8.

Re-using data from editions

Elena Spadini and José Luis Losada Palenzuela

Re-using content, re-using data: new forms of an old practice

Using the content of an edition in scientific research is not something new.[1] For centuries, historians, literary scholars, philosophers, to name just a few, have used scholarly editions to access the sources on top of which to build scholarship. Thus, it won't be a surprise that in the survey by Franzini et al. (2019) data re-use appears among the most requested features of editions;[2] or that in the MLA *Statement on the Scholarly Edition in the Digital Age* (2016) re-use is mentioned as an elemental character of the edition.[3] The difference, however,

1 Parts of this chapter were written in the context of the following projects: 'El teatro áureo en colaboración: textos, autorías, ámbitos literarios de sociabilidad y nuevos instrumentos de investigación (tac)', PID2020-117749GB-C22 and 'Computational Text Reuse Detection in Literary Texts', BPIDUB.29.2022.

2 '[O]ne way to alleviate the negative sense of frustration conveyed by these user responses might be to reconcile data reuse, licensing, image availability, and comprehensive documentation – the four most requested features – to the extent possible and to more clearly state motivations, objectives, and intended audience' Franzini et al. (2019).

3 '[W]here possible, it [the edition] should attend to possibilities of sampling, reuse, and remix, supporting approaches to the formation and curation of the edition such as reconstructing and documenting instances of texts and textual change over time, like algorithmic construction and reconstruction (with possible extensibility, including external data)' MLA Committee on Scholarly Editions (2016).

between these remarks and the long-standing use of editions content is to be found in the changed medium: re-use takes a different aspect when the user is not only a scholar, who extracts knowledge from the edition, but also a machine programmed by a scholar, that further processes (analyses, transforms, merges and so on) existing data.

In a digital paradigm, the re-use of editions data is enabled by good practices of research data management governed by the FAIR principles.[4] The last of them, the letter R in FAIR, indicates that data must be 'Re-usable'. The reasons why editors should care about FAIRness is summarised by Susanna-Assunta Sansone and Barend Mons: enabling re-use could 'facilitate data sharing and collaborations; increase the visibility of research and can lead to more citations; improve the transparency, reliability, and replication of research; prevent data loss. And thereby: maximise potential from data assets; maximise research impact' (Deutz et al. 2020). In line with this statement, everyone would probably agree that stories of re-use are stories of success. A fortunate edition is, for example, *The Proceedings of the Old Bailey*, listing on their website not less than 15 projects and resources that re-use their data (Emsley et al. 2018).

In what follows we study data re-use in scholarly editing, providing insights into the current panorama and imagining future developments. This chapter is not a state of the art on the topic but proposes concrete cases to exemplify re-use practices and a few suggestions to improve the re-usability of editions. We will focus on the re-use of data, leaving aside as much as possible the re-use of code and

4 FAIR principles suggest that data management should address 'Findability', 'Accessibility', 'Interoperability' and 'Reuse' (FAIR). Data are findable when they have a persistent identifier and when relevant metadata are exposed in search engines and research data catalogues. They are accessible if they are stored in appropriate repositories, if they can be retrieved using standard technical procedures and if there is documentation on how to retrieve them. They are interoperable if they can be exchanged and used in different applications and systems. They are re-usable if they conform to community standards and are well documented.

models, which would require a separate inquiry. As briefly mentioned above, a fundamental distinction in re-using print and digital editions is the type of consumption: human consumption, when the user extracts information from the data to be re-used in a noncomputational context, such as writing an article or taking unstructured notes for a dictionary; and machine consumption, when data is re-used in a computational context, for example for further annotation or compiling with external data sets. In this chapter, we will only pursue the second, that is machine-actionable re-use. Furthermore, the re-usability of editions data is to be considered within the framework of research data management; but an in-depth analysis of topics such as licensing and documentation are out of scope for this chapter.

Data re-use scenarios

The data of editions are potentially re-usable in many ways. To list some examples, the text of the edition (documentary, diplomatic, critical and so on) may be re-used for text analysis in literary, linguistics, historic research and other disciplines, as well as in the context of scholarly editing, for collation with new witnesses and inclusion in a larger corpus or as training data. The description of archival documents may be integrated into catalogues. The entities records may be re-used for prosopography and gazetteers.

The four re-use scenarios described below address some of these possibilities, discussing different types of data and purposes of re-use in concrete cases from real-life or fictitious projects.

Search multiple data sets with an authority record

Enriching data with references to authority records (such as VIAF, ISNI and the authority files provided by national libraries) is a common practice in digital scholarly editing. The following example shows how to make use of this additional information to link and re-use data from multiple data sets.

Our fictitious case study is a research project on the classicist Karl August Böttiger (1760–1835), for which we want to re-use existing data. We know that information about him can be found both in the *Carl-Maria-von-Weber-Gesamtausgabe. Digitale Edition* (*WeGA*)[5] and in *correspSearch*, a web service aggregating metadata of scholarly editions of letters (Dumont et al. 2021).[6] First, we look into*WeGA*. The Search functionality gives access to the biography of Karl August Böttiger, as well as to his GND entry. The *WeGA* API can then be used to export the biographical information for re-use:

```
curl -L -H "Accept: application/tei+xml" https://weber-gesamtausgabe.de/en/gnd/118824775
```

This query, performed using curl,[7] asks for the information related to a person identified through the GND number (in this case, "118824775") in the XML/TEI format (here, "application/tei+xml").[8] For reasons of space, we won't copy here the XML result, but the same is available at https://weber-gesamtausgabe.de/en/A000194.html#bs-tab-XMLPreview.

We then turn to *correspSearch* and its API. The same GND number can be used here to retrieve the data about Karl August Böttiger. The two data sets are not overlapping, even if the *WeGA* letters are indexed in *correspSearch*, because *WeGA* not only contains letters and *correspSearch* includes the correspondences of Böttiger with persons other than Carl Maria von Weber.

The *correspSearch* API query is the following:

5 *Complete Works of Carl Maria von Weber. Digital Edition*, http://weber-gesamtausgabe.de/A070006 (Version 4.6.1 of September 30, 2022).

6 *correspSearch. Briefeditionendurchsuchen und vernetzen,* https://correspsearch.net.

7 A command line tool for transferring data using URLs, https://curl.se/.

8 The API allows to retrieve the data in different formats (XML/TEI, JSON-LD, BEACON, HTML, TXT).

https://correspsearch.net/api/v1.1/tei-xml.xql?correspondent=http://d-nb.info/gnd/118824775

The query retrieves letters in which a person has the role of correspondent ("?correspondent=") and the person is identified through a GND URL ("http://d-nb.info/gnd/118824775").

In this example, we showed how persistent identifiers and authority records are key to retrieve data for re-use: the curated and rich data that an edition exposes are more difficult to re-use if the data is kept in silos instead of being connected to external structured knowledge. The link to authority records and the possibility to call them in APIs enables scholars to retrieve data from multiple sources using a single standard identifier, instead of a different internal identifier for each of the data sets.

Editions data in dictionaries

In this re-use scenario, we focus on the references to scholarly editions within dictionaries, leaving aside the presence of dictionaries or linking to them within editions.[9] The integration of scholarly editions and dictionaries, and especially historical dictionaries, is not something new nor bound to the digital medium. Although historical dictionaries are, and have been already for some decades, predominantly electronic (see, for example, the *Oxford English Dictionary*[10] and the *Tesoro della Lingua Italianadelle Origini*[11]), the scholarly editions referenced in them are almost only print publications.

9 On this second aspect, see for example *The Online Froissart* at https://www.dhi.ac.uk/onlinefroissart (Ainsworth and Croenen 2013) or *eBalzac* at https://www.ebalzac.com. From the technical point of view, an LOD- compliant solution is proposed in Tittel et al., 2018. For the connection of scholarly editions and linguistic resources, see Franzini, 2019.

10 *OED Online*. September 2022. Oxford University Press. https://www.oed.com/view/Entry/100528.

11 http://tlio.ovi.cnr.it/TLIO.

The reasons for this are certainly manifold, including the fact that digital scholarly editions, as other digital scientific outputs, are not yet fully integrated into the academic ecosystem.[12] There may also be technical reasons, since for many scholars digital editions are just websites, that is, ephemeral resources. Even when scholars are able to access the edition data, versioning (for which see below) might get in the way of stable referencing. In the previous example, we mentioned external persistent identifiers as an important component of a knowledge graph. Here too, persistent identifiers are central. Referencing a text section (a sentence, a paragraph or a word in context) within a dictionary is only possible if there is a way to identify it in a stable manner. The use of persistent identifiers is a step in the direction of FAIR data, addressing their Findability.

The Distributed Text Services (DTS) Specifications, inspired by the Canonical Text Service (for which see below), have been developed to this end,[13] as a standard way to access texts in XML/TEI. The DTS API provides three end points (collection, document and navigation), which allow you to reference texts at different scales, from a portion of a single document to an ensemble of documents. The DTS Specifications may prove very useful for referencing, from within a dictionary, the occurrence of a term in a scholarly edition.

DTS is meant to offer a standard solution for texts, very much as the International Image Interoperability Framework (IIIF) provides it for referencing images and their parts. Ad hoc solutions are also possible. An example is the references to the *Documents linguistiques galloromans* (Glessgen 2016) into the *Dictionnaire éty-*

12 The more or less standardised practice of peer review prior to publication, ensured by academic publishers, and review after publication are probably key factors here, as well as the difficulty to fully appreciate and evaluate digital research products by nondigital humanists.

13 'Publishers of digital text collections can use the DTS API to help them make their textual data Findable, Accessible, Interoperable and Reusable (FAIR)' https://distributed-text-services.github.io/specifications/.

mologique de l'ancien français (*DEAF*)[14] (see Figure 8.1). The challenges are not only technical, as semantic mapping is complex and time-consuming. To achieve this goal the two projects have been working together for years and positively contaminate each other: 'Notre idée est donc celle d'un lien vivant entre texte et dictionnaire, où ce dernier ne prévoit pas seulement des hyper-liens vers le texte mais peut avoir un impact sur les choix éditoriaux' (Glessgen and Dallas 2019, 237; cf. Tittel 2018). Another example is the *Dictionary of Old Norse Prose* (*ONP*),[15] integrating references to the occurrence of each term in the digitised editions of 437 works preserved in 4.807 manuscripts, available in the same environment of the dictionary (*ONP Dictionary* and *ONP Reader*),[16] as well as in external resources (see section 'Word in other corpora' at the bottom of each entry, also featuring manuscript references with image segments when available, for example, https://onp.ku.dk/onp/onp.php?o51596).

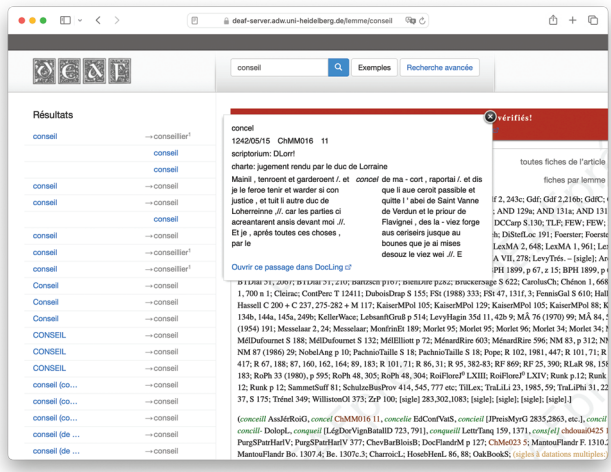

Figure 8.1. The DEAF entry for 'conseil'. In the smaller window, the preview of the occurrence of the term in ChMM016 from the *Documents linguistiques galloromans* (*https://deaf.hadw-bw.de/lemme/conseil*).

14 http://www.deaf-page.de/.

15 https://onp.ku.dk/onp/onp.php.

16 It is noteworthy that the ONP Dictionary implements the e-lexis API, a standard protocol for accessing dictionaries, facilitating linking from the editions to the dictionary.

In this re-use scenario, we pursue the connection of scholarly editions and linguistic resources by looking at historical dictionaries. In terms of technical infrastructure, permanent identifiers and APIs are always central to enable data re-use. Workflows for semantic mapping, on the contrary, should be established on a case-by-case basis, and shared (as in Tittel 2018) to move towards a certain standardisation.

Detecting intertextuality in drama

Edition data is re-usable in the context of literary studies. Distant reading often requires a large amount of data for statistical analysis and machine learning techniques. In some of the approaches to computational literary studies, the corpus to be analysed need not be composed of scholarly editions. In stylometry, for example, it has been proven that noise produced by OCR inaccuracies does not invalidate the result (Franzini et al. 2018). In other cases, such as the detection of text re-use, the type of texts in the corpus has consequences on the results.

The detection of text re-use has been and is widely used for the study of intertextuality and attribution, both before and after the advent of computers. This re-use scenario deals with the case of Spanish Golden Age theatre, in which the corpus is composed of hundreds of plays. Many of them are reworked versions of contemporary texts and have been written collaboratively by two or more authors (Matas Caballero 2017; Hirschfeld 2004). Textual transmission is active, and the witnesses preserve many variants. For these reasons and because the goal is to obtain fine-grained results (the re-use of a literal string of characters between a source text and a target text),[17] working with critical editions may yield different

17 Intertextuality and text re-use are important subjects within the Information Retrieval field. Considering the different methods (n-gram matching, TF-IDF, sequence alignment algorithms and so on), tools (Tracer, Passim and so on) and text re-use definitions (allusive, quotation, paraphrase and so on) is beyond the scope of this chapter.

results from working, for example, with a diplomatic transcription of a single witness. In the case of Golden Age theatre, though, it is not easy to build a homogeneous corpus to apply computational methods of text re-use detection, due to the variety of sources and the scarce availability of FAIR data: we can access mainly nineteenth-century critical editions (available in HTML) and modern critical editions (in PDF), plus a few examples of digital scholarly editions (providing structured data for access and download).

As said, the analysis of parallel texts is a long-standing method to discern authors and their contributions, and this is particularly relevant for the many works written collaboratively by multiple authors. As an example, we can examine Agustín Moreto's contribution in *Oponerse a las estrellas*, attributed to Juan de Matos Fragoso, Antonio Martínez de Meneses and Moreto himself. We look for parallel texts in this play and in a selection of 500 other plays of Spanish Golden Age Drama, including Agustín Moreto's *El parecido* (Losada Palenzuela 2022). The latter is available in at least three digital versions: (1) a 2018 print and digitised scholarly edition,[18] (2) a reworked neoclassical version, available in HTML,[19] and (3) the digital facsimile, OCRed, of a seventeenth-century witness within a drama anthology.[20] One verse line, 'el precepto de mi padre' appears in both *El parecido* and *Oponerse a las estrellas* (see Figure 8.2). Significantly, within the analysed corpus, this verse also appears in two other works of Moreto's single authorship, *Eneas de Dios* and *El lindo don Diego*. To detect the parallel texts, it is crucial to choose among the three available versions of *El parecido*, since the parallel line only appears in the 2018 scholarly edition (1) and in the digital facsimile (3), but not in the eighteenth-century version (2).

18 (Moreto 2008). A printed critical edition, also available in pdf format on the website www.moretianos.com.

19 Available on the Biblioteca Virtual Miguel de Cervantes (BVMC) https://www.cervantesvirtual.com/nd/ark:/59851/bmc639m7.

20 Available on the Biblioteca Virtual Miguel de Cervantes (BVMC) https://www.cervantesvirtual.com/nd/ark:/59851/bmc2f873.

Scholarly editions also ensure compliance with certain standards, defined within the field of Textual Criticism. For the Golden Age Spanish theatre, the norm is to modernise the spelling in the established critical text. Thus, if a witness transcription is part of the corpus along with a scholarly edition, many occurrences of re-use would be overlooked due to orthographic differences (among the most common, the use of long 's', the alternance 'u'/'v', 'x'/'j' and 'g'/'j', the double 's'). We can, for example, look at the opening verses in the play *El parecido*, which read *'Mi aluedrio dexo preſo / desta paſsion rigurosa, / no vi muger mas hermoſa'* in (3) and *'Mi albedrío dejo preso / desta pasión rigurosa, / no vi mujer más hermosa'* in (1). The orthographic differences have an impact on the detection of parallel texts. There are of course several ways to avoid the influence of these differences before and during the application of text re-use algorithms (automatic modernisation with regex, elimination of punctuation, k-shingling and so on). But it is important to consider that all these decisions affect the results.

In this re-use scenario, we focus on the importance of working with scientifically established texts for performing certain types of computational analysis. The availability of editions in different formats facilitates the re-use by providing authoritative resources for different re-use cases.

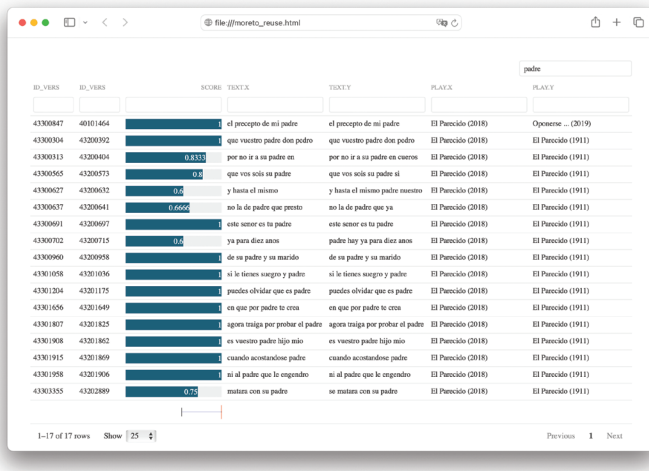

Figure 8.2. Example of comparison between (1) and (2) using the software Tracer (Büchler et al. 2014; Jänicke et al. 2014). Source: Authors.

Digital editions and gazetteers

In this re-use scenario, we focus on the re-use of editions and of gazetteers beyond academy through *ToposText*, a 'collection of ancient texts and mapped places relevant to the history and mythology of the ancient Greeks'.[21] The intended audience of *ToposText* goes beyond the scholarly community and includes students, tourists (thanks to a downloadable application for using while visiting the places mentioned in the texts) as well as the general public interested in the dissemination of scientific contents. To our purpose, *ToposText* is an example of re-use from multiple digital resources, including scholarly editions from the *Perseus Digital Library* and named entities (places) from the *Pleiades* project.

Perseus[22] is a well-known and long-standing Digital Humanities project, providing access to out-of-copyright editions of classical texts within a digital library with a robust infrastructure (Lang 2018). One of the pillars of this infrastructure is the implementation of the Canonical Text Services (CTS) protocol to serve texts and their passages, which greatly facilitate the re-use of *Perseus* data (Smith 2009). The library contains digital versions (XML/TEI encoded) of print scholarly editions from which the apparatus has been stripped out. This is in line with the *ToposText* policy: '[texts] have been stripped of footnotes and other scholarly apparatus, … ToposText is not a substitute for the most recent scholarly edition of a given work."[23] *Pleiades* is a community-built gazetteer of ancient places.[24] Because of the many ways to access *Pleiades* data (namely the Pleiades Places API, Pleiades CSV data for GIS, the RDF dump[25]) and of its open licence, it is an ideal resource

21 ToposText, version 3.0, Aikaterini Laskaridis Foundation <https://topostext.org>.

22 Perseus Digital Library, <http://www.perseus.tufts.edu/hopper>.

23 'The project', Topostext,<https://topostext.org/the-project>. On consequences of stripping out the critical apparatus from editions, see Pierazzo (2016).

24 Pleiades, <https://pleiades.stoa.org>. On digital gazetteers, see Berman et al. (2016).

25 Available at http://api.pleiades.stoa.org/, https://atlantides.org/downloads/pleiades/gis/, https://atlantides.org/downloads/pleiades/rdf/.

for re-use and remix. The *TopostText* project exploits this potential by offering classical texts with annotated places, which can be visualised on a map (see Figure 8.3).

The example of *TopostText* shows another way to re-use data from scholarly editions, enabled by suitable technical infrastructures and licensing policies. The re-use between editions and gazetteers functions in both directions: gazetteers can be used to enrich editions and editions can be used to enrich gazetteers. This fruitful dialogue is also at the heart of other Digital Humanities approaches, such as literary mapping (Cooper, Donaldson and Murrieta-Flores 2016; Losada Palenzuela 2019).

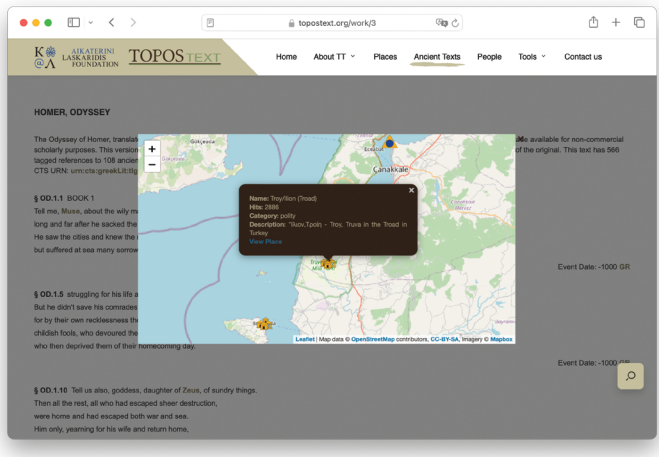

Figure 8.3. The *Odyssey* in Topostext, with the location of Troy as an overlay map (https://topostext.org).

Scholarly editions at the crossroads of disciplines

The content of a scholarly edition is a mine of information for many disciplines, such as history, linguistics, literary history and criticism, philosophy and more: all those fields whose primary sources or object of study are text documents, considered in their many aspects as the linguistic code, the textual entity, the work, the document

(Sahle 2013, 45–9). The content of a print scholarly edition – that is, what is to be found in the text, apparatus, introductions, editorial notes and so on – as it is today, is the result of centuries of dialogue between disciplines and reflects the current trends (for example, the attention to the materiality of documents). Digital editions, on the contrary, have just started in the last decades to assume the role of a carrefour for different disciplines, and the negotiation process is ongoing: should the critical or diplomatic text be enriched with part of speech tags? Should statistical methods for authorship attribution be part of the edition? These legitimate questions, and many more of the kind, will probably be answered case by case and it is difficult to anticipate future tendencies. The current panorama suggests that the digital incunabula phase of the edition is not finished yet:[26] so far, the content of a digital edition is in most cases very similar to the content of a print edition.

In addition to the remediation of content, digital scholarly editions should face the challenge of defining the technical infrastructure to support the dialogue between disciplines and reinforce their role as carrefours. Considering the re-use scenarios sketched above, we can provide suggestions to improve the re-usability of editions data. Although these suggestions are simple and in no way revolutionary, they are still not widely implemented: provide the data in multiple formats (at least XML/TEI, but also TXT, JSON and JSON-LD, CSV are suitable formats); provide multiple access to the data (API,[27] data dump, single resource download); implement internal persistent identifiers and enrich data with external persistent identifiers; offer documentation for users to make sense of the data and understand how they have been collected and generated.

The re-use audience is potentially very large and it is difficult to anticipate possible re-use scenarios. We think, though, that the

26 The first occurrence of the term 'digital incunabula' is probably to be found in Tolva, 1995.

27 An analysis of the current state of the art concerning APIs in scholarly editing is undergoing (Spadini and Losada Palenzuela 2023).

variety of formats and multiplicity of end points is a true advantage. In terms of formats, for example, the requirements when re-using data for their integration in a larger edition corpus (for example, the complete works of X, for which XML/TEI would be the format of choice) are different from the requirements when re-using data for a text analysis task (for which TXT or CSV is generally suitable). Of course, not all edition projects can be expected to have the resources to implement solutions for different kinds of re-use. As mentioned before, a minimal setting for enabling re-use is to provide data dumps or single resources data in XML/TEI: this is the choice of many edition projects that make their data available.

Versioning is an open issue in scholarly editing and in data re-use, and one that apparently clashes with our suggestion of multiplying the access points to data. The projects *The Proceedings of the Old Bailey* and *Registres de la Comédie-Française* exemplify this tension, since both signal to the user that data retrieved through the API and from the data dumps 'might represent slightly different versions'.[28] The problem, however, is not that the data is exposed in different ways, but that the open-endedness of digital editions comes with certain disadvantages, among them the 'perpetual beta status' (Gengnagel 2017). Moreover, the different versions of the data, of the software and sometimes even of the model, makes a complete replicability over time impossible to achieve.

Replicability is not only relevant for external validation, typically through peer-review: within an edition project itself or in a follow-up project, data can be used for replication and maintenance, or for

28 *The Proceedings of the Old Bailey*, https://www.oldbaileyonline.org/static/Data.jsp: 'Much of the OBAPI documentation is also useful for understanding the XML files, as the API is based on the same data. However, please note that sometimes the files on ORDA and the API might represent slightly different versions of the data.' *Registres de la Comédie-Française*, https://hack.cfregisters.org/en/receipts/database.html: 'That dump file is a snapshot of the database from a point in time, and is not the most recent version of the live database; to access the most recent data, please refer to the REST API.'

re-use purposes other than the edition itself. This is the case, for example, of visualisations based on editions data; or of a new web application to be built on top of existing data, when an 'old' website is becoming technically or visually obsolete. Re-using one's own data may be as difficult as re-using others', if proper access to the data or documentation are lacking.

To conclude, we think that tackling data re-use is relevant in the field of scholarly editing: enabling re-use can be considered an intrinsic quality of digital editions and improves the research impact. In scholarly editing, as in any act of communication and scientific publishing, defining the audience can help in designing a fitting resource. Re-use is just another way of consuming the data, re-users just another type of users, and the re-use audience another type of audience. We hope that the concrete examples of re-use scenarios presented in this chapter will help editors to consider data re-use and how to facilitate it, from an early stage of the edition planning. Enabling re-use does not mean to confine scholarly editing among the so-called auxiliary sciences, a label used for disciplines that were considered ancillary to history (and literary studies or philosophy), such as palaeography, epigraphy, diplomatics and, more recently, information technology. On the contrary, it means to strengthen its pivotal position at the crossroads of Humanities disciplines.

References

Ainsworth, P. and Croenen, G., eds. 2013. *The Online Froissart*. Version 1.5. Sheffield: HRIOnline. http://www.dhi.ac.uk/onlinefroissart.

Berman, M. L., Mostern, R. and Southall, H. 2016. *Placing Names: Enriching and Integrating Gazetteers*. The Spatial Humanities. Indiana University Press.

Büchler, M., Burns,P. R., Müller, M., Franzini, E. and Franzini, G. 2014. 'Towards a Historical Text Re-Use Detection.' In *Text Mining*, edited by Biemann, C. and Mehler, A. Springer International Publishing. https://doi.org/10.1007/978-3-319-12655-5_11.

Complete Works of Carl Maria von Weber. Digital Edition. Last modified July 1, 2024. http://weber-gesamtausgabe.de/A070002.

Cooper, D., Donaldson, C. and Murrieta-Flores, P. 2016. *Literary Mapping in the Digital Age*. Taylor & Francis.

Deutz, D. B., Buss, M. C. H., Hansen, J. H., Hansen, K. K., Kjelmann, K. G., Larsen, A. V., Vlachos, E. and Holmstrand, K. F. 2020. 'How to FAIR: A Website to Guide Researchers on Making Research Data More FAIR.' Zenodo. https://doi.org/10.5281/zenodo.3712065.

Dumont, S., Grabsch, S. and Müller-Laackman, J. 2021. 'CorrespSearch – Briefeditionen Vernetzen (2.0.0) [Webservice].' https://correspSearch.net.

Emsley, C., Hitchcock, T. and Shoemaker, R., eds. 2018. 'Digital Projects Using Old Bailey Online Data.' In *Old Bailey Proceedings Online*. March 2018. https://www.oldbaileyonline.org/static/Projects.jsp.

Franzini, G. 2019. 'Towards Connecting Scholarly Editions to Corpora in the LiLa (Linking Latin) Knowledge Base of Linguistic Resources.' Zenodo. https://doi.org/10.5281/zenodo.3613371.

Franzini, G., Kestemont, M., Rotari, G., Jander, M., Ochab, J. K., Franzini, E., Byszuk, J. and Rybicki, J. 2018. 'Attributing Authorship in the Noisy Digitized Correspondence of Jacob and Wilhelm Grimm.' *Frontiers in Digital Humanities* 5. https://www.frontiersin.org/articles/10.3389/fdigh.2018.00004.

Franzini, G., Terras, M. and Mahony, S. 2019. 'Digital Editions of Text: Surveying User Requirements in the Digital Humanities.' *Journal on Computing and Cultural Heritage* 12 (1): 1–23. https://doi.org/10.1145/3230671.

Gengnagel, T. 2017. 'The 'Beta Dilemma' – A Review of the Faust Edition.' *RIDE – A Review Journal for Digital Editions and Resources* 7. https://ride.i-d-e.de/issues/issue-7/faustedition/.

Glessgen, M., ed. 2016. *Documents linguistiques galloromans. Édition électronique*. https://www.rose.uzh.ch/docling/.

Hirschfeld, H. A. 2004. *Joint Enterprises: Collaborative Drama and the Institutionalization of the English Renaissance Theater*. University of Massachusetts Press.

Jänicke, S., Geßner, A., Büchler, M. and Scheuermann, G. 2014. 'Visualizations for Text Re-Use.' In *2014 International Conference on Information Visualization Theory and Applications (IVAPP)*, edited by Laramee, R.S, Kerren, A. and Braz, J. IEEE.

Lang, S. 2018. 'Review of Perseus Digital Library – RIDE.' *RIDE – A Review Journal for Digital Editions and Resources* 8. https://ride.i-d-e.de/issues/issue-8/perseus/.

Losada, P. and Luis, J. 2019. 'Mapeado digital de lugares en la novella

bizantina española.' *Artnodes. Journal of Art, Science and Technology* 23 (1): 72–8. http://dx.doi.org/10.7238/a.v0i23.3222.

–. 2022. 'Análisis cuantitativo de la reutilización textual en las comedia scola borradas (Moreto).' Zenodo. https://doi.org/10.5281/zenodo.7271369.

Matas Caballero, J., ed. 2017. *La comedia escrita en colaboracióne nel teatro del Siglo de Oro*. Ediciones Universidad de Valladolid.

MLA Committee on Scholarly Editions. 2016. 'MLA Statement on the Scholarly Edition in the Digital Age.' MLA. https://www.mla.org/content/download/52050/file/rptCSE16.pdf.

Moreto, A. 2008. *Comedias de Agustín Moreto. Segunda Parte de Comedias. Volumen VII*. Edited by Santos, H. B., Castillo, L. R. and Trambaioli. Edition Reichenberger.

Pierazzo, E. 2016. 'Modelling Digital Scholarly Editing: From Plato to Heraclitus.' In *Digital Scholarly Editing: Theories and Practices*, edited by Driscoll, M. J. and Pierazzo, E. Open Book Publishers. https://doi.org/10.11647/OBP.0095.03.

Sahle, P. 2013. *Digitale Editionsformen – Teil 3: Textbegriffe Und Recodierung: Zum Umgang Mit Der Überlieferung Unter Den Bedingungen des Medienwandels*. Schriften Des Instituts Für Dokumentologie Und Editorik. Books on Demand.

Smith, N. 2009. 'Citation in Classical Studies.' *Digital Humanities Quarterly* 3 (1). http://www.digitalhumanities.org/dhq/vol/3/1/000028/000028.html.

Spadini, E. and Palenzuela, J. L. L. 2023. 'Scholarly Digital Editions: APIs and Re-use Scenarios.' In *Collaboration as Opportunity (DH2023)*. Graz, Austria. https://doi.org/10.5281/zenodo.7271447, https://doi.org/10.5281/zenodo.8140728 and https://zenodo.org/record/8107582.

Tittel, S. 2018. 'Historical Corpus and Historical Dictionary: Merging Two Ongoing Projects of Old French by Integrating Their Editing Systems.' In *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*, edited by Krek, S., Čibej, J., Gorjanc, V. and Kosem, I. Ljubljana University Press, Faculty of Arts.

Tittel, S., Bermúdez-Sabel, H. and Chiarcos, C. 2018. 'Using RDFa to Link Text and Dictionary Data for Medieval French.' In *Proceedings of the 6th Workshop on Linked Data in Linguistics*. Myazaki, Japan. http://lrec-conf.org/workshops/lrec2018/W23/pdf/10_W23.pdf.

Tolva, J. 1995. 'The Heresy of Hypertext: Fear and Anxiety in the Late Age of Print.' http://www.ascentstage.com/papers/heresy.html.