is nevertheless that Zenodo had its origins in the "hard" sciences, as revealed by its basis in CERN, but use by other researchers is nevertheless rapidly increasing.

A different approach is provided by GitHub which is a private company that has been used for some time now to publish code and, increasingly, other forms of data and even text-based publications.[12] The Git model explicitly encourages the principle of "commit [or publish] early, commit often," an approach that stands strongly in contrast to the traditional approach of publishing only a "final" version after a long period of careful reflection, writing, and checking. Indeed, an interesting example here is the parallel operation of GitHub (and GitLab), Zenodo, and the Software Heritage Archive. Summarizing somewhat crudely, the model is that one more or less continually publishes to GitHub or an instance of GitLab, with an implicit expectation but no guarantee that this will remain available for the foreseeable future. At specific moments, when the authors decide that the content has reached a sufficiently stable point, it can be automatically harvested by Zenodo where it receives a DOI and is published in an archived form for the long term. At the same time, all the small updates that are made to GitHub are also automatically harvested by the Software Heritage Archive, where again they are assigned long-term identifiers which can be resolved to URLs in much the same way as DOIs. This then results in three separate copies of the material, hosted on three distinct infrastructures, two of which have permanent identifiers and are designed along FAIR and archival principles, and the third intended more for day-to-day use. This seems to bring the advantage of three different forms of archiving and publication, for three different purposes, as well as the improved likelihood of at least one of these remaining available. However, it in turn poses further questions such as which version, if any, is "the" definitive one.

Questions arise here also around the use of public or private services for publication and archiving. On the one hand, one may wish to avoid private companies due to concerns around the

control and ownership of data, the guarantee that data will remain freely accessible for research, and the perception that public money means at least theoretical accountability to the public and more specifically to voters, as opposed to the very large technology companies which increasingly give the impression of being accountable to no one. On the other hand, many custom-built local setups lead to fragmentation of data, the difficulty of finding content, and the risk of having to duplicate content to ensure visibility or even compliance with different and sometimes conflicting requirements of funding bodies, home institutions, and so on. The question of sustainability is more complex: even very large multinational corporations can still go bankrupt or (perhaps more likely) decide that a given service is no longer profitable and so terminate it at short notice. However, much the same can be observed in public institutions, as funding priorities and indeed governments change over time, such as the Arts and Humanities Data Service in the United Kingdom which began in 1996 but stopped receiving deposits in 2008 and was decommissioned entirely in 2017.[13]

Peer Review

Peer review is sometimes considered one of the pillars of scholarship: the fact that the arguments and findings of a scholar or a group of scholars must be validated by their own peers has been in place since the seventeenth century and constitutes arguably the single most important method for building trust and knowledge. However, digital publications have challenged this assumption in many ways, arguing that digital scholarship does not need it, or that it should be substituted by commentaries, or by online social practices, or proposing new, open, and more transparent forms of validation and so on. Borgman (2007, 84) has broadened the question by discussing how legitimization is built in digital form and maintains that while the traditional forms of validation ("authority, quality control, certification, registration in the scholarly priority and trustworthiness") remain substantially the same, "new technologies are enabling these functions to be accomplished in new ways." For example, in February 2015 Matthew Jockers (2015a) published a contribution on his blog maintaining that his toolset (called the "Syuzhet package") was able to analyze plots of novels and thanks to this he had been able to detect the existence of six or seven novel plot shapes; this claim was picked up by several online journals and made a bit of a sensation. However, weeks after this publication, one scholar, Annie Swafford, started to point out several issues with the package; several other scholars followed her lead and by the beginning of April, Jockers declared the "requiem" of this part of the package (2015b). As pointed out by Jockers himself, it is worthwhile noting how these flaws were discovered within just a few weeks, after more than daily online contributions on blog posts and tweets, while offline such discussion would have taken "years to unfold." Digital scholarship can indeed be fast, and scholarly validation can be reached within a few weeks and a few hundred tweets, a fact that might prompt one to suggest that formal peer review is finished.

The alleged obsolescence of peer review is based on considerations that it was invented in times of scarcity, namely when the limited physical space in scholarly journals forced one to choose what to print. It can then be argued that, once those constraints are removed, there is no reason why contributions should be rejected, since scholarly contributions will go through a natural selection, with good contributions being cited and used, while bad contributions will be substantially ignored. Ford (2008) reflects on the mechanism of building scholarly authoritativeness and concludes that this is achieved through a form of accountability that is established on a disciplinary basis, namely that not all disciplines build their trustworthiness in the same ways. This process is indeed multifold,

in the humanities as in any other sector, but for the most part it relies on the way we "allow" some content to be present in the scholarly discourse. The publications of gray literature in various types of disciplinary, national, or even private repositories are challenging this assumption (for which see above), since what we "allow" to be published here is indeed "whatever their authors think fit." However, while many scholars are equipped to evaluate the quality of publications, official or not, this may not apply to students or less experienced researchers, and least of all by those with no expertise in the field, with the risk that unfinished or "half-cooked" material may be cited and trusted beyond its worth. These considerations have and still are fueling a general distrust of online publications, with the results that some people may see their career expectations being jeopardized by the fact of having published online.

The scholarly community has reacted to the issues in different ways. Several professional organizations such as the MLA (2012) have published series of guidelines for assessors and career panelists on how to evaluate online resources. In France, the National Scientific Research Council (CNRS) has sponsored several tools to improve the quality of online publications, favoring the reaction of digital and hybrid publishing houses, a hub of open access online journals and a national pre-print repository now used for the assessment of careers and research centers, forcing the latter then to self-regulate.[14] Some other professional organizations have developed forms of accreditation such as badges that can be added to websites, or lists of approved or otherwise certified sites, examples here including the Medieval Electronic Scholarly Alliance (MESA) and NINES projects, or the Medieval Academy of America's database of digital resources, which includes publication of the standards for evaluation and criteria for inclusion.[15] The situation is changing rapidly, then, but a long road still remains.

Citation, Credit, and Intellectual Property

One important question that arises here is how to properly manage credit and attribution, particularly when so many different people are contributing to publications in many different ways and at different levels. One may well argue that this is not a new problem: after all, we are increasingly recognizing that the print model favors the "headline" author but ignores entirely the contribution of all the other people who have also influenced and sometimes even directly written the text, such as the editors, copyeditors, typesetters, publishers, research assistants, librarians and archivists, artists or draftspeople, and more. Nevertheless, the fact remains that better systems for credit and attribution are necessary, particularly since an increasing number of early-career positions are focused on producing this source of content for digital projects, and it is precisely these people for whom proper attribution and visibility is the most important. It is therefore imperative that everyone but especially those in senior positions do all that is possible to ensure that credit is properly given, both within projects and when external content is being harvested and re-used in any way. Doing so is not as straightforward as one might hope, however: for instance, proper crediting and attribution also requires proper citability, since the one is impossible without the other. As we have seen above, citing texts is complicated when the text itself is fluid, and although systems such as CTS are significant contributions here, many difficulties remain including not only that these systems need to be implemented in practice. Similarly complex are questions of intellectual property, copyright, and licensing. On the one hand, it may be very unclear who owns the intellectual property, and this can be a significant barrier to the forms of publication discussed here. The principle is becoming increasingly widespread that cultural heritage and even information should be free to all, and one

can certainly dispute the practices of some publishing houses which have fees and revenues that seem disproportionate to their services. Indeed, the increasing role of open access publishing has been implicit throughout this chapter but not expressly addressed. However, publishers continue to play a very important role and are likely to do so for the foreseeable future, since authors normally have neither the desire nor the aptitude to fulfill these tasks, and their labor must of course be paid for like all others. How to manage this in a world of constant re-use, re-publication, mashups, and so on is by no means clear.

Citation, credit, and intellectual property are all complicated by these new forms of publication. If a transcript is prepared by one person in one project, then elaborated with different layers of markup potentially by different people in different projects, then integrated into a collection by someone else and finally processed and analyzed by someone else, then properly crediting all these people requires significant care and effort and assumes that this history of use and re-use is properly documented and available, something that is by no means always the case. In principle, one can imagine some form of tracking, whereby authorship in its different senses is labeled in a fine-grained way and re-use and re-publication of data can thereby be followed, quantified, and rewarded, whether that reward is in terms of payment, career progression, citation metrics, or so on. The technology here would be relatively straightforward, but the details are complex and potentially very problematic, so a great deal of thought and care would be required.

CONCLUSIONS

The discussion here touches on just some of the important issues that have arisen due to the changing nature of publication. Many others remain which are no less pressing or difficult and which can only be hinted at here. Overall, then, the changing nature of publication has in many ways moved the center of authority and control, since publishers and authors can no longer predict the forms in which their material will be accessed. This situation has led Vitali-Rosati to the elaboration of the theory of "editorialisation" (2018), in which he claims that authors and editors in a digital space have ceased to be central to the dissemination of knowledge and have been replaced by APIs and content management systems. While this conclusion may feel a bit extreme, it is certainly true that nobody can predict the way any given content, including a digital scholarly edition, will be accessed and consumed, and if this could be true to some extent also for print publications, the digital format has taken this affordance very much further. The many lives and embodiments of a text prove its relevance and impact, for instance, but the risk of some or all of one's publication being taken out of context and bent to new, unexpected meanings is indeed something that can make a scholar very uncomfortable. It is clear that we need to reflect on this and indeed all these points, and potentially to take action if we want to have a role in the publication of our content, and in the forms of dissemination and use of the work of the editor and the edited texts.

NOTES

1. The Functional Requirement for Bibliographic Records (FRBR) is a model developed by the International Federation of Library Associations and Institutions (IFLA) which is intended to describe and facilitate the retrieval of bibliographic objects. Currently, the last version available from the IFLA website is dated 2008: see http://www.ifla.org/VII/s13/frbr/. For the TEI, see further http://www.tei-c.org.

2. See https://web.archive.org/.
3. The set of metadata developed by the Dublin Core Metadata Initiative is perhaps the most used format of metadata of the web: see https://dublincore.org/. The Metadata Encoding & Transmission Standard (METS) is one of many metadata standards developed, distributed and maintained by the Library of Congress: see https://www.loc.gov/standards/mets/.
4. The "FAIR" principles state that data should be Findable, Accessible, Interoperable and Reusable: see, for example, https://www.go-fair.org/fair-principles/.
5. The formal specification for the CTS URN is available at https://cite-architecture.github.io/ctsurn_spec/. Similar to the CTS is the Distributed Text Service, DTS: the details are not relevant to this discussion, but for more see https://distributed-text-services.github.io/specifications/.
6. See http://iiif.io/.
7. One very broad example is the OpenAIRE initiative (https://www.openaire.eu), and for a very specific one see OCR-d (https://ocr-d.de).
8. See further Jupyter, https://jupyter.org.
9. This figure is obtained by searching GitHub for files with extension:ipynb (https://github.com/search?&q=extension%3Aipynb, but note that this search requires a GitHub account). A search for repositories containing "ipynb" where the language is "Jupyter Notebook" gives approximately 2,600 results (https://github.com/search?l=Jupyter±Notebook&q=ipynb&type=Repositories); this search is open to those without accounts but gives only repositories (collections of files), not individual notebooks.
10. The first part of this section, including the terms "prêt-à-porter" and "haute couture," draws heavily on Pierazzo (2019).
11. This and the other citations on Zenodo that follow are from https://about.zenodo.org.
12. Summarizing somewhat crudely, git is the name of software which can be used to manage distributed copies of files; GitHub is a commercial service and website which can be used to host and publish content managed via git. Also of interest is GitLab, another commercial service which is very similar to GitHub but can be used either as a service or installed on an institutional or other server. See further https://git-scm.com, https://github.com and https://about.gitlab.com.
13. The former site is http://www.ahds.ac.uk, which is currently occupied by a placeholder page. For the original plan of the AHDS see Greenstein and Trant (1996), and for general process of decommissioning, applied to a significant number of sites, see Smithies et al. (2019).
14. Open Edition is a publicly funded digital publishing infrastructure for scholarship; it publishes both Freemium and Open Access books as well as a large number of peer-reviewed journals (https://www.openedition.org). HAL-Archives Ouvertes is a nationwide, pluridisciplinary open archive for all sorts of scholarly outputs (https://hal.archives-ouvertes.fr/).
15. See https://mesa-medieval.org, https://nines.org and http://mdr-maa.org, respectively.

REFERENCES

Blanke, Tobias, E. Pierazzo, and P. A. Stokes. 2014. "Digital Publishing Seen from the Digital Humanities." *Logos* 25: 16–27. https://doi.org/10.1163/1878-4712-11112041.

Boot, Peter. 2009, *Mesotext: Digitised Emblems, Modelled Annotations and Humanities Scholarship*. Amsterdam: Pallas Publications.

Borgman, Christine L. 2007. *Scholarship in the Digital* Age: *Information, Infrastructure and the Internet*. Cambridge, MA: MIT Press.

Budrick, Anna, Johanna Drucker, Peter Lunenfeld, Todd Presner and Jeffrey Schnapp. 2012. *Digital Humanities*. Cambridge, MA: MIT Press.

Canart, Paul. 2006. "La Paléographie est-elle un art ou une science?" *Scriptorium* 60 (2): 159–85.

Derolez, Albert. 2003. *The Palaeography of Gothic Manuscript Books from the Twelfth to the Early Sixteenth Century*. Cambridge Studies in Palaeography and Codicology, 9. Cambridge: Cambridge University Press.

Eggert, Paul. 2019. *The Work and the Reader in Literary Studies: Scholarly Editing and Book History*. Cambridge: Cambridge University Press.
Flanders, Julia and Fotis Jannidis, eds. 2019. *The Shape of Data in Digital Humanities: Modeling Texts and Text-Based Resources*. London: Routledge.
Ford, Michael. 2008. "Disciplinary Authority and Accountability in Scientific Practice and Learning." *Science Education* 92 (3): 404–23. https://doi.org/10.1002/sce.20263.
Greenstein, Daniel and Jennifer Trant. 1996. "Arts and Humanities Data Service." *Computers & Texts* 13. http://users.ox.ac.uk/~ctitext2/publish/comtxt/ct13/ahds.html.
Jockers, Matthew L. 2015a. "Revealing Sentiment and Plot Arcs with the Syuzhet Package," "Some Thoughts on Annie's Thoughts … about Syuzhet," "Is that your Syuzhet Ringing?" https://www.matthewjockers.net//?s=Syuzhet.
Jockers, Matthew L. 2015b. "Requiem for a Low Pass Filter." https://www.matthewjockers.net/2015/04/06/epilogue/.
Knuth, D. E. 1984. "Literate Programming." *The Computer Journal* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.
McCarty, Willard. 2005. *Humanities Computing*. Basingstoke: Palgrave Macmillan.
MLA (The Modern Language Association). 2012. "Guidelines for Evaluating Work in Digital Humanities and Digital Media." https://www.mla.org/About-Us/Governance/Committees/Committee-Listings/Professional-Issues/Committee-on-Information-Technology/Guidelines-for-Evaluating-Work-in-Digital-Humanities-and-Digital-Media.
Moretti, Franco. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso. http://www.loc.gov/catdir/toc/ecip0514/2005017437.html.
Pierazzo, Elena. 2015. *Digital Scholarly Editing. Theories, Models and Methods*. London: Routledge.
Pierazzo, Elena. 2019. "What Future for Digital Scholarly Editions? From Haute Couture to Prêt-à-Porter." *International Journal of Digital Humanities* 1: 209–20. https://doi.org/10.1007/s42803-019-00019-3.
Smithies, James, Carina Westling, Anna-Maria Sichani, Pam Mellen, and Arianna Ciula. 2019. "Managing 100 Digital Humanities Projects: Digital Scholarship and Archiving in King's Digital Lab." *Digital Humanities Quarterly* 13 (1). http://www.digitalhumanities.org/dhq/vol/13/1/000411/000411.html.
Stokes, Peter A. 2015. "Digital Approaches to Palaeography and Book History: Some Challenges, Present and Future." *Frontiers in Digital Humanities* 2 (5). https://doi.org/10.3389/fdigh.2015.00005.
Swafford, Annie. 2015. "Problems with the Syuzhet Package," "Continuing the Syuzhet Discussion," and "Why Syuzhet Doesn't Work and How We Know." *Anglophile in America: Annie Swafford's Blog*. https://annieswafford.wordpress.com/category/syuzhet/.
Vitali-Rosati, Marcello. 2018. *On Editorialization: Structuring Space and Authority in the Digital Age*. Amsterdam: Institute of Network Cultures. https://papyrus.bib.umontreal.ca/xmlui/handle/1866/19868.
Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 160018. https://doi.org/10.1038/sdata.2016.18.