RESEARCH-ARTICLE

The Semantic Digital Edition of Aldo Moro's Writings: A Workflow Supporting Data Sharing and Replicability

SEBASTIAN BARZAGHI, University of Bologna, Bologna, BO, Italy

ALESSIO PALMERO APROSIO, University of Trento, Trento, TN, Italy

FRANCESCO PAOLUCCI, University of Bologna, Bologna, BO, Italy

FRANCESCA TOMASI, University of Bologna, Bologna, BO, Italy

SARA TONELLI, Bruno Kessler Foundation, Trento, TN, Italy

MARIALAURA VIGNOCCHI, University of Bologna, Bologna, BO, Italy

View all

Open Access Support provided by:

Bruno Kessler Foundation

University of Bologna

University of Trento

Published: 21 June 2025
Online AM: 26 March 2025
Accepted: 30 December 2024
Revised: 11 October 2024
Received: 22 April 2024

The Semantic Digital Edition of Aldo Moro's Writings: A Workflow Supporting Data Sharing and Replicability

SEBASTIAN BARZAGHI, Department of Cultural Heritage, University of Bologna, Bologna, Italy
ALESSIO PALMERO APROSIO, Psychology and Cognitive Science, University of Trento, Trento, Italy and Digital Humanities, Fondazione Bruno Kessler, Trento, Italy
FRANCESCO PAOLUCCI, University of Bologna, Bologna, Italy
FRANCESCA TOMASI, Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy
SARA TONELLI, Digital Humanities, Fondazione Bruno Kessler, Trento, Italy
MARIALAURA VIGNOCCHI, AlmaDL, University of Bologna, Bologna, Italy
FABIO VITALI, Department of Computer Science and Engineering, University of Bologna, Bologna, Italy

Digital editions have been long recognized as significant scholarly outputs, reflecting a tradition dating back to computational philology and evolving to encompass comprehensive literary and scientific knowledge management on the web. However, debates still persist around how to use data models, technical tools, and interface design to implement and document such resources so as to make them as replicable and usable as possible. Given these premises, this article focuses on presenting a semantic digital edition workflow, and, specifically, examines three crucial aspects: the integration of Semantic Web technologies for enriching textual representation; the development of a web-based markup tool that enables domain experts to produce crowdsourced markup; and the production of comprehensive technical documentation to ensure replicability. To do so, the article details how the workflow was applied to create a digital edition of Aldo Moro's works within a project funded by the Italian Ministry for Cultural Heritage and Activities and for Tourism. The article outlines the workflow stages, including survey and transcription, data modeling, markup and metadata insertion, web design, and documentation, underscoring the importance of formalized methods in developing digital scholarly editions. By adhering to established editorial and technological standards, the edition offers a nuanced exploration of Moro's writings, leveraging semantic technologies to provide a platform for scholarly engagement and exploration on the web.

CCS Concepts: • **Information systems** → **Web data description languages**; • **Applied computing** → **Document management**; **Digital libraries and archives**;

Additional Key Words and Phrases: Digital Edition, Digital Humanities, Web Application

This work was supported by Ministero della cultura (D.M. 17/10/2016—Edizione nazionale delle opere di Aldo Moro).
Authors' Contact Information: Sebastian Barzaghi, Department of Cultural Heritage, University of Bologna, Bologna, Italy; e-mail: sebastian.barzaghi2@unibo.it; Alessio Palmero Aprosio (corresponding author), Psychology and Cognitive Science, University of Trento, Trento, Italy and Digital Humanities, Fondazione Bruno Kessler, Trento, Italy; e-mail: a.palmeroaprosio@unitn.it; Francesco Paolucci, University of Bologna, Bologna, Italy; e-mail: francesco.paolucci.93@gmail.com; Francesca Tomasi, Department of Classical Philology and Italian Studies, University of Bologna, Bologna, Italy; e-mail: francesca.tomasi@unibo.it; Sara Tonelli, Digital Humanities, Fondazione Bruno Kessler, Trento, Italy; e-mail: satonelli@fbk.eu; Marialaura Vignocchi, AlmaDL, University of Bologna, Bologna, Italy; e-mail: marialaura.vignocchi@unibo.it; Fabio Vitali, Department of Computer Science and Engineering, University of Bologna, Bologna, Italy; e-mail: fabio.vitali@unibo.it.

ACM 1556-4711/2025/6-ART34
https://doi.org/10.1145/3725534

ACM Reference format:
Sebastian Barzaghi, Alessio Palmero Aprosio, Francesco Paolucci, Francesca Tomasi, Sara Tonelli, Marialaura Vignocchi, and Fabio Vitali. 2025. The Semantic Digital Edition of Aldo Moro's Writings: A Workflow Supporting Data Sharing and Replicability. *ACM J. Comput. Cult. Herit.* 18, 2, Article 34 (June 2025), 22 pages.
https://doi.org/10.1145/3725534

1 Introduction

By looking at the Digital Humanities projects listed on the web site of the European Association for Digital Humanities,[1] it is clear that "digital edition" is one of most widely used keyword adopted for describing web applications associated with Digital Humanities projects, representing in fact one of the main attested scholarly outputs in the domain. It is sufficient to examine two important catalogues to confirm this trend: the *Catalogue of Digital Editions*[2] [Franzini et al., 2016] lists 337 web sites, and the *Digital Scholarly Editions* catalogue mentions 843 projects[3] [Sahle, 2016].

Digital editions can be thought, since the very first experiments, as the expression of the highest hermeneutical action in a project-oriented approach [Burnard et al., 2006]. Indeed, building editions through digital applications has a long tradition, starting from the very first applications of technologies to humanities during the 1960 (what was called computational philology [Perilli et al., 1995]) and the first experiments on collation and automatization of the stemma codicum [Froger, 1968; Irigoin, 1979], to the possibility to manage the whole lifecycle of a textual tradition. In particular, from the late 1990s [Mordenti, 2001; Orlandi, 1998], editions and literary works were rethought as being part of a digital environment and browsable on the web [McGann, 2001]. Creation, description, modeling, annotation, and dissemination are the pillars of the contemporary model adopted for digital editions, and they need to be intended and managed as actions of a cycle (see the Tadirah model).[4] Digital editions can be then cast as the result of a complex process [Fiormonte, 2003], based on choices and steps [Daquino et al., 2019].

Several discussions about models, tools and applications, design of the interface, and service development have engaged scholars throughout the years [De Blasi, 2020; Pierazzo, 2016; Sahle, 2016], also debating about the lack of user-oriented applications for managing scholarly activities [Pierazzo, 2019; Robinson, 2005]. More recently, a huge attention has been paid to the Semantic Web environment and, in general, the graph-oriented perspective [Spadini et al., 2021].

In this article, we do not aim to address this debates, but we would like to present the entire workflow of a (semantic) digital edition, focusing in particular on three aspects, that we consider relevant in a contemporary digital edition: (1) the *adoption of Semantic Web technologies* (data modeling) in order to add knowledge to a highly enriched textual representation; (2) the creation of a web application, serving as a markup tool, enabling crowdsourcing activities and hiding the code for non-expert users; (3) the production of a *full set of technical documents* describing the entire process of the creation of the edition, in order to guarantee the maximum replicability of the whole edition in an Open Science perspective.

Our use case concerns the works written by Aldo Moro (1916–1978), a prominent Italian politician and member of the Christian Democratic party, serving twice as prime minister. His works represent the perfect case study to apply this model, in compliance with the tradition of philological processes, but with the innovation given by the digital ecosystem for producing (creation), managing (description and annotation), and publishing (dissemination) the edition with the added value of Semantic technologies.[5]

[1] https://eadh.org/projects.
[2] https://dig-ed-cat.acdh.oeaw.ac.at/.
[3] https://www.digitale-edition.de/exist/apps/editions-browser/index.html.
[4] https://vocabs.dariah.eu/tadirah/en/.
[5] https://aldomorodigitale.unibo.it/.

This digital edition is the main outcome of the project on the National Edition of Aldo Moro's works, which has been funded by the Italian Ministry for Cultural Heritage and Activities and for Tourism to promote, preserve, and disseminate the statesman's written works and provide a cultural resource available to the public. The edition is meant to establish a new standard for national and international research on political communication.

Aldo Moro has been selected for this National edition because his legacy is mainly focused on the last period of his political activity, while his early contributions to Italian cultural and political life have been quite disregarded [Torresi, 2021]. Indeed, he was kidnapped in 1978 by the militant far-left organization known as Red Brigades and murdered after 55 days of captivity, following the government's refusal to negotiate. This tragic event has heavily influenced historical debate on his role in Italian politics, with most studies and debates focusing on the years around his death, including conspiracy theories. One of the goals of the edition is therefore to give equal visibility to the documents published throughout Moro's life, starting from his early writings as a journalist in 1932. Furthermore, the aim is to cover different genres, not only the well-studied captivity letters and the official documents issued by Aldo Moro, but also his monographs and the drafts of the lectures he gave as a Law professor, which remained unpublished before this edition.

Overall, the edition focuses on providing the scientific analysis of the written texts, political speeches, and legal works of Aldo Moro, offering historical introductions, comments, and reconstructing the text in a form as close as possible to the author's intent. The edition also includes general interpretative introductions and a critical essay preceding the volumes, accounting for previous editions and tracing the history of each writing or group of writings. Another goal is to provide, in some volumes, a dual system of notes: one for those of the author and one for those of the critical edition.

All these aspects have been taken into account to create, for a wide audience, a born digital (semantic) edition of Aldo Moro's heterogeneous works, providing scholars with a platform to annotate texts through a shared conceptual model and to explore the edition through a web interface, enriched with browsing features.

A last detail about the research team. The edition has been developed within the AlmaDL center of the University of Bologna, the unit devoted to managing, promoting, and disseminating digital collections, guaranteeing at the same time long-term preservation of data produced by scholars.[6] The technical aspects have been committed to the DH.arc research center of the University of Bologna,[7] with the collaboration of the Digital Humanities group at Bruno Kessler Foundation in Trento,[8] but several other partners have been involved in the project.[9]

To conclude, this article aims to investigate the history of this digital edition from the design to the publication and is organized as follows: an overview of some of the most significant digital editions that served as sources of inspiration for the edition's development, design, and implementation is given in Section 2. The workflow steps that were used to develop the edition are listed in Section 3. The works included in the edition are introduced in Section 4, along with the editorial guidelines that were adhered to in order to get them ready for digitization. Section 5 describes the data modeling process and the pre-existing ontologies that were reused to represent the data in the edition in a semantic, machine-readable format. The semi-automatic annotation of the documents using a web application specifically designed for this task and their open access release are explained in Section 6. The online platform of the Edition for interactive exploration is covered in Section 7. Our conclusions are provided in Section 8, which summarizes the project, highlights its advantages and disadvantages, and suggests potential future developments.

2 Related Works

As shown by a preliminary benchmark analysis carried out on a representative sample of digital scientific editions [Barzaghi, 2021a], different approaches can be adopted to build such editions depending on various aspects,

[6] https://sba.unibo.it/it/almadl.
[7] https://centri.unibo.it/dharc/en.
[8] https://dh.fbk.eu/.
[9] https://aldomorodigitale.unibo.it/credits.

including the target audience, the documentation, the textual representation of the primary sources and the quality of their presentation, the underlying data model, the navigation mechanisms, the metadata, the formats of the documents, the granularity of text and data identification and citation, the licenses used to access the information, and additional features such as data visualization and APIs. In the remainder of this section, we provide examples of several digital scholarly editions that, based on these factors, offered us thought-provoking perspectives that we tried to incorporate into the creation of our own edition.

One of the first examples we would like to mention is the Key Documents of German-Jewish History[10] [Dickow-Rotter and Burckhardt, 2022], a bilingual digital edition, aimed at both specialists and a general audience, that uses a selection of sources (also known as "key documents") to shed light on important aspects of Hamburg's Jewish history from the early modern period to the present. Digital facsimiles and transcripts with a high-level quality representation are available for consultation and are accompanied by historical interpretations and background information. Additionally, details on the origins of the source material, historical responses to it, and scholarly debates are included in the visualization of the single document. Moreover, the edition offers different ways to approach the texts, for example, via an online exhibition, a timeline, a specific topic, or a map. However, in contrast to this wealth of data and tools for exploration, the edition is missing a more refined and precise faceted search functionality based on metadata.

1914-1918-online[11] [Daniel et al., 2014] is an international collaborative project that aims to create a multi-perspective, freely accessible digital edition of written records about the First World War. Non-linear access to the edition's content is enabled through creative navigation strategies based on Semantic Media Wiki technology. Furthermore, the edition offers various download formats (including RDF/XML), additional data visualizations (e.g., interactive maps) to navigate the corpus of documents, and a wide range of metadata describing each document, although it lacks thorough technical documentation and a more meaningful representation of textual features.

Epistolario De Gasperi[12] [Tonelli et al., 2020] is an extensive project that aims to gather, digitize, and publish a corpus of letters sent and received by the famous Italian statesman Alcide De Gasperi in a digital edition that is openly accessible to the public. In addition to its accurate representation of texts, neat presentation quality, and precise identification mechanisms, the edition is distinguished by a comprehensive faceted search and an intuitive navigation system. However, it still lacks a comprehensive technical documentation, the possibility to download a work in multiple formats, and licensing information.

Vespasiano da Bisticci's Letters[13] [Tomasi et al., 2020] is the digital edition of a series of manuscript letters that were sent to and received by Vespasiano da Bisticci (1421–1498), an Italian humanist and librarian, over a time period spanning from 1444 to 1497. Their content focuses mostly on manuscripts and codex trade with other notable people of that time. The edition showcases many relevant features of these manuscripts: materials, copyists, costs, but also names of Latin, Greek, and humanistic authors and texts in order to facilitate the navigation, study, and conservation of such documents, as well as to contextualize them in a semantically rich and metatextual framework. It is characterized by semantic indices and in-depth documentation, as well as descriptions based on comprehensive metadata and its quality of presentation. However, it still lacks search functionalities, identification mechanisms for the single letters, and additional downloadable formats.

David Livingstone (1813–1873), the well-known Victorian explorer, missionary, and abolitionist, left behind a verbal, visual, and material legacy that is published in Livingstone Online[14] [Fitch, 2017]. With its extensive collection of high-resolution manuscript images and transcriptions, this digital edition aims to contextualize the

[10] https://jewish-history-online.net/.
[11] https://encyclopedia.1914-1918-online.net/home.html.
[12] https://www.epistolariodegasperi.it/.
[13] https://projects.dharc.unibo.it/vespasiano/.
[14] http://www.livingstoneonline.ucl.ac.uk/.

various historical and cultural environments Livingstone traversed through significant scholarly collaboration and research. The edition also distinguishes itself by providing an advanced search engine, a comprehensive citation-ready documentation, transparent licensing guidelines, and multiple download formats for every resource. Its lack of indexes and its copious documentation, however, result in an oversized information architecture that can make navigation confusing for new users.

Burckhardt Source[15] [Evans et al., 2017] offers open access to previously unpublished correspondence of the Swiss historian Jacob Burckhardt. The digital edition includes letters written by about 400 authors, most of which were members of the European intellectual elite and sheds light on the social and cultural issues of the period of 1843–1897 in European history. The collection, which contains approximately 1,100 documents, offers two textual representations for every letter: the "philological version," which displays transcriptions along with the choices and variations made by the authors, and the "semantic edition," which includes semantic annotations. Other notable features include a complete set of bibliographic metadata for each letter, and a thorough search mechanism based on facets, as well as a clear navigation system, indexes, and technical documentation. However, it still lacks multiple downloadable formats and clear license information.

The digitized records of the Old Bailey Proceedings from 1674 to 1913 are available for full access through the Proceedings of the Old Bailey Online[16] [Hitchcock and Shoemaker, 2006]. This digital edition includes biographical data on approximately 2,500 people who were executed at Tyburn, as well as more than 190,000 trials, each accompanied by digital images of the original pages and richly annotated XML transcriptions. It also contains the project history, specifics on trial-related materials, and background information on law and history. Users can perform different searches and browse results based on keywords and categories, as well as produce statistical information on the fly and even use the edition API to interrogate, download, and work programmatically with the edition's data.

In general, high-quality editions typically attract a diverse and composite audience formed by domain experts and generic users alike. Documentation about the contextual background, workflow, data, and technical implementation is generally regarded as a significant editorial and scholarly product by itself, although in many cases it is often incomplete or loosely organized. While metadata play a crucial role and are generally well considered, there is room for optimization, particularly in enhancing interlinking among edition contents and related web resources. Many digital editions recognize the potential of digital technologies for textual representation, favoring XML-TEI as the golden standard, while the use of RDF and other data models remains limited. Permanent identification and citation are viewed as important but rarely implemented aspects. Alternative approaches in communicating scholarly insights embedded in digital editions, such as storytelling and data visualization, are underutilized and often project-dependent when implemented. Information architecture tends to emphasize search systems and is often enhanced with functionalities powered by Boolean operators or faceted navigation. Indexes are commonly employed tools for intuitive content navigation. A notable challenge is the absence of standardized formats for publishing, downloading, and sharing edition data and contents. While digital editions are generally open for consultation and developed with Open Source software, they often adhere to restrictive policies concerning Open Access and reuse.

Taking into account the aforementioned characteristics of digital scholarly editions, we propose an approach that prioritizes a workflow supported by semantic technologies. Our formalized and documented method not only facilitates the process of collecting and publishing knowledge but also provides a transparent narrative of our editorial journey that guarantees the maximum potential for replicating our process also with other digital materials, ensuring the production and maintenance of a valuable resource for scholarly exploration and cultural preservation.

[15] https://burckhardtsource.org/.
[16] http://www.oldbaileyonline.org/.

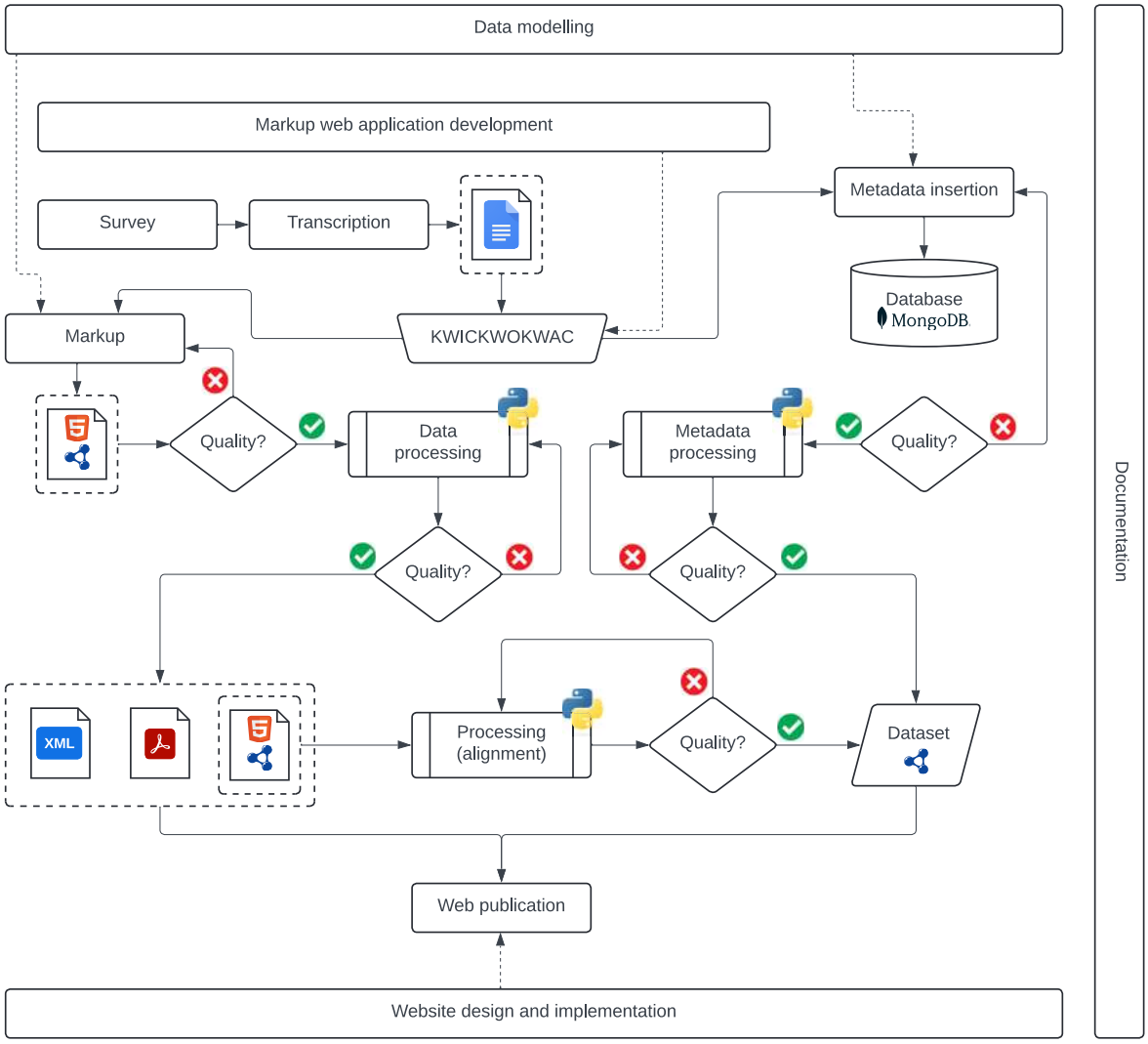

Fig. 1. A diagram which illustrates the general workflow for designing and developing the Edition and its materials.

3 General Workflow

First of all, we describe the different steps of our process leading to Aldo Moro's digital edition. Developing a digital edition benefits from the explicit definition of a publishing workflow that, starting from potentially unordered physical documents, ends with a fully fledged digital platform, so that all steps related to dealing with both technical and managerial issues are encompassed (in Figure 1, a diagram of the general workflow). An extensive summary of each major phase involved in this large-scale project is described as follows:

Survey and Transcription. The survey phase included an analysis of existing digital scholarly editions to identify best practices and issues in past projects. This was followed by a selection phase, during which the set of documents to published in the Edition were selected by a committee of history scholars according to scientific criteria, such as completeness, authorship, and publicity. Part of the works had been already published in the past on paper,

while others were still unpublished. Following this selection, researchers with a background in history and/or archival studies digitally acquired the documents which were saved in `.docx` format, to prepare them to be further processed, as illustrated in Section 4. The acquisition included scanning and performing OCR on the original documents, followed by a revision phase aimed at fixing errors and cleaning the data.

Development of the Data Model. The data model—designed and developed to organize the products of markup, metadata insertion, and publication—was used to represent the structural, textual, and contextual aspects of the documents. Special attention was given to capturing semantic information related to people, places, organizations, bibliographic references, and quotations mentioned within the documents, to set up mechanisms to efficiently navigate the corpus. The model, detailed in Section 5, was realized by reusing open and widely used standards, considering the preferences expressed by the Scientific Committee leading to the development of the Edition.

Markup, Metadata, and Scripts Definition. This phase consisted of, on one hand, the markup of the entities mentioned within the documents, bibliographic references, and quotations; on the other, the insertion of bibliographic metadata identified during the survey phase. To carry out these operations, researchers used KwicKwocKwac [Barzaghi et al., 2024], a tool specifically developed to gather digital documents, convert them into web pages, produce markup, insert metadata, link mentioned entities to authority records already existing on the Web, and perform other operations to make documents semantically enriched and publishable on the Edition web site. Further details about this process can be found in Section 6.1. Moreover, textual data processing was realized by running a series of scripts to execute code refactoring, conversion to other publishing formats, formalization of metadata into semantic assertions, and data integration. These activities are also examined in following sections of the article, namely Sections 6.2, 6.3, and 6.4.

Web Design. The development of the Edition web site was based on a preliminary competitive audit of a representative sample of major digital editions currently existing on the Web, according to a series of quality criteria identified in the scientific literature. Starting from a comparative study, we designed the information architecture of the web site. This process involved the creation of a site map and a series of low-fidelity representations (wireframes) of its main pages, to plan its interface, structure, and contents. The development of the Edition web site, further explained in Section 7, makes use of versatile and efficient frameworks, based on criteria such as modularity, accessibility, and responsive design.

Documentation. The Edition is the result of a complex series of processes. At each phase, a parallel documentation process related to the other steps (choices about data modeling, architecture, coding, etc.) was carried out, to make the processes and results as open and accessible as possible. The documentation, which records the project development through text, images, and technical reports published on both Zenodo and AMS Acta, is freely available on the Edition's web site.[17]

4 Transcription Guidelines

The bulk of Moro's works comprises two main sections, each with a distinct focus and organizational structure. The first section, titled *Scritti e Discorsi*, is dedicated to the religious, journalistic, and political writings, speeches, and interviews of Moro. This section is presented in four volumes, chronologically arranged to correspond to Moro's major life stages. The volumes within this section, excluding *Gli anni giovanili*, are each divided into two parts or tomes, providing further thematic and chronological articulation. The second section, named *Opere Giuridiche* , is centered on Moro's academic writings during his roles as an assistant, lecturer, full professor of criminal law, and appointee professor of philosophy of law. Similar to the first section, *Opere Giuridiche* is organized into four volumes that follow a chronological order.

Throughout the editing and transcription processes, the primary criterion emphasized was the fidelity to the original sources. Several specific guidelines were followed to ensure this. Additionally, graphical signs, such as asterisks, employed for paragraph division were replicated to closely resemble the original format. Attribution

[17] https://aldomorodigitale.unibo.it/about.

challenges related to unsigned journalistic writings were addressed through a comprehensive philological and archival survey. Reliable criteria, including witness verification and alignment between content and style, were applied to resolve these attribution issues. Missing conjunctions, if unavoidable, were inserted within square brackets during the transcription process.

The use of italics and other typographical styles was preserved in alignment with the author's original intent. The author's signature, when present, was retained in its original form, and if there was no explicit attribution, a signature was not added. Each text was preceded by a summary abstract containing source information, with unpublished texts including archival identifiers following standards set by national archives. The author's usage of uppercase and lowercase letters was preserved, even when it deviated from uniformity criteria. The application of capital letters for general concepts in the researchers' commentary was minimized to essential instances. Moro's critical commentary was faithfully transposed, with his notes enclosed in square brackets during transcription to distinguish them from researchers' comments.

Editorial interventions were limited to rectifying potential typos, minor oversights, and necessary integrations. In the critical apparatus, missing or imprecise references to provenance were indicated. Titles were retained in their original form whenever possible, and if a title had been added later, it was enclosed in square brackets.

Witness information was derived from collation with printed originals, manuscripts, and typewritten documents, maintaining fidelity to the sources. Additionally, words that had fallen into disuse but remained grammatically correct were preserved in the text to retain historical accuracy.

5 Data Modeling

Once the transcription is made, the following crucial activity is to move to the conceptualization level, by choosing the appropriate models to be used in producing the semantic representation of the Edition. The data modeling process requires to focus on the main ontologies useful to represent textual, contextual, and bibliographic information related to Moro's works (a full description of the resources reused and produced during this process is available online [Barzaghi, 2021b]).

A series of models belonging to the **Semantic Publishing and Referencing (SPAR)** suite—such as FRBR-aligned Bibliographic Ontology[18] [Peroni and Shotton, 2012], Bibliographic Reference Ontology[19] [Di Iorio et al., 2014], Citation Counting and Context Characterization Ontology[20] [Di Iorio et al., 2014], Discourse Elements Ontology[21] [Constantin et al., 2016], Document Components Ontology[22] [Constantin et al., 2016], **Publishing Roles Ontology (PRO)**[23] [Peroni et al., 2012], **Publishing Status Ontology (PSO)**[24] [Peroni et al., 2012]—were used to represent bibliographic entities and their features. Some **Ontology Design Patterns (ODPs)** [Gangemi and Presutti, 2009], such as Time Interval pattern[25] and Time-indexed Value in Context pattern,[26] were reused as well to express temporal parameters. Other widely reused models, like Dublin Core Metadata Terms (DCTerms)[27] [Sugimoto et al., 2002], Friend Of A Friend vocabulary[28] [Brickley and Miller, 2007], and **Simple Knowledge Organization System (SKOS)**[29] [Miles and Pérez-Agüera, 2007], were reused to represent generic utility classes and characteristics, such as names, agents, and classifications. Additionally, three SKOS-based controlled

[18] http://purl.org/spar/fabio.
[19] http://purl.org/spar/biro.
[20] http://purl.org/spar/c4o.
[21] http://purl.org/spar/deo.
[22] http://purl.org/spar/doco.
[23] http://purl.org/spar/pro.
[24] http://purl.org/spar/pso.
[25] http://www.ontologydesignpatterns.org/cp/owl/timeinterval.owl.
[26] http://purl.org/spar/tvc.
[27] http://purl.org/dc/terms.
[28] http://xmlns.com/foaf/0.1.
[29] http://www.w3.org/2004/02/skos/core.

```
<body prefix="deo: http://purl.org/spar/deo/ dcterms: http://purl.org/dc/terms/ fabio: http://purl.org/spar/fabio/">
    <p class="paragraph" data-counter="2" id="p-2">Il momento essenziale nel quale si esprime, o dovrebbe esprimersi, la
        volontà di rinnovamento della <span about="https://w3id.org/moro/enoam/data/141012/v1/mention-5"
            class="mention organization" id="mention-5" property="dcterms:references"
            resource="https://w3id.org/moro/enoam/data/democrazia-cristiana" typeof="deo:Reference">Democrazia
            Cristiana</span>, pur sempre fedele alla sua funzione storica, è il congresso del partito. Ma il dibattito
        stenta ad avviarsi come frenato da molti equivoci e timori. Purtroppo temiamo che, per volontà dell'attuale
        maggioranza relativa<sup><a href="#curatornote-2" id="curatornote-ref-2">[2]</a></sup>, ci si trovi di fronte
        proprio a quel congresso di ratifica, che fu additato come il secondo tempo di una peraltro infelice operazione
        politica<sup><a href="#curatornote-3" id="curatornote-ref-3">[3]</a></sup>, invece che a quel congresso creativo
        che sembrava, all'inizio, auspicato da tutti. Per fare la nuova maggioranza occorre aprirsi e discutere. Ed
        invece la vecchia maggioranza resta chiusa, silenziosa ed indifferente, preclusiva oggi, in questo momento
        decisivo, come lo fu ieri. È una grave responsabilità che ci si assume. Che se poi qualcuno pensasse ad
        annullare l'impegno del congresso, la cosa sarebbe ancora più grave, tenuto conto che la prospettiva di un
        congresso aperto, offerta dal Segretario on. <span about="https://w3id.org/moro/enoam/data/141012/v1/mention-7"
            class="mention person" id="mention-7" property="dcterms:references"
            resource="https://w3id.org/moro/enoam/data/mariano-rumor" typeof="deo:Reference">Rumor</span><sup><a
                href="#curatornote-4" id="curatornote-ref-4">[4]</a></sup>, costituì fondamento politico della
        formazione del governo<sup><a href="#curatornote-5" id="curatornote-ref-5">[5]</a></sup>. Noi – ha concluso
        l'on. Moro – faremo quel che è possibile da parte nostra, onestamente, con intenti costruttivi. Ma non basta la
        nostra buona volontà. Occorre la volontà di tutti, per fare della <span
            about="https://w3id.org/moro/enoam/data/141012/v1/mention-6" class="mention organization" id="mention-6"
            property="dcterms:references" resource="https://w3id.org/moro/enoam/data/democrazia-cristiana"
            typeof="deo:Reference">Democrazia Cristiana</span> un partito libero e aperto, un partito di eguali al
        servizio della Democrazia e del Paese.</p>
```

Fig. 2. A snippet of an example written in RDFa-HTML, encoding information about mentions in the text.

vocabularies were developed to describe the information related to Aldo Moro's roles he carried out in his lifetime, the subjects of its writings, and their types: Moro's Roles Vocabulary,[30] Moro's Subjects Vocabulary,[31] and Moro's Types Vocabulary.[32]

The documents published in the Edition were modeled by reusing these semantic models and were expressed through the **Resource Description Framework in Attributes (RDFa)** [Adida et al., 2007], an RDF serialization which allows adding structured data to HTML documents through a series of specific HTML attributes. Each document (see Figure 2 for an example) is characterized by:

- —basic bibliographic information, such as the license, the **Digital Object Identifier (DOI)**, the respective page on the Edition's web site, and the bibliographic citation;
- —a series of mentions to people, places, and organizations, bibliographic references to works cited by Aldo Moro, and quotations;
- —information about mentioned entities, such as attested forms, authority control, and controlled forms of people's names;
- —a series of notes that make up the critical commentary made by both researchers and Aldo Moro himself;
- —a series of paragraphs, uniquely identified and numbered, containing the document's text.

The text of each work is contained in both paragraphs and notes. Mentions, bibliographic references, and quotations are paragraph elements that were marked up semi-automatically by researchers using the KwicKwocKwac web application. Mentions and notes are modeled as text fragments, enclosed by a series of tags contained in the <body> element. Each is characterized by a series of attributes that reproduce RDF syntax within the HTML code: `@about` (the subject); `@property` (the predicate); and either `@resource` (the entity object) or `@content` (the value).

[30] http://purl.org/moro/voc/roles/.
[31] http://purl.org/moro/voc/subjects/.
[32] http://purl.org/moro/voc/types/.

Moreover, each represents an instance of a certain ontological class, which can be either doco:TextChunk (quotation), biro:BibliographicReference (bibliographic reference), or deo:Reference (mention). This information is expressed by using the @typeof attribute. If a represents a mention or a bibliographic reference, the values of its @property and @resource attributes are respectively dcterms:references and the mentioned entity's URI. Notes are represented as a series of <li> tags contained in an <ol> element, characterized by a @id attribute with a value expressing the type of the notes it contains ("curatorNotes" for notes made by researchers, and "moroNotes" for notes produced by Moro himself). Similar to the elements described above, each <li> element represents an instance of an ontological class (in this case fabio:Comment), characterized by the same @about, @property, and @resource attributes. The values of the @property and @resource attributes, in this case, are dcterms:creator and the note author's URI (either the researcher who curated the document or Aldo Moro, in accordance with the note's type).

Mentioned entities (people, organizations, places, and bibliographic resources) are collected within the <head> element. Each mentioned entity is modeled as a <meta> element representing an RDF statement about that entity, by using the same attributes mentioned above (@about, @property, and @resource or @content). Each entity is expressed as an instance of person (foaf:Person), organization (foaf:Organization), place (dcterms:Location), or bibliographic resource (fabio:Expression), by using the @typeof attribute. Each entity is also identified with a natural language label by using the @property and @content attributes with rdfs:label and a string of text (e.g., a name, a title, or a citation) as their respective values. Moreover, each entity can be aligned with the respective record on Wikidata to enforce authority control and link it to external information by using the @property and @resource properties with owl:sameAs and the corresponding Wikidata entity's URI. Finally, if an entity is a person, it is characterized by a controlled form of its name ("[Family name], [Name]") by using the @property and @content attributes with skos:prefLabel and a string of text containing the controlled form.

The information encoded in the documents through RDFa syntax was added to an RDF dataset as well, by taking the content of the statements and enriching it with more relations between entities and converting them into Turtle syntax. Indeed, having a separate dataset alongside the RDFa documents offers several advantages. First of all, a separate RDF dataset that consolidates data from different documents into a single location can facilitate integration with other datasets. By doing so, it also simplifies the processes of performing queries and analyzing their results. In general, an RDF dataset can be easily shared independent of the documents. This promotes data sharing and interoperability, allowing others to access and utilize the structured data for different purposes. Last, having a separate RDF dataset can improve the overall system performance in terms of querying and processing data to be visualized in the edition as elements that can be navigated and interacted with.

In Figure 3, we illustrate the entities and relationships that make up the information contained in the dataset.

The bibliographic resource, understood as intellectual content (fabio:Expression), is made up (frbr:part) by instances of four classes that represent mentions (deo:Reference), bibliographic references (biro:BibliograficReference), footnotes (fabio:Comment), and quotations (doco:TextChunk). The Expression is related to a Manifestation, which is the bibliographic resource understood as the digital document that was marked up by researchers and published on the Edition (fabio:Manifestation), through the frbr:embodiment property. An instance of fabio:Comment is created by (dcterms:creator) an instance of foaf:Person. An instance of biro:BibliographicReference refers to an Expression (fabio:Expression) through the biro:references property. The textual content of mentions, references, footnotes, and quotations is expressed through the c4o:hasContent property.

Bibliographic metadata related to the Edition's works were modeled by reusing well-known conceptual schemas to build a solid and extensible foundation for the knowledge base. As shown in Figure 4, metadata are distributed between three FRBR levels (Work, Expression, and Manifestation) that are interlinked through their respective properties: a Work is realized (frbr:realization) through an Expression, which in turn is embodied (frbr:embodiment) in a Manifestation. Work and Manifestation are directly linked through the fabio:hasManifestation property.

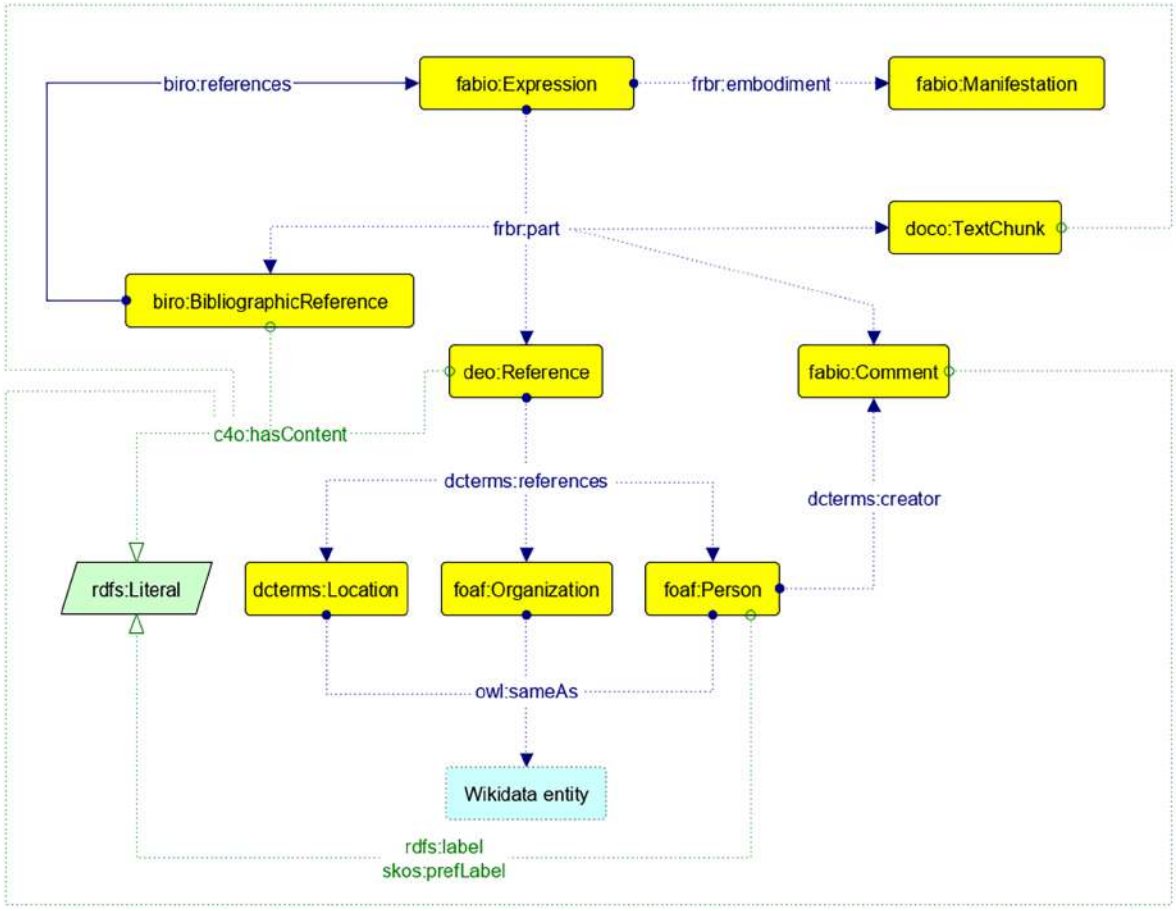

Fig. 3. The visual diagram shows the relations between mentions, bibliographic references, notes, mentioned entities, and bibliographic resources.

The metadata at the Work level (`fabio:Work`) are as follows:

- the title (`dcterms:title`);
- the author (`dcterms:creator`);
- the subject (`dcterms:subject`, whose range is an instance of the `msv:Subject` class, extracted from the controlled vocabulary of subjects of Moro's works);
- the author's role, understood as bibliographic metadata contextualized in a certain period and expressed through PRO. According to PRO, a role in time (`pro:RoleInTime`) is a situation that describes a role (specifically for the Edition, an instance of `mrv:Role`, a class extracted from the controlled vocabulary of Moro's roles) that an agent (Aldo Moro) holds for a certain period (`ti:TimeInterval`) and in relation to a certain context (`fabio:Work`).

The metadata at the Expression level (`fabio:Expression`) are as follows:

- the researcher who curated the document (`dcterms:contributor`, whose range is an instance of `foaf:Person`);
- the abstract describing it (`dcterms:abstract`);

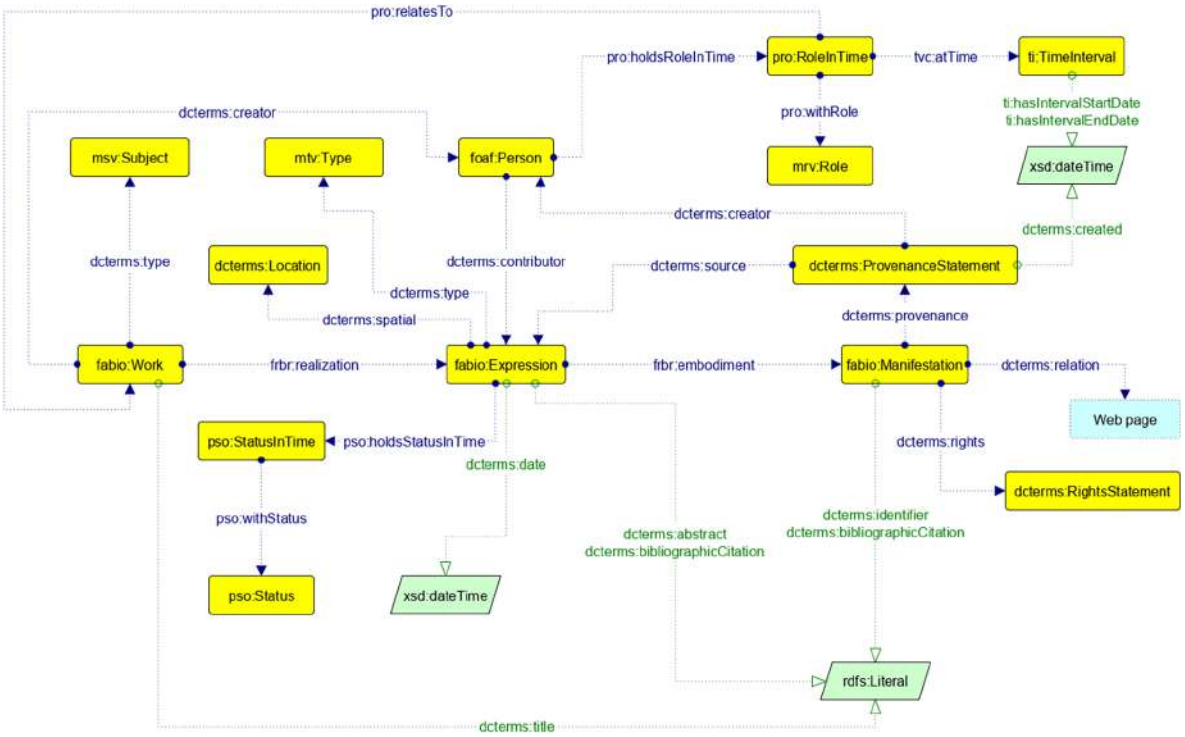

Fig. 4. The visual diagram illustrates the relation between the conceptual levels of a Moro's work and the metadata that describes its bibliographic context.

- —its type (`dcterms:type`, whose range is an instance of `mtv:Type`, a class representing the top concept of the controlled vocabulary of types of Moro's works);
- —the date of creation or first publication (`dcterms:date`);
- —the place of the event described in it (`dcterms:spatial`);
- —the publication status, expressed by using PSO, according to which a status in time (`pso:StatusInTime`) is a situation that describes a state (an instance of `pso:Status`) held by a publishable entity (the Expression, in this case);
- —additional notes (`skos:note`).

The metadata at the Manifestation level (`fabio:Manifestation`) are as follows:

- —the identifier (`dcterms:identifier`), converted into a string containing the document's DOI during the document processing phase;
- —the bibliographic sources (*dcterms:provenance*), treated as metadata of the single work with *dcterms:ProvenanceStatement*, a class that expresses a provenance statement related to any entity. Each instance of provenance statement is characterized by a series of properties:
  - —the researcher who created it (`dcterms:creator`);
  - —its creation date (`dcterms:created`);
  - —its textual content (`c4o:hasContent`), which is the metadata value originally inserted by the researcher;
  - —the source that provides the theoretical foundation of said statement (`dcterms:source`, whose range is an instance of `fabio:Expression`). The instance of `fabio:Expression` that indicates the document source

is the bibliographic resource identified by the textual content of the statement, repeated also as a citation to the work itself (`dcterms:bibliographicCitation`).

During the data processing phase, additional information was added to the Manifestation level:

— the license (`dcterms:rights`): works published on the Edition are distributed under the Creative Commons Attribution-Non-Commercial 4.0 International (CC BY-NC 4.0),[33] which allows one to freely access, use, reproduce, distribute, transmit, communicate, and show them as well as produce and distribute derivative works with any directly or indirectly commercial use strictly forbidden on the condition that the authors' moral rights are safeguarded and that the original sources are adequately acknowledged and cited;
— the textual string of the bibliographic citation (`dcterms:bibliographicCitation`);
— the link to the work's web page on the Edition's web site (`dcterms:relation`).

6 Annotation Process

6.1 Markup

To prepare Aldo Moro's works to be published on the Edition web site, we developed a web editor to semi-automatically generate markup for the correct indexing and processing of the digital versions of Moro's texts. The web application, called *KwicKwocKwac*,[34] provides rapid and complete tools to markup texts according to a parameterized set of features and generates visualizations of them using **"KeyWord in Context" (KWIC)**, **"KeyWord out of Context" (KWOC)**, and **"KeyWord Alongside Context" (KWAC)**, three standard methods for visualizing concordance lines in literary studies,[35] possibly first discussed in [Luhn, 1960].

The main features of the application are as follows:

— Rapid markup of intertextual elements, e.g.: people, organizations, places, bibliographic references, and quotations; the list is totally parameterized and can be expanded and modified at will;
— Expansion of marked up elements to all instances of the same entity throughout the text;
— Automatic generation of concordance lines, alphabetically ordered lists of entities that are mentioned in the text, with an indication of the textual context surrounding that mention, using one of the KWIC, KWOC, or KWAC approaches;
— Automatic and manual entity disambiguation (especially in presence of different spellings, partial persons' names, indirect mentions, and so on);
— Data reconciliation with authority lists such as Wikidata[36] to further disambiguate marked entities and link them to web resources, to ensure validity and quality control;
— Metadata insertion, to describe the historical and bibliographic context of each work.

Intertextual elements that are marked up by researchers belong to two main categories: mentions and references. Mentions have been established, for this project, to point to People (e.g., "El Maghrabi"), Organizations (e.g., "FAO"), and Places (e.g., "Libia"); references for this project are either Bibliographic references or Quotations (following the data model described in Section 5). Intertextual elements, once identified, are recorded in the HTML markup and used for further operations:

— searching, marking, and grouping mentions that refer in the same way to the same entity (e.g., all instances of the word "Libia");

[33] https://creativecommons.org/licenses/by-nc/4.0/.
[34] https://github.com/sanofrank/KwicKwocKwac.
[35] The name of the tool is also a play on words on the Italian names of Huey, Dewey, and Louie, the nephews of Donald Duck, which in Italy are called Qui, Quo, and Qua.
[36] https://www.wikidata.org/.

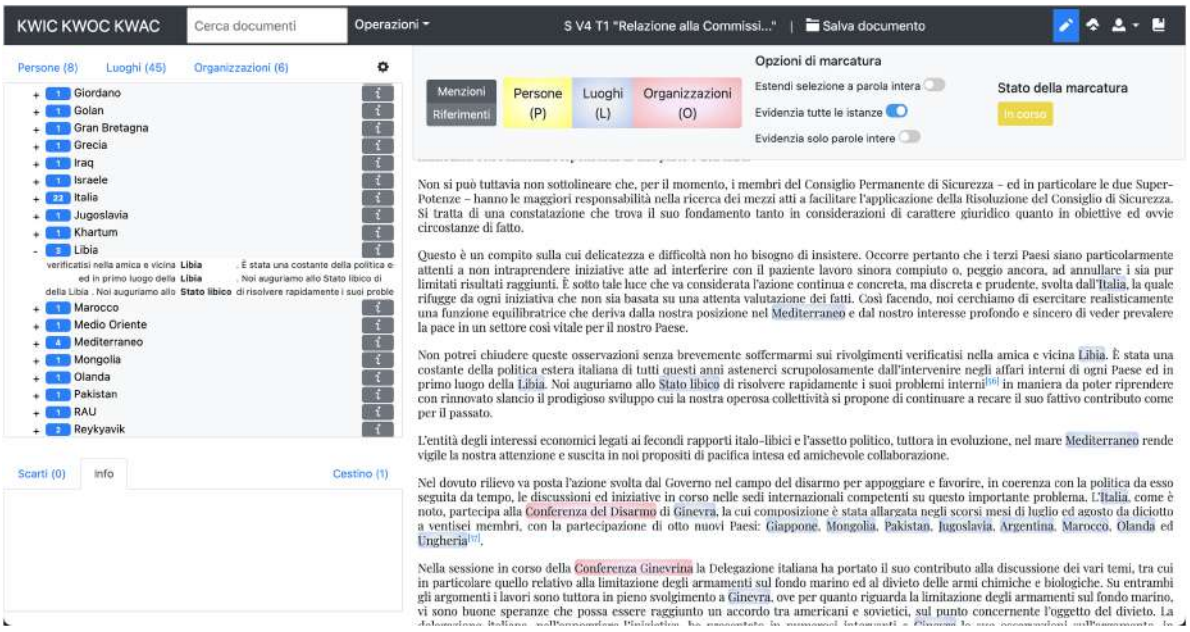

Fig. 5. Joining the mentions "Libia" and "Stato libico" as one single entity.

— joining different groups of mentions that refer in different ways to the same entity (e.g., see Figure 5, where "Libia" and "stato libico" in the original texts are grouped together since the two wordings clearly refer to the same entity);
— deleting incorrectly marked up mentions;
— complementing entities with additional information, such as a label (e.g., "NATO" for the grouping of the mentions of "N.A.T.O.," "North Atlantic Treaty Organization," "OTAN," and "Organizzazione del Trattato dell'Atlantico del Nord") and sort value (in Italian, it is customary to show person names with their first name first (e.g., "Giulio Andreotti"), but to sort them according to their surname first (e.g., "Andreotti, Giulio");
— associating each entity (after grouping) to the Wikidata and Treccani identifiers if available. KwicKwocKwac, as shown in Figure 6, connects to the Wikidata endpoint using its REST API[37] and looks for all entities matching the name of the given entity, suggesting to the users the retrieved matches and allowing them to choose the best fit (if any). The corresponding Wikidata or Treccani ID[38] is then permanently associated with the entity as retrieved.

For each work, the final result of the markup process is represented by an HTML document enriched with semantic information about the markup generated by the researcher.

6.2 Metadata Addition

As shown in Figure 7, the web application allowed researchers to add bibliographic metadata to each document they worked on during the markup phase by filling in a form connected to a dedicated MongoDB database (the model is based on what already described in Section 5). In particular, the metadata inserted include:

[37] https://www.wikidata.org/wiki/Wikidata:REST_API.
[38] https://www.wikidata.org/wiki/Property:P3365.

Fig. 6. Choosing the Wikidata identifier of a named entity.

—Document number: a three-digit number ranging between 001 and 999, to be inserted in chronological order (e.g., the number 001 will be assigned to the first document, 002 to the second, and so on);
—Author's role: the role held by Aldo Moro when the document was created/published;
—Researcher curator: the full name of the researcher who transcribed, commented, marked, and assigned metadata to the document;
—Abstract: the document description prepared by the researcher;
—Document type: one or more categories the document belongs to;
—Document subject: one or more categories the document contents belong to;
—Document status: the indication whether the document was published or not;
—Bibliographic reference/Archival identifier: one or more provenance indications of the document (reference to the editorial source, if published; archival identifier, if unpublished);
—Place: the name of the location where the facts described in the document take place, if applicable;
—Date: the date (day-month-year, or month-year, or year) of the first publication (if published) or creation (if unpublished);
—Additional notes: generic information that could not be integrated immediately in the metadata model (such as the author's signature, undefined sources, and so on).

For each work, the final result of the process of metadata insertion is represented by a JSON document stored in the external database and uniquely assigned to that work.

6.3 Data Reworking and Control

Following the completion of markup and metadata insertion processes, the HTML documents and corresponding metadata stored in the database underwent additional refinement through a series of Python scripts to optimize their presentation in the Edition. These scripts included the following functions: refactoring document code, rectifying errors, standardizing resource URIs, generating TEI documents that closely mirror the source code while incorporating bibliographic metadata in the document header, producing PDF documents with pagination and visualization controlled by a CSS stylesheet, converting metadata into an RDF dataset written in Turtle syntax and organized according to the standards established in the data modeling phase and illustrated in the Section 5, integrating RDFa markup into the dataset and forming a complete knowledge base of the Edition, and associating a DOI to each document.

6.4 Data Publication

The works that at present are published in the Edition are 1631, distributed in a total of 11 volumes organized around two main sections (as described in Section 4). Each digital work is licensed under a Creative Commons

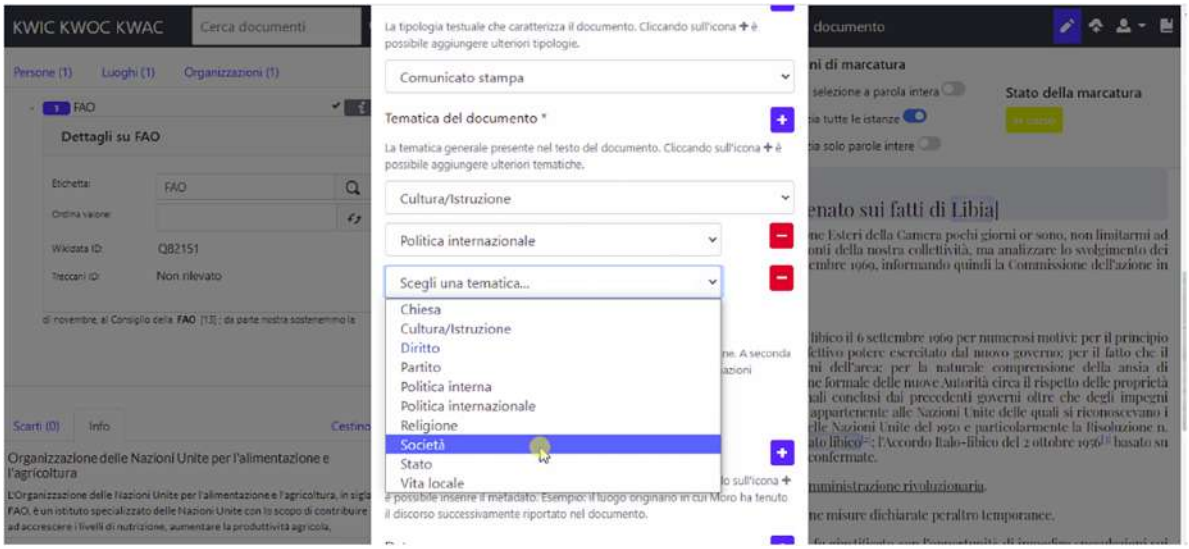

Fig. 7. An example of metadata insertion in KwicKwocKwac.

Attribution-Non-Commercial 4.0 license[39] and can be downloaded in three main formats (HTML-RDFa, XML-TEI, and PDF), so as to be easily shared and ease further scholarly analyses. The knowledge base of the Edition is the main deliverable of the data semantification process and has been published openly as a data dump.[40]

One of the main objectives of the Edition is documenting the work phases of a project that required multiple different skills and the adoption of specialized methodologies, standards, and tools. Indeed, the workflow of the Edition represents an expansive scientific process. In addition to the aforementioned resources, its outputs include other deliverables, such as the code and the dataset of the competitive audit, or the images used during the design process and the content strategy. All these resources produced during the process are available on Zenodo as well and listed on the Edition web site.[41]

7 Web site Structure and Online Interface

As mentioned in Section 2, the Edition web site development was based on a preliminary audit of a selected sample of digital editions available on the Web. The evaluation considered a series of quality criteria recognized in the scientific literature. In particular, it was articulated in three main phases:

—Content analysis of several digital editions, using as main reference the Criteria for Reviewing Scholarly Digital Editions, version 1.1, edited by Patrick Sahle in collaboration with Georg Vogeler and the members of Institut für Dokumentologie und Editorik (IDE);[42]
—Data processing, so as to extract relevant information;
—Review of the data processing results and consideration of the models that can be used as a reference.

The audit results were reported in a technical report [Barzaghi, 2021a] that offers a high-level overview of the results obtained by interpreting gathered data, computing and visualizing the process in detail, and suggesting

[39] https://creativecommons.org/licenses/by-nc/4.0/.
[40] https://zenodo.org/doi/10.5281/zenodo.5592156.
[41] https://aldomorodigitale.unibo.it/about/docs/results#reports-section.
[42] https://www.i-d-e.de/publikationen/weitereschriften/criteria-version-1-1/.

the digital edition models to consider with respect to the project goals. Based on the audit results, we designed the information architecture of the Edition web site,[43] whose content is organized in four core sections:

(1) *Aldo Moro*: a section presenting an overview of Aldo Moro's personal story;
(2) *The Works*: a section dedicated to exploring the Edition through three main navigation systems (i.e., browsing by content, semantic indexes, and faceted search);
(3) *The Itineraries*: a section dedicated to visualizing data about some meaningful aspects of Aldo Moro's life;
(4) *The Project*: a section dedicated to documenting the whole process of designing, developing, and publishing the Edition, including information about the editorial principles followed by researchers, the conceptual models used to represent data, the treatment of the digitized works, the results, and more.

The Works and *The Projects* are the two views that contain most materials. By clicking on *The Works* tab in the menu, three navigation options are displayed, as shown in Figure 8. The first option to navigate the edition is through key information that was previously indexed. Indeed, by clicking on *Indexes*, further options are shown to the user to navigate the document collection starting from lists of *People*, *Places*, *Organizations*, or going through the *Bibliography*. Each of these options follows specific visualization strategies. For instance, *Places* can be accessed starting from a map, while *People* and *Organizations* are listed alphabetically together with the number of occurrences in the document collection. Search in these tabs can be further refined by defining a specific time span of interest or a portion of the corpus. If, from the view in Figure 8, we click on *Contents*, the collection can be explored by following its editorial structure, for example, by selecting the different sections and subsections of the digital edition. The different documents can be opened, each with its metadata, abstract, provenance, and citation. Critical notes can be accessed too when available. The third navigation option is *Advanced search*, giving users the possibility to perform searches through one or more query terms (document title and/or keywords), also combined with filters related to the year of publication, author's role, document type and more.

The page of the individual work is organized into three main sections. The first section contains navigation and data dissemination functionalities (such as downloading in multiple formats) and basic identifying information (title, abstract, sources, and bibliographic citation). The second section contains the text (Figure 9) displaying a series of mentions that can be clicked to show additional information taken from WikiData and related to the mentioned entity. Other clickable elements are note reference pointers, which when clicked open modals containing the respective notes written by Moro or the curators, and a table of contents that collects the mentioned entities of the text and makes them highlightable in different colors, also offering the possibility to activate a reading mode to completely hide the mentions and make the text more readable. The third section contains the remaining bibliographic metadata associated with the document (Figure 10).

The platform was implemented using Vue, a versatile framework to build user interfaces, together with Vuetify, a library that implements Google's Material Design. Both choices, providing strong accessibility (a11y) features on a code and design level, comply with three fundamental principles: Clear, Robust, and Specific.[44] The actual interactions between interface and final user depend on Node.js, a widely known framework trusted by developers due to its high performance, functions extensibility, and security.

7.1 Geographical Entities

As described in Section 6.1, the named entities annotated in the documents were semi-automatically linked to the corresponding record in WikiData. Through this link, some additional information about the cited entity is available. In particular, for geographical entities, one can get the exact coordinates (longitude and latitude). We therefore included a map in the platform, where users can search and filter geographical terms that were found in the corpus. Since the number of distinct geographical entities is very big (568 items), a map containing all the

[43] https://aldomorodigitale.unibo.it/.
[44] https://m2.material.io/design/usability/accessibility.html.

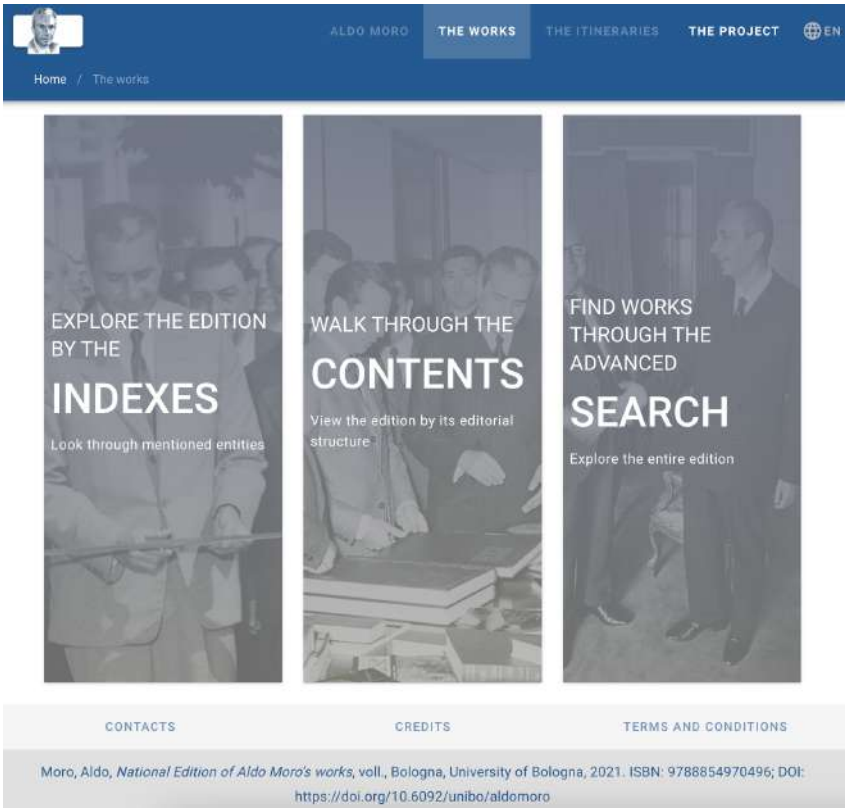

Fig. 8. Three navigation possibilities available under *The Works*.

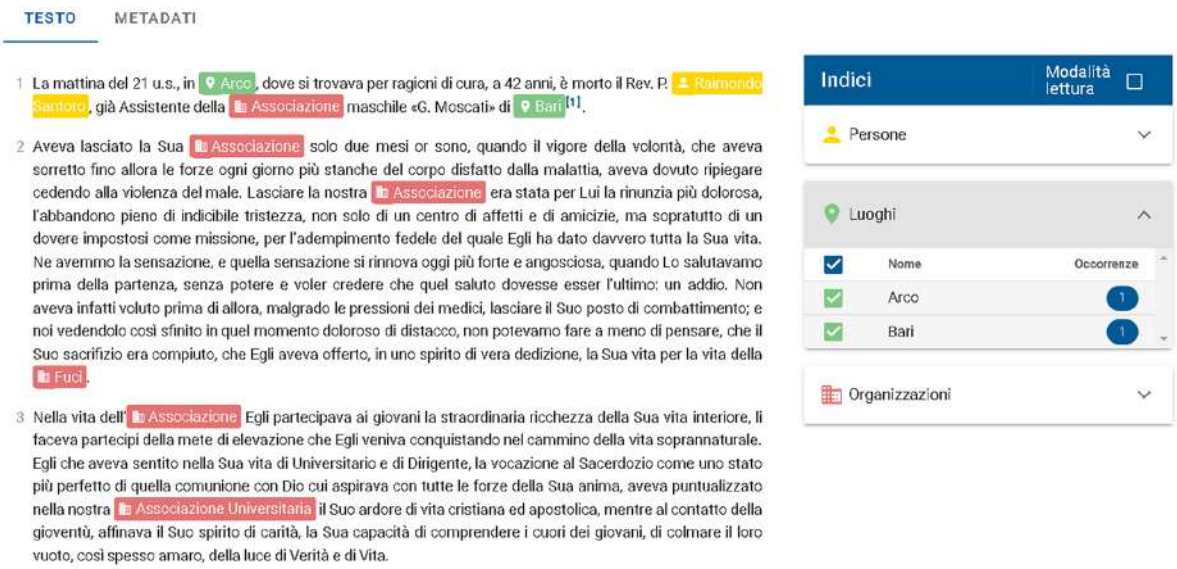

Fig. 9. The second section of the page of a work in the Edition.

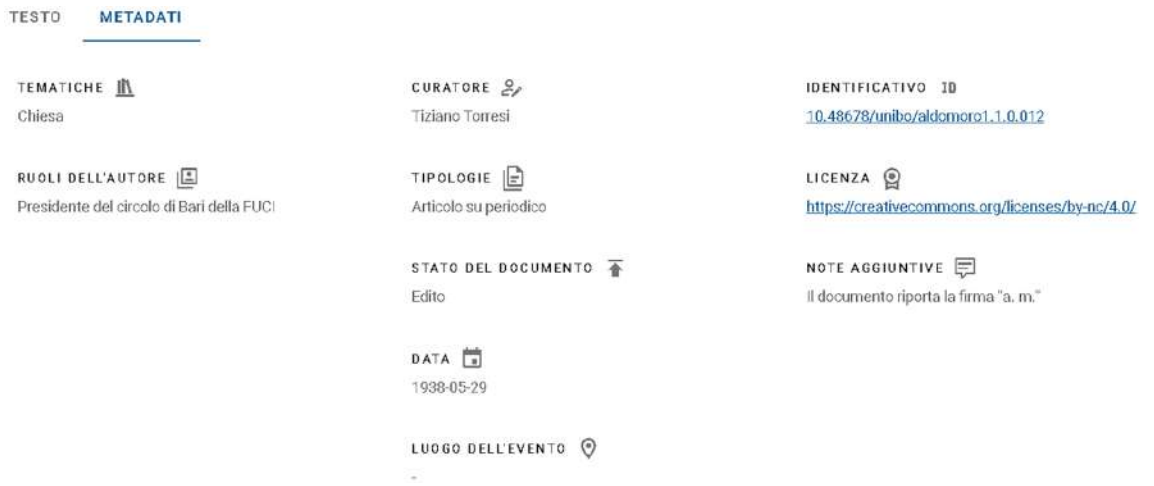

Fig. 10. The third section of the page of a work in the Edition.

Fig. 11. The three zoom levels available for the geographical entities navigation.

dots at once would be difficult to surf, therefore we decided to provide three different zoom levels to let users have a better navigation experience, as displayed in Figure 11:

- On the first level (purple dots), when all the world is visible in the canvas, only continents, big regions, and seas are visible. These include, for instance, "Nord America" (North America), "Oceano Atlantico" (Atlantic Ocean), and so on.
- The second level (blue dots), when a single continent is included, contains countries, medium-sized regions, and big bodies of water; for example, "Spagna" (Spain), "Scozia" (Scotland), or "Eufrate" (Euphrates).
- The last level (orange dots), with the maximum level of zoom, contains small regions and cities, such as "Firenze" (Florence), "Sicilia" (Sicily), "Berlino Est" (East Berlin).

In all the described levels, the size of the dots depends on the number of times that location is contained in the corpus. The list of dots can also be filtered by document category and time span.

7.2 Keywords

When dealing with large corpora, it is difficult to search through the text and group documents using key concepts. Most of the times, this task is done manually, therefore the needed effort is proportional to the size of the resource.

Fig. 12. Keywords filter.

To tackle this challenge, we use **Keyphrase Digger (KD)** [Moretti et al., 2015], a tool to perform automatic keyphrase extraction and clustering. Given a document, KD derives a ranked and weighted list of single words and multi-token expressions, which represent the most important concepts mentioned in the text, based on a statistical analysis combined with position information of the term(s) in the document.

Compared to single word searches, multi-token expressions can better cover complex concepts, for instance, "aumento dei prezzi" (price increase), "campagna elettorale" (electoral campaign), and so on. A total of 4,622 concepts have been extracted from the corpus. This list can be used as an additional filter to search documents that deal with a specific topic. Figure 12 shows the interface where keywords can be selected from a menu in alphabetical order.

8 Conclusion and Future Work

In this article, we detailed the entire process of conceptualizing, developing, and implementing a digital scholarly edition along with its features as a workflow supported by semantic technologies. In particular, we described how this workflow was applied to create the National Edition of Aldo Moro's works, a digital edition encompassing both published and unpublished texts of the esteemed Italian statesman. Currently, the edition stands as a web-based scientific resource for experts and the public alike to delve into Aldo Moro's life and work through exploration, visualization, and download of a data-rich corpus of texts. It boasts 1,631 fully digitized documents, enriched with bibliographic metadata encoded in RDF, identified with DOIs, and available for download in various formats, including RDFa-HTML and TEI-XML. The use of open standards during the data modeling process ensures data reusability for further processing, analysis, and study. By leveraging constructs such as SPAR ontologies and ODPs to structure the data, the edition offers a rich and extensible knowledge base about the context surrounding Aldo Moro's endeavors. In order to enrich the texts with such semantic information, we developed a markup tool that allows non-experts to inject semantically meaningful information in their documents and that can be eventually adapted so as to be used in other projects as well. The full range of activities involved in this workflow have been recorded in both the edition itself and in other scholarly deliverables that were published in open repositories (such as Zenodo and AMS Acta) to foster data sharing and replicability.

While our work provides valuable insights, it is not without limitations. Indeed, the edition is still a work in progress, with several volumes and sections yet to be incorporated. Further iterations are required for digitization, encoding, and uploading of the entire corpus of texts. Moreover, rigorous quality checks must be conducted to ensure the integrity of the new material and to evaluate its potential impact on web site performance and search functionalities.

Furthermore, certain sections within the edition (such as "The Itineraries") need additional development efforts aimed at implementing better data visualization strategies. "The Itineraries" section, indeed, holds particular significance as it aims to offer insights into Aldo Moro's activities throughout his life, providing valuable contextual information for understanding his contributions and the overall themes he dealt with in his political activities. A possible inspiration could come from projects like the Library of Digital Latin Texts [Huskey, 2020] for innovative forms of data visualization of texts and their metadata. Finally, in light of the continuously growing volume of information within the digital edition, it is necessary to consider transitioning the existing knowledge base from its current single dump file format to a more adaptable and scalable storage solution, such as Blazegraph, to accommodate a substantially larger volume of data with ease, without compromising performance or usability.

References

Ben Adida, Mark Birbeck, Shane McCarron, and Ivan Herman. 2007. RDFa Core 1.1. Retrieved from https://www.w3.org/TR/2015/REC-rdfa-core-20150317

Sebastian Barzaghi. 2021a. Competitive audit for designing and developing the National Edition of Aldo Moro's works. DOI: https://doi.org/10.5281/zenodo.5184721

Sebastian Barzaghi. 2021b. *Data Modelling in the National Edition of Aldo Moro's Works*. Technical Report. Zenodo. DOI: https://doi.org/10.5281/zenodo.5524746

Sebastian Barzaghi, Francesco Paolucci, Francesca Tomasi, and Fabio Vitali. 2024. KwicKwocKwac, a tool for rapidly generating concordances and marking up a literary text. arXiv:2410.06043. Retrieved from https://arxiv.org/abs/2410.06043

Dan Brickley and Libby Miller. 2007. *FOAF Vocabulary Specification 0.9*. Technical Report. FOAF Project. Retrieved from http://xmlns.com/foaf/spec/20070524.html

Lou Burnard, Katherine O'Brien O'Keeffe, and John Unsworth. 2006. *Electronic Textual Editing*. Modern Language Association of America.

Alexandru Constantin, Silvio Peroni, Steve Pettifer, David Shotton, and Fabio Vitali. 2016. The document components ontology (DoCO). *Semantic Web* 7, 2 (2016), 167–181.

Ute Daniel, Peter Gatrell, Oliver Janz, Heather Jones, Jennifer D. Keene, Alan Kramer, and Bill Nasson. 2014. 1914–1918-online. International Encyclopedia of the First World War. Introduction.

Marilena Daquino, Francesca Giovannetti, Francesca Tomasi. 2019. Linked data per le edizioni scientifiche digitali. Il workflow di pubblicazione dell'edizione semantica del quaderno di appunti di paolo bufalini. *Umanistica Digitale* 7 (2019), 49–75.

Margherita De Blasi. 2020. Paola italia, editing duemila. PER una filologia dei testi digitali. *Annali-Sezione Romanza* 62, 1 (2020), 287–294.

Angelo Di Iorio, Andrea Giovanni Nuzzolese, Silvio Peroni, David M. Shotton, and Fabio Vitali. 2014. Describing bibliographic references in RDF. In *SePublica 2014*, 49–60.

Sonja Dickow-Rotter and Daniel Burckhardt. 2022. Chapter 2 traces of Jewish Hamburg: A digital source edition of German-Jewish history. In *Writing the Digital History of Nazi Germany—Potentialities and Challenges of Digitally Researching and Presenting the History of the Third Reich, World War II, and the Holocaust*. Julia Timpe and Frederike Buda (Eds.), De Gruyter, Berlin, Boston, 17–38. DOI: https://doi.org/10.1515/9783110714692-002

Emilie Oléron Evans, Susanne Müller, and Costanza Giannaccini. 2017. Cultural Constellations: Burckhardtsource.org. *Open Library of Humanities* 3, 1 (2017), 7. DOI: https://doi.org/10.16995/olh.158

Domenico Fiormonte. 2003. *Scrittura e Filologia Nell'era Digitale*. Bollati Boringhieri.

Samantha Kathryn Fitch. 2017. Rediscovering livingstone: Livingstone online. In *Digital Initiatives Symposium*.

Greta Franzini, Simon Mahony, and Melissa Terras. 2016. *A Catalogue of Digital Editions*. Open Book Publishers.

Dom J. Froger. 1968. *La critique des textes et son automatisation*.

Aldo Gangemi and Valentina Presutti. 2009. Ontology design patterns. In *Handbook on Ontologies*. SteffenStaab and RudiStuder (Eds.), Springer, Berlin, 221–243. DOI: https://doi.org/10.1007/978-3-540-92673-3_10

Tim Hitchcock and Robert Shoemaker. 2006. Digitising history from below: The old bailey proceedings online, 1674–1834. *History Compass* 4, 2 (2006), 193–202. DOI: https://doi.org/10.1111/j.1478-0542.2006.00309.x Retrieved from https://compass.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1478-0542.2006.00309.x

Samuel J. Huskey. 2020. Scholarly digital editions: A wise investment for scholars and institutions. *Digitale Altertumswissenschaften: Thesen Und Debatten zu Methoden Und Anwendungen* 4 (2020), 43–54.

Jean Irigoin. 1979. La pratique des ordinateurs dans la critique des textes. *Revue D'Histoire Des Textes* 8 (1979), 337–340.

Hans Peter Luhn. 1960. Key word-in-context index for technical literature (kwic index). *American Documentation* 11, 4 (1960), 288–295.

Jerome J. McGann. 2001. *Radiant Textuality: Literature after the World Wide Web*. Palgrave.

Alistair Miles and José R. Pérez-Agüera. 2007. SKOS: Simple knowledge organisation for the web. *Cataloging and Classification Quarterly* 43, 3-4 (2007), 69–83. DOI: https://doi.org/10.1300/J104v43n03_04

Raul Mordenti. 2001. *Informatica e Critica Dei Testi*. Bulzoni.

Giovanni Moretti, Rachele Sprugnoli, and Sara Tonelli. 2015. Digging in the dirt: Extracting keyphrases from texts with KD. In *Proceedings of the Second Italian Conference on Computational Linguistics (CLiC-It '15)*, 198–203.

Tito Orlandi. 1998. Ripartiamo dai rDiasistemi. *I Nuovi Orizzonti Della Filologia, Ecdotica, Critica Testuale, Editoria Scientifica e Mezzi Informatici Elettronici, Convegno Internazionale 27-29 Maggio 1998. Atti Dei Convegni Lincei*, 87–101.

Lorenzo Perilli. 1995. *Filologia computazionale*. Accademia Nazionale dei Lincei.

Silvio Peroni and David Shotton. 2012. FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. *Journal of Web Semantics* 17 (2012), 33–43. DOI: https://doi.org/10.1016/j.websem.2012.08.001

Silvio Peroni, David Shotton, and Fabio Vitali. 2012. Scholarly publishing and linked data: Describing roles, statuses, temporal and contextual extents. In *Proceedings of the 8th International Conference on Semantic Systems*, 9–16.

Elena Pierazzo. 2016. *Digital Scholarly Editing: Theories, Models and Methods*. Routledge.

Elena Pierazzo. 2019. What future for digital scholarly editions? From Haute Couture to Prêt-à-Porter. *International Journal of Digital Humanities* 1, 2 (2019), 209–220.

Peter Robinson. 2005. Current issues in making digital editions of medieval texts—Or, do electronic scholarly editions have a future? *Digital Medievalist* 1 (2005). DOI: https://doi.org/10.16995/dm.8

Patrick Sahle. 2016. What is a scholarly digital edition? *Digital Scholarly Editing: Theories and Practices* 1 (2016), 19–39.

Elena Spadini, Francesca Tomasi, and Georg Vogeler. 2021. *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, Vol. 15. BoD–Books on Demand.

Shigeo Sugimoto, Thomas Baker, and Stuart L. Weibel. 2002. Dublin core: Process and principles. In *Digital Libraries: People, Knowledge, and Technology*. Ee-Peng Lim, Schubert Foo, Chris Khoo, Hsinchun Chen, Edward Fox, Shalini Urs, and Thanos Costantino (Eds.), Springer Berlin, Berlin, 25–35.

Francesca Tomasi, Marilena Daquino, and Sebastian Barzaghi. 2020. Vespasiano da bisticci, lettere. *Knowledge Base* 2020 (2020). DOI: https://doi.org/10.6092/unibo/amsacta/6852

Sara Tonelli, Rachele Sprugnoli, Giovanni Moretti, Stefano Malfatti, and Marco Odorizzi. 2020. Epistolario De gasperi: National edition of De gasperi's letters in digital format. *IX Convegno Annuale AIUCD* (2020), 253–259.

Tiziano Torresi. 2021. Nota Storico-Critica. Edizione Nazionale Delle Opere di Aldo Moro, Scritti e Discorsi, Gli Anni Giovanili (1932-1946). DOI: https://doi.org/10.48678/unibo/aldomoro1.1.0.note

Received 22 April 2024; revised 11 October 2024; accepted 30 December 2024