International Journal of Information Science & Technology – iJIST, ISSN : 2550-5114
Vol. 9 - No. 1 - March 2025

Implementing Collaborative Digital Scholarly Editions: Insights from *Bellini Digital Correspondence*

Salvatore Cristofaro[1], Angelo Mario Del Grosso[2], Laura Mazzagufo[3], Pietro Sichera[1], Daria Spampinato[3]

[1]*CNR Istituto per il Lessico Intellettuale Europeo e Storia delle Idee*
Roma, Italy
Email: salvatore.cristofaro@cnr.it, pietro.sichera@cnr.it

[2]*CNR Istituto di Linguistica Computazionale "A. Zampolli"*
Pisa, Italy
Email: angelomario.delgrosso@cnr.it

[3]*CNR Istituto di Scienze e Tecnologie della Cognizione*
Catania, Italy
Email: laura.mazzagufo@istc.cnr.it, daria.spampinato@cnr.it

***Abstract*—This paper presents the work carried out within the *Bellini Digital Correspondence* project. Specifically, it outlines the activities aimed at improving the harmonization process, regarding the production of collaborative digital scholarly editions. The study proposes a preliminary semantic approach to the harmonization process, exploiting an ontology-based mapping perspective. The article also illustrates the functionalities and practicalities of the NormaTEI tool, designed to streamline harmonization processes within the *Bellini Digital Correspondence* project and applicable to similar initiatives.**

***Index Terms*—Cooperative Digital Scholarly Edition, Schema-matching, Harmonization, NormaTEI, Bellini Digital Correspondence**

I. Introduction

Digital scholarly editing is a scientific activity that involves both the formal modeling and the digital representation of a piece of text conveyed by one or more documents (commonly referred to as primary sources or witnesses). These documents, along with their corresponding text, hold significant value from a humanistic perspective. To achieve this goal, descriptive markup languages, such as XML, are employed to produce the editorial outcome (i.e., the Digital Scholarly Edition, hereinafter referred to as DSE). Over time, shared and community-driven XML schemata have been defined to establish best practices and ensure interoperability, as promoted by the Text Encoding Initiative (XML/TEI).[1] Distributed technologies and Web accessibility make it possible to collaboratively create DSEs (see section II). At this point, syntactic and semantic issues arise when encoding the text resources coherently. Indeed, harmonizing/normalizing XML/TEI documents, encoded by different scholars, poses significant challenges (see section III).

Methodologies and tools can be employed to address the harmonization process of encoding from a syntactic perspective [1]. In the scientific literature, this challenge is also addressed in the field of knowledge representation and management [2]. Diffing algorithms have been developed to identify divergent locations between XML documents, both in terms of their encoded structure and content [3]. Furthermore, mapping strategies between formal schemata and conceptual categories have been explored in order to align the structure of various XML documents [4].

Indeed, merely adopting the TEI guidelines is not sufficient to expedite the representation of a textual phenomenon: the diverse solutions permitted by XML/TEI may vary in expression but be equivalent in terms of content. The various syntactic structures can be mapped to a single encoding model. However, this mapping process is much more challenging from a semantic standpoint. Indeed, it is not feasible to formally define the semantic value of a given phenomenon using only natural language. The scientific validation of the encoding documents enables us to explore additional methodological solutions that can be better addressed through an ontological formalization of textual representation, such as an OWL-based ontology leveraging *description logics* [5] (see section IV). In this contribution, we posit that the harmonization process of a collaborative DSE would greatly benefit from the exploitation of semantic technologies, such as formal ontologies. We will demonstrate some strategies implemented to automate, to the greatest extent possible, the process of encoding analysis within the context of a digital scholarly project named *Bellini Digital Correspondence* (see section V).

The encoding phase of this project followed the typical methodology for collaborative DSEs in educational contexts, as detailed in sections II and V-A. The collaborative work often provided a variety of encoding choices for the textual phenomena found in the epistolary texts. To address this multiplicity of options and facilitate structural harmonization

[1] https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html

(see section V-B), the research team designed and developed the NormaTEI tool (described in section VI). Section VII will showcase the use of NormaTEI within the *Bellini Digital Correspondence* project, reporting also quantitative evaluation regarding the harmonization process applied to the XML/TEI files of the letters.

II. Digital Scholarly Editing

The scholarly editing process "forces" scholars into a hermeneutical circle [6]. This approach involves critical choices that start reading the text and lead to potential interpretations. These interpretations are then refined iteratively by revisiting the text itself for new insights, perpetuating this process indefinitely. While this approach has deep scientific roots, contemporary scholars have also access to a wealth of tools and technologies that were unavailable in the past.

A DSE is the result of a process of text modeling that overcomes the constraints inherent in previous print technology. Indeed, textual representation no longer relies on typographic conventions (implicit, symbolic, and mostly arbitrary), but instead relies on explicit and formal meta-annotation defined by international standards (such as XML/TEI). One of the most distinctive features about making cutting-edge DSEs lies in their cooperative and collaborative nature. This possibility paves the way for creating the edition by harnessing the work of multiple collaborators editing simultaneously on the text.

This new editing model is, by its nature:

1) collaborative (involving different parties working on the same resource and sharing the same activities),
2) cooperative (involving different parties working on multiple resources and in different activities),
3) distributed (taking place in different locations and at different times).

This articulated process creates a complex system where the curator's critical evaluation of the text interacts with the edition curator's choices, further enriched by the contributions of a broad community of encoders. This synergy is the added value of a product that embodies the nature of the digital environment, constantly evolving and improving.

The collaborative and cooperative properties of a DSE require additional efforts to ensure consistency and coherency across the whole edition. To maintain rigorous scientific standards, collaborative editing necessitates that all updates, corrections, and edits adhere to the established model and encoding criteria. Additionally, preliminary examination by the editorial team is crucial. Given that looking at individual encoding interventions can be time-consuming and challenging, this process becomes particularly difficult without the aid of customized tools that can also track changes and accountability in history and authorship (*blame*[2]).

The availability of a formal model and a standard vocabulary is not sufficient to fully automate the encoding process. Even with valid editions, there are various semantic nuances and interpretive intentions, all of which are correct, but they result in different markup solutions [7], [8].

Encoding normalization is a well-documented issue in the literature, as it involves several levels of interpretation:[3]

> "*TEI encoding offers a wide variety of markers and often more than one marker for the same phenomenon: it is therefore necessary for each editorial project to choose which textual phenomena to annotate and how, which ones to ignore and why. This choice is not mechanical; on the contrary, it is the formal answer to the scientific research questions developed by the project; for example, in the case that the text in question contains abbreviations, will it be necessary to encode them or expand them silently? TEI offers at least five different ways to expand abbreviations, each of which has slightly different functionality and purpose* [...]" [9]

III. Schema and encoding normalization

Textual encoding harmonization – also referred to as normalization – constitutes a scholarly activity aimed at restructuring the various encoding solutions to ensure the highest degree of uniformity and coherence throughout the entire edition. The harmonization process is a complex activity due to the possibility of expressing the same phenomena syntactically and lexically using different XML structures while maintaining TEI conformance.[4] Indeed, the TEI guidelines afford the editor a wide variety of possible solutions, making it challenging to constrain the definitive schema definition during the early stages of the modeling process. Consequently, a textual resource encoded using XML/TEI guidelines is not inherently predetermined in terms of its structure and markup labels. Conversely, it is the outcome of a strong interpretative act made by the editor. For example, the TEI vocabulary provides multiple ways to encode the name of a person, such as using the `<rs type="person">` element, or the `<name type="person">` element, or the `<persName>` element. The same considerations can be applied to several other phenomena, including the encoding of footnotes (in the `<back>` section or inline) and generic cross-references (`<ref>` or `<ptr>`). A dual case occurs when quite different phenomena can be encoded using the same element with different attributes. Some examples can be found in the usage of the `<div>` element or the usage of the `<gap>` element.

In projects that focus on manuscript sources, the usage of the `<gap>` element[5] has been particularly notable. The `<gap>` element is used in conjunction with the

[2]For instance see the command `git blame` at https://git-scm.com/docs/git-blame.

[3]The original quoted text is in Italian; consequently, the English translation has not been verified by the original authors. It is merely our interpretation of the text.

[4]https://tei-c.org/release/doc/tei-p5-doc/en/html/USE.html#CF

[5]In the TEI guidelines, `<gap>` "indicates a point where material has been omitted in a transcription, whether for editorial reasons described in the TEI header, as part of sampling practice, or because the material is illegible, invisible, or inaudible" (https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-gap.html).

`@reason` attribute (describing the cause of omission by selecting values among {"illegible", "lost", "deleted", "editorial"}) and with the description of the extent of the missing text, which can be expressed in three different ways: 1) `@extent` attribute 2) `@quantity` and `@unit` attributes 3) `@atLeast`-`@atMost` and `@unit` attributes.

```
<gap reason="illegible"
   quantity="5" unit="char"/>
<gap reason="lost"
   atLeast="4" atMost="5" unit="char"/>
<gap reason="deleted"
   atLeast="4" atMost="5" unit="char"/>
<gap reason="editorial"
   extent="2 pages"/>
```

Listing 1. Combination of texttgap elements with their attributes

Listing 1 demonstrates the usage of the <gap> element to encode the following scenarios:

1) illegible text but present in the primary source (precise textual span);
2) lost text that is no longer present in the primary source (range textual span);
3) deleted text that is present in the primary source (range textual span);
4) text that is legible in the primary source but not transcribed by the editor (precise textual span).

Simultaneously, the <gap> element may be utilized by various encoders with different attributes to encode the same phenomenon (e.g., concerning the description of the extent of the missing text, as mentioned above), without compromising TEI conformance.

Indeed, distributed work for collaborative DSEs faces challenges with heterogeneity in representational choices for the same textual phenomena or classes of phenomena. This heterogeneity can be mitigated by defining constraints and data types within the encoding schemata, as well as by strictly following the separation between data representation and data presentation. [10]. Another example of a common textual phenomenon in DSE texts is the abbreviation, which is encoded using the <abbr> tag. Once again, combinations of tags and attributes allow for a wide choice of encoding solutions (Listing 2).

```
<abbr>Sign:</abbr>

<abbr type="title">Sign:</abbr>

Sign<ex rend=":">or</ex>

<choice>
   <abbr>Sign:</abbr>
   <expan>Signor</expan>
</choice>

<choice>
   <abbr>Sign<am>:</am></abbr>
   <expan>Signor</expan>
</choice>

<choice>
   <abbr>Sign<am>:</am></abbr>
   <expan>Sign<ex>or</ex></expan>
</choice>
```

Listing 2. Combination of `abbr` and `expan` elements

To address these kinds of ambiguities, the use of XML diffing algorithms[6] is not entirely applicable for several reasons:

- the XML documents describing different textual units can be highly dissimilar, making direct comparison challenging;
- the same value could be correct for tags with different paths;
- the same values can be correct for different tags, identical attributes on different tags, or regardless of the attribute name.

For this reason, we chose to pursue two separate directions, addressing both the syntax and semantics of the TEI encoding through an *ontology mapping* strategy.

IV. Ontology Mapping Perspective

From an ontological point of view, a TEI tag could be conceptually described as a *class* C coupled with a number of *properties* $P_1, P_2, \ldots, P_n$ each of which represents either: (a) the fact that the tag owns a given attribute (*property of type 1*); or (b) the fact that the tag is an *immediate descendant* (i.e., a *child*) or the *immediate ancestor* (i.e., the *father*) of another tag within the *tree structure* of an XML/TEI document (*property of type 2*).[7] TEI tags can own different (combinations of) attributes each ranging over a specified set of possible values which leads to many different types of *tag instances*. Moreover, a tag can have various immediate descendants (nested tags) but the immediate ancestor of a tag is unique. For a concrete example, let us consider, once again, the case of the tag <gap>. Disregarding kinship relationships (i.e., properties of type 2), let us make further the simplifying assumptions that there are only two mutually exclusive types of <gap> instances, where any <gap> instance either involves the two attributes `@reason` and `@extent` or it involves the four attributes `@reason`, `@atLeast`, `@atMost`, and `@unit`, with the *literal sets* {"illegible", "lost" "deleted", "editorial"}, {"char"} and {"2 pages"} as the sets of possible values for the attributes `@reason`, `@unit`, and `@extent`, respectively, and with the attributes `@atLeast` and `@atMost` ranging over integer numbers (see Listing 1 above). Such a simplified scenario easily leads to the ontological representation depicted in Fig. 1, which we refer to below as the OntoGap ontology.

Notice that the *data properties* `hasReason`, `hasExtent`, `hasUnit`, `hasAtLeast` and `hasAtMost` (property of type 1) correspond to the attributes `@reason`, `@extent`, `@unit`, `@atLeast` and `@atMost`, respectively. Moreover, each data property is a *functional property*, meaning that an attribute can acquire only one value per single instance. The *main class* `GAP` is the *disjoint union* of the two classes `GAP1` and `GAP2`

[6]This remark concerns both structural and content control algorithms, based on the distinction made in literature [1].

[7]Notations and terminology used in this section follow that of [5], up to minor details.

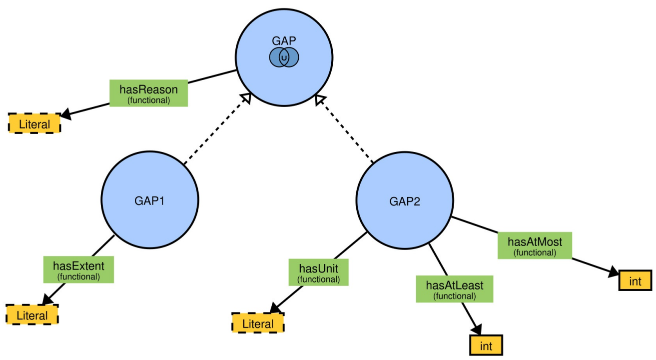


Fig. 1. Ontological representation of a simplified variant of the tag `<gap>`.

which represent the two different types of instances of the tag `<gap>`. (Note here that disjointness of classes `GAP1` and `GAP2` implements mutual exclusiveness of `<gap>` instance types). According to the above simplifying assumptions about the set of possible values of the attribute `@reason` (resp., `@unit` or `@extent`), the class `GAP` (resp., `GAP1` or `GAP2`) is a *sub-class* of the *inverse image* of the literal set {“illegible”, “lost” “deleted”, “editorial”} (resp., {“2 pages”} or {“char”}) under the data property `hasReason` (resp., `hasExtent` or `hasUnit`).[8] The class `GAP` (resp. `GAP1`) is the *domain* of the data property `hasReason` (resp. `hasExtent`), whereas data properties `hasUnit`, `hasAtLeast` and `hasAtMost` all have the class `GAP2` as their domain. [9]

Observe that the case of `<gap>` instances of the form

```
<gap reason="illegible" quantity="n"
    unit="char"/>
```

(see Listing 1) could be covered similarly as above, namely, one introduces a further sub-class `GAP3` of `GAP`, disjoint from `GAP1` and `GAP2`, and adds the property `hasQuantity` (with `GAP3` as its domain) to mimic the attribute `@quantity`.

For the sake of completeness we report in Fig. 2 the formal representation in *OWL Functional-style Syntax* of the (main parts of) the OntoGap ontology.[10]

Basically, a formal ontology is intended to describe the knowledge about properties of, and relationships between *individuals* of a given *universe of the discourse*. A peculiar characteristic of formal ontologies (at least nowadays) is that one can *reason* over them, inferring new knowledge, automatically. Starting from explicitly stated facts – the *asserted facts*, or *axioms* – about the *entities* of the universe of the discourse (i.e., the properties of the individuals and the relationships between them), new facts – the *inferred facts* – about such entities can be “mechanically” deduced to hold by means of logical rules of inference.[11]In the case of our OntoGap ontology describing the universe of the instances of the tag `<gap>`, we have (e.g.) the asserted fact that `GAP` is the disjoint-union of `GAP1` and `GAP2`. Then, when we reason over the ontology, from this fact we can deduce that: **(1)** the classes `GAP1` and `GAP2` are disjoint (i.e., they cannot share anything in common); and, in particular, **(2)** both `GAP1` and `GAP2` are sub-classes of `GAP`, namely, everything in `GAP1` is in `GAP`, and, similarly, everything in `GAP2` is in `GAP` as well. Therefore, once we assert within the ontology the assumed fact that all instances in `GAP` have the fixed literal set {“illegible”, “lost” “deleted”, “editorial”} as the range of possible values of the attribute `@reason`,[12] during the reasoning process this assertion “logically propagates” to all instances in `GAP1` and `GAP2`, which then also turn out to be bound to have {“illegible”, “lost” “deleted”, “editorial”} as the range of possible values of the attribute `@reason`, since `GAP1` and `GAP2` “become” sub-classes of `GAP`:[13] and this would lead to a *semantic inconsistency* (i.e., contradiction) if a *formal assertion* such as g `hasReason` v were eventually added to the ontology,[14] where g is an instance of either of the two classes `GAP1` or `GAP2`, and v is a value not included in the set {‘illegible”, “lost” “deleted”, “editorial”}

[8]Note that, in Fig. 1, these *ontological restrictions* are not actually shown. This is inherent to the OWL visualizer used to create the image in the figure, namely, WebVOWL (https://github.com/VisualDataWeb/WebVOWL). More details about the ontology can be found in the formal representation reported Fig. 2.

[9]Actually, the condition that class `GAP1` (resp., `GAP2`) is the domain of data property `hasExtent` (resp., `hasUnit`) is not fully essential and could be replaced by the weaker condition that the intersection of the domain of `hasExtent` (resp., `hasUnit`) and the class `GAP` is a sub-class of `GAP1` (resp., `GAP2`), and similarly for the domain of data properties `hasAtLeast` and `hasAtMost`; moreover, the requirement that `GAP` is the domain of the data property `hasReason` could be entirely dropped. Observe also that, by adopting an alternative ontological representation, there would be no need to use different (disjoint) sub-classes of `GAP` to represent instance types; in fact, besides the main class `GAP`, one could introduce: (a) a class C_A to represent each attribute A of the tag `<gap>`; (b) a single property `hasAttribute` with domain `GAP` and with range the union of the C_As, with the intended meaning that relationship $\texttt{hasAttribute}(\texttt{GAP}, C_A)$ expresses that A is an attribute of `<gap>`. Mutual exclusiveness of `<gap>` instance types can then be enforced by means of appropriate *class axioms*. For instance, the fact that no `<gap>` instance can involve both the two attributes `@extent` and `@unit` corresponds to the axiom that, for no x in `GAP`, it is the case that, for some y in the intersection of classes $C_{\texttt{@extent}}$ and $C_{\texttt{@unit}}$, the relationship $\texttt{hasAttribute}(x, y)$ holds true.

[10]OWL Functional-style Syntax is a human-readable, text-based formal representation system for Web Ontology Language (OWL), defined by the “OWL 2 Web Ontology Language, Structural Specification and Functional-Style Syntax” W3C Recommendation. https://www.w3.org/TR/owl2-syntax/

[11]The asserted facts constitute the *Semantic Knowledge Base* of an ontology (usually identified with the ontology itself). Let us stress that, at the formal level, the asserted facts, as well as the inferred facts, are intensionally represented by such designated *well-formed syntactic expressions* built up according to the formation rules of a given formal representation system (e.g., OWL-style Function Syntax); and in fact, it is at this very formal level of syntactic representations that the deduction of the inferred facts from the asserted facts actually occurs, with the collection of the syntactic expressions representing the asserted facts being expanded by means of an algorithm (i.e., mechanically) to the corresponding collection of syntactic expressions representing the inferred facts. This algorithm is (in essence) what is commonly called *a reasoner*.

[12]Recall that this corresponds to the ontological restriction that the class `GAP` is a sub-class of the *inverse image* of the literal set {“illegible”, “lost” “deleted”, “editorial”} under the data property `hasReason`.

[13]And thus (e.g.) there would be no need to explicitly state this fact, as it is inferred.

[14]A formal assertion is the syntactic expression representing a given asserted fact (cf. footnote 11). The formal assertion g `hasReason` v represents (intuitively) the fact that the instance g includes the attribute `@reason` with value v.

```
FunctionalDataProperty(:hasAtLeast)
DataPropertyDomain(:hasAtLeast :GAP2)
DataPropertyRange(:hasAtLeast xsd:int)

FunctionalDataProperty(:hasAtMost)
DataPropertyDomain(:hasAtMost :GAP2)
DataPropertyRange(:hasAtMost xsd:int)

FunctionalDataProperty(:hasExtent)
DataPropertyDomain(:hasExtent :GAP1)

FunctionalDataProperty(:hasReason)
DataPropertyDomain(:hasReason :GAP)

FunctionalDataProperty(:hasUnit)
DataPropertyDomain(:hasUnit :GAP2)

SubClassOf(:GAP DataSomeValuesFrom(:hasReason
                                        DataOneOf("illegible" "lost" "deleted" "editorial")))

DisjointUnion(:GAP :GAP1 :GAP2)

SubClassOf(:GAP1 DataSomeValuesFrom(:hasExtent DataOneOf("2 pages")))

SubClassOf(:GAP2 DataSomeValuesFrom(:hasUnit DataOneOf("char")))
```

Fig. 2. OWL Functional-style Syntax representation of the OntoGap ontology.

(e.g., v might be a *dummy value*, say v = "undefined", designated to signal the lack of the attribute @reason; i.e., g hasReason "undefined" means that instance g does not actually include the attribute @reason.[15])

Similarly, the asserted facts that **(3)** the domains of data properties hasExtent and hasUnit are the classes GAP2 and GAP1, respectively, coupled with the inferred fact that these classes GAP2 and GAP1 are disjoint (cf. **(1)** above), would lead to a semantic inconsistency as well, if one adds an instance g in the class GAP, with g participating to both properties hasExtent and hasUnit; indeed, from this, one would deduce (by **(3)** above) that g is an instance of both GAP2 and GAP1, which contradicts the inferred disjointness of GAP2 and GAP1.

During the process of creation of a digital edition, if we expand the ontology by adding formal assertions representing the adopted choices of TEI encodings (such as the choice of including within a <gap> instance a particular attribute value), then, the detection of such semantic outcomes as above (i.e., ontology inconsistencies) could ultimately be interpreted as signaling the occurrence of an error in the TEI encoding of a document. In fact, by leveraging the deductive mechanisms of formal ontologies, combined with the data analysis tools offered by NormaTEI (see Section VI), we argue for the possibility of developing a unified, and at-some-extent automated methodology to facilitate collaborative work during the creation of DSEs.

V. Bellini Digital Correspondence

Some of the challenges related to the harmonization of text encoding have been extensively investigated in conceiving and implementing *Bellini Digital Correspondence* (*BDC*), a DSE comprising a corpus of letters, which are autographs of Vincenzo Bellini (1801-1835). The corpus, preserved at the Belliniano Civic Museum in Catania, Italy, consists of a collection of 40 letters (35 codicological units) digitally captured in 111 facsimile images. The material encompasses various types of resources, including letters, letter drafts, and cards, each with its own specific characteristics from a structural, textual, functional, and physical perspective. Although the 40 letters constitute a small part of Bellini's correspondence collection of texts, they also represent a particularly significant portion of it, which mainly spans the last six years of the Maestro's biographical and creative arc. The *BDC* project, which aims to provide a philological web platform[16] for exploring Bellini's texts, was conceived and developed by the National Research Council of Italy (CNR).[17] At the same time, *BDC* is part of the BellinInRete project,[18] aimed at analysing, organising and renewing the use of the museum's heritage. BellinInRete also saw the realisation of a new exhibition,

[15]Notice in fact, in regard to this point, that, due to *Open World Assumption* (OWA) adopted in ontological reasoning (at least in the context of the Semantic Web), the *negation* of a relationship $x\mathscr{R}y$ between two individuals x and y needs to be explicitly asserted within an ontology as it does not automatically follows from the absence of $x\mathscr{R}y$ in the ontology; thus, if our OntoGap ontology contains no assertion of the form g hasReason w, we cannot actually infer that g does not include the attribute @reason.

[16]http://bellinicorrespondence.cnr.it. In this particular context, the OntoBelliniLetters ontology (http://bellinicorrespondence.cnr.it/ontologia/) was defined in alignment with the principles of Linked Open Data and Semantic Web, aiming to improve interoperability of the Bellini epistolary corpus preserved in the Belliniano Civic Museum of Catania.

[17]https://www.cnr.it/en

[18]http://bellininrete.istc.cnr.it/

designed by the CNR, in which the tangible and intangible heritage is narrated through a multi-channel and immersive itinerary aimed at both specialists and occasional museum visitors. The virtual museum tells the story of Vincenzo Bellini through the sensations evoked by multimedia technologies: the museum tour guides visitors through the main periods of the Maestro's life, imagining '*A life in four acts*'. The narration relies on audiovisual technologies, with evocative Bellini music pieces as protagonists, transporting visitors to immersive scenic settings embedded within various theaters of the time: a *proscenium*, a *boccascena*, a *box* and a *foyer*. Moreover, the latter can access the edition through a touchscreen placed within the museum tour.

The approach adopted is similar to that of other editorial initiatives involving epistolary text corpora.[19] Among the projects most closely related to *BDC*, particularly notable are the digital edition of letters from the Flemish literature magazine "Van Nu en Straks"[20] [12], [13] and the German project WeGa[21] [14], which concerns the digital edition of works and writings by Carl Maria von Weber. Some remarkable DSE projects in the Italian scholarly scene include the digital edition of the *Letters of Vespasiano da Bisticci*[22] [15], [16] and the digital edition of Alcide De Gasperi's correspondence[23] [17].

The corpus has been encoded following the TEI guidelines, leveraging both the transcription of the text and the availability of the original facsimile images (image-based edition). The Web-based presentation of the letters has been accomplished by customizing the second version of Edition Visualization Technology software (EVT)[24] (see Fig. 3).

As for the encoding model, the management of metadata involves the description of the correspondence, which is recorded using the `<correspDesc>` tagset within the `<profileDesc>` section of the `<teiHeader>`. Moreover, the elements and the attributes defined within the TEI Manuscript Description (module 10 of the TEI) and the `<facsimile>` tagset (module 11 of the TEI) have been employed to describe the textual phenomena in the primary source and their relationship with the regions of interest in the corresponding facsimile. The work includes both diplomatic and interpretive edition levels, which have been based on the publication by Seminara (2017) [18].

A. The Educational Context

The encoding activities concerning Bellini's letters were carried out within an educational framework, where students[25] contributed to the *BDC* project, fostering integration between teaching, technical, and scientific activities.

The process thus implemented brought forth new challenging issues regarding the management of a cooperatively made editorial project. In particular, the validation of the encoding required the development of a new methodological approach to handle the different "valid" encoding solutions arising from the heterogeneity of the primary source.

For example, the final encoding schema and the evaluation methodology have to consider the presence of different letters on the same codicological units (e.g., missives LL1.23.I and LL1.23.II).[26] In such a case, the modeling issue involves the choice between two possible encoding solutions:

1) the representation of the codicological divisions and
2) the representation of the text divisions.

The model established by the research team favored the second option, while students often preferred to encode by adopting the first option.[27]

B. Methodology

BDC is an open-source and open-access project, the result of a multidisciplinary effort that implements a distributed, cooperative and collaborative model for making scholarly digital editions.[28] This model is generic and methodologically valuable. It has provided us with the opportunity to develop digital tools aimed at standardizing the text representation towards a rigorous textual model. These tools have been employed to select, align, and correct the encoding choices made by different contributors. The result is a consistently encoded corpus, encompassing textual, paratextual, and extratextual phenomena conveyed by the primary sources.

Moreover, these tools have broad applicability and can be directly utilized in other DSE projects. The *BDC* experience has raised general methodological concerns in the field of Digital Humanities, particularly emphasizing the still open issue regarding the interoperability and comparability of DSEs [19]. In fact, both the data and the process developed in the *BDC* project adhere to the so-called FAIR principles [20].[29]

In light of the above, the digital edition of the Bellini corpus is modeled as a *distributed process* that encompasses several stages. Note that these involve the participation of groups of students engaged in collaborative text transcription activities, where each student is assigned the task of transcribing a letter (or part of a letter) in XML/TEI. The process begins with the selection of the available primary sources; then, the relevant

[19] Additional similar initiatives are discussed in [11].

[20] http://ctb.kantl.be/project/dalf/index.htm

[21] https://weber-gesamtausgabe.de/en/Index, and in particular the "Guidelines for the Edition of Weber's Letters, Diaries and Documents" webpage (https://weber-gesamtausgabe.de/en/Project/Editorial_Guidelines_Text.html), and the paragraph "Weber's Correspondence and other Letters included" on the "Project Description" webpage (https://weber-gesamtausgabe.de/en/Project/Project_Description.html).

[22] https://projects.dharc.unibo.it/vespasiano/

[23] https://epistolariodegasperi.it/

[24] http://evt.labcd.unipi.it/

[25] Text Encoding class at the Pisa University, Digital Humanities Degree.

[26] The encoding takes into account both the material aspect of the document (the two letters are located on the same folio and thus are identified with the identifier "23") and the textual significance (in the event that two separate letters share the same folio, they are identified with the roman numbers "I" and "II").

[27] In practice, this resulted in the production of an XML/TEI document for each missive, independently of the layout of the letters on the sheet, whereby the students' work brought together in a single file the encoding of any *multiple* texts, which could be attributed to different missives but contained in the same codicological unit.

[28] A more substantial reflection on the method is available in [11].

[29] FAIR stands for *Findable, Accessible, Interoperable, Reusable*.

Fig. 3. *BDC*- Letter LL1.15 in the Edition Visualization Technology tool.

musical and codicological information to be subsequently added to the transcriptions of the texts are identified and collected. In order to facilitate collaborative work, an encoding model was created and instantiated on some exemplar letters. Due to the flexibility of the XML/TEI vocabulary and the various encoding strategies, the textual models are never definitive solutions. Afterward, the XML/TEI documents contain lists of musical terms, citations of Bellini's works in the corpus, bibliographical references, and named entities (persons, places, terms). These materials, along with the high-resolution facsimile, were made available for collaborative encoding activities. At this stage, students propose their own encoding for the assigned letters.

The encoded letters, consisting of XML documents and various accompanying files for web visualization, undergo a semi-automatic evaluation. The encoded corpus faces formal and scientific validation to ensure the correctness of the data.

The final validation phase focuses on harmonizing the encoding model. This phase is essential due to the multiple encoding solutions generated during the distributed work. It includes:

- the revision and/or recording of multiple textual and paratextual phenomena in the digital document;[30]
- the completion of lists of named entities and other notable data;
- the encoding of links to authoritative resources and repositories available on the Web according to Linked Open Data practices (e.g., VIAF,[31] GeoNames,[32] RISM[33]);
- the review of regions of interest from facsimile sources.

[30]The digital scholarly editing process is described in detail in [21].

VI. NormaTEI

To support the harmonization phase, the *BDC* team decided to develop a tool – NormaTEI[34] – to control XML/TEI-encoded phenomena in a comprehensive and organic way across all resources under consideration. NormaTEI is designed to collect the encoded data present in the XML documents altogether.

The software is based on a relational database, implemented by using the 4D platform. 4D is a Rapid Application Development (RAD) for DBMS that integrates a SQL database engine, a proprietary programming language and a web server. The Linux-Apache-MySQL-PHP (LAMP) architecture brought all together in a single tool. In addition to these features, the tool also offers others technologies such as a native SVG support,[35] DOM and SAX commands for parsing and writing XML/TEI files, and a PHP interpreter.[36]

[31]The Virtual International Authority File, https://viaf.org
[32]https://www.geonames.org
[33]Répertoire International des Sources Musicales, https://rism.info
[34]https://doi.org/10.5281/zenodo.7703053
[35]SVG support by 4D was used for the development of another tool within *BDC* project to manage the image zones of interest associated with an XML/TEI file. The tool name is ZoneRW (https://zenodo.org/doi/10.5281/zenodo.5599509).
[36]NormaTEI is compatible with Windows 10, Windows 11, Windows Server versions from 2012 R2 to 2022, macOS versions from Big Sur (11) to Ventura (13) (the latest releases for each version).

The goal of NormaTEI is to allow the user to analyze the tags and the attributes (with their values) of a group of documents in XML/TEI format. The tool can be used with any type of SGML markup file. When NormaTEI is opened, an *import window wizard* asks to specify the folder from which the XML/TEI files to be analyzed should be imported. Note that, by default, NormaTEI selects the last used folder.

To perform the import operation of XML/TEI files, an algorithm has been developed that exploits the 4D DOM commands. The implemented algorithm returns the names of the TEI elements (tags), the tag attributes and their values.

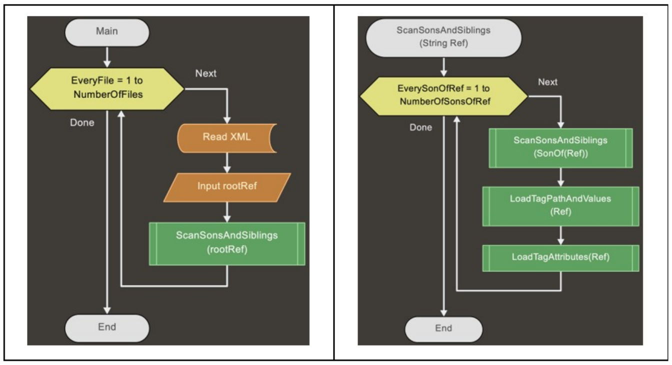


Fig. 4. Flowchart showing the algorithm implemented within the NormaTEI tool.

For each XML/TEI file to be analyzed, the algorithm (Fig. 4) creates a pointer to the root of the tree structure of the file (the `<TEI>` tag) and then launches the *ScanSonsAndSiblings* procedure. This procedure takes as a parameter the pointer to a node (initially the root of the tree structure) and is called recursively for each child of that node. During each iteration, the procedure stores the current tag and its value, its full path, as well as its attributes along with their respective values. This provides all the data and information for the entire XML/TEI tree for each file.

At the end of the import operation, NormaTEI displays the analyzed data in the form of a table (the *data table*), where, for each XML tag, the columns contain (see Fig. 5):

- the name of the tag;
- the file where that tag is present;
- the full path of the tag;
- the value of the tag (i.e., its content);
- the attributes of the tag;
- the values of the attributes of the tag;
- the *unique path* of the tag, namely, the variant of the full path of the tag where each node `N` in the path is followed by the specification of the position `p`, enclosed in brackets, at which `N` occurs as a child of its immediate ancestor, from left to right; i.e., `N` is the `p`-th child of its immediate ancestor. For instance, the unique path `/TEI[1]/text[3]/body[2]/` specifies that the node `body` is the second child of the node `text` preceding it; the node `text` is the third child of the node `TEI`, which, in turn, is the first (and only) child of the root.

Note that the data table contains a row for each attribute of a tag; thus, if a tag has (e.g.) three attributes, then there are three different rows corresponding to the tag: one row for each attribute. Observe also that, by clicking on the border between the table column headers you can resize them.

From the inspection of the imported data, a number of similarity checks can be performed. For instance, one can checks whether:

- tags with the identical paths have identical or similar values;
- tags with identical names, and identical or similar depth levels (i.e., with at most a one-level swing in the hierarchy) have identical paths;
- attributes with identical names record identical values;
- identical attributes referring to the same tag have identical values;
- tags with equal hierarchy have the same attributes;
- attributes with identical values have identical names.

Through its search interface, NormaTEI allows to carry out column-wise search operations within the data table cells to retrieve information for (e.g.) data analysis purposes; for each column of interest, it is enough to specify in the yellow *search box* located at the top of the column the text to be searched for within the cells of that column (see Fig. 5).

NormaTEI uses *incremental searching*; so, as one updates the content of the search box, by inserting or deleting characters, the search results are correspondingly updated as well.

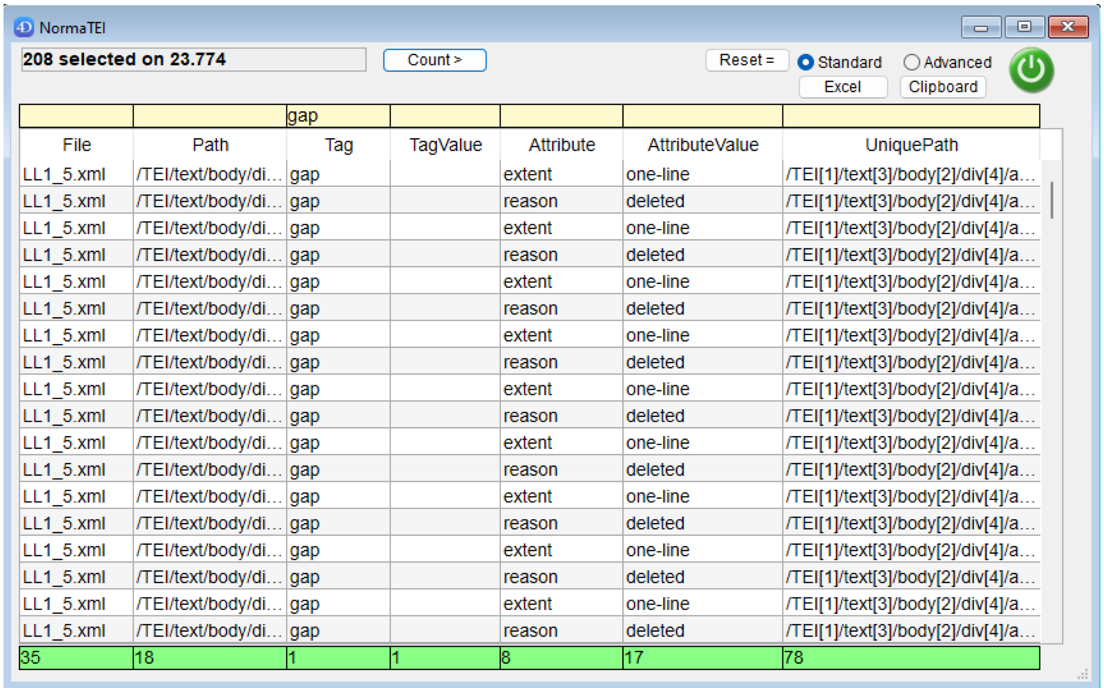
As a result of a search operation, the following scenarios may occur: (i) if a tag has no attributes, a row with empty attribute data is shown; (ii) if a tag has a single attribute, then this attribute is shown with its value; (iii) if a tag has multiple attributes, one row is shown for each attribute. The number of distinct values for each column is shown at the bottom of the NormaTEI panel.

The result of a search overwrites the content of the data table cells; however, the cells can be reset to their original content (i.e., the data loaded when XML/TEI files are imported into NormaTEI), if desired.

SectionVI-A describes some complex, and particularly useful search operations exploiting advanced search tools available in NormaTEI.

NormaTEI allows the export of data (tags, path, depths, attributes, values) into Excel spreadsheets, and, in addition to the information obtained from the search interface (cf. Fig. 5), one can even inspect the frequency of occurrences of encoded phenomena under different counting metrics; for instance, one can count the number of occurrences, within the selected corpus, of all instances of a given tag name that include a given attribute name (or attribute value), and the number of XML/TEI files containing these occurrences, besides listing the files where the occurrences are not present. To this end, a summary panel can in fact be accessed that reports:

- the total number of occurrences of a phenomenon;



File	Path	Tag	TagValue	Attribute	AttributeValue	UniquePath
		gap				
LL1_5.xml	/TEI/text/body/di…	gap		extent	one-line	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		reason	deleted	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		extent	one-line	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		reason	deleted	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		extent	one-line	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		reason	deleted	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		extent	one-line	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		reason	deleted	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		extent	one-line	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		reason	deleted	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		extent	one-line	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		reason	deleted	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		extent	one-line	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		reason	deleted	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		extent	one-line	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		reason	deleted	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		extent	one-line	/TEI[1]/text[3]/body[2]/div[4]/a…
LL1_5.xml	/TEI/text/body/di…	gap		reason	deleted	/TEI[1]/text[3]/body[2]/div[4]/a…
35	18	1	1	8	17	78

Fig. 5. The search interface of NormaTEI.

- the number of XML/TEI files where a phenomenon is present at least once;
- the XML/TEI files where a phenomenon is either present or absent.

Occurrence counting can be performed either by attribute (as illustrated in Fig. 6) or by tag. Consider a scenario where the chosen corpus only contains one instance of the `<ab>` tag, as shown in Listing 3.

If we count occurrences by attribute, the value would be 6 (representing the number of attributes actually present within this single tag occurrence). Conversely, if we count occurrences by tag, the resulting value would be 1 (since there's only one occurrence of the `<ab>` tag itself).

```
<ab n="ab_02" next="#LL1.10_ab_01_1v"
 part="I" rend="first_line_indented"
 type="parag" xml:id="LL1.10_ab_01_1r">
   ...
</ab>
```

Listing 3. XML fragment of an `ab` element with its attributes.

A. Advanced Data searching in NormaTEI

As mentioned above, NormaTEI allows to perform even complex search operations within the data table cells.

Besides the string S to be searched for (the *search string*), the following *search criteria* can be specified (see Fig. 7):

- **Contains**: retrieves the cells C such that S is a *factor* of the content of C;[37]
- **Starts with**: retrieves the cells C such that S is a *prefix* of the content of C;[38]
- **Ends with**: retrieves the cells C such that S is a *suffix* of the content of C;[39]
- **Equals**: retrieves the cells C such that S is identical to the content of C;
- **Doesn't contain**: retrieves the cells C such that S is not a factor of the content of C;
- **Match RegEx**: the search string S is evaluated as a regular expression.

It is also possible to perform more advanced *cascading searches* within the data corpus currently loaded into NormaTEI (see Fig. 8). In fact, once a given search operation O is performed, one can select a specific superset S of the *result-set* of this operation O (see below);[40] then, a superset S' of the result-set of a subsequent search operation O' can even be selected and combined with S by means of different boolean set operations, to obtain more refined search results; specifically, S and S' can be combined to form their union $S \cup S'$, their intersection $S \cap S'$ or their difference $S \setminus S'$, which consist of (i) the results that are included in S as well as those included in S'; (ii) the results included in both S

[37]For the sake of completeness we recall that, given two strings S_1 and S_2:
- S_1 is a factor of S_2, if there are strings X and Y such that $S_2 = XS_1Y$;
- S_1 is a prefix of S_2, if there is a string X such that $S_2 = S_1X$;
- S_1 is a suffix of S_2, if there is a string X such that $S_2 = XS_1$.

[38]See footnote 37.

[39]See footnote 37.

[40]By the result-set of a search operation it is meant, here, the whole set of table rows containing the search results obtained immediately after the execution of the operation. Then, the *current result-set* is the result set of the operation just performed.

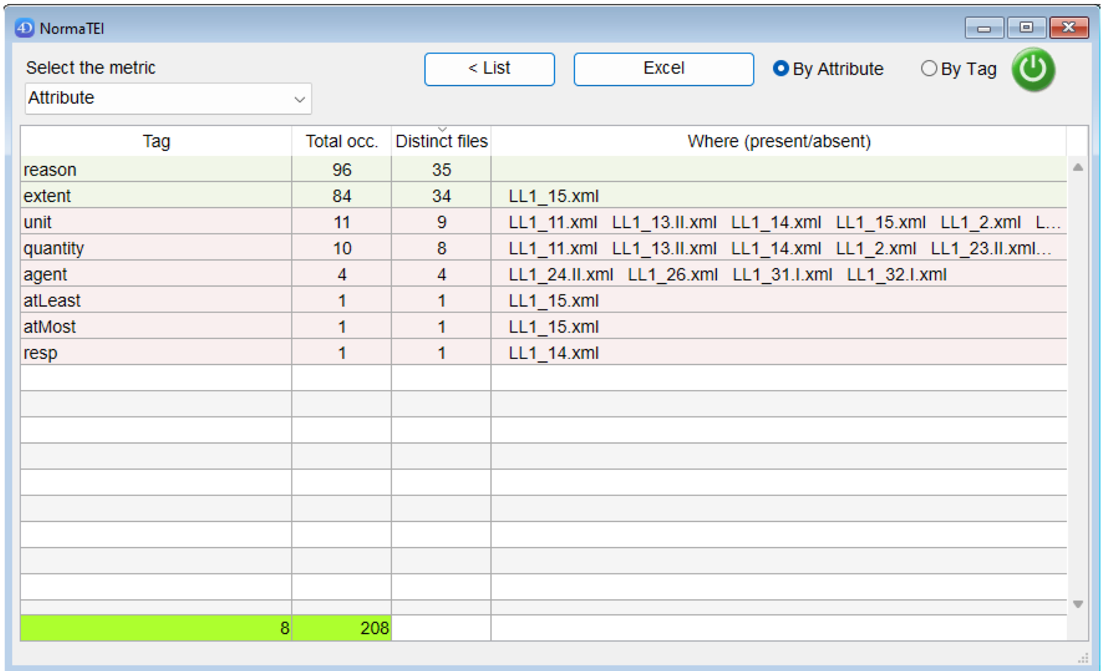


Fig. 6. Metrics count panel in NormaTEI.

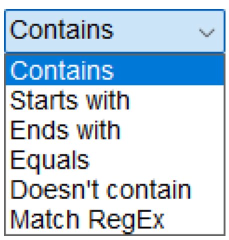


Fig. 7. Advanced search criteria of NormaTEI.

and S'; and, (iii) the results included in S but not included in S', respectively. This process can be accomplished through the Advanced Search panel of NormaTEI (see Fig. 8), by clicking on the "Use" button that allows the user to choose the search results to be included within the superset S of the current result-set that is being selected, which are displayed in the lower part of the panel (see Fig. 8).

The "Use" button asks the user two questions within a *wizard window*:

- **What should I use?**: this question instructs NormaTEI about the search results to be included within S; two answer options are possible:
  - **The selection**: includes precisely all results from the result-set of the current search operation;
  - **Column values**: includes the set of rows corresponding to the whole data corpus that match the values contained in the selected column of the current result-set. A classic case related to this option occurs when, after the searching whether a certain characteristic is present in the data corpus, one would carry out subsequent searches involving exactly the XML/TEI files where the characteristic is present: in this case it is enough to select the column "File" of the result-set.
- **What should I do?**: this question instructs NormaTEI about the boolean set operation that is to be used (see above); i.e., union, intersection or difference. The following answers are possible:
  - **Add**: the union operation is used;
  - **Remove**: the difference operation is used;
  - **Get common**: the intersection operation is used.

At this point it is possible to carry out a new search on the whole corpus, or, by pressing "Search among results", a search can be carried out within the select data (cf. the set S above), reported in the lower part of the NormaTEI panel.

VII. Normalization: a Case Study

A normalized encoding is necessary both to ensure the homogeneity of the edition and to enable an effective transformation of the edition into conforming documents compatible with the visualization tools.

The usefulness of tools like NormaTEI lies in two aspects: 1) it is possible to know in which documents a phenomenon is not present in order to check and analyze why it has no occurrences; 2) it is possible to record phenomena in order to check and analyze why the frequency of attestation is high or low. In both cases, the tool allows scholars to find errors and encoding patterns, or to check why a phenomenon is present on a large part of the corpus but not within the corpus as a whole.

The use of the NormaTEI tool within the *BDC* project has highlighted the substantial halving of the number of different

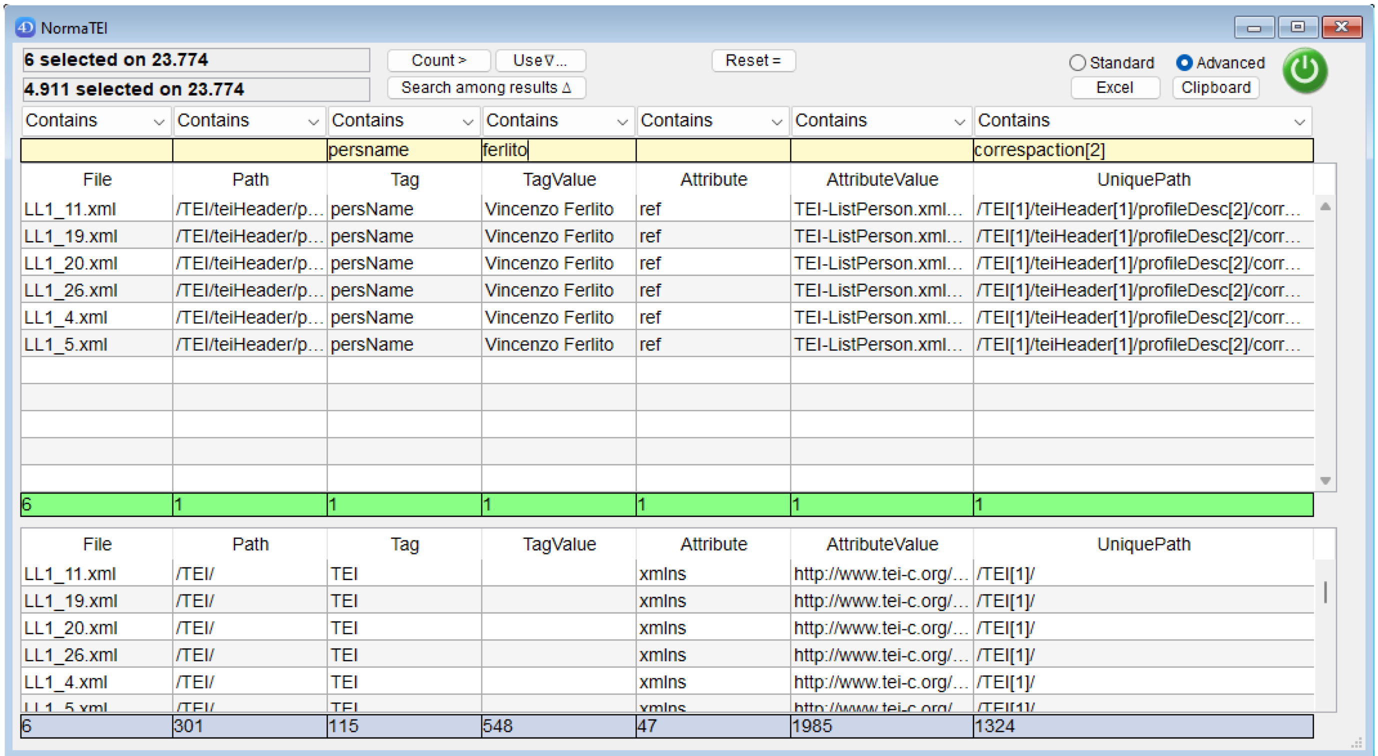


Fig. 8. Results of cascading searches in NormaTEI. In this case, the purpose is to research a specific subcorpus of the epistolary, consisting of letters sent to a specific recipient (Vincenzo Ferlito).

XML/TEI paths (1148) noted in the initial encoding of the corpus, compared to the tally made after the harmonization phase (594).

In the particular case of the <gap> element and its frequency data, it is interesting to note that the number of distinct paths in the original corpus (93) is much the same as in the normalized corpus (96). Even with regards to the attributes, they are the same and in equal amounts, but with a different distribution, as shown in Table I.

TABLE I
ATTRIBUTES OF <GAP> TAG

Normalized Corpus			*Original Corpus*		
attribute	**occurs**	**letters**	**attribute**	**occurs**	**letters**
reason	96	35	reason	92	26
extent	84	34	extent	32	12
unit	11	9	unit	18	9
quantity	10	8	quantity	7	4
agent	4	4	agent	7	4
atLeast	1	1	atLeast	3	2
atMost	1	1	atMost	3	2

Also, it should be noted that, within the normalized corpus, the six most used paths cover more than a quarter (i.e., 26) of the total number of path occurrences. Moreover, the first two most used paths are present in 19 and 15 letters, respectively; a total of 30 letters involve at least one of these two paths. In contrast, the six most used paths in the original corpus cover one seventh (i.e., 12) of the total occurrences. The aforementioned statistics suggest that the harmonization process improved the uniformity and the general encoding coherence of textual phenomena without producing substantial adjustments, acting only on the distribution of encoding the phenomena rather than on their nature. In fact, as we see in Table II, the attributes used in the original corpus are more than three times the attributes adopted in the corpus after the normalization process.

TABLE II
@REASON ATTRIBUTE VALUES FOR <GAP> TAG

Normalized Corpus			*Original Corpus*		
attribute	**occurs**	**letters**	**attribute**	**occurs**	**letters**
illegible	60	33	illegible	20	6
deleted	31	9	deleted	3	1
lost	5	5	lost	3	3
			editorial	34	4
			missing	16	9
			cancelled	9	6
			ink blot	2	1
			omissis	2	1
			omit-in-orig	2	1
			absent	1	1

The harmonization process has resulted in a marked streamlining of the variants used for the <gap> and <abbr> phenomena, as we can see in Table III. The variants of <gap> were halved, while those of <abbr> were reduced by a third.

TABLE III
DIFFERENT PATHS OF <GAP> AND <ABBR> TAGS

Encoding feature	*Normalized corpus*		*Original corpus*		Ratio
	occurs	letters	occurs	letters	
<gap>	18	35	36	37	0.5
<abbr>	26	35	81	39	0.3

VIII. Conclusions and Further Works

Collaborative DSEs present novel methodological challenges related to schema harmonization, resource interoperability, and both syntactic and semantic matching/mapping issues. In this paper, we have introduced a positional perspective to ensure semantic validation through an ontology-based mapping approach. This approach facilitates the alignment between textual phenomena encoded using XML markup vocabularies (such as XML/TEI) and their corresponding representations described by means of OWL technology. Furthermore, applying semantic-based schema harmonization could enhance the capabilities of the NormaTEI tool developed for the *Bellini Digital Correspondence* project. The software already developed but still in progress facilitates the editorial team in harmonizing collaboratively encoded documents. The proposed model is versatile and can be applied to other similar encoding initiatives. NormaTEI is currently released and reusable, and we plan to enhance the ontological description to test it on a subset of relevant TEI elements.

References

[1] M. Cuculovic, F. Fondement, M. Devanne, J. Weber, and M. Hassenforder, "Semantics to the rescue of document-based xml diff: A jats case study," *Software: Practice and Experience*, vol. 52, no. 6, pp. 1496–1516, 2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/spe.3074

[2] J. Euzenat and P. Shvaiko, "The Matching Problem," in *Ontology Matching*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 25–54. [Online]. Available: https://doi.org/10.1007/978-3-642-38721-0-2

[3] B. Jang, S. Park, and Y.-g. Ha, "A stream-based method to detect differences between XML documents," *Journal of Information Science*, vol. 43, no. 1, pp. 39–53, 2017, publisher: SAGE Publications Sage UK: London, England.

[4] P. Ciancarini, A. D. Iorio, C. Marchetti, M. Schirinzi, and F. Vitali, "Bridging the gap between tracking and detecting changes in XML", *Software: Practice and Experience*, vol. 46, no. 2, pp. 227–250, 2016, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/spe.2305. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/spe.2305

[5] S. Rudolph, "Foundations of description logics," in *Reasoning Web. Semantic Technologies for the Web of Data - 7th International Summer School 2011, Galway, Ireland, August 23-27, 2011, Tutorial Lectures*, ser. Lecture Notes in Computer Science, A. Polleres, C. d'Amato, M. Arenas, S. Handschuh, P. Kroner, S. Ossowski, and P. F. Patel-Schneider, Eds., vol. 6848. Springer, 2011, pp. 76–136.

[6] G. Rockwell and S. Sinclair, *Hermeneutica: Computer-Assisted Interpretation in the Humanities*, ser. Ebook Central. MIT Press, 2016. [Online]. Available: https://books.google.it/books?id=mzbfCwAAQBAJ

[7] P. Robinson, "Some principles for making collaborative scholarly editions in digital form," *Digital Humanities Quarterly*, vol. 11, no. 2, 2017. [Online]. Available: http://www.digitalhumanities.org/dhq/vol/11/2/000293/000293.html

[8] F. Ciotti, "A realistic theory of textuality and its consequences on digital text representation," in *What is text, really? TEI and beyond*, ser. TEI Conference Book of Abstracts. Graz: TEI, 2019. [Online]. Available: https://gams.uni-graz.at/o:tei2019.100

[9] T. Mancinelli and E. Pierazzo, *Che cos'è un'edizione scientifica digitale*. Roma: Carocci Editore, 2020.

[10] J. Cummings, "Opening the book: data models and distractions in digital scholarly editing," *International Journal of Digital Humanities*, vol. 1, no. 2, pp. 179–193, Jul. 2019. [Online]. Available: https://doi.org/10.1007/s42803-019-00016-6

[11] A. M. Del Grosso, E. Capizzi, S. Cristofaro, M. R. De Luca, E. Giovannetti, S. Marchi, G. Seminara, and D. Spampinato, "Bellini's Correspondence: a Digital Scholarly Edition for a Multimedia Museum," *Umanistica Digitale*, vol. 3, no. 7, Jan. 2019. [Online]. Available: https://umanisticadigitale.unibo.it/article/view/9162

[12] E. Vanhoutte, and R. Van den Branden, "Presentational and Representational Issues in Correspondence Reconstruction and Sorting," in *Electronic Scholarly Editing–Some Northern European Approaches.* A Special Issue of *Literary and Linguistic Computing*, M. Dahlström, E. S. Ore, and E. Vanhoutte F. Unwalla, Eds., vol. 19/1 (2004), pp. 45–54.

[13] E. Vanhoutte, and R. Van den Branden, "Describing, Transcribing, Encoding, and Editing Modern Correspondence Material: a Textbase Approach," in *Literary and Linguistic Computing*, vol. 24/1 (2009), pp. 77–98. [Online]. Available: https://doi.org/10.1093/llc/fqn035

[14] J. Veit "Die Codierung digitaler Briefeditionen mit TEI P5. Konzepte und Probleme am Beispiel der Carl-Maria-von-Weber-Briefausgabe," in *Digitale Edition zwischen Experiment und Standardisierung. Musik – Text – Codierung*, P. Stadler, and J. Veit, Eds., vol. 31 (2009), pp. 217–232. [Online]. Available: https://doi.org/10.1515/9783110231144.217

[15] M. Barbera, F. Meschini, C. Morbidoni, and F. Tomasi, "Annotating digital libraries and electronic editions in a collaborative and semantic perspective," in *Proceedings of the 8th Italian Research Conference on Digital Libraries (IRCDL 2012) - Revised Selected Papers. Communications in Computer and Information Science (CCIS) 354*, Heidelberg, Germany, Springer, 2012, pp. 45–56.

[16] F. Tomasi, "L'edizione digitale e la rappresentazione della conoscenza. Un esempio: Vespasiano da Bisticci e le sue lettere," in *Ecdotica*, vol. 9 (2012), pp. 264–286.

[17] S. Malfatti, "The Digital Edition of Alcide De Gasperi's Correspondence. Some Reflections on an Ongoing Project," in *JLIS.It*, vol. 11/1 (2020), pp. 89–105. [Online]. Available: https://doi.org/10.4403/jlis.it-12599

[18] V. Bellini, *Carteggio*, G. Seminara, Ed. Leo S. Olschki, 2017.

[19] D. Schmidt, "Towards an Interoperable Digital Scholarly Edition," *Journal of the Text Encoding Initiative*, no. 7, Nov. 2014. [Online]. Available: http://jtei.revues.org/979

[20] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data*, vol. 3, pp. 160 018 EP –, 2016. [Online]. Available: https://doi.org/10.1038/sdata.2016.18

[21] S. Pellino, P. Sichera, A. M. Del Grosso, and D. Spampinato, "Dalla codifica alla fruizione: l'edizione digitale Bellini Digital Correspondence," in *Proceedings della XI conferenza nazionale AIUCD-Culture digitali. Intersezioni: filosofia, arti, media*, F. Ciracì, G. Miglietta, and C. Gatto, Eds. Alma Mater Studiorum-Università di Bologna (Bologna, ITA), 2022, pp. 163–168. [Online]. Available: http://amsacta.unibo.it/6848/1/Proceedings_AIUCD2022.pdf