

LEXICON PHILOSOPHICUM

International Journal for the History of Texts and Ideas

NICOLA RUSCHENA

Notes on Machine-readable Sources for the History of Philosophy

ABSTRACT: In the last decades, a variety of digital resources has become available to historians of thought, enabling them to increasingly rely on digitised sources for their researches. Numerous projects developed infrastructures designed for the retrieval, collection and analysis of historical and critical sources. It is difficult to map the landscape of such resources due to the diversity of their aims, scopes and treated subjects, and the lack of shared standards in their contents and designs makes them difficult to compare. Yet, this variety of architectures seems to mirror the plurality of research approaches that characterises historiography of thought. This contribution discusses four types of resources: author-centred collections, thematic collections, generalist infrastructures and repositories of bibliographic and archival records. By relying on the notion of machine-readability as developed within the FAIR guidelines, these types of infrastructures are presented as resources available, to different degrees of accessibility, to historians of thought to answer diverse research needs.

KEYWORDS: Digital Humanities; Digital Research Infrastructures; Digital History of Thought; Distant Reading; Online Archives

1. INTRODUCTION

Archival and librarian resources that are made available online can play a major role in determining conditions of work of historians.

At the very least, digitised editions and catalogues provide historians and students with means for the consultation of many different works and allow them to access larger collections of archival references for their research. Moreover, some resources give access to great quantities of information that are consistently stored in more or less structured forms, thus providing representations of specific publishing or institutional contexts. These contexts themselves can be objects of historiographical inquiry, pursued by means of the analysis of the corpora of individual sources.¹

1. Though less famous than Franco Moretti's distant reading of catalogues (Moretti 2000), there are some examples of the study of wide historical contexts represented by specific archival resources also in the field of the history of philosophy: see, for instance, Bonino & Tripodi 2020, de Bolla 2013, de Bolla *et al.* 2020, Petrovich & Buonomo 2018, Petrovich 2022.



Machine-readability is a critical aspect of digital resources, since it heavily conditions their usefulness for consultation and research. Machine-readable digitisation of documents and books allows to apply digital tools of research and analysis to their content, e.g. by searching for specific expressions in the digitised copy of a book, or by analysing the frequency in the use of some specific terms in – say – a collection of English books from the 18th century.

Moreover, machine-readability gains extraordinary relevance for managing large collections of sources as wholes, which can be sorted, ordered and explored for retrieving relevant historical information only by means of digital tools.

However, building digital resources for archival consultation and historical research is a demanding process. Original sources have to be collected and digitised by transcription or scanning, bibliographic metadata have to be recorded in consistent formats, pre-existing archives have to be rebuilt in order to ensure consolidation of data that are already available with new ones. Indeed, digitisation projects may require significant funding for extended periods; they involve heterogeneous teams for the general design of the project (scope and intended use of the resource to be built, technologies, etc.) and need the essential contribution of both librarians and computer scientists. The former are necessary for cataloguing and for the digitisation procedures, while the latter (often provided by private companies) allow to set up the infrastructures for storing and accessing digitised material.

The evident usefulness of digital resources led both institutions and private companies to the creation of several digital resources for the history of thought. Moreover, the transition to digital means in administrative work determined the availability of a great variety of registries, indexes and logs in digital – and, to various extents, machine-readable – form. Nonetheless, in many departments of philosophy one could expect most historians to be far more expert about the locations of documents in physical archives and libraries dispersed in a variety of countries, rather than about the resources granting access to digitised documents of historical interest. There is quite a variety of digital resources resulting from different digitisation projects, focusing on specific kinds of publications, or on bibliographic and archival materials. Yet, relying extensively and systematically, though clearly not exclusively, on such digital infrastructures is still a relative novelty for historians of philosophy, especially in comparison with the longevity of other traditional means and resources for historiographical work.

The discussion of resources and infrastructures in the following pages does not aim at being an exhaustive survey. Rather, resources are selected as representative of the variety of available means to access historical documents and data for both study and research. Some occasional remarks on the genesis of specific projects are meant to provide some reference to the contexts of digitisation processes that might hopefully be of use while searching for other specific resources.

Historical contexts are often characterised by peculiar dynamics in the production and conservation of sources. Moreover, different historiographical methodologies can rely on different sets of sources, focusing on some collection of published works, correspondences, bibliographical and archival catalogues, institutional and administrative

documental material and so on. Apparently, there seems to be no unique standard for comparing and evaluating the usefulness, for research purposes, of differently-conceived resources. Yet both their content and their suitability for different research tasks can be specifically discussed.

At a descriptive level, it is possible to classify resources according to their content: individual resources are characterised either chronologically, geographically, thematically or according to their author etc. Moreover, they store and make available, in the form of more or less structured data, different kinds of documents (archival records, digital reproductions of books, indexes of specific kinds of sources, bibliographies) by relying on different infrastructures.

At a functional level, different technical characteristics of infrastructures entail different possibilities in the use of data for both study and research purposes. Given the plurality of purposes and approaches that may be proper to historiographical inquiry, rating digital resources with respect to their historiographical potential per se would seem the most sterile and empty exercise, at least while maintaining some trust towards historians' skill and flair. Moreover, although the notion of machine-readability is central in the evaluation and discussion of different resources, it seems too broad to differentiate them effectively. Indeed, it would be difficult to evaluate machine-readability of resources as wholes, because different kinds of data can be provided by the same resource in formats that are more or less machine-readable. Moreover, machine-readability broadly conceived usually encompasses so-called machine-actionability, i.e. the degree of possibility to interact with data by means of automatic digital tools, or "a continuum of possible states wherein a digital object provides increasingly more detailed information to an autonomously-acting, computational data explorer".²

In light of these elements, comparison of multiple machine-readable collections should, in principle, be aware of what elements can effectively be used in digitally-aided browsing and analysis, and to what extent.

A common framework encompassing some of these aspects is provided by the FAIR principles for data management,³ designed to serve as guidelines for the development of digital archival resources. The framework allows finer-grained differentiation among the manifold aspects of machine-readability, to be assessed according to four principles: findability, accessibility, interoperability and reusability of resources. Given the general purpose of the framework, FAIR guidelines do not prescribe implementation choices, nor do they "suggest any specific technology, standard or implementation solution".⁴

As it can be easily imagined, specific digitisation projects have favoured the digitisation of different features of documents, thus satisfying the principles to varying degrees. For example, some resources might grant open access to large quantities of data, which would be made highly accessible, while nonetheless relying on inconsistent indexing systems which compromise findability of data in the resource. Moreover,

2. Wilkinson *et al.* 2016: 3.

3. *Ibid.* (principles: p. 4, "box 2"), "FAIR Principles".

4. *Ibid.*: 5.

(partial) satisfaction of different principles may easily overlap. Indeed, consistency in the reliance on data formats is often crucial in determining interoperability and in many cases the need for consistency favours the choice of widely used, standard formats with high descriptive power,⁵ which in turn play an important role in reusability.

Noticeably, the development of most digital infrastructures available for research in the history of philosophy preceded the formulation of FAIR principles. Indeed, the FAIR framework has been designed in order to provide a comprehensive synthesis of the needs that emerged from a community defined by a shared general problem, i.e. the reliance on large digital collections of research and historical data. The development of many digital infrastructures dedicated to the management of both cultural heritage and research data have represented much of the experience and results from which the proposal of FAIR principles has been drawn from. A thorough assessment of infrastructures' compliance to FAIR principles would surely represent a significant part of any project improving on existing infrastructures. Nonetheless, the principles are designed to be guidelines for the design of infrastructures rather than a standard post-factum evaluation of existing resources.

For these reasons, comments on functional aspects of infrastructures will be formulated consistently with the FAIR data framework, but they will focus on selected features that appear as particularly relevant for historians of philosophy.

2. AUTHOR-CENTRED RESOURCES

Philosophers' works tend to qualify particularly well as subjects of digitisation projects, probably representing safe bets on the impact that digitisation efforts may have on research and study activities. At least to some extent, historians of philosophy are used to rely on author-based digital collections, since it is commonly accepted that individual authors' productions represent legitimate and proper objects of research in the history of thought. Moreover, philosophers' productivity has often been impressive, so that machine readable versions of their works can be of great use to specialists for research and teaching activities.

Under this view, the digital edition of some author's collected works provides access to a disciplinarily-autonomous collection without the need of relying on sophisticated reflections concerning either field demarcation or possible biases in the collection of sources (other than those normally required to critical editions). As a digital collection or corpus, the digital edition of the collected works of a philosopher is exhaustive (to the varying extent to which collected works can be exhaustive) of a clearly delimited historiographical scope.

5. xml schemas came to be powerful standards for reproduction of both textual content (TEI-compliant schemas) and bibliographic data (xmlrdf, xmlmarc); JSON (or JSONL) files are often used for sharing more heterogeneous data, especially if they are expressly aggregated and provided to the user via web services. That is the case for JSTOR's infrastructure for dataset building ("Constellate").

As stated above, a number of author-centred digital collections is available online, provided by both private companies and academic institutions. IntelLex Past Masters⁶ (henceforth: PM) is a prominent example of resource developed by a private company, offering access to full texts of many author-based digital collections of works and correspondences.⁷ Some of PM's collections are digital editions made anew (as in the cases of Locke's works and selected correspondence, and Descartes works⁸), while the majority of collections are digitised versions of printed critical editions (as in the cases of the Clarendon edition of the Locke's correspondence and of the New York University Press's works of Darwin⁹). PM provides high-quality digital or digitised editions to institutions subscribing to its services, which are accessible through academic proxy authentication. Unfortunately, subscriptions to PM are a matter of private commercial negotiation, leading to significant variation in the accessible corpora depending on institutional affiliations.

Texts in PM are available either as full texts or as xml files encoded according to the standards of the Text Encoding Initiative (xml-TEI).¹⁰ Reliance on TEI encoding allows for the conservation of valuable paratextual features such as page numbers and critical apparatus, thus ensuring reliability of the collections for professional use.

Some high-quality digital critical editions have been produced as the result of specific publicly-funded projects. Such resources are not developed for commercial use and thus grant open access to research and teaching material regardless of institutional affiliations. That is the case of the critical edition of Nietzsche's works and correspondence,¹¹ which is made available as the *Digitale Kritische Gesamtausgabe Werke und Briefe* (eKGWB)¹² through the web application Nietzsche Source.¹³ The eKGWB is completed by the diplomatic digital edition of archived manuscripts *Digitale Faksimile-Gesamtausgabe* (DFGA).¹⁴ Texts in the eKGWB are available in plain text and in xml-TEI, while DFGA's scans of the manuscripts can be downloaded as jpg images.

The eKGWB has been claimed to be "one of the first digital editions suitable to be used and cited in academic research".¹⁵ Similar to PM, texts in the eKGWB are available in plain text and in xml-TEI format (while DFGA's scans of the manuscripts can be downloaded as jpg images). Concerning the technical means of reproduction of texts,

6. "Intelex Past Masters".

7. "Past Masters – Authors".

8. Respectively, Rooks 1995 and Gombay 2001.

9. Respectively, De Beer 1976-1989 and Barrett & Freeman 2010.

10. TEI: Text Encoding Initiative.

11. Noticeably, two distinctly printed critical works are joined in one digital edition: Colli & Montinari 1967-, pertaining to Nietzsche's works, and Colli & Montinari 1975-1984, pertaining to his correspondence.

12. "Digitale Kritische Gesamtausgabe von Nietzsches Werken und Briefen (eKGWB) – nietzschesource".

13. "Nietzsche Source – Home". For a thorough description of the resource see D'Iorio 2010.

14. "Digitale Faksimile-Gesamtausgabe Nietzsches (DFGA) – nietzschesource".

15. D'Iorio & Barbera 2011.

the two resources provide text at the same qualitative level, entailing a similar degree of interoperability of the texts contained in the two resources. Nonetheless, overall usefulness of eKGBW for research seems greater than that of PM's editions. First, being visibly built by and for researchers, Nietzsche source web app allows for the completion of tasks such as advanced searches throughout the corpus and the storage of multiple results as data for reuse in further analyses. Second, eKGBW allows scholars to switch effectively between digital and printed critical editions. Works, sections etc. are indexed according to the architecture and nomenclature of the printed editions, thus ensuring findability of data and their reusability for bibliographic documentation: indeed, the resource can be cited directly with stable urls composed of the standard abbreviations for bibliographic references. Furthermore, post-print philological corrections to the printed critical editions are integrated and marked out in eKGBW along with previous versions.¹⁶

Noticeably, digital critical editions require great efforts and suitable means. The possibility of their realisation is grounded as well on the availability of systematic scholarly endeavours devoted to critical collection of materials, as it is for the case of eKGBW and the printed works and correspondence of Nietzsche. The "Œuvres complètes de Voltaire" of the Oxford Voltaire Foundation,¹⁷ for instance, are expected to be made available in high-quality digital format (with the name "Digital Voltaire"¹⁸) by means of the infrastructure that is being developed by the Voltaire Lab:¹⁹ "Digital Enlightenment" (DE). This resource will host both "Digital Voltaire" and "Digital d'Holbach", a born-digital critical edition of d'Holbach's works.²⁰

Until the publication of the digital editions in the Digital Enlightenment platform, full text of the production of d'Holbach and Voltaire have been openly available through two dedicated resources, namely "Tout d'Holbach" and "Tout Voltaire",²¹ which are hosted by the Division of the Humanities and the libraries of the University of Chicago, in the context of the French-American academic consortium ARTFL.²² In spite of the lack of critical annotations and of any machine-readable means to include paratextual features, these resources can be of great use for exploratory research: indeed, the searching interface allows for both exact and proximity searches in the corpus or in its (pre-determined) chronological sectioning. Results of searches can be temporarily stored and exported for analysis and comparison. Digitised texts are indexed by

16. "Nietzschesource – Documentation".

17. Voltaire, 1968-2022.

18. "Digital Voltaire – Voltaire Foundation".

19. "Voltaire Lab – Voltaire Foundation". The digitized critical edition will be available as "Digital Voltaire".

20. "Digital d'Holbach – Voltaire Foundation". See further documentation about the DE project at "Digital Enlightenment – Voltaire Foundation".

21. "Tout d'Holbach", "Tout Voltaire". Other author-centred collections in French are available through ARTFL's portal (e.g. Rousseau, Robespierre), many of which are openly accessible. See "Public Databases – The ARTFL Project".

22. "The ARTFL Project". ARTFL's "Tout Voltaire" has been extensively relied upon in the works of the Voltaire lab.

referencing to printed sources, thus ensuring a certain degree of findability, although reference to the exact printed edition of the source is often incomplete.²³

3. THEMATIC COLLECTIONS

Historical documents and books are also made available by means of resources collecting digitised material that is deemed to be relevant for specific historical contexts or topics. Indeed, a number of digital thematic collections can be accessed online and digitally analysed, although their availability is obviously dependent on the scholarly status of the specific topics and on the degree of digital literacy of researchers who might engage with such resources. In short, the investment of efforts and funds required by the development of such resources needs to bear the promise of fruitful future research or teaching activities.

In line with this principle, research groups and institutions operating in long-lasting traditions such as classical studies have committed quite early to the building of research-suitable corpora, by compiling a number of resources collecting ancient texts. Most famously, this is the case of the “Thesaurus Linguae Graecae” (TLG), probably the largest collection of texts in ancient Greek, from Homer to the fall of Constantinople in 1453.²⁴ A dedicated search engine has been developed and the collection has been available online since 2001, allowing for the parcelling of sub-corpora on authorial or chronological basis. TLG additionally allows researchers to cite passages of digitised primary sources with standard bibliographic referencing systems.

While TLG requires individual or institutional subscription to its services, there are openly accessible infrastructures which encompass open-access portions of TLG, such as the Perseus Digital Library, which provides open access to considerable digital collections of ancient sources in Latin and Greek.²⁵ Collections often include many works originally included in TLG and then shared with other infrastructures. Much of digitised open-access classical texts are being consistently included in the on-going “Open Greek and Latin” project (OGL), which provides browsing tools online and well-documented repositories of texts belonging to various Greek and Latin collections in xml-TEI format.²⁶ OGL greatly improved the overall reliability of digitised classic

23. A – conceptually – simple way to overcome this issue would be for the texts to reference the ISBN or other stable identifier of the so-called “manifestation” of the intellectual production, accordingly to Functional Requirements for Bibliographic Records (IFLA Study Group 2009), thus ensuring higher levels of findability and reusability.

24. “TLG – About”. The programme started in 1972 at the University of California and began digitisation by first tackling the issue of ancient Greek text encoding.

25. “Greek and Roman Materials – Perseus Project”. Perseus started in 1983 as a Tuft-based open access project on digitised classical texts, rapidly expanding and involving other library – or university – based digitisation projects.

26. “Open Greek & Latin”. Corpora are searchable online by means of Scaife viewer (“Scaife Viewer – About”), but they are also available and in online repositories documenting the work done on the specific collections joined in OGL (e.g. “GitHub – PerseusDL/Canonical-Latin-Lit” and “GitHub – OpenGreekAndLatin/First1KGreek”).

texts: documents in OGL are tagged with bibliographic annotations that can be easily extracted and they are provided with stable identifiers.²⁷ Moreover, researchers can finely tune their tools for analysis and reuse openly available resources, since OGL's corpora can be easily browsed as open access datasets without relying solely on the versatility of browsing and searching tools provided by online resources.

Classical studies surely benefit from high-quality resources that are not available for every field of research nor for every historical period. There are nonetheless scholarly efforts in other fields focused on specific historical periods and traditions of thought.

Thanks to the perceived historical relevance of the enlightenment, a considerable quantity of sources from 18th century British and French printed production have been digitised aiming at a research-level standard of reliability.²⁸ That is the case of multiple digitised collections compiled and made available by the aforementioned ARTFL project for open consultation through the PhiloLogic4 web application.²⁹ Along with author-centred resources, larger contextual collections are hosted in ARTFL infrastructure,³⁰ such as Gale's Eighteenth Century Collection Online³¹ (ECCO), which is a corpus of digitised books printed in the British Isles during the 18th century. ECCO's full text (in TEI-compliant xml format), provided by the Text Creation Partnership (TCP)³² by means of OCR software, can be openly searched online, but the error-rate in the transcription of the collection is difficult to assess (like for various other resources in the ARTFL environment).

Among scholarly infrastructures, one of the most complete and versatile thematic resources is the open-access *Édition Numérique Collaborative et CRitique de*

27. Inspection of xml files reveals many book-in-hand annotations made by librarians. These include important metadata such as editors, authors, titles etc., but often they lack any indication of serial identifiers (e.g. ISBN) of the original physical source. This makes some nuanced aspects of findability and reusability particularly evident: digital texts in OGL are very effectively indexed as digital objects, so that text can be automatically retrieved at the level of lines; yet, although such texts are digital copies of physical books referenced in catalogues with specific identifiers, the xml files do not provide references to cataloguing identifiers, so that the digitally retrieved text cannot automatically reference the physical book it was digitised from.

28. See for instance Burrows, Roe, & Baker 2020 for extensive discussion of the (Anglo-American-made and partly French-funded) digital resources concerning the enlightenment developed in the chronological interval 2006-2016 ca.

29. "PhiloLogic4 – The ARTFL Project" is the digital tool for discovery and search of full texts available by means of ARTFL infrastructure.

30. "Public Databases – The ARTFL Project".

31. "Eighteenth Century Collections Online". The resource can be accessed through different infrastructures, some of which seem outdated. The collection has been enriched in several batches and it is not clear if such enrichments are included in all accessible infrastructure providing access to ECCO. The most up-to-date resources are Gale's Primary Sources service and the aforementioned PhiloLogic4. Arguably, Gale's service is newer and better performing, although it requires institutional or individual subscription. For thorough discussion of ECCO's contents see Tolonen, Mäkelä & Lahti 2022. For a representation of ECCO's arc of development see Gregg 2020. For an example of research in the history of philosophy relying on ECCO see de Bolla 2013.

32. "Text Creation Partnership".

l'Encyclopédie de Diderot, de D'Alembert et de Jaucourt (1751-1772) (ENCCRE).³³ The digital edition and the infrastructures for accessing and browsing it are the result of a large-scale effort shared among multiple French academic and research institutions. Scanned pages of the *Encyclopédie* and their text-rendered xml (manual) transcriptions can be browsed and downloaded with no apparent losses of paratextual features. The corpus can be segmented and searched, authors and entries are indexed by relying on stable identifiers and the resource provides a dictionary of contributors to the *Encyclopédie*. In addition to representing a versatile and reliable digital resource for consultation and research on the *Encyclopédie* as a primary source, ENCCRE is also designed to serve as a field-specific digital venue for research and documentation: the infrastructure has dedicated sections for secondary academic literature about the *Encyclopédie*, for on-going research within the ENCCRE environment (e.g. work in progress about attributions of anonymous entries to specific authors), and for specific annotations on the ENCCRE (e.g. annotated textual passages originally published in other historical sources and then reused in the *Encyclopédie*).

Some thematic resources, nonetheless, can be of scholarly interest even when lacking the investments needed for providing users with full text of historical documents. That is the case of resources such as the “Dictionnaire des Journalistes” and the “Dictionnaire des Journaux”,³⁴ which respectively index journalists and periodicals in France between 1600 and 1789. The projects started out as printed catalogues edited by Jean Sgard³⁵ and moved completely online since 2011. In 2009, the “Dictionnaires” have been paired by a digital library, the “Gazetier Universel”,³⁶ providing openly accessible external links for consultation of digitised copies of periodicals. Although data in the “Dictionnaires” are not thoroughly structured, thus lacking high degrees of actual machine-readability, the exhaustiveness of the resource and the availability of a digital library with permanent links to proper bibliographic records allow researchers to develop procedures for semi-automatic data collection.

Effective indexing of collected records can be achieved by consolidating the records with external dataset which stores stable identifiers of retrieved data. In the case of personal records, for instance, Wikidata³⁷ provides identifiers for a great number of authors among historical people recognised with some varying degree of ‘notability’. Although Wikidata’s identifiers may overlap or duplicate some personal record they can play an important role in the indexation of datasets. Issues such as the variations in the spelling of authors’ names may represent significant obstacles for the consistent collection of biographic or authorship data. Effective indexing of a good number of Europe-

33. “ENCCRE – Édition Numérique Collaborative et CRitique de l’Encyclopédie”.

34. Respectively, “Liste Des Journaux | Dictionnaire Des Journaux” and “Liste Des Journalistes | Dictionnaire Des Journalistes”.

35. Sgard, Gilot & Weil 1976 and Sgard 1991. Noticeably, the Voltaire Foundation is publisher of the *Dictionnaire des Journaux*.

36. “Le Gazetier Universel”. A separate digital library is dedicated to revolutionary press: “Le Gazetier Révolutionnaire”.

37. “Wikidata”.

an modern authors can nonetheless be achieved by relying on the Thesaurus provided by the Consortium of European Research Libraries.³⁸ The Thesaurus collects authors' personal files which list recorded spelling variations of the name and match authors with stable personal identifiers provided by national or international library systems.

4. GENERALIST INFRASTRUCTURES

A number of digital infrastructures for consultation and analysis of generalist historical collections have been made available online. Digital collections of historical correspondences are an example of generalist infrastructures focused on a specific kind of historical document. Although letters are often included in author-centred digital editions, extensive collections of correspondences among authors have been compiled in order to provide insights on the network of communication among historical actors (including those whose collected letters are unlikely to be the object of publishing enterprises). Access to digital texts is often dependent on the availability of edited correspondences that have been printed and then digitised, while it is more difficult to provide textual transcription of handwritten letters which have never been published. "Early Modern Letters Online"³⁹ (EMLO), for instance, is a general-scope digital library collecting correspondences among modern authors, developed by a cross-disciplinary team at the Bodleian Libraries. EMLO allows for advanced research among metadata of letters (e.g. by specifying author, recipient, places, mentioned people etc.) and full texts (when available). This single catalogue collects a variety of documents stored in different resources or recorded directly in EMLO itself. Stable identifiers are assigned to letters and correspondents, and digitised copies of letters are referenced when available through partnering resources.⁴⁰ Noticeably, the access to the catalogue (and to many digitised texts) is open, although the reuse of large parts of the archive and the automatic collection of data from the resource are subjected to authorisations by EMLO. Correspondents indexed in EMLO are tagged as either organisations (e.g. the Royal Society) or people, and gender of the latter is specified. This allowed for the identification of specific sub-corpora in EMLO, collecting metadata (and full text, when available) provided by several different digitisation projects: "Women's Early Modern Letters Online"⁴¹ (WEMLO) and "The Royal Society Early Letters"⁴² are two examples of sub-corpora that have been individuated by EMLO's collaborators, but similar criteria could guide historians as well in collecting relevant material for their research.

38. "CERL Thesaurus".

39. "Early Modern Letters Online".

40. Noticeably, EMLO allows to browse metadata of collections with restricted access, such as the semi-commercial project Electronic Enlightenment (EE) ("Electronic Enlightenment – Letters & Lives Online") developed by the Bodleian libraries at Oxford since 2008. EE's catalogue cannot be accessed through EE's portal without subscription, while EE's items are listed in EMLO along with metadata.

41. "Women's Early Modern Letters Online – WEMLO".

42. "The Royal Society Early Letters – EMLO".

The most prominent generalist digital collections are often the result of large-scale digitisation projects, often pursued by national or international agencies. Arguably, the goals of these projects are defined by the needs of heritage conservation rather than by those of research. Nonetheless, while the availability and quality of machine-readable reproduction of text vary among individual items, bibliographic metadata are often rich and perfectly suitable for research purposes. Moreover, the general scope of such collections allows researchers to challenge criteria for field demarcation which may bias the collections of ‘philosophical’ works. Limiting the examples to European-made infrastructures, the digitisation of modern intellectual materials and catalogues has been particularly extensive among national projects. In Germany, Deutsche Forschungsgemeinschaft (DFG) funded extensive digitisation projects involving major German academic libraries (Herzog August Bibliothek in Wolfenbüttel, Berlin Staatsbibliothek, Bayerische Staatsbibliothek), digitally cataloguing and collecting both incunabula and materials printed or published during 16th, 17th and 18th century (INKA, VD16, VD17, VD18).⁴³ Many different projects have converged in the German Central Archive of Digitised Press (ZVDD) digital library,⁴⁴ which provides open access to bibliographic records in German catalogues, allowing for century sub-selection. Authors and works are richly indexed by relying on multiple stable identifiers, while the digitised copies of printed books – at varying but usually low degrees of machine-readability of content – that are made available by various libraries are linked to records, allowing for the direct consultation of contents and for the collection of findable and reusable references.

5. BIBLIOGRAPHIC AND ARCHIVAL RECORDS

Digital historical catalogues by themselves represent useful research resources for the preliminary review of available sources, for contextualisation of scholarly literature and for the evaluation of the scope of research results. Indeed, generalist digital short titles catalogues of historical material have been compiled in several European librarian consortia, resulting in the development of both national and international digital catalogues. Furthermore, similar catalogues often constitute the ground for further projects devoted to the digitisation of contents. That is the case of the early version of the English

43. Respectively, “Inkunabelkatalog INKA”, “VD 16 Digital Catalog of Printed Works of the 16th Century Published in German Speaking Countries of the Bavarian State Library – MDZ”, “VD17 – Das Verzeichnis Der Im Deutschen Sprachraum Erschienenen Drucke Des 17. Jahrhunderts”, “VD18 – Das Verzeichnis Der Im Deutschen Sprachraum Erschienenen Drucke Des 18. Jahrhunderts”. Noticeably, incunabula catalogues may require specific expertise with this kind of sources, which are often indexed with specific archival tools devoted to thorough material descriptions. A number of catalogues of preserved incunabula are provided at national and international level, e.g. Italian “Manus Online – OPAC SBN” and British “ISTC (Incunabula Short Title Catalogue)”.

44. “Zentrales Verzeichnis Digitalisierter Drucke”. Non-textual printed sources such as maps and written music are excluded from ZVDD’s collection.

Short Title Catalogue (ESTC),⁴⁵ which then indexed printed books in the British Isles up until 1700: a selection of the collected works has been digitised and manually transcribed by the aforementioned Text Creation Partnership. The resulting resource, “Early English Books Online”⁴⁶ (EEBO) is a large, geographically-bound and high-quality textual corpus of early-modern works; noticeably, human transcription adds significant value if EEBO is compared with ECCO, which contains automatically-retrieved text.

As it often happens with large-scale projects involving multiple institutions (along with their local projects) across tens of years, catalogues and digital reproductions of printed materials can be accessed through different infrastructures by means of multiple research tools, so that the coextensiveness of scope is not always easy to ascertain. In the German case, along with the aforementioned discovery tool of ZVDD, the discovery tool K10 Plus⁴⁷ allows for search in catalogues and digital reproductions of a large number of historical printed books preserved in German libraries. Although records are indexed with consistent identifiers, thus ensuring retrievability of documents, users can seldom assess overlapping among the databases of different resources. Something similar happens as well in international collaborative efforts: it is not clear if the Universal Short Title Catalogue (USTC)⁴⁸ does include the records collected in different national or regional catalogues, such as the Short Title Catalogue of the Netherlands⁴⁹ (STCN) or the German regional catalogues accessible through K10 Plus. This seems to hold for other infrastructures provided by international consortia such as the Consortium of European Research Libraries⁵⁰ (CERL): the coverage of the Heritage of Printed Books⁵¹ (HPB) with respect to the aforementioned resources and catalogue is very difficult to assess.⁵²

Digital collections of sources can also be developed as scholarly bibliographies aiming at the collection of works or publishing venues belonging to specific academic fields. This is the case of the Philosophers’ Index⁵³ (PI) and of the International Philosophical Bibliography of the university of Louvain⁵⁴ (IPB), which both claim to index books and papers in philosophy and related fields from the beginning of 20th century

45. After a cyber-attack to the IT infrastructures of the British library ESTC is entirely available through a temporary hosting: “ESTC (Temporary)”. ESTC’s scope has enlarged, while previous versions of ESTC encompassed printed works only until 1700.

46. “Early English Books Online”.

47. “K10plus – Kooperationsprojekt BSZ und GBV”.

48. “Universal Short Title Catalogue (USTC) – About”.

49. “Short-Title Catalogue Netherlands”.

50. “CERL Resources”.

51. “Heritage of the Printed Book Database (HPB)”.

52. Noticeably, HPB can be searched by means of a dedicated K10 Plus web page which allows for sectioning results on the basis of library provenance, but systematic description of resources coverage does not seem to be available for the moment.

53. “Philosopher’s Index – The Premier Online Philosophy Database”.

54. “International Philosophical Bibliography – Répertoire Bibliographique de La Philosophie (IPB)”.

(the resources are available behind individual or institutional subscription). In this case as well, a comparison regarding coverage and an evaluation of overlapping is not straightforwardly available.

Finally, in some cases digitised archival records of career-related historical documents are available. Academic training and career can be particularly relevant on large-scale reconstructions of 19th and 20th century philosophical production.⁵⁵ Digital archival repositories of enrolments in specific university departments on national scale as well as of doctoral degrees granted in specific field can be of great use of contemporary historians. Concerning 19th-century French academic philosophy, for instance, the “Ès Lettres” project⁵⁶ (EL) indexes all doctoral theses in the humanities (actually, “ès lettres”, as the curricular path of academic training in philosophy). EL grants open access to records of theses indexed by means of stable identifiers provided by the Bibliothèque Nationale de France (BnF), which also offers online consultation of a number of digitised theses. Personal records of authors of doctoral theses are available and do reference stable identifiers. Similarly, although behind institutional or personal subscription, ProQuest web portal (PQ) provides access to metadata of PhD theses defended in the 20th century, allowing for sectioning on the basis of academic field. Although PQ’s sources are not as thoroughly indexed as EL’s, metadata are usually sufficient for chronological, geographical and institutional individuation of both theses and authors.

REFERENCES

- Online Resources* (last access February 22, 2024)
 CERL Resources, <https://www.cerl.org/resources/main>
 CERL Thesaurus, https://www.cerl.org/resources/cerl_thesaurus/main
 Constellate, <https://constellate.org>
 Digital d’Holbach – Voltaire Foundation,
<https://www.voltaire.ox.ac.uk/digital-enlightenment/digital-dholbach/>
 Digital Enlightenment – Voltaire Foundation, <https://www.voltaire.ox.ac.uk/digital-enlightenment/>
 Digital Voltaire – Voltaire Foundation,
<https://www.voltaire.ox.ac.uk/digital-enlightenment/digital-voltaire/>
 Early English Books Online, <https://quod.lib.umich.edu/e/eebogroup/>
 Early Modern Letters Online, <http://emlo.bodleian.ox.ac.uk/home>
 Eighteenth Century Collections Online,
<https://www.gale.com/intl/primary-sources/eighteenth-century-collections-online>
 Electronic Enlightenment Letters & Lives Online, <https://www.e-enlightenment.com/index.html>
 ENCCRE – Édition Numérique Collaborative et CRitique de l’Encyclopédie,
<http://enccre.academie-sciences.fr/encyclopedie/>
 Ès Lettres Les Thèses de Doctorat Ès Lettres En France Au XIXe Siècle – Version Bêta,
<https://eslettres.bis-sorbonne.fr/?db=eslettres&website&id=18050&pageid=18046#>
 ESTC (Temporary), <https://estc.printprobability.org/>
 FAIR Principles, <https://www.go-fair.org/fair-principles/>
 GitHub – OpenGreekAndLatin/First1KGreek: XML Files for the Works in the First Thousand Years of Greek Project, <https://github.com/OpenGreekAndLatin/First1KGreek>

55. See for instance Bonino & Tripodi 2020.

56. “Ès Lettres Les Thèses de Doctorat Ès Lettres En France Au XIXe Siècle – Version Bêta”.

- GitHub – PerseusDL/Canonical-LatinLit: XML Canonical Resources for Latin Literature, <https://github.com/PerseusDL/canonical-latinLit>
- Heritage of the Printed Book Database (HPB), <https://www.cerl.org/resources/hpb/main>
- Inkunabelkatalog INKA, <https://www.inka.uni-tuebingen.de/>
- Intelix Past Masters, <https://www.nlx.com/home>
- International Philosophical Bibliography – Répertoire Bibliographique de La Philosophie (IPB), <https://uclouvain.be/en/research-institutes/isp/ipb.html>
- ISTC (Incunabula Short Title Catalogue), <https://data.cerl.org/istc/>
- K10plus – Kooperationsprojekt BSZ und GBV, <https://www.bszgbv.de/services/k10plus/>
- Le Gazetteur Révolutionnaire, <https://gazetier-revolutionnaire.gazettes18e.fr/>
- Le Gazetteur Universel, <https://gazetier-universel.gazettes18e.fr/ressources-numeriques-sur-la-presse-ancienne>
- Liste Des Journalistes | Dictionnaire Des Journalistes, <https://dictionnaire-journalistes.gazettes18e.fr/>
- Liste Des Journaux | Dictionnaire Des Journaux, <https://dictionnaire-journaux.gazettes18e.fr/>
- Manus Online – OPAC SBN, <https://manus.iccu.sbn.it>
- Nietzsche Source – Home, <http://www.nietzschesource.org/>
- Open Greek & Latin, <https://www.opengreekandlatin.org/>
- PhiloLogic4 – The ARTFL Project, <https://artfl-project.uchicago.edu/philologic4>
- Philosopher’s Index – The Premier Online Philosophy Database, <https://philindex.org/>
- ProQuest Dissertations & Theses, <https://about.proquest.com/en/dissertations/>
- Public Databases – The ARTFL Project, <https://artfl-project.uchicago.edu/content/public-databases>
- Scaife Viewer – About, <https://scaife.perseus.org/about/>
- Short-Title Catalogue Netherlands, https://www.cerl.org/resources/stcn/stcn_en
- Text Creation Partnership, <https://textcreationpartnership.org/>
- The ARTFL Project, <https://artfl-project.uchicago.edu/>
- The Royal Society Early Letters – EMLO, <http://emlo-portal.bodleian.ox.ac.uk/collections/?catalogue=royal-society-early-letters>
- TLG – About, <http://stephanus.tlg.uci.edu/tlg.php>
- Tout d’Holbach, <https://artflsrv03.uchicago.edu/philologic4/tout-d-holbach/>
- Tout Voltaire – Voltaire Foundation, <https://www.voltaire.ox.ac.uk/voltaire-lab/tout-voltaire/>
- Tout Voltaire, <https://artflsrv03.uchicago.edu/philologic4/toutvoltaire/>
- Universal Short Title Catalogue (USTC) – About, <https://www.ustc.ac.uk/about>
- VD16Digital-CatalogofPrintedWorksofthe16thCenturyPublishedinGermanSpeakingCountriesofthe Bavarian State Library – MDZ, <https://www.digitale-sammlungen.de/en/collections/210/about>
- VD17 – Das Verzeichnis Der Im Deutschen Sprachraum Erschienenen Drucke Des 17. Jahrhunderts, <https://kxp.k10plus.de/DB=1.28/>
- VD18 – Das Verzeichnis Der Im Deutschen Sprachraum Erschienenen Drucke Des 18. Jahrhunderts, <https://vd18.k10plus.de/DB=1.65/>
- Voltaire Lab – Voltaire Foundation, <https://www.voltaire.ox.ac.uk/voltaire-lab/>
- Wikidata, https://www.wikidata.org/wiki/Wikidata:Main_Page
- Women’s Early Modern Letters Online – WEMLO, http://emlo-portal.bodleian.ox.ac.uk/collections/?page_id=2595
- Zentrales Verzeichnis Digitalisierter Drucke, <https://www.zvdd.de/startseite/>

Studies

- Barrett, P. H. and Freeman, R B. (eds.) 2010. *The Works of Charles Darwin*, 29 vols., New York, New York University Press.
- Bonino, G. and Tripodi, P. 2020. “Academic Success in America: Analytic Philosophy and the Decline of Wittgenstein”, *British Journal for the History of Philosophy*, 28/2, pp. 359-392, <https://doi.org/10.1080/09608788.2019.1618789>
- Burrows, S., Roe, G. and Baker, K. M. (eds.) 2020. *Digitizing Enlightenment: Digital Humanities and the Transformation of Eighteenth-Century Studies*, Liverpool, Liverpool University Press.

- Colli, G. and Montinari, M. (eds.) 1967. *Nietzsche. Werke. Kritische Gesamtausgabe*, 45 vols. Berlin & New York, De Gruyter.
- Colli, G. and Montinari, M. (eds.) 1975-1984. *Nietzsche Briefwechsel. Kritische Gesamtausgabe*, 24 vols., Berlin & New York, De Gruyter.
- D'Iorio, P. 2010. "The Digital Critical Edition of the Works and Letters of Nietzsche", *The Journal of Nietzsche Studies*, 40/1, pp. 70-80, <https://doi.org/10.5325/jnietstud.1.40.0070>
- D'Iorio, P. and Barbera, M. 2011. "Scholarsource: A Digital Infrastructure for the Humanities", in *Switching Codes*, edited by T. Bartscherer, Chicago, University of Chicago Press, pp. 61-88.
- De Beer, E. S. (ed.) 1976-1989. *The Correspondence of John Locke*, 8 vols., Oxford, Oxford University Press.
- De Bolla, P. 2013. *The Architecture of Concepts: The Historical Formation of Human Rights*, Fordham, Fordham University Press
- De Bolla, P., Jones, E., Nulty, P., Recchia, G. and Regan, J. 2020. "The Idea of Liberty, 1600-1800: A Distributional Concept Analysis", *Journal of the History of Ideas* 81/3, pp. 381-406, <https://doi.org/10.1353/jhi.2020.0023>
- Gombay, A. (ed.) 2001. *Oeuvres complètes de René Descartes* ("Past Masters"), Charlottesville (Va), IntelLex Corporation.
- Gregg, S. H. 2021. *Old Books and Digital Publishing: Eighteenth-Century Collections Online*, Cambridge, Cambridge University Press, <https://doi.org/10.1017/9781108767415>
- IFLA Study Group on the Functional Requirements for Bibliographic Records, 2009. *Functional Requirements for Bibliographic Records*. n.d.: International Federation of Library Associations and Institutions (IFLA).
- Moretti, F. 2000. "The Slaughterhouse of Literature", *MLQ: Modern Language Quarterly*, 61/1, pp. 207-227, <https://muse.jhu.edu/pub/4/article/22852>
- Petrovich, E. and Buonomo, V. 2018. "Reconstructing Late Analytic Philosophy. A Quantitative Approach", *Philosophical Inquiries*, 6/1, <https://doi.org/10.4454/philing.v6i1.184>
- Petrovich, E. 2022. "Acknowledgments-Based Networks for Mapping the Social Structure of Research Fields. A Case Study on Recent Analytic Philosophy", *Synthese* 200/3, p. 204, <https://doi.org/10.1007/s11229-022-03515-2>
- Rooks, M. C. (ed.) 1995. *The Philosophical Works and Selected Correspondence of John Locke* ("Past Masters"), Charlottesville (Va), IntelLex Corporation.
- Sgard, J. (ed.) 1991. *Dictionnaire Des Journaux, 1600-1789*. Paris & Oxford, Universitas & Voltaire Foundation.
- Sgard, J., Gilot, M. and Weil, F. (eds.) 1976. *Dictionnaire Des Journalistes, 1600-1789*, Grenoble, Presses Universitaires de Grenoble.
- Tolonen, M., Mäkelä, E. and Lahti, L. 2022. "The Anatomy of Eighteenth Century Collections Online (ECCO)", *Eighteenth-Century Studies* 56/1, pp. 95-123, <https://doi.org/10.1353/ecs.2022.0060>
- Voltaire, 1968-2022. *Œuvres complètes de Voltaire*, 205 vols., Oxford, Voltaire Foundation.
- Wilkinson, M. D., Dumontier, M., Jan Aalbersberg, I., Appleton, G., Myles, A., Baak, A., Blomberg, N., et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship", *Scientific Data*, 3 /1, art. n. 160018, <https://doi.org/10.1038/sdata.2016.18>

Nicola Ruschena
North-West Italian Philosophy PhD Program (FINO)
nicola.ruschena@unito.it
ORCID: 0009-0008-0817-5733