

Burckhardtsource.org. Eine semantische digitale Edition der Briefe an Jacob Burckhardt

Burckhardtsource.org. A semantic digital edition of the correspondence to Jacob Burckhardt

Susanne Müller
Scuola Normale Superiore / ERC
P.zza Cavalieri 7, Pisa, ITALIEN
E-Mail: susanne.muller@sns.it, Internet: <http://www.burckhardtsource.org>

Francesca Di Donato
Scuola Normale Superiore / ERC
P.zza Cavalieri 7, Pisa, ITALIEN
E-Mail: francesca.didonato@sns.it, Internet: <http://www.burckhardtsource.org>

Zusammenfassung:

Dank des ERC Advanced Grant Project EUROCORR sollen die Briefe europäischer Intellektueller an den Schweizer Kulturhistoriker Jacob Burckhardt gesammelt, erschlossen und auf der für das Vorhaben geschaffenen Plattform Burckhardtsource.org ediert werden. Dieser Artikel versteht sich als eine Einführung und Beschreibung in das Projekt; im letzten Teil wird die Neuigkeit dieser Plattform, d.h. die Möglichkeit eines semantischen Kommentars vorgestellt.

Abstract:

Burckhardtsource.org is the platform created by EUROCORR ERC Advanced Grant Project, which aims at mapping and producing a critical edition of the extensive correspondence of European intellectuals with the Swiss cultural historian Jacob Burckhardt over a period of more than half a century. This article is an introduction to EUROCORR project and a description of the platform Burckhardtsource.org. The final section is dedicated to the most innovative aspect of the platform, that is the advanced tool to semantically annotate transcriptions.

1. EUROCORR Project

The purpose of the EUROCORR¹ project is to publish in open access a critical edition of the extensive correspondence of European intellectuals with the Swiss historian and art historian Jacob Burckhardt (1818–1897) over a period of more than half a century. While Burckhardt's letters have already been published in 10 volumes,² the other side of the dialogue has not been disclosed yet. The correspondence to Jacob Burckhardt, with its over 300 authors, represents a valuable source that bears witness to a historical period characterised by important cultural changes such as: the advance of industrial culture; the development of the modern concept of democracy; the establishment of art history as an autonomous discipline, separate from history or archaeology. Basel and Burckhardt turn out to be a nodal point where animate debates converge. Completing the correspondence means to reconstruct one of the most important written exchange

¹ The ERC-project, Advanced Grant EUROCORR, Grant Agreement n. 249483, is coordinated by Maurizio Ghelardi (Scuola Normale Superiore, Pisa), the research group involves 6 people: 3 Humanists, 2 developers and 1 'digital humanist' (see the Team page on www.burckhardtsource.org).

² Jacob Burckhardt, Briefe, vollständig und kritisch bearbeitete Ausgabe mit Benützung des handschriftlichen Nachlasses hergestellt von Max Burckhardt, 11 Bde., Basel 1949–1994.

of the 19th century and to make it available for further researches.³ Indeed, in spite of their documentary value, letters are used by researchers for proving thesis or to develop new research questions only if they are edited. The reason lies in the specific complexity of this source that is dispersion of the letter corpus, conservation situation, palaeography etc.).

With critical edition we refer to an accurate version of the text with an apparatus both analysing the source material and recording all editorial interventions, while semantic annotations enhance the understanding of the text content. Its function follows two principles: to keep the connection to the past, and at the same time, to look forward to the future. On the one hand an edition aim is to reconstruct the document in the most accurate way, on the other the text representation has to be up to date and should offer many potentialities for further researches.⁴

To align with these purposes EUROCORR seeks to answer scholarly questions in philology and in edition by using computer technologies. It is obvious that different resources and tools require methodological reflection and even if the digital edition follows the principles of the traditional philology, the result will inevitably offer new perspectives. Indeed, the use of computational methods enhances the analysis of the data, improves transparency and entail the possibility to check and verify the procedure of editing; in other words, a digital edition fulfils the needs of a critical approach so that each reader can pursue his own path.⁵

Both aspects of digital publishing – producing a critical edition and reproducing the source in a suitable manner – are determined by the text type. In our specific case we deal with letters belonging to the 19th century or with biographical and cultural documents written under specific socio-historical circumstances. The corpus, whose reconstruction represents the editor's first task, comprises about 300 senders and about 1000 preserved letters.

The project has three principal goals:

- a) to produce a trustworthy critical Edition, whose texts are philologically exact and easy to consult ;
- b) to create a digital platform (Burckhardsource.org) that interrelates all texts and data and offers the connective resources to and from the World Wide Web;
- c) to provide long term use and preservation.

A letter is a historical record and is therefore treated as such, so the project's philological approach observes a materialistic idea of the text: the transcription and the constitution of the text are as close as possible to the manuscript and they do justice to the single document. Original spellings (and misspellings) are kept: editors don't emend, nor smooth. This decision is even more fundamental because in the 19th century spelling rules didn't exist yet.⁶ Nevertheless, when constituting and editing the text, the editor's task is both to recognize peculiar signs and to understand and take into consideration the author's intentions.⁷

³ In the near future we hope to be able to complete the digital edition with the letters written by Jacob Burckhardt.

⁴ @@@@

⁵ Cfr. M. N. Smith, *Electronic Scholarly Editing*, in: *A Companion to Digital Humanities*, ed. S. Schreibman, R. Siemens, J. Unsworth, Oxford 2004, p. 306-322. P. Sahle, *Zwischen Mediengründenheit und Transmedialisierung. Anmerkungen zum Verhältnis von Edition und Medien*, in: *editio*, 2010 (24), S. 23–36.

⁶ A single author often uses different graphemes for the same phoneme, e.g. for the double –ss–.

⁷ Of course all editorial decisions are documented in the critical apparatus and the rules for transcription are shown in the *Transcription and Encoding Compendium*, see <http://wiki.burckhardtsource.org>.

2. Burckhardtsource.org

To meet the principle of transparency and to satisfy every possible query Burckhardtsource.org shows three distinct texts:

- a) facsimile
- b) diplomatic transcription
- c) edited text

The platform provides the reproductions of the manuscripts thus showing its effective materiality. The quality of the facsimiles and the zoom function allow an easy reading of the original. Since scholars and researchers don't always need to examine the originals, one of the main purposes of the visualisation is to protect and preserve the manuscripts.

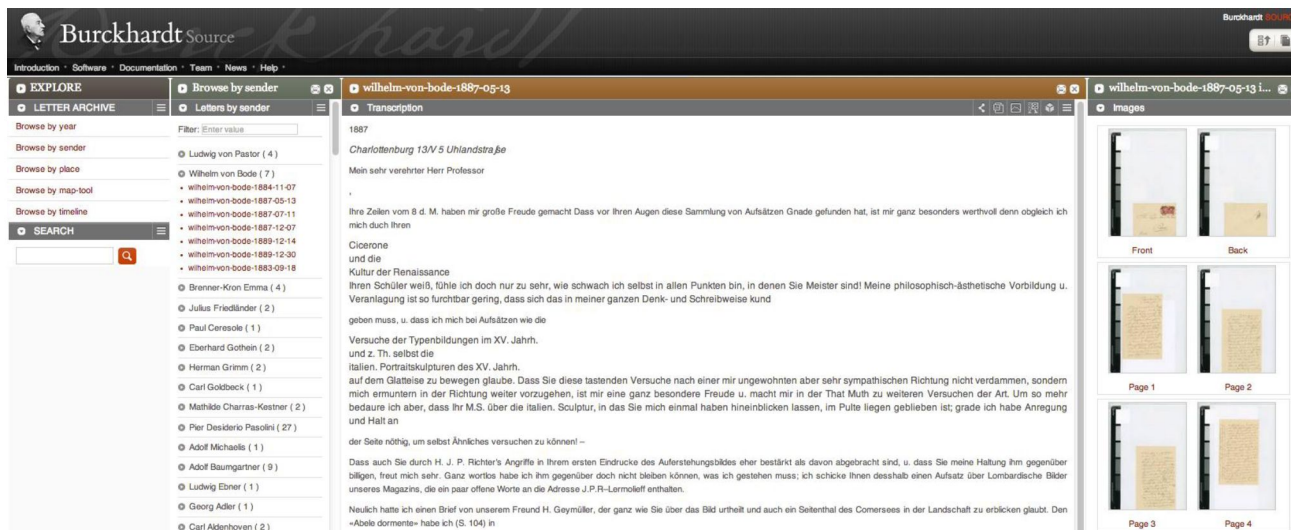


Fig. 1 Burckhardtsource platform:

a view showing the explore and browsing options, a letter transcription and facsimile images

The diplomatic representation offers a line per line visualization with the author corrections, afterthoughts and deletions, while the philological apparatus attests both writer's corrections and any editorial decision. Making the transcription and encoding procedures intelligible means to provide scholars with a check tool and to confer them an active part.

The edited text in html (on the platform also available in XML) represents a continuous text illustrating the author's (virtual) fair copy (clear-text).⁸ This version corresponds to the default mode of visualization and contains the semantic annotations, thanks to which the letter is contextualized in its socio-historical environment. Indeed, on Burckhardtsource.org any single letter is connected with all information levels, i.e. metadata and semantic annotations, as well as with the rest of the correspondence to Burckhardt.

While a printed edition offers to the reader an unchangeable text, representing different versions of the same letter emphasizes the editorial process and its reversibility. The multi-faceted representation of Burckhardtsource.org shows a new concept of critical edition: the idea is to leave the path of one single fixed text and to propose several possible ways of seeing and understanding.

In addition, thanks to the adoption of Linked data technologies, texts can share information and results with other open platforms. The edition goes beyond the mere accumulation and simple classification or management of data. In other words, as Burckhardtsource.org aims at

⁸ The encoding of the transcription follows the Text Encoding Initiative (TEI) - P5 standards, albeit in some specific cases it was necessary to adapt the guidelines to the structural and morphological specific needs of a letter.

representing the opposite of a “monolithic” and fixed edition, letters won’t be closed nor self-referential.

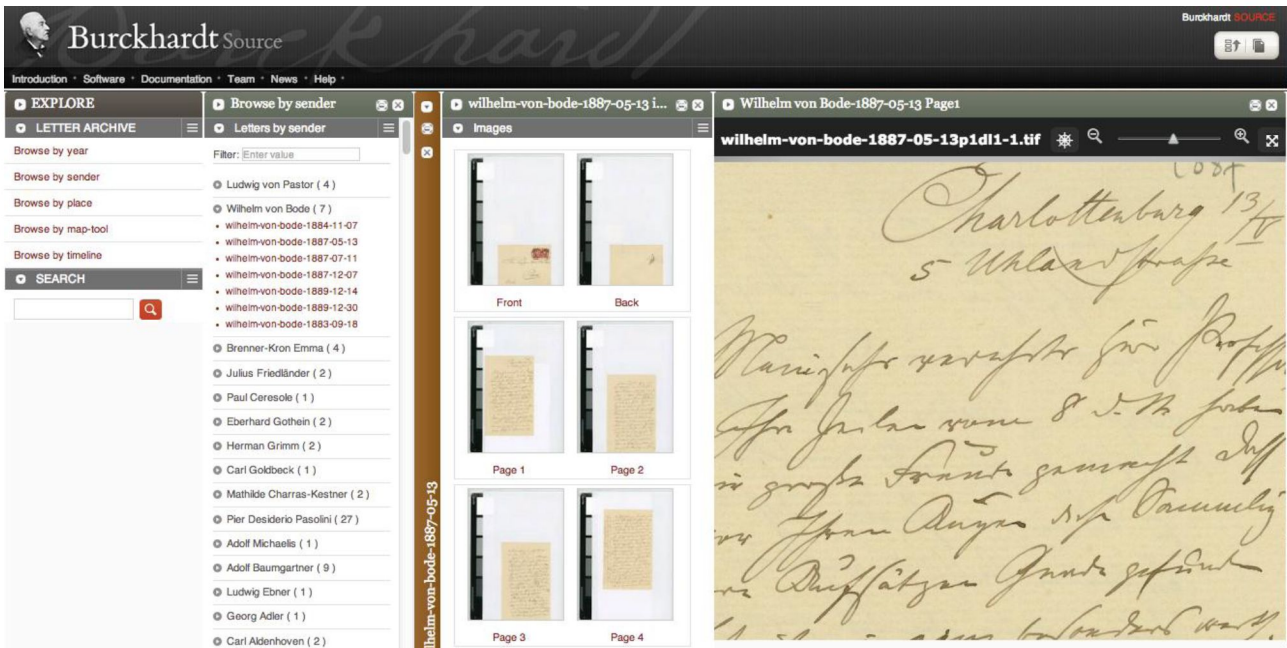


Fig. 2 On the right box, a manuscript view

Burckhardtsource.org metadata section describes the physical characteristics of the text support and any letter-specific element like envelope or attachments. It indicates place and compilation date, conservation status and the provenance (Archive or Institution). Furthermore the platform also indicates the right holders of the manuscripts as well as any earlier edition. Through the metadata we may reconstruct the correspondence, and provide each letter with an accurate description of its circumstances (e.g. is this letter a response to an epistle of B. and/or did B. ever answered to it?). Finally, a short abstract summarises the content of the letter.

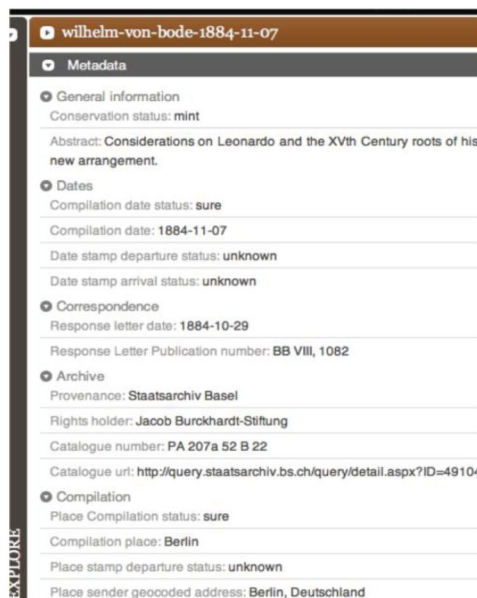


Fig. 3 Metadata section on Burckhardtsource.org platform

3. Semantic Annotation

The most innovative aspect of Burckhardtsource.org platform lies in the possibility of semantically annotate the transcriptions. The aim of semantic annotations is to give meaning and structured information to text parts. If we consider texts and editions like a store room of knowledge which we can share with others, it is important to use norm-data⁹ or to structure the information in a standard way, using Linked Data technologies.

Our commentary will be limited to four types of entities: persons, places, artworks and publications quoted by Burckhardt's correspondents.

At the present time persons, places, and works of art cited in the letters have been annotated. The annotation process will continue until the end of the project.

The choice to annotate these four classes of entities depends on particular research questions. For example, as the meaning of a letter is strictly related to the relationship between sender and receiver, to understand it it's essential to know as much as possible about the identity and the biography of Burckhardt's correspondents.

From a technical point of view, all the entities chosen from the texts build the vocabularies of the research group. When possible, entities are extracted from already existing Linked Data providers (such as Freebase, Dbpedia, etc); when missing, editors have enriched the existing providers with new entries. This part of the work requires an accurate research, that is preliminary to the annotation of the platform contents. For example, localization or attribution of an artwork is very often a difficult task for scholars, since all the information about it (metadata) might have been possibly changed over time or because the artwork is lost.

The annotation task is performed through Pundit, an open source research tool that allows to create semantic annotations on web pages¹⁰. Thanks to Pundit, the research group fastens external links and makes the concepts public and sharable.

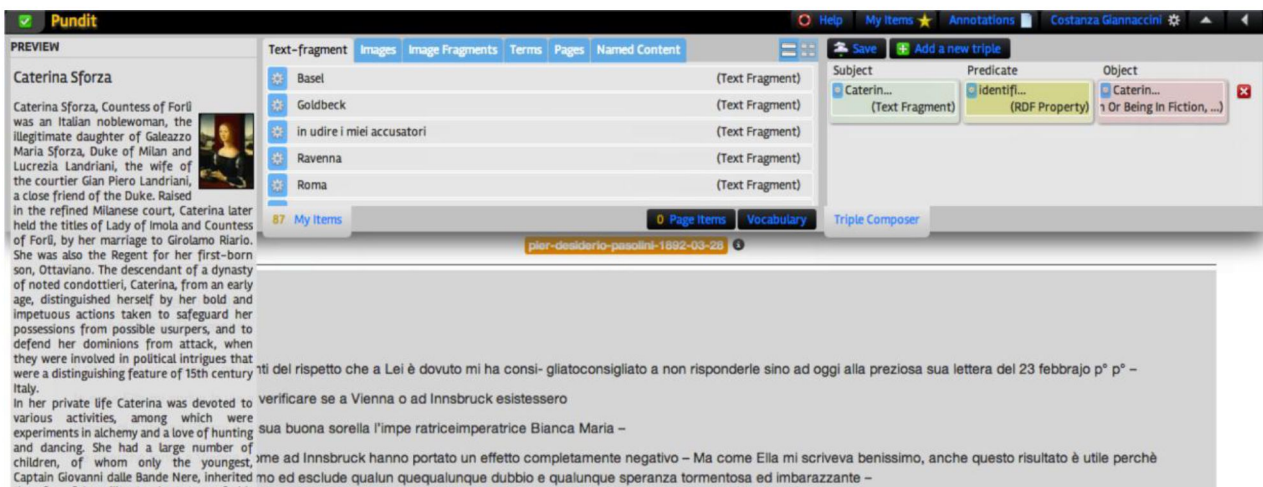


Fig. 4 The process of annotating

In the example above the annotation concerns a text fragment containing a person, Caterina Sforza. On the right side is shown the triple composer: through a predicate ("identifies") a text fragment is there associated to the right "Caterina Sforza" item which is chosen from a specific and recognised vocabulary, such as Freebase or Dbpedia. On the left side of Pundit interfaces is shown a preview of the entity on Freebase. Predicates are customized depending on specific research needs, and vary according to the typology of concepts annotated. Annotations are then directly visible on the web page.

⁹ For persons, for instance, we used the standardised data of the Deutsche National Bibliothek.

¹⁰ See <http://www.thepund.it>.

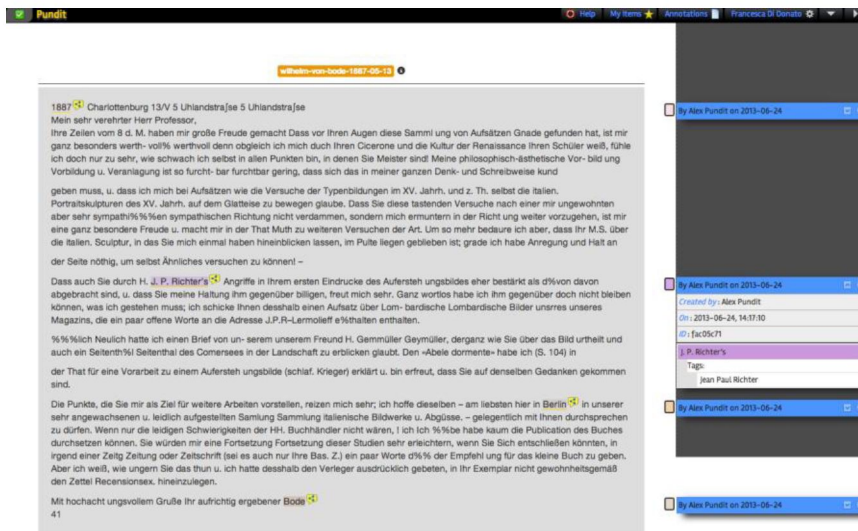


Fig. 5 Annotations in a web page

Ad hoc visualizations have been developed in order to meet Scholars needs. A first one shows a graph of Burckhardt's correspondents and persons cited in the letters (relations types can be visually distinguished). Nodes dimension depend on the volume of letters sent and on citations received in letters written by others.

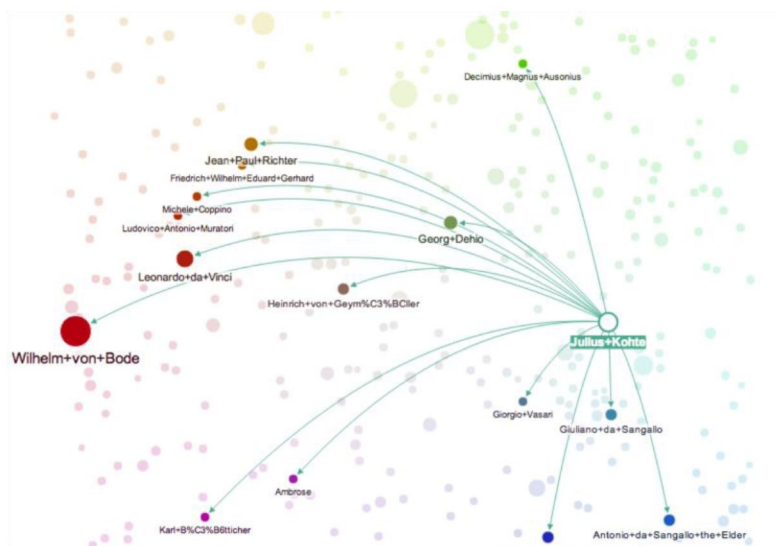


Fig. 6 Graph of persons in the correspondence to Jacob Burckhardt

A second visualization sets the letters on a timeline. While the timeline at the bottom locates the document in the context of the whole correspondence, on the right side of the page the reader immediately finds places and persons commented in the letter. By clicking on them, the commentary will be disclosed.

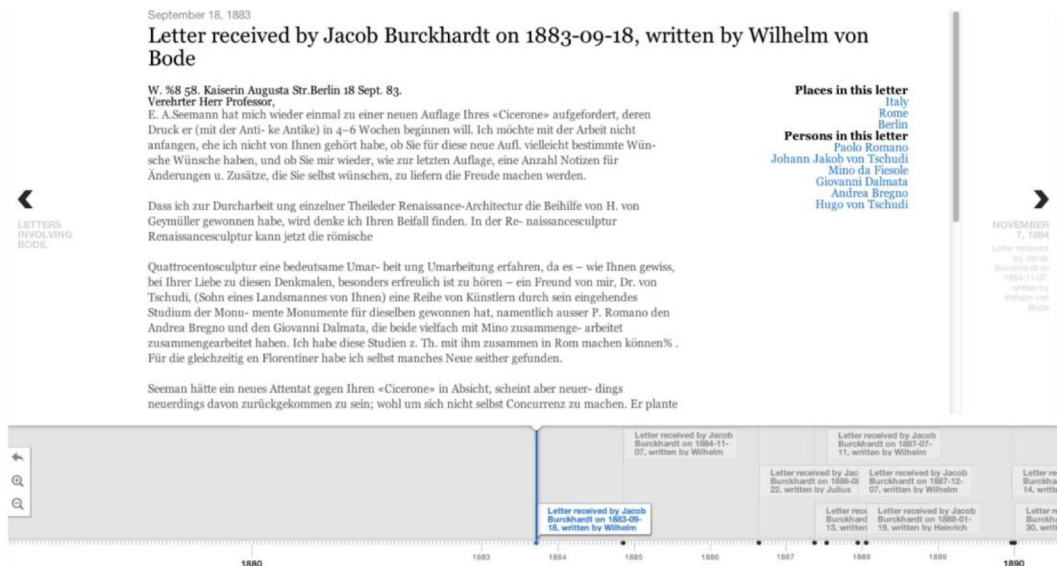


Fig. 7 Letters on a timeline

It is clear then that to fully exploit Linked Open Data potentialities it would be necessary a wider adoption of such standards by Libraries, Archives and Museum – while their closure prevents research projects such as EUROCORR from obtaining all the benefits that LOD technologies make possible.

Although the benefits of adopting Linked Open Data standards and tools are evident, entities that we expected to annotate in order to answer to EUROCORR research questions are often missing in Linked Data providers databases. For example, EUROCORR scholars made specific researches for about 600 persons, most of which were absent in LOD database and have then been added to Freebase, thus enriching the Web of Data with new entities. As for the artworks we met the same difficulties: actually only very few public art museums (such as the British Museum) put their data at web-user's disposal in the right formats. This is a delay that European public institutions should fill as soon as possible, in order to promote the enhancement of our cultural heritage.