

magazén

International Journal
for Digital and Public Humanities

e-ISSN 2724-3923

Vol. 4 – Num. 1 June 2023



Edizioni
Ca' Foscari

e-ISSN 2724-3923

magazén
International Journal
for Digital and Public Humanities

Editors-in-chief
Franz Fischer
Diego Mantoan

Edizioni Ca' Foscari - Venice University Press
Fondazione Università Ca' Foscari
Dorsoduro 3246, 30123 Venezia
URL <https://edizionicafoscari.unive.it/en/edizioni/riviste/magazen/>

magazén

International Journal for Digital and Public Humanities

Semestral Journal

Editors-in-chief Franz Fischer (Università Ca' Foscari Venezia, Italia) Diego Mantoan (Università degli Studi di Palermo, Italia)

Associate editor Barbara Tramelli (Libera Università di Bolzano, Italia)

Managing editor Elisa Corrà (Università Ca' Foscari Venezia, Italia)

Advisory board Ben Brumfield (Brumfield Labs, Texas, USA) Stefano Campana (Università di Siena, Italia) Maria Luisa Catoni (Scuola IMT Alti Studi Lucca, Italia) Thomas Cauvin (C2DH, Université du Luxembourg) Gregory Crane (Tufts University, USA) Andreas Fickers (C2DH, Université du Luxembourg) Erma Hermens (Fitzwilliam Museum, Cambridge, UK) Karin Leonhard (Universität Konstanz, Deutschland) Serge Noiret (European University Institute, Italy) Tito Orlandi (Accademia dei Lincei, Roma, Italia; Hiob Ludolf Centre for Ethiopian and Eritrean Studies, Hamburg, Deutschland) Chiara Ottaviano (Cliomedia Officina, Torino, Italia) Jussi Parikka (Aarhus University, Denmark) Sebastian Federico Ramallo Asensio (Universidad de Murcia, España) Gino Roncaglia (Università della Tuscia, Italia) Charlotte Roueché (King's College London, UK) Patrick Sahle (Bergische Universität Wuppertal, Deutschland) Chiara Zuanni (Karl-Franzens-Universität Graz, Österreich) Joris van Zundert (Huygens Instituut, Nederland)

Editorial board Federico Bernardini (Università Ca' Foscari Venezia, Italia) Federico Boschetti (Università Ca' Foscari Venezia, Italia) Elisa Corrà (Università Ca' Foscari Venezia, Italia) Stefano Dall'Aglio (Università Ca' Foscari Venezia, Italia) Stefania De Vincentis (Università Ca' Foscari Venezia, Italia) Holger Essler (Università Ca' Foscari Venezia, Italia) Tiziana Mancinelli (Università Ca' Foscari Venezia, Italia) Paolo Monella (Sapienza Università di Roma, Italia) Deborah Paci (Università di Modena e Reggio Emilia, Italia) John Pavlopoulos (Athens University of Economics and Business, Greece) Barbara Tramelli (Libera Università di Bolzano, Italia)

Head office Università Ca' Foscari Venezia | Venice Centre for Digital and Public Humanities | Dipartimento di Studi Umanistici | Palazzo Malcanton Marcorà | Dorsoduro 3484/D, 30123 Venezia, Italia | magazen@unive.it

Publisher Edizioni Ca' Foscari | Fondazione Università Ca' Foscari | Dorsoduro 3246, 30123 Venezia, Italia | ecf@unive.it

© 2023 Università Ca' Foscari Venezia

© 2023 Edizioni Ca' Foscari for the present edition



Quest'opera è distribuita con Licenza Creative Commons Attribuzione 4.0 Internazionale
This work is licensed under a Creative Commons Attribution 4.0 International License

[ve]dph

Venice Centre for
Digital and Public
Humanities



Certificazione scientifica delle Opere pubblicate da Edizioni Ca' Foscari: tutti i saggi pubblicati hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima sotto la responsabilità del Comitato scientifico della rivista. La valutazione è stata condotta in aderenza ai criteri scientifici ed editoriali di Edizioni Ca' Foscari.

Scientific certification of the works published by Edizioni Ca' Foscari: all essays published in this issue have received a favourable opinion by subject-matter experts, through an anonymous peer review process under the responsibility of the Advisory Board of the journal. The evaluations were conducted in adherence to the scientific and editorial criteria established by Edizioni Ca' Foscari.

Table of Contents

RELATIONS

- Building Relations and Enhanced Relationality
as the Backbone of Methodologies in the Digital
and Public Humanities**
Franz Fischer, Diego Mantoan, Barbara Tramelli 7
- Articulating Intra- and Intertextual Relationships
in the Fragment Collection
Working with the Digital Edition of Giacomo Leopardi's *Zibaldone***
Silvia Stoyanova 13
- Public History e metadata
Da fonti a oggetti informativi.
Diffusione, didattica, co-creazione**
Manfredi Scanagatta 43
- La 'Galassia MQDO': un concetto di filologia
tradizionale, digitale, sostenibile**
Martina Venuti, Angelo Mario Del Grosso, Federico Boschetti,
Luigi Tessarolo, Alessia Prontera, Dylan Bovet,
Gianmario Cattaneo, Valeria Melis 71
- La ricerca sulle ultime lettere della Resistenza
dagli anni Cinquanta a oggi**
Igor Pizzirusso 121
- Designing for Audience Engagement
Exploring the Use of Online Metrics in the GLAM Sector**
Ellen Charlesworth, Claire Warwick,
Leonardo Impett, Andrew M. Beresford 145



Relations

edited by Franz Fischer, Diego Mantoan, Barbara Tramelli

Building Relations and Enhanced Relationality as the Backbone of Methodologies in the Digital and Public Humanities

Franz Fischer

Università Ca' Foscari Venezia, Italia

Diego Mantoan

Università degli Studi di Palermo, Italia

Barbara Tramelli

Libera Università di Bolzano, Italia

1 **Sharing and Caring as a Driver for Better Humanistic Knowledge**

With still no end in sight to the catastrophic conflict in Ukraine, its possible solution hinges on the creation of a new world order, hence pointing at the necessary improvement or, at least, at the development of different international relations: among the traditional East-West divide, as well as between the hyper-industrialised North and the global South. It may seem pretentious that a scholarly enterprise in the Digital and Public Humanities can act upon such superior forces, but in our understanding the contribution of academics working in

This introduction paper was mutually agreed on by the Authors who acted as editors of *magazén's* issue 1, volume 2023, with the precious support of the Journal Manager Elisa Corrà and the extended help of the journal's Editorial Board.

this field is showing how open participation is a principle leading to a more stable and shared balance, in knowledge building as well as in democratic consent. When we set out to determine this year's topic, we were indeed convinced that the concept of "Relations" constituted an intrinsic characteristic of Digital and Public Humanities that we should highlight and thematise in a global discourse (Fiormonte, Chaudhuri, Ricaurte 2022). Our intention for the present issue was to draw particular attention to the public aspects of research endeavours in this field, given that successful projects hold firm to the principle of audience involvement from their very inception, rather than having public interaction just as a late side-effect of scholarly work.

Hence, the present issue focuses on the concept of "Relations" and how they are implemented, operationalised, and analysed as interdependencies, links, and connections in the various practices of Digital and Public Humanities scholarship. Be it through the construction of simple relational databases or by means of complex correlations of data, materials, immaterial aspects, and publics, the interdisciplinary field of Digital and Public Humanities truly strives on building relationships. Far from shutting themselves up in a lonesome ivory tower, scholars in this domain are prone to bridging experiences between different disciplines, interconnecting with diverse audiences for research and dissemination purposes, and linking computational models to cultural manifestations (Marty 2010). Relations maintain a very humanistic character, as they form the immaterial structure onto which human societies are based. Indeed, transferred into a technological setting, the principle of relationship-building can be found as the core feature of interconnected data, authority files, user-centred design, usability, user-experience, audience interaction, and many more aspects that rapidly became an integral part of Digital and Public Humanities (King, Stark, Cooke 2016).

2 Successful Relationships as the Result of Broad and Novel Collaborations

The very act of building relations or putting things into relationship - be they technology and culture, scholars and audiences, data and materials - thus form the backbone of scholarly projects that came to define this research field. In Digital Textual Scholarship, scholars can see the text as a hub, whose relations point in many directions (Drucker 2003). They can explore, formalise, visualise, and process its connections with paratexts, other texts, physical documents, places, people, concepts as well as with readers. In Digital and Public History, it is possible to establish connections between sources, events, places, dates, or people, through a database, a network analysis map, or more generally the semantic web (Sternfeld 2011).

In Public History, also the relationship between the research and its public is crucial: the active or passive participation of the audience affects questions, methodology and research outcomes in many different ways. Relations between archaeological contexts and spaces are key factors to decipher and investigate stratigraphic sequences and past landscapes in virtual reconstructions (extended matrix approach, archeoBIM, etc.) and GIS studies (viewshed analysis, least-cost path analysis, etc.) (Dell'Unto, Landeschi 2022). At the same time, the integration of different remote sensing techniques and expertises often provide significant results. In Digital and Public Art History, subjects may address the connection between GLAM studies and the creative industries sector as well as the artistic productions, considering: the relation between public space and art historical environment, including different techniques of virtual reconstruction (3D modelling, gaming, VR360), the implementation of digital collections adopting the IIF framework (iconography, diagnosis, descriptions), and multimedia solutions for data collection and accessibility (Barranha, Henriques 2021). The implication of digital technologies in the preservation, management and interpretation of historical, cultural, and archaeological contexts is a way to make knowledge more accessible for different audiences. The engagement of society in participatory ways highlights the importance of creating and cherishing dialogue, cohesion, identity, sustainable development, and - hopefully - peace.

3 Of the Many Modes of Building and Presenting Relational Resources

The authors chosen for this first issue interpret the topic of “Relations” at different levels, from the scholarly practice of creating internal and external references across a digital research project (Stonayova, Pizzirusso), to the necessity of connecting digital resources and audiences from different domains (Venuti et al.), to the more theoretical reflection on relations between resources and users in the web (Scanagatta, Charlesworth et al.). What emerges from all papers is the on-going effort to build platforms and resources that aim to become collaborative workspaces for sharing research results and the essential precondition of building digital communities of scholars and practitioners to achieve this goal.

Our first contributor, Silvia Stonayova, presents the digital edition of the *Zibaldone* by Giacomo Leopardi as exemplary case of creating such relational resources; she underlines the necessity to relate all the different fragments of this iconic text and ultimately stresses the need to establish semantic relationships in the text, in order to allow researchers to generate new information and to conduct text analysis with digital methods.

Manfredi Scanagatta in the second contribution, analyses the role of metadata within information digitisation processes with attention to the issue of accessibility and the possible use of historical sources by the general public. He supports his discussion with an in-depth analysis of three case studies, observing how participatory processes in creating digital archives and metadata favour a renewed relationship between historical sources and the public.

The third article by Martina Venuti, Angelo Mario Del Grosso, Federico Boschetti, Luigi Tessarolo, Alessia Prontera, Dylan Bovet, Gianmario Cattaneo and Valeria Melis offers a discussion on the latest developments of the so-called “*Musisque Deoque Galaxy*” (MQDQ) both from the point of view of the tools created to allow the integration of new texts and to enable new analyses and from the point of view of sustainability of the technological framework for a digital ecosystem of knowledge. MQDQ aims at serving as a reference model within the national and European landscape for the creation of relationships and synergies between digital text corpora offering ideal opportunities for philological and literary studies.

The fourth author of this issue, Igor Pizzirusso, presents the database *Ultime lettere di condannati a morte e di deportati della Resistenza italiana* (Last Letters of Death Row Inmates and Deportees of the Italian Resistance) and its constant growth over time. Among its merits he highlights the possibility of faithfully reconstructing the original state of these autographs in their social and cultural context.

The final paper by Ellen Charlesworth, Claire Warwick, Leonardo Impett and Andrew M. Beresford discusses two UK-based online responses to the national lockdowns in 2020, namely by the British Museum and the Linen Hall Library. The authors investigate the discrepancy between the data and the narrative, examining how institutional aims inform the way audiences are incorporated in online metrics such as ‘views’, ‘likes’, and ‘shares’. The two case studies show how cultural heritage organisations can engage with online audiences, and how a proper contextualisation of these metrics is essential to better understand how people relate to digital content from the GLAM sector and how the COVID pandemic affected their practices of public engagement.

4 Of Hard Work and Entering Global Classification Systems

As this issue opens the fourth year in a row of our scientific venture, we are pleased to announce that the international relations, hard work, and novel collaborations inaugurated along this editorial path already bore considerable fruits. Indeed, *magazén* has just been accepted for SCOPUS, since the Scopus Content Selection & Advisory Board (CSAB) approved our application for coverage. The reviewers attested that “this is a well organised journal publishing material that is likely to interest SCOPUS users”, as it “consistently includes articles that are academically sound and relevant to an international academic or professional audience in the field”.

We cannot but share our satisfaction and pride with the many scholars working to make *magazén* possible, but the SCOPUS classification also encourages us to dig deeper in the domain of Digital and Public Humanities, as well as making some changes to our editorial project. At this point we feel that we have reached a novel phase and broader audience, which allows us to completely open the journal to our scholarly community. In the first four years, by publishing thematic issues we were trying to establish our journal as a new platform that fosters an international and open debate about the fundamentals of our scholarly field. From now on, however, we believe we can move past specific subjects and methodological topics, thus truly highlighting recent challenges and cutting-edge experiences in our interdisciplinary domain. In order to do so, together with the Editorial Board we resolved to stay open for single paper proposals from international scholars as well as for covering entire issues with guest editors. Thus, *magazén* will look even more like a place for sharing and arguing such as in the public house at the time of the Venetian Republic, from which we derived our journal title *magazén* (Tassini 1970, 364-5).

Given our recent SCOPUS classification, we particularly wish to express our gratitude to all scholars and experts involved over the last four years in the making of this *magazén*: our Advisory Board members, all published authors, the uncountable peer reviewers, the strenuous members of the editorial board, as well as our publisher's team. We will cherish this splendid relationship.

Bibliography

- Barranha, H.; Henriques, J.S. (eds) (2021). *Art, Museums, Digital Cultures: Rethinking Change*. Lisbona: IHA/NOVA FCSH e maat.
- Dell'Unto, N.; Landeschi, G. (2022). *Archaeological 3D GIS*. London: Routledge.
- Drucker, J. (2003). "The Virtual Codex from Page Space to E-Space". <http://www.philibiblon.com/drucker>.
- Fiormonte, D.; Chaudhuri, S.; Ricaurte, P. (eds) (2022). *Global Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press.
- King, L.; Stark, J.F.; Cooke, P. (2016). "Experiencing the Digital World: The Cultural Value of Digital Engagement with Heritage". *Heritage & Society*, 9(1), 76-101. <https://doi.org/10.1080/2159032X.2016.1246156>.
- Marty, P.F. (2010). "An Introduction to Digital Convergence: Libraries, Archives, and Museums in the Information Age". *The Library Quarterly: Information, Community, Policy*, 80(1), 1-5. <https://doi.org/10.1086/648549>.
- Sternfeld, J. (2011). "Archival Theory and Digital Historiography: Selection, Search, and Metadata as Archival Processes for Assessing Historical Contextualization". *The American Archivist*, 74(2), 544-75. <https://doi.org/10.17723/aarc.74.2.644851p6gmg432h0>.
- Tassini, G. [1863] (1970). *Curiosità Veneziane: Ovvero origini delle denominazioni stradali di Venezia*. Venezia: Filippi Editore.

Articulating Intra- and Intertextual Relationships in the Fragment Collection Working with the Digital Edition of Giacomo Leopardi's *Zibaldone*

Silvia Stoyanova

Arizona State University, USA

Abstract Leopardi's *Zibaldone* presents a compelling study of semantic relations between textual fragments because the author's interpretive annotations form complex intra- and intertextual networks that elude close reading. The *Digital Zibaldone* project has built the foundation of a semantic digital edition to reconstruct the discourse order of Leopardi's intellectual inquiries and his reception of fictional and scholarly works, while offering readers some tools for their analysis. Its larger research framework aims to create synergy between the hermeneutic project of the fragment genre and digital methods of knowledge representation.

Keywords Giacomo Leopardi. *Zibaldone*. Digital editions. Fragment genre. Semantic networks. Intratextuality. Intertextuality. Linked data. Humanities ontologies. Graphesis.

Summary 1 Introduction. – 2 The *Digital Zibaldone* Project: Editorial Models and Research Frameworks. – 2.1 Semantic Web Technologies in the Personal Research Archive. – 2.2 Activating the Relational Power of Ideas: Fragments and Thought-Images. – 3 Reconstructing the *Zibaldone*'s Intra- and Intertextual Networks. – 3.1 Developing the Foundation of a Semantic Digital Edition. – 3.2 Aligning the Semantic Layers of Intratextual References and Thematic Indexes. – 3.3 Disambiguating Bibliographic Sources with Authority Databases. – 4 Synthetic Frameworks for Exploring the Semantic Relationships of the Fragments. – 4.1 Navigation Settings and Paragraph Information Box. – 4.2 Thematic Index Aggregates and Related Headings. – 4.3 Network Graph Generation and Visualisation. – 5 Conclusion.



Peer review

Submitted 2023-03-13
Accepted 2023-05-19
Published 2023-08-02

Open access

© 2023 Stoyanova | 4.0



Citation Stoyanova, S. (2023). "Articulating Intra- and Intertextual Relationships in the Fragment Collection. Working with the Digital Edition of Giacomo Leopardi's *Zibaldone*". *magazén*, 4(1), 13-42.

DOI 10.30687/mag/2724-3923/2023/01/001

La scienza della natura non è che scienza di rapporti. Tutti i progressi del nostro spirito consistono nello scoprire i rapporti. [...] è manifesto che colui che ignora una parte, o piuttosto una qualità una faccia della natura, legata con qualsivoglia cosa che possa formar soggetto di ragionamento, ignora un'infinità di rapporti.
(Giacomo Leopardi, *Zibaldone*, 1836)¹

To great writers, finished works weigh lighter than those fragments on which they work throughout their lives.
(Walter Benjamin, *One-Way Street*, "Standard Clock")

1 Introduction

Giacomo Leopardi's *Zibaldone* offers a compelling case study of semantic relations between textual fragments because its non-sequential discourse, explicitly interlinked textuality, and the number and complexity of its authorial intra- and intertextual references challenge the semantic organisation of the codex and the cognitive process of close reading. The text amounts to c. one million words of notes and reflections written in several languages on an encyclopaedic variety of topics and punctuated with quotes and bibliographic information, while the manuscript abounds in marginal and interlinear additions, underlining, and references connecting related fragments. Leopardi uses date divisions to mark a circumscribed reflection and adopts page and paragraph numbers to locate the targets of his references. He also indexed the fragments thematically at the paragraph level, using different heading categories that denote topic specificity. His indexing method follows a principle of "logica combinatoria" (Cacciapuoti 2001, 252) which ascribes each fragment to multiple headings and makes cross-references between headings. Given the magnitude of this reference system (the text has over 6,000 explicit and implicit internal references, and the two main indexes list over 10,000 paragraphs), the reader's task of following the intersections of these relationships becomes a logistical impasse, as experienced first-hand by the author who left many of his projects at the editorial stage of annotation.²

¹ All quotations from the *Zibaldone* are from its digital edition (Stoyanova, Johnston 2011-23) and its English translation (Leopardi 2013b). In subsequent quotations from the text the title is abbreviated to *Zib.* and the page numbers refer to those marked by Leopardi in the manuscript.

² In letters to his editor Stella from September 13 and November 22, 1826, Leopardi describes the difficulty of harvesting his knowledge from his "immenso volume manoscritto, o scartafaccio" (immense manuscript volume or notebook; Author's transl.)

Leopardi's structured close analysis of his fragment collection can take advantage of the synergistic application of computational methods for knowledge representation in order to articulate the complex aggregate of semantic relations established one at a time from different phenomenological perspectives.³ Engaging with the text in a digital research environment would bring definition to its networked discourses and make Leopardi's thought more accessible to scholars, students, and the general public. This methodological approach to the analysis of the text will allow readers to follow the development of arguments, as a concept accretes signifying value through its network of references in different thematic contexts; to gain insight into Leopardi's scholarly practices of taking notes, citing sources, making and qualifying connections in the process of interpretation; to explore the historical, linguistic, disciplinary, and cultural contexts of his thought through his intertexts. What is more, in addressing these challenges of semantic organisation through the research framework of the fragment genre in its contention with its codex medium, the *Digital Zibaldone* project proposes that the affordances of a digital workspace could significantly advance the critical, aesthetic, and performative engagement with the constellation hermeneutic methodology of this textual phenomenon.⁴

as one of scale, order, and arrangement, which he was hoping to confront with his alphabetical index (Leopardi 2013a, 1334). Arguably (cf. Acanfora 1989, 89-90), in rendering explicit the intertwining of conceptual fields and the vast scope of the collection, the process of indexing discouraged the author from proceeding with his projects.

3 D'Intino describes the *Zibaldone's* genre as a "kind of 'laboratory notebook'", in which "to the despair of readers and interpreters, Leopardi returns again and again to the same questions, looking at them from different, often divergent, angles" (Leopardi 2013b, 79).

4 The constellation metaphor is used by Benjamin and Adorno to describe the fragment's resistance to instrumental reductionism in the hermeneutic process: "by gathering concepts round the central one that is sought, they attempt to express what that concept aims at, not to circumscribe it to operative ends" (Adorno 2003, 165-6). Such constellations of ideas tend to "structure the philosophy of an author from the inside" (Desideri 2020, 12), and struggle to find narrative form.

2 **The *Digital Zibaldone* Project: Editorial Models and Research Frameworks**

2.1 **Semantic Web Technologies in the Personal Research Archive**

The digital remediation of the *Zibaldone* did not begin with the aim to produce a digital edition of the text but with an experiment to model Leopardi's notebook as a digital editing environment in response to its function as a personal knowledge base, and so, to support Leopardi's methods of analysis and interpretation – a model that can be extended to and is informed by other representatives of the literary-philosophical fragment genre. The TEI data model adopted at the project's onset has allowed us to articulate the stratified composition of the text, which is fundamental for representing the conceptual relationships between its semantic units; to link Leopardi's references with their targets within the text, as well as between the *Zibaldone* and its indexes, and to describe their typologies; to define named entities (persons, works, places) with references to authority databases; to describe manuscript layout features, some of which contain key indications for determining the degrees of relatedness between the fragments. These markup decisions have been dictated by the need to define the semantic networks of the collection and parameters for assigning attributes of relatedness to their edges. Our editorial objectives to reconstruct the intratextual network of thematic relationships, to identify Leopardi's sources through his bibliographic references, and to categorise the modalities in which he employs them, naturally align with the graph data model and the affordances of semantic web technologies to support the interpretative core of humanities research by

discovering, articulating, and visualizing relationships that may otherwise be difficult to investigate. (Usiskin et al. 2019, 191)

Semantic ontologies, linked open data, and network graph visualisations can serve as informational, analytical, exploratory, and narrative tools for bringing Leopardi's fragments into epistemic structures of higher semantic orders – those implicit in the author's discrete records of semantic annotation, those that readers could generate in research queries, and those inferred by machines.

The critical and practical efforts to bring about a paradigm shift in the modelling of digital scholarly editions by broadening the philological focus on the “text-as-document” towards modelling the “text-as-work” (Robinson 2013, 110) in order to “express and support the properties [...] that construct its meaning” (van Zundert 2016, 104),

and towards the “text-as-network” (Witt 2018, 221) to reflect the dominant activities of textual editors as “identifying and describing relationships between identifiable text parts” (223), broadly support our editorial model of Leopardi’s notebook as a digital workspace for text analysis and interpretation based on formalising the relationships between the various parts of the text. The digital assertive edition proposed by Vogeler to “use the information conveyed in the text as structured data” (2019, 318) further describes our focus on the thematic contents of the text and on representing its named entities as linked open data. In the objective to reconstruct the intertextuality of the *Zibaldone*, the project takes advantage of the semantic scholarly edition model advanced by Tomasi, and especially as employed in Bufalini’s *Notebook* (Daquino et al. 2020) which maps the semantic relationships between the authors and works discussed by Bufalini onto the notebook’s fragmented textuality and links them to the web of data. Semantic ontologies for intertextual relationships can formalise Leopardi’s critical engagement with other texts,⁵ and the publication as linked open data of the several thousand quotes and bibliographic references to works and their authors, editors, translators, and publishers, allows to build a comprehensive tableau of his sources by enhancing them with information from authority databases.

A congenial editorial model, in particular for representing the conceptual networks of Leopardi’s intratextual references and thematic indexes, is offered by Wittgenstein’s *Nachlass* which similarly records the author’s editorial annotations for structuring his collection of remarks by means of indexing systems indicating his “*action intentions*” (Pichler 2021, 197; italics in the original).⁶ The *Zibaldone* shares with the *Nachlass* its “‘crisscross’ knowledge” (Pichler et al. 2021, 60) characterised by “concept dynamics, vagueness, multiperspectivism, [...] competing and contentious knowledge claims” (59). The Wittgenstein ontology maps the conceptual organisation of the fragments through philosophical subjects with subclasses, modeling different types of arguments and “chaining [them] into greater debates” (66). The “simple index structure, proven in the Gutenberg era” (Pichler, Zöllner-Weber 2013, 703), is employed as a main element of the Wittgenstein ontology to denote philosophical topics or “issues” (704), and in our case would correspond to Leopardi’s in-

⁵ Besides the methodology of the Bufalini edition, the INTRO ontology (<https://github.com/BOberreither/INTRO>) for intertextual relationships in literary studies can be adopted for our purposes.

⁶ Wittgenstein marked his notes with different symbols, numbers, and letters for the reordering of the collection – a task demanding a digital editing environment with tools for filtering the fragments semantically, which the interface of the Wittgenstein Archive at the University of Bergen (WAB) affords through interactive dynamic presentation and semantic faceted search and browsing.

dexes for the *Zibaldone* and their expansion by editors of print editions and by the project *Lessico Leopardiano*.⁷ The scope of an ontology for representing the conceptual structure of the author's own records of text analysis could be extended through collaborative expert participation to include different modalities of argument. Their clustering into convergent and divergent "perspectives" – to borrow another term from the Wittgenstein ontology (Pichler et al. 2021, 69), would allow to represent, for example, the dialectical development of key binary notions in Leopardi's thought. Besides the classification of thematic categories and arguments, the definition of typologies for the references and index headings, which our project infers from the author's formal patterns of composition, would allow us to establish parameters for evaluating the degree of relatedness inherent in the intratextual relationships of the collection. This data of rich authorial analysis of semantic relatedness between text fragments then could be formalised for purposes of natural language processing.⁸

Semantic web technologies thus can assist scholars to analyse the relational structure of complex datasets which lack a formal discourse organisation but – as is the case of the fragment genre – present with copious indications for their conceptual modelling. Conversely, the research queries that necessitate their computational formalisation in the first place could inform the design of ontologies and of digital workspaces to lend definition to the process of interpretation itself. As Pichler et al. (2021) suggest, in order to advance the integration of semantic ontologies with the discourse of the humanities, ontology development needs to account for "non-shared conceptualization and contradiction between viewpoints," to "reason over multiperspectivism, contentiousness, conceptual vagueness," to "computationally map the dynamism of human knowledge" (71). Bradley and Pasin (2017) have proposed that integrating ontologies with a tool for supporting scholarly interpretation, such as *Pliny*,⁹ could help scholars bring emerging ideas from the note-taking and annotation stages towards clarity and formality, while acknowledging the limitations of existing graph-based data models for the "in-

7 The project for a Lexicon of Leopardi's concepts initiated by Bellucci and D'Intino at Sapienza University is a continuing collaboration of the community of Leopardi scholars.

8 Although the application of NLP methods is beyond the editors' scope and expertise, such an approach could complement Leopardi's close analysis by comparing the virtual hyperlinks between fragments and their thematic tagging with the co-occurrences of key terms. Conversely, his analytical annotations could be formalised as an algorithm for computing semantic relatedness between text fragments, as in Yazdani, Popescu-Belis (2013, 178) who evaluate the "real-world knowledge" of hyperlinks in Wikipedia articles in relation to their lexical similarity.

9 The project has been retired but can be consulted at <https://www.kcl.ac.uk/research/pliny-project>.

completeness and ambiguity” (2017, § 88) inherent in the process of interpretation. In its function as a personal knowledge base, the unbound codex of the *Zibaldone* accommodates the capture and storage of observations, while Leopardi’s analysis through referencing and indexing corresponds to the intermediate stage defined in *Pliny* as organising the notes “into a collection of interrelated concepts, categories, or topics” (§ 27) – a process which requires a technology to support the dynamic sorting, retrieving, aggregating, and re-arranging of research material on the scale of thousands interconnected fragments. While *Pliny* is a promising theoretical model of bringing ontologies to bear on the process of constructing interpretation, the design of such a workspace would benefit from case studies of intellectual notebooks which dwell at the project stage and employ organisational expedients to simulate the dimensions of a digital editing environment. The digital remediation of the notebook genre furthermore necessitates “synthetic frameworks” (Drucker 2013, § 38) in order to convey “fragmented and correlated points of view that resist self-evident reification” (§ 39). Drucker’s proposal to develop the design of interpretative humanities interfaces to

create environments that are constellationary, so that diagrammatic relations can be used to re-order familiar conventions through acts of generative, performative engagement (§ 39)

resonates with the exigency of the fragment collection for a workspace to support its hermeneutic practices. The creation of a computational model of the semantic organisation of the *Zibaldone* thus also entails the design of a research platform to configure its multiperspectival discourse, and therefore can take advantage of and offer explorative ground for modelling authoring environments to support modular, fluid, networked, multimodal scholarship. Whereas “literal translations of a book into a non-book-oriented medium” (van Zundert 2016, 104) abound among digital scholarly editions, and “digital simulation of print monographs” (Spence 2018, 471) among emerging digitally-mediated long-form scholarly publications, some fragment collections emphatically fail to become books in an endeavour to convey the constellationary order of their phenomenological discourse.

2.2 Activating the Relational Power of Ideas: Fragments and Thought-Images

Leopardi's *Zibaldone*, Wittgenstein's *Nachlass*, Coleridge's *Notebooks*, Valéry's *Cahiers*, Benjamin's *Passagenwerk*, Joubert's *Carnets*, Novalis' *Brouillon*, among others, are kindred texts that capture the "infinitely resonant interconnectedness or relationality" (Gifford 1998, 303). The hermeneutic methodology of fragment writers magnifies the networked structure of meaning by repeatedly returning to phenomena from a multiplicity of contexts in an extended temporal and textual space:¹⁰

The same problems, regularly resurgent, are forever reformulated and reviewed, giving wider associations, sharper definitions, renewed perspectives and contexts of significance. (Gifford 1998, 46)

Their expansive perspective, which the Jena Romantics describe as "potentiation",¹¹ raises the signifying power of ideas and renders increasingly scrupulous its articulation – the arranging of words in a sentence and the circumscribing of concepts in a discourse framework. The *Zibaldone* fragments contend with the challenge of semantic organisation already at the level of syntax which is laden with relative clauses, lists of synonyms and qualifiers, as reflected in the manuscript's interlinear and marginal additions¹² – some integrated with pointers into a sentence while others positioned in parallel, before Leopardi confronts their relationality at the paragraph level with his combinatory methods of cross-referencing and thematic indexing. As the open space of the notebook expands in volume, contexts, and scale of observation, the correlations between fragments become convoluted, though they can gain form and clarity in a mental image described by Leopardi as

colpo d'occhio che scuopre in un tratto le cose contenute in un vasto campo, e i loro scambievoli rapporti.¹³

10 The *Zibaldone* has 4,526 manuscript pages written over the span of fifteen years, and with intense continuity during five. Valéry, who perhaps exemplifies the most unremitting practice of the fragment, wrote c. 28,000 pages of notes over the course of more than fifty years.

11 In his theory and practices of the fragment, "F. Schlegel frequently uses the verb 'to potentiate' (*potenzieren*) to describe the process of recontextualization that continually raises an artwork's semantic potentiality to a higher power" (Hall 2009, 426). The notion is likewise common in the writings of Novalis.

12 As D'Intino explains, Leopardi's tendency "to stretch his discourse – at times almost to the breaking point" is motivated by an ambition "to seize an idea, a concept, a proposition" (Leopardi 2013b, 102).

13 'The ability to take things in at a glance that discovers the things contained in a vast field, and their reciprocal relationships' (*Zib.* 1854).

The simultaneous comprehension of the interrelations between many concepts which had been observed one at a time, fosters their qualitative re-evaluation and enhances the observer's cognitive capacity:

Quindi è che scoprendo in un sol tratto molte più cose ch'egli non è usato di scorgere a un tempo, e d'un sol colpo d'occhio discernendo e mirando una moltitudine di oggetti, ben da lui veduti più volte ciascuno, ma non mai tutti insieme (se non in altre simili congiunture), egli è in grado di scorgere con essi i loro rapporti scambievoli, e per la novità di quella moltitudine di oggetti tutti insieme rappresentantisegli, egli è attirato e a considerare, benché rapidamente, i detti oggetti meglio che per l'innanzi non avea fatto, e ch'egli non suole; e a voler guardare e notare i detti rapporti. ¹⁴

This superior cognitive experience prompts the thinker to want to look at the new relationships and to take note of them, so that the mental image can translate into interpretative discourse -

purch'ei sia capace di ben esprimere i propri concetti, ed abbia bene e chiaramente e distintamente presenti le cose allora concepite e sentite. ¹⁵

Although it is conditioned by the application of close attention, such augmented perception cannot be purposefully attained; it is fleeting and its recollection is unreliable:

E forse esso stesso dopo quel punto, non crede più alle verità che allora avea concepite e trovate, cioè o non si ricorda, o non vede più con egual chiarezza, i rapporti, le proposizioni, i sillogismi, e le loro concatenazioni che l'avevano portato a quelle conseguenze. ¹⁶

¹⁴ 'Hence, in discovering all at once many more things than he would ordinarily be accustomed to notice at one time, and in discerning and seeing at a single glance a multitude of objects, each of which he has seen individually on many occasions but never all together (apart from in similar circumstances), along with them he is able to see all their reciprocal relations, and as a result of the novelty of this multitude of objects that presents itself to him all together, he is led to consider these objects, albeit fleetingly, better than he has done before this time, and better than he is used to doing, and to want to look at and note these relationships' (*Zib.* 3269-70).

¹⁵ 'Provided he has the things conceived and felt in that moment clearly and distinctly present before him' (*Zib.* 3271).

¹⁶ 'And perhaps he himself after that point no longer believes in the truths that he has then grasped and discovered, that is, either does not remember or he no longer sees with the same clarity the relationships, the propositions, the syllogisms, and their concatenations that had brought him to those conclusions' (*Zib.* 1976).

As the thinker endeavours to translate the conceptual order of relations into narrative, the synoptic image of the mind is simulated by visual metaphors and graphic configurations. Thus, for example, Walter Benjamin's "flashlike image" is a "mode of cognition" (Weigel 2015, 347) very similar to Leopardi's synoptic flash-like glance:

It stands for a way of knowing that in an instant can illuminate an entire situation. [...] a faculty in which an 'enhanced presence of mind' (*gesteigerte Geistesgegenwart*)¹⁷ is coupled with an involuntary mode of seeing. The flashlike cognition is something that befalls the person unintentionally. (Weigel 2015, 348)

Moreover, for Benjamin,

the tension of his writing becomes pictorially evident, almost to the extent of drawing itself, of composing itself in an image: in the figure-of-thought. (Desideri 2020, 14)

Indeed, the discursive methods that fragment collections employ to convey a conceptual image as "interrelatedness that is not stated but rather shown" (Brand 2004, 41) often rely on the spatial and visual semiotics of cross-references, index cards and indexes, folders, classificatory tables, catalogues, outlines, diagrams, page layout, underlining, geometrical figures, colours, connecting lines, shading, cutting and sticking pieces of paper, montage.¹⁸

The research framework of the *Digital Zibaldone* project explores how the phenomenological method of the fragment collection, of relating disparate ideas as they arise in an "intention-less" state of "immersion" (Benjamin 1998, 36) and their gathering into semantic constellations that resist an instrumental order of discourse, could be conveyed through the computational aggregation of the fragments' metadata and their visualisation as semantic networks, thus replicating the inferences drawn by the flash-like image to support the interpretative process it engenders. By rendering the fragments readily accessible for recall and re-arrangement, their configurations could be examined anew with greater clarity and discernment. Beyond analysis, the visual hermeneutics of graphesis has potential to further mediate the interpretative process into discourse and create rhetorical forms that embody, rather than describe, the connecting

¹⁷ Weigel cites from Benjamin's *Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit*.

¹⁸ For a discussion on some of the organisational expedients of different fragment collections see Marx et al. 2015; Gifford 1998; Kinloch 1996.

threads of interpretation.¹⁹ In a basic sense, the research framework for the *Digital Zibaldone* recalls Engelbart's vision for augmenting human intellect with a computer system that would make explicit the "types of network relationships among concepts" (Engelbart, English 1968, 398; quoted in Barnet 2013, 1325), which are embedded in language, and so would facilitate the human ability to

externalize [our] thoughts in the concept structures that are meaningful outside; moving around flexibly, manipulating them and viewing them. (Engelbart 1998, quoted in Barnet 2013, 1461-2)

3 Reconstructing the *Zibaldone's* Intra- and Intertextual Networks

3.1 Developing the Foundation of a Semantic Digital Edition

The *Digital Zibaldone* builds on a rich editorial history of scholarly print editions and some digital remediations of the text.²⁰ While editorial priority has been to reconstruct the author's conceptual organisation of the fragment collection and facilitate its further analysis, rather than to produce a diplomatic edition which had been accomplished by the CD-ROM edition (Ballerini, Ceragioli 2009), at this mature phase of development the divergencies from a fully diplomatic edition are few.²¹ The transcription of the *Zibaldone* with its authorial thematic indexes and the encoding of their structural elements in XML according to the TEI *P5 Guidelines* are based on the CD-ROM edition, the facsimile edition (Peruzzi 1989-94), and on partial examination of the manuscript at the archives of the Biblioteca Nazionale di Napoli. The *Roma* tool of the TEI consortium was used to create a custom schema for the project, and was gradually modified with the addition of new elements. Based originally on the TEI Lite customisation, a custom schema was produced and exported as an ODD customisation and as a RELAX NG schema used to validate

¹⁹ Drucker's scholarship on visual epistemology and hermeneutic visualisation in the humanities offers a critical framework for remediating the visual metaphors of cognition employed by Leopardi and other practitioners of the fragment genre.

²⁰ The editorial methodology, bibliography of sources, a list of errata corrected in referenced editions, a user's guide, and related publications are made available on the project's website. As the interface undergoes continued development, tutorials will be created to illustrate the affordances of the research platform.

²¹ The transcription does not include deleted text, markers for the distribution of the lines on a page and for the breaking of words over two lines, and a few common abbreviations have been normalised silently to facilitate string search and corpus analysis.

the project's TEI files. The markup distinguishes significant layout features of the manuscript and charts the stratified and interlinked composition of the text with date divisions, marginalia, inline and interlinear additions, the targets of intra- and intertextual references, distinctions between heading categories of the indexes, some of the ink variations, several types of underlining. The edition also identifies quotes and their language, titles of works, miscellaneous bibliographic information, persons, and places, and provides editorial notes for corrections, normalisations, and additions, such as target specifications for Leopardi's verbal references, with the number of encoded items being on the scale of 100,000. Besides the four authorial indexes, the dataset includes three editorial ones for persons, works, and places, with corresponding Wikidata and VIAF identifiers, biographical and bibliographic data, links to online encyclopaedia information and to the text of the works (of the edition cited by Leopardi, if available) on Google Books, the Internet Archive, Wiki-source, and other open access digital resources.

For reasons of editorial expediency and limitations of resources for project development, the dataset has been hosted as a MySQL database with the provision to transfer it to a graph database once the TEI markup is completed, which was accomplished in the Spring of 2023. The database is used to store a search index and records of entries contained within the document with associated dates; it also contains tables storing representations of elements and element relationships in the form of nodes and edges, allowing for the generation of graphs and other metadata. The graph database transition will allow users to harvest the data more quickly and explore the networks between encoded entities dynamically in the context of the entire graph. The next phase of development will translate into RDF triples those entities encoded in the TEI that are relevant for defining the relationships between the textual segments, namely all of Leopardi's targets of references (notes written in the margin and inline, paragraphs, paragraph sequences, and index headings), his date divisions, his mentions of works, persons, and places. Existing ontologies would allow to adequately describe the intertextual relationships based on the bibliographic references,²² but the main challenge in defining further the intratextual relationships is the need to lend semantic value to the edges in order to reconstruct the thematic networks of the fragments according to a hierarchy of relatedness. The properties of these relationships can be inferred from the layout of the text and Leopardi's methods of composition, and have been described in the TEI.

²² Besides the *Wittgenstein Ontology*, *Paolo Bufalini's Notebook* and the *Dante Hypermedia Network* are excellent models in Italian Studies; the *Intertextual Relationship Ontology* (INTRO) for literary studies will also be explored.

The excerpt below [fig. 1] of encoding a *Zibaldone* date division consisting of a single paragraph (p. 1155,3) gives an example of some of the distinctive features of the text which constitute semantic entities. As can be seen in the facsimile, Leopardi uses date markers in parentheses at the end of a paragraph to mark his occasions of composition and to denote a circumscribed reflection as the main building block of the text. In this example, the date marker of the previous paragraph (visible on the first line of the facsimile excerpt) has the same date of June 11, 1821 but each of these is a separate date division. This function of the date division to mark a semantic unit is reinforced by the cases of thematic discontinuity between passages separated by identical date markers.²³ In the TEI, the date label replicates Leopardi's date marker and has a range to reflect Leopardi's use of compound dates (for example, "14.-15. Luglio. 1823") or of multiple date markers written in the body of the same paragraph adding layers to the same reflection,²⁴ as well as of the time period of composition of the undated fragments (mostly in the first 100 pp.). These correlations between theme, paragraph, and chronology allow to compute thematic continuities and discontinuities across the collection.

Although Leopardi does not mark the paragraphs, they are clearly indented and he counts them in directing his intratextual references, as with one of the references here to the first paragraph on p. 1212 ("V. p. 1212. capoverso 1"). Each paragraph has been assigned an `xml:id` which serves as the target of most intratextual and index references. In identifying the targets, close reading was employed to ensure coherence with Leopardi's count or when he does not specify a paragraph, as in the first reference on p. 1155,3 to p. 1117 ("Alla p. 1117") which is part of the first paragraph beginning on p. 1116. This process is sometimes facilitated by reciprocal references, especially when there are multiple references in the same paragraph added in the margin or inline.

23 One example are the four different date divisions on July 4, 1822 which altogether consist of five paragraphs indexed under twelve unique themes from the 1827 index, giving a glimpse of the diverse subjects occupying the thinker that day.

24 Among the most anomalous in this respect is the short paragraph 2664,3 which has seven date markers written over three days.

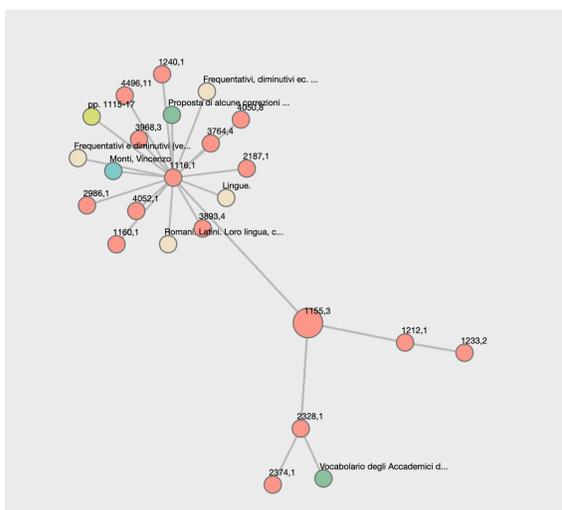


Figure 2
Digital Zibaldone,
two-generations
network of p. 1155,3

Since the various kinds of additions often contain cross-references, as in this example, and those in the margin are sometimes reference targets (with an xml:id), they are encoded as notes with a location (interlinear, inline, margin, foot), which allows to represent statistically and explore Leopardi's method of text analysis and semantic annotation in terms of chronology and perspective – whether there are correlations between establishing connections during the composition of the text (for example, when references are written in the middle of a sentence), as continuation of previous fragments, during re-reading, during the indexing stage, between fragments with reciprocal references and those that are only outgoing or incoming. For the purpose of generating such statistical data, the manuscript was consulted to record ink variations for notes containing paragraph references which are not always evident in the black and white images of the facsimile editions. The inline notes are typically written after or near the date marker, but sometimes there are multiple consecutive inline and marginal notes and their chronological layering can only be determined by the differences in pen and ink. Thus, the last reference in this example (to p. 2328) is in a separate inline note and marked with a different handwriting style from the rest of the text, but previous editions either do not make a distinction or represent both consecutive references (p. 1212 and p. 2328) as written in the same inline addition. This information can also be inferred from the fact that both fragments are “subordinate” to p. 1155,3 and the outgoing references on p. 1155,3 were likely written at the time of their composition.

3.2 Aligning the Semantic Layers of Intratextual References and Thematic Indexes

The document analysis and TEI markup have benefited from Leopardi's rigorous attention to form in composing and editing the text. He adopts two main layers of annotation for navigating the fragments according to their semantic order: the map of intratextual references which are employed to indicate their generative relationship, to record associations during composition as ideas are elaborated further, and to establish them in retrospect; and the indexes which add thematic metadata to the associations and thereby exponentially increase the density of the network. For instance, p. 1329,3 has no direct intratextual references, but it is listed under several headings from the 1827 Index which thus relate it to 69 paragraphs in an extended thematic network on the subjects of grace, naturalness, affectation, etc. [fig. 3]. In the network visualisation clearly emerge those paragraphs between the several clusters which have most themes in common with the central node. The significance of the number of paragraphs listed under each index heading for determining the hierarchy of thematic foci of a fragment is evident in this paragraph which has no less than six different headings²⁵ but "Sgraziataggine" (Gracelessness) would be most applicable because it has no other references. By overlaying the reference network within the text with that of the index headings, we can compute the extent to which direct references overlap with those in an index theme, so as to better define their weight in the network. Within these two authorial methods of text analysis, the structural organisation of the collection suggests additional parameters for determining a hierarchy of semantic relatedness which will serve the modelling of the *Zibaldone* ontology.²⁶

²⁵ The index headings "Pudore. Verecondia" and "Verecondia. Pudore" are redundant, as they list the same two paragraphs.

²⁶ The first categorisation of these relationships was presented at the *Open Annotation Collaboration Workshop Phase II* (Chicago, March 2011) and has been developed in greater detail than can be discussed here.

15. Luglio 1821.

①

←Affettazione. ←Grazia. ←Natura e Fortuna; provvidenza ed arte.

←Naturalezza. ←Pudore. Verecondia. ←Sgraziataggine.

←Verecondia. Pudore.

[1329,3] Non è mai sgraziato un fanciullino che si vergogna, e parlando arrossisce, e non sa stare nè operare nè discorrere in presenza altrui. Bensì un giovane poco pratico del buon tratto, e desideroso di esserlo, o di comparirlo. Non è mai sgraziata una pastorella che non sa levar gli occhi, trovandosi fra persone nuove, nè ha la maniera di contenersi, 1330 di portarsi ec. Bensì una donna, egualmente o anche meno timida, e più istruita, ma che volendo figurare, o essere come le altre in una conversazione, non sappia esserlo o non abbia ancora imparato. Così lo sgraziato non deriva mai dalla natura (anzi le dette qualità naturali, sono graziose sempre ec. ec.), ma bensì frequentemente dall'arte, e questa non è mai fonte di grazia nè di convenienza, se non quando ha ricondotto l'uomo alla natura, o all'imitazione di essa, cioè è alla disinvoltura, all'inaffettato, alla naturalezza ec. E fondamento necessario dell'arte, è quasi sempre questo. Farci disimparare quello che già sapevamo senza fatica, e toglierci quelle qualità che possedevamo naturalmente. Poi con grande stento, esercizio, tempo, tornarci a insegnare le stesse cose, e restituirci le stesse qualità, o poco differenti. Giacchè quella modestia, quella timidezza, quella vergogna naturale ec. si trova bene spesso in molti, non più naturale, chè l'hanno perduta, ma artificiale, chè mediante l'arte appoco appoco e stentatamente l'hanno ricuperata. (15. Luglio 1821.).

«15 Luglio 1821.

16. Luglio 1821.»

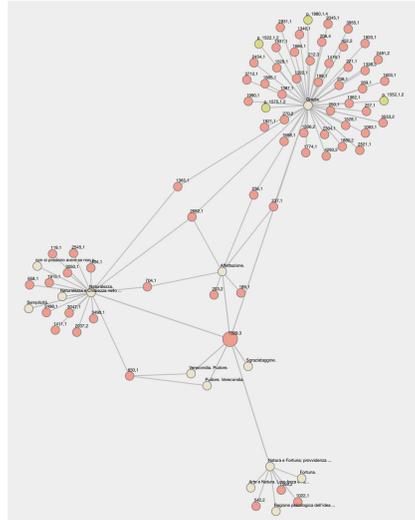


Figure 3 Digital Zibaldone, p. 1329,3 with information box showing the index themes under which it is listed and a network graph of its index headings along with their paragraph references. <https://digitalzibaldone.net/?go=1329%2C3>

3.2.1 Target Directedness and Chronology

There is a significant number of paragraphs that could be described as strictly (no other references) or partially subordinate to their targeted fragment. This typology of explicit directedness would have the highest weight in a network, and, as discussed in the previous section, can be inferred from the location of the references. When the data of the intratextual network is compared with that of the thematic indexes, these relationships can be evaluated further, such as whether the subordinate paragraph shares index themes with the parent, in order to examine how Leopardi constructs his arguments and potentially identify unique combinations of thematic perspectives. The greater part of numeric references establishes non-hierarchical relationships of two typologies: those written in the middle of a sentence or paragraph and those written at the end or as inline or marginal notes. These could be evaluated statistically and through close reading to explore whether references written in the middle of a paragraph refer to fragments of relative spatial and chronological proximity and whether their mode of analogy differs in its logic from that of references written in the margins. While some subordinate references are reciprocal, many fragments have only outgoing or incoming references. Although authorial intentionality may not play a role in lending value to reciprocity, the directedness of the edges is

a main category that could be evaluated in terms of semantic relatedness. Incoming references are surely easy to miss in reading the text sequentially since they may be written much later, therefore on the website they appear in the paragraph information window which aggregates first-generation references.

3.2.2 Target Specificity

Besides referencing numbered locations and on occasion (102 refs.) the titles of index headings (i.e., “see my theory on pleasure”), Leopardi directs his references with verbal indications to the following or preceding paragraphs or with variations of the phrase “as I have said elsewhere”. These indefinite targets are assigned numbered paragraphs with editorial responsibility,²⁷ but while those in the immediate vicinity can be treated at the same semantic level as defined targets, the editorial additions identifying possible targets would have weaker value. The impact of distinguishing between explicit and implicit authorial references is evident in the large number of editorial targets which more than doubles the specified targets in the *Zibaldone*. Leopardi’s vagueness and the fact that these references often repeat or come as a sequence of consecutive and non-consecutive targets, also suggest that they are directed at a more general level of discourse and address a topic more comprehensively, comparable to the lists of paragraphs under the index headings. The alignment of intratextual references and indexes allows to generate statistics on the frequencies of thematic subjects pertaining to the editorial references and to the paragraph sequences respectively, and so to further define their relevance in the network.

3.2.3 Target Granularity

While the dominant level of target segmentation both in the main text and in the indexes is the paragraph, some references point to smaller units, namely additions in the margins which contain either a reciprocal reference or the relevant text, and in a few cases the lines on a page, suggesting strong relevance. Conversely, some references address paragraph or page sequences (i.e. “p. 2171,1-p. 2177,1”), and the indexes also reference clusters of non-consecutive paragraphs on a page (i.e. “p. 4101,1.2.5.7.8”). A more general relationship combin-

²⁷ The Pacella edition (Leopardi 1991) and the English translation (Leopardi 2013b) have been consulted for suggestions on many of the verbal references and included with a responsibility tag referring to the respective edition.

ing granularity and specificity is found in the authorial and editorial sequences with open-ended targets (a page number and the following). These aggregates are encoded as sequences and listed as individual paragraphs.

3.2.4 Parallel Targets

The relatedness among successively written references to non-consecutive fragments (“See p. 2826 and the following thought, and p. 3037”) is evident in a reading context,²⁸ however it could be lost in the computational processing if they are simply three edges among many of the node that lists them, and specifying their degree of semantic relatedness is especially useful for long paragraphs with numerous targets, as in this example of p. 2928,2 which has 19 authorial outgoing references.

A similar principle of relatedness is employed by the lists of references in the index headings which are thus related by virtue of their thematic metadata. As discussed above, these relationships are generally weaker than those established by the direct intratextual references, and to some extent they would overlap. We can qualify them further based on the layout of the headings. The main alphabetical index of 1827 demarcates three types of semantic relationships which tend to correspond to their number of paragraphs: headings, sub-headings, and references to headings written on index cards (abbreviated in the scholarly literature as PR) which Leopardi did not copy because of their excessive length. Another level of semantic relatedness is established by Leopardi’s references between index headings in 20% of those in the 1827 index. These can be categorised further according to the typology of relation they indicate, among which that of synonymy, antonymy, hypernymy, or between persons. The most general categories are offered by the headings of the so-called PNR index which distribute 2,533 targets over eight topics suggesting books titles, such as a treatise on human passions and qualities, a theory of the arts and letters, a handbook of practical philosophy, linguistic scholarship. The two short partial indexes, in which Leopardi experiments with indexing methods, can sometimes be redundant but also add extensive thematic descriptions of the fragments.

28 In this example the consecutive references are also explicitly cross-referenced with each other, but this is not always the case.

3.3 Disambiguating Bibliographic Sources with Authority Databases

The complexity of thematic relations in the *Zibaldone* acquires another dimension through the several thousand references to texts and authors with which Leopardi engages and which are often at the centre of his inquiry. The TEI encoding identifies bibliographic references, titles, persons, quotes, and places as elements relevant to the bibliographic networks, and integrates Wikidata and/or VIAF identifiers for persons, works, and places [fig. 4].²⁹ This information is aggregated in editorial indexes which include links to the record in the authority files; an online encyclopaedia record; references to their paragraph and index locations; dates of birth and death for the persons; authors and editors for the works. Each title record also links to open access resources on the web that have the full text of the work, in order to create a virtual library for consulting these and, when possible, Leopardi's own editions. Additional bibliographic information included in Leopardi's reference, such as volumes, subtitles, page numbers, date and place of publication, and publisher, will eventually be integrated from the spreadsheet used for preparing the data.

```
<person xml:id="Q3541151" role="author" sex="0" sortKey="Manfredi">
  <persName>Manfredi, Eustachio</persName>
  <persName xml:lang="it">Eustachio Manfredi</persName>
  <persName xml:lang="en">Eustachio Manfredi</persName>
  <link type="wikidata" target="https://www.wikidata.org/wiki/Q3541151"/>
  <link type="biography" target="https://it.wikipedia.org/wiki/Eustachio_Manfredi"/>
  <link type="1827index" target="ind436" rend="Manfredi. Sua Lirica."/>
  <link type="VariaIndex" target="var45" rend="Giudizio sopra la lirica del Manfredi."/>
  <link type="ref" target="p28_2" rend="Manfredi"/>
  <link type="ref" target="p28_2" rend="Manfredi"/>
  <link type="ref" target="p28_3" rend="Manfredi"/>
  <link type="ref" target="p4264_1" rend="Manfredi"/>
  <birth when="1674" when-custom="+1674-09-20T00:00:00Z"/>
  <death when="1739" when-custom="+1739-02-15T00:00:00Z"/>
</person>
```

Figure 4 Digital *Zibaldone* dataset. TEI encoding of a person record

Whereas Leopardi's bibliographic references often contain very specific information about the editions he is using, many quotes are not explicitly marked and there are numerous cases of intertextual derivation (quotes within quotes, references within references) which can be difficult to define because he frequently employs secondary literature and reference materials: anthologies, lexicons, articles in

²⁹ More than a hundred persons and works are not present in either database, and Wikidata ids will be created for them. The editorial process was based on reconciliation of the data in OpenRefine and entailed manual verification of all records and identification for many that were not suggested by the software.

journals translated from other journals, critical editions of classical authors by multiple editors, book reviews. Disambiguating the relationships of citation through an ontology in order to articulate and visually represent their extended chains of derivation becomes indispensable, especially for references to the works of ancient authors which have been transmitted through encyclopaedias, biographers, and editors.

The difference in the presentation of intratextual relationships between the narrative style of a print edition's apparatus embedded in the sequential reading of the text and the modular presentation of the digital environment which disambiguates and abstracts each record from its narrative stream in order to chart its relationships across the collection, can be of significant impact to the reader of a fragmented text which lends itself to numerous entry points and selective reading. Even at the basic hypertext interface, the identification of a work from Leopardi's abbreviations and incomplete bibliographic references and the immediate access to its text already present in a synthetic form a lot of implicit and distributed information. The complete reconstruction of Leopardi's network of intertextual references through linked data and semantic ontologies would create an augmented reading experience, in which their relations can be queried at various levels of granularity according to multiple parameters across and beyond the dataset. Wikidata, for example, allows to extract properties for persons and works, which can be harvested to create comprehensive charts with the genre of the work, its language and date of inception, and for persons their occupation, dates of birth and death, language, gender, country of origin, field of work, etc., and thereby guide various research questions.

4 Synthetic Frameworks for Exploring the Semantic Relationships of the Fragments

The *Digital Zibaldone* website interface offers navigation and re-search menus which allow access to the relational organisation of the fragments through several synthetic frameworks, without requiring data processing competencies.

4.1 Navigation Settings and Paragraph Information Box

The research platform presents the text by date divisions with a calendar menu, and employs Navigation Settings allowing a modular selection of manuscript markers and editorial additions to suit different reader purposes – quotes, for example, are italicised and can be easily distinguished in reading the text; the additions are integrated according to their indicated position in line of text but their markers can be shown or hidden; rhetorical underlining can be shown only, etc. One of these options is an information box linked to each paragraph displaying its first-generation references with incoming and outgoing paragraphs, index themes, persons and works mentioned, to show at a glance the topics of that fragment, and to direct the reader to its most relevant paragraphs, to Leopardi's thematic indexes, and to the persons and works indexes [fig. 5].

The screenshot displays the 'Giacomo Leopardi's Zibaldone di pensieri' digital research platform. The interface includes a search bar at the top right and a navigation menu with options for HOME, ZIBALDONE, INDEXES, and PEOPLE. On the left, the 'Navigation Settings' panel is expanded, showing options for 'Show All / Hide All', 'Manuscript Annotations' (with checkboxes for 'Additions interlinear', 'Marginalia attached', 'Footnote', and 'Unattached'), 'Editorial Annotations' (with checkboxes for 'Emendations', 'Correction', 'Implicit reference', 'Paragraph number', 'Quote', 'Rhetorical underlining', and 'Direct references'), and 'People'. The main content area shows a date range from 28 Nov. 1820 to 1. Dicembre 1820, with a selected date of 29. Nov. 1820. Below this, there are navigation links for '←363,1', '←411,1', and '→Illusioni'. A paragraph of text is displayed, starting with '[362.1] L'Essai sur l'indifférence en matière de Religion, Capo 7, verso la fine, dice. Da una dozzina di pagine nasce un culto indigente ai cui di essi. Quindi, quanto maggiore il numero dei doppi che una setta ha conservato, tanto maggior vita e romore e grandezza ha il suo culto. E vedo in quello che segue perché fa al mio proposito. Questa osservazione di fatto si può addurre fra le tante altre (in conferma di quello ch'io dico) che senza illusioni di cui l'uomo sia persuaso, non c'è vita né azione, giacché l'uomo 363 non opera senza persuasione, e se la persuasione non è illusoria, ma viene dalla ragione, l'uomo non opera, perché la ragione non lo persuade ad operare, anzi ne lo distoglie, e lo getta nell'indifferenza. La pompa e la vita del culto senza una persuasione della sua necessità, doverosità, importanza, non ha potuto durare. Limitate le credenze, allargato il dubbio, allargata la ragione e l'indifferenza, e la (peccata) speculazione delle cose, il culto è svanito, laddove si mantiene presso i cattolici, i quali ne conservano tutte le basi, cioè tutti i dogmi, le credenze ec. tanto relative ad esso culto, quanto generalmente alla Religione. Se non ch'egli va languendo anche tra noi, sia nel fatto, sia nell'impressione e l'effetto che produce, e il modo e l'animo con cui è considerato, e veduto o esaguito: e ciò in proporzione dei progressi dell'incertezza, o diminuzione della fede, perché non si può dar gran cura, né coltivare, né promuovere, né esser molto affetti e toccati, da quello che si considera come poco importante, e che non è in relazione colla nostra opinione. (29. Nov. 1820.)'. To the right of the text is a calendar grid for November 1820, with the 29th highlighted. Below the text, there are more navigation links for '←28. Nov. 1820.' and '1. Dicembre 1820.'.

Figure 5 Digital Zibaldone, p. 362,1. Navigation Settings and Information Box. <https://digitalzibaldone.net/?go=362%2C1>

4.2 Thematic Index Aggregates and Related Headings

Each index heading page [fig. 6] aggregates a list with hyperlinks to its paragraph references and to any sub-themes and referenced themes; the text of all referenced paragraphs (not displayed entirely); a statistical chart of the frequencies of the other index headings that also list these paragraph nodes; and a first-generation network graph (not displayed). The limitations of the MySQL database are most evident in this interface which slows down the dynamic harvesting of the data; alternatively, for the longer lists, such as PR and PNR, the statistical data is displayed as pre-generated images. A future addition to the related themes frequency statistics will be a list with links to the paragraphs that each of these headings have in common.

The screenshot displays the 'Giacomo Leopardi's Zibaldone di pensieri' digital research platform. The main heading is 'Affettazione. Affectation.' with a list of manuscript annotations (189, 1 203, 2 236, 1 237, 1 705 1329, 11329, 3 2682, 1). The text discusses the concept of 'affettazione' as a form of artificiality or affectation, contrasting it with naturalness. It references various literary and philosophical sources, including Montesquieu and Leopardi's own work. The interface includes a navigation sidebar with settings for manuscript annotations, editorial annotations, and direct references. A 'Related Themes' section on the right lists other thematic areas like 'Teorica delle arti, lettere ec. Parte pratica, storica ec. (pnr) (6)', 'Grazia. (1827) (4)', and 'Natura e Fortuna; provvidenza ed arte. (1827) (3)'. The page also features a search bar and navigation links for Home, Zibaldone, Indexes, and People.

Figure 6 Digital Zibaldone, Index of 1827, "Affettazione. Affectation".
<https://digitalzibaldone.net/theme/ind3>

4.3 Network Graph Generation and Visualisation

At a more comprehensive level of aggregation the web platform allows to generate the network graph of first and second-generation relations of any intratextual references (paragraph, note, sequence), index themes, persons, and works, or a combination thereof, through a Node ID query by using their xml:id. While the interface for this query is under development at the time of writing, it is accessible at <https://digitalzibaldone.net/?a=graph> and was used to automatically generate the graphs in figures 2, 3, 7, and 8. The data can also be exported for custom manipulation in GEXF, GraphML, CSV, and JSON formats. A future functionality of this interface is a custom selection of types of nodes, which will allow users to pre-define the extent of the network according to their research criteria. The computational and diagrammatic challenge is to filter and arrange the interrelations according to a hierarchy of relevance in order to be able to follow intersecting paths through the extent of the network that is meaningful. For instance, including all the references from the more general PNR and PR index headings or some reference works, such as Forcellini's *Lexicon* (the title with highest frequency), may be unfeasible to represent coherently and irrelevant at the level of paragraph nodes.³⁰

Figure 7 shows a second-generation network of a single paragraph, 2619,2 (large red node in the centre), which includes paragraphs, paragraph ranges, index themes, persons and works mentioned [fig. 7]. The graph allows to quickly identify points of connection among the thematic clusters for further exploration: 1570,1 (to the left of the main node) is not directly related to the main node (in fact, it lacks direct paragraph references altogether, as can be seen from its info box), but it is the connecting point between two of the main node's index themes. Similar is the case of 3661,1 (to the right of the main node) which shares a direct incoming reference and an index theme with the main node, while checking its info box or extending the network to another generation shows that its immediate links are limited to these two, thus reinforcing its potential relevance to the main node.

30 This is evident in many examples on linguistics: 2688,1 has a single outgoing reference, two authorial and four editorial incoming ones, but is inscribed in one very general PNR theme, one large PR and one 1827 index sub-theme, which altogether produce 1375 nodes and 1730 edges. In this case focusing on the sub-heading and the references from the text would produce more meaningful results.

5 Conclusion

The design of the interface and the refinement of the dataset cannot be completely developed without feedback from user experience with extensive scholarly research on the text. Conversely, in reconstructing the semantic relationships in the *Zibaldone* and determining the parameters of a research query, a researcher should be able to generate exploratory statistical information about the features of the dataset, or conduct text analysis with digital methods independent of the authorial interpretation, such as topic modelling or sentiment analysis. The next phase of the project will develop towards building the semantic edition and facilitating the mediation between development and user community. The completed dataset will be made available in a public repository; the editor will publish onsite tutorials and research studies illustrating the usability of the research platform; the website will integrate a user annotation environment to allow for dynamic engagement with the text which extends the author's semantic annotations. The platform thus aims to become a collaborative workspace for sharing the results of research queries and expanding the critical apparatus of the edition, for example with secondary bibliography on the text linked to the referenced fragments. The objective to afford collective annotation that can target the collection at various levels of semantic granularity and to create a platform for scholarly communication is motivated by the processual and open-ended nature of Leopardi's discourse, and has been informed in particular by the notions of "knowledge site" (Shillingsburg 2006), "fluid, co-operative and distributed editions" (Robinson 2004), "dynamic contextualization" (D'Iorio 2016), "reusable third-party annotations" (Koolen, Boot 2020), among other scholarly contributions to the design of a "social edition" (Siemens et al. 2012; Price 2016).

Many levels of discursive mediation are needed for the methods of close and distant reading to productively inform one another, but the fragmented discourse of the *Zibaldone* lends itself precisely to a performative reading that requires both: in the process of configuring its constellations of insights – perhaps arrested by a *colpo d'occhio* – we also cultivate the phenomenological attention that drives its hermeneutic project. And as we endeavour to render our datasets machine-readable and interoperable, the design of inclusive interfaces to augment our scholarly and readerly primitives demands equal attention. This vision for the *Digital Zibaldone* can only heed Leopardi's own aspiration for mediating his encyclopaedic personal knowledge base to the general public:

Ora son dietro ad ordinare i materiali della Enciclopedia. Spero che sarà un'Opera che si farà leggere per forza ad ogni sorta di persone. (Leopardi 2013a, 1345)³¹

Bibliography

- Acanfora, S. (1989). "Indice e indicizzazione dello *Zibaldone*". Peruzzi, E. (a cura di), *"Zibaldone di Pensieri"*, edizione fotografica dell'autografo con gli indici e lo schedario, vol. 1. Pisa: Scuola Normale Superiore, 69-95.
- Adorno, T. (2003). *Negative Dialectics*. Transl. by E.B. Ashton. New York: Continuum.
- Ballerini, M.; Ceragioli, F. (a cura di) (2009). *"Zibaldone di Pensieri" di Giacomo Leopardi*. Bologna: Zanichelli. CD-ROM.
- Barnet, B. (2013). *Memory Machines: The Evolution of Hypertext*. London; New York: Anthem. Kindle Edition.
- Benjamin, W. (1989). *Reflections*. Ed. by P. Demetz. Transl. by E. Jephcott. New York: Schocken Books.
- Benjamin, W. (1998). *The Origin of German Tragic Drama*. Transl. by J. Osborne. London; New York: Verso.
- Bradley, J.; Pasin, M. (2017). "Fitting Personal Interpretation with the Semantic Web: Lessons Learned from Pliny". *DHQ*, 11(1). <http://www.digital-humanities.org/dhq/vol/11/1/000279/000279.html>.
- Brand, R. (2004). "Schlegel's Fragmentary Project", *Epoché*, 9(1), 37-52. <https://doi.org/10.5840/epoche20049113>.
- Cacciapuoti, F. (2001). "La scrittura dello *Zibaldone* tra sistema filosofico ed opera aperta". F. Foschi et al. (a cura di), *Lo "Zibaldone" cento anni dopo: composizione, edizioni, temi = Atti del X Convegno internazionale di studi leopardiani* (Recanati-Portorecanati, 14-19 settembre 1998). Firenze: Olschki, 249-56.
- Daquino, M. et al. (2020). *Paolo Bufalini, "Appunti" (1981-1991)* [Semantic Scholarly Digital Edition]. Digital Humanities Advanced Research Centre (/DH.arc). <https://doi.org/10.6092/unibo/amsacta/6415>.
- Desideri, F. (2020). "The Autographic Stance. Benjamin, Wittgenstein and the Re-Shaping of the Philosophical Opus. About Manuscripts, Fragments, Schemes, Sketches and Annotations". *Aisthesis* 13(2), 9-15. <https://doi.org/10.13128/Aisthesis-12086>.
- D'Iorio, P. (2016). "Digital Scholarly Edition. Interface Issues". Italia, P.; Bonsi, C. (a cura di), *Edizioni Critiche Digitali. Digital Critical Editions*. Roma: Sapienza Università Editrice, 37-52. <https://doi.org/10.13133/9788893770033>.
- Drucker, J. (2013). "Performative Materiality and Theoretical Approaches to Interface". *DHQ*, 7(1). <http://www.digitalhumanities.org/dhq/vol/7/1/000143/000143.html>.
- Engelbart, D. (1995). "The Strategic Pursuit of Collective IQ". Paper presented at *The Brown/MIT Vannevar Bush Symposium* (12 October 1995). <https://dougengelbart.org/content/view/258/>.

31 "Now I am after putting in order the materials for the Encyclopaedia. I hope that it will become a Work that naturally will be read by all kinds of people" (Author's transl.). Leopardi shared this objective in a letter to his editor Stella on July 13, 1827, two days after starting the indexing of the *Zibaldone*.

- Engelbart, D.; English, W. (1968). "A Research Center for Augmenting Human Intelligence". *Proceedings of the AFIPS 1968 Fall Joint Computer Conference* (CA, 9 December 1968). Washington: Spartan Books, 33, 395-410.
- Gifford, P. (1998). "Thinking-Writing Games of the 'Cahiers'". Gifford, P.; Stimpson, B., *Reading Paul Valéry: Universe in Mind*. Cambridge: Cambridge University Press, 35-52. <https://doi.org/10.1017/cbo9780511553929.003>.
- Hall, M. (2009). "Friedrich Schlegel's Romanticization of Music". *Eighteenth-Century Studies*, 42(3), 413-29.
- Kinloch, D. (1996). "Reading and Writing in Joubert's *Carnets*". *The Modern Language Review*, 91(2), 342-54. <https://doi.org/10.2307/3735016>.
- Koolen, M.; Boot, P. (2020). "Facilitating Reusable Third-Party Annotations in the Digital Edition". Nantke J.; Schlupkothen, F. (eds), *Annotation in Scholarly Editions and Research*. Berlin; Boston: De Gruyter, 177-99. <https://doi.org/10.1515/9783110689112-009>.
- Leopardi, G. (1989-94). *Zibaldone di pensieri: edizione fotografica dell'autografo con gli indici e lo schedario*. A cura di E. Peruzzi. Pisa: Scuola Normale Superiore.
- Leopardi, G. (1991). *Zibaldone di pensieri: edizione critica e annotata*. A cura di G. Pacella. Milano: Garzanti.
- Leopardi, G. (2013a). *Tutte le poesie, tutte le prose e lo "Zibaldone"*. A cura di L. Felici; E. Trevi. Roma: Newton Compton.
- Leopardi, G. (2013b). *Zibaldone*. Edited by M. Caesar; F. D'Intino. Transl. by K. Baldwin et al. New York: Farrar, Straus and Giroux. Kindle Edition.
- Marx, U. et al. (eds) (2015). *Walter Benjamin's Archive. Images, Texts, Signs*. Transl. by E. Leslie. London; New York: Verso. Kindle Edition.
- Oberreither, B. (2023). *INTRO, an Intertextual Relationships Ontology for literary studies*. <https://github.com/BOberreither/INTRO>.
- Pichler, A. (2021). "Complementing Static Scholarly Editions with Dynamic Research Platforms: Interactive Dynamic Presentation (IDP) and Semantic Faceted Search and Browsing (SFB) for the Wittgenstein Nachlass". *Selected papers from the 9th CLARIN Annual Conference 2020* (5-7 October 2020), 180, 194-207. <https://doi.org/10.3384/ecp18021>.
- Pichler, A.; Zöllner-Weber, A. (2013). "Sharing and Debating Wittgenstein by Using an Ontology". *Literary and Linguistic Computing*, 28(4), 700-7. <https://doi.org/10.1093/lc/fqt049>.
- Pichler, A. et al. (2021). "Crisscross ontology: Mapping Concept Dynamics, Competing Argument and Multiperspectival Knowledge in Philosophy". Ciraci, F.; Fedriga, R.; Marras, C. (a cura di), *Quaderni di 'Filosofia'. Filosofia digitale*. Milano; Udine: Mimesis, 59-73.
- Price, K. (2016). "Social Scholarly Editing". Schreibman, S.; Siemens, R.; Unsworth, J. (eds), *A New Companion to Digital Humanities*. Chichester: Wiley, Blackwell, 137-49. <https://doi.org/10.1002/9781118680605.ch10>.
- Robinson, P. (2004). "Where We Are with Electronic Scholarly Editions, and Where We Want to Be". *Jahrbuch für Computerphilologie*, 5, 125-46.
- Robinson, P. (2013). "Towards a Theory of Digital Editions". *Variants*, 10, 105-31.
- Shillingsburg, P. (2006). *From Gutenberg to Google: Electronic Representations of Literary Texts*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CB09780511617942>.
- Siemens, R. et al. (2012). "Toward Modelling the Social Edition: An Approach to Understanding the Electronic Scholarly Edition in the Context of New

- and Emerging Social Media”. *Literary and Linguistic Computing*, 27(4), 445-61. <https://doi:10.1093/lc/fqs013>.
- Spence, P. (2018). “The Academic Book and Its Digital Dilemmas”. *Convergence*, 24(5), 458-76. <https://doi.org/10.1177/1354856518772029>.
- Stoyanova, S.; Johnston, B. (2011-23). *Digital Zibaldone*. <http://digital-zibaldone.net>.
- Usiskin, J.; Walde, C.; Winter, C. (2019). “Ontologies for Digital Humanists”. Crompton, C.; Lane, R.J.; Siemens, R. (eds), *Doing More Digital Humanities. Open Approaches to Creation, Growth, and Development*. London: Routledge, 184-98. <https://doi.org/10.4324/9780429353048-14>.
- van Zundert, J. (2016). “Barely Beyond the Book?”. Driscoll, M.J.; Pierazzo, E. (eds), *Digital Scholarly Editing: Theories and Practices*. Cambridge: OpenBook, 83-106. <https://doi.org/10.11647/OBP.0095.05>.
- Vogeler, G. (2019). “The ‘Assertive Edition’”. *International Journal of Digital Humanities*, 1, 309-22. <https://doi.org/10.1007/s42803-019-00025-5>.
- Weigel, S. (2015). “The Flash of Knowledge and the Temporality of Images: Walter Benjamin’s Image-Based Epistemology and Its Preconditions in Visual Arts and Media History”. Transl. by C.T. Smith, C. Kutschbach. *Critical Inquiry*, 41, 344-66. <https://doi.org/10.1086/679079>.
- Witt, J.C., (2018). “Digital Scholarly Editions and API Consuming Applications”. Bleier, R. et al. (eds), *Digital Scholarly Editions as Interfaces*, Norderstedt: Books on Demand GmbH, 219-47.
- Yazdani, M.; Popescu-Belis, A. (2013). “Computing Text Semantic Relatedness Using the Context and Links of a Hypertext Encyclopedia”. *Artificial Intelligence*, 194, 176-202. <https://doi.org/10.1016/j.artint.2012.06.004>.

Public History e metadata Da fonti a oggetti informativi. Diffusione, didattica, co-creazione

Manfredi Scanagatta

Università degli Studi di Modena e Reggio Emilia, Italia

Abstract The dematerialisation of information and the development of Information Systems combined with the use of flexible metadata standards enable a new accessibility for the sources. Matching this new accessibility with the bottom-up historiographic production promoted by Public History enables the development of new forms of history teaching and the generation of information objects through participatory metadata creation with a community of reference. Through the analysis of three different cases, this article brings out the importance of metadata for the organisation, archiving and use of historical sources in a Public History perspective.

Keywords Metadata. Digital archives. Information systems. Public history. History didactics. Information objects.

Sommario 1 Introduzione. – 2 Da fonti a oggetti digitali. – 3 Metadata e serendipità, i metadati nella ricerca sul Web generalista. – 4 Metadata, accesso telepistemico e didattica. – 5 Oggetti informativi e metadattazione partecipata. – 6 Conclusioni.



Edizioni
Ca' Foscari

Peer review

Submitted 2023-02-28
Accepted 2023-05-30
Published 2023-08-02

Open access

© 2023 Scanagatta | © 4.0



Citation Scanagatta, M. (2023). "Public History e metadata. Da fonti a oggetti informativi. Diffusione, didattica, co-creazione". *magazén*, 4(1), 43-70.

1 Introduzione

Interesse di questo lavoro è provare a porre alcune riflessioni di natura metodologica e epistemologica, riguardo al possibile ruolo dei metadata all'interno di progetti di Public History.

Per proporre un ragionamento che non riguardi unicamente questioni di natura teorica ci avvaleremo di alcuni casi di studio che ci aiutino se non a comprendere, almeno a porci le domande corrette riguardo al ruolo dei processi di digitalizzazione dell'informazione nelle pratiche di ricerca e diffusione delle fonti.

Dal momento che la digitalizzazione dell'informazione richiede lo sviluppo di nuove pratiche per la gestione, contestualizzazione e conservazione delle fonti digitali, la disciplina che più di altre lavora in modo approfondito e accurato su questi temi è sicuramente l'archivistica informatica. All'interno di un contesto di dematerializzazione dell'informazione è determinante il ruolo dei metadata, che sono necessari per la realizzazione di un oggetto digitale e per assicurare alle fonti una conservazione e un ordinamento coerenti e affidabili anche all'interno di archivi digitali.

Da una prospettiva di Public History i metadata assumono un interessante valore anche nei processi di ricerca e ci consentono di immaginare pratiche di diffusione delle fonti attraverso un nuovo livello di intermediazione, che può avere interessanti ricadute anche all'interno della didattica della storia. Inoltre, grazie a strumenti open source come Omeka S è possibile applicare i modelli logici di OAIS (Open Archival Information System) per realizzare degli oggetti informativi tramite una metadatazione partecipata, così da poter coinvolgere una comunità di riferimento in un processo di significazione dei documenti.

L'approccio dell'articolo è di natura interdisciplinare.

In prima istanza ci confronteremo con le riflessioni di filosofi quali Luciano Floridi e Maurizio Ferraris che ci interrogano sul valore re-ontologizzante dei processi di digitalizzazione dell'informazione (Ferraris, Painsi 2018) e che ci invitano a riflettere sulle società Iperstoriche (Floridi 2017).

Partire da questo contesto di riferimento ci è utile per procedere nell'analisi di alcuni casi studio relativi a progetti di Public History dove i metadata assumono un valore centrale.

Da storici e public historians confrontarsi con l'analisi filosofica e informatico-archivistica dei processi di digitalizzazione delle fonti consente di porsi alcune domande di natura metodologica. È fondamentale per le nostre discipline interrogarsi su quale sia oggi e possa essere domani la relazione tra fonti, storici e pubblico, in un ambiente - quello digitale - all'apparenza disintermediato, ma che nella realtà dei fatti grazie all'utilizzo ricorsivo di metadata può portare a una nuova forma di intermediazione delle fonti.

L'aspetto più interessante e rilevante della Public History non si manifesta solo nell'elaborare nuove pratiche di divulgazione del contenuto storico, quanto piuttosto nello sviluppare innovativi processi per coinvolgere il 'pubblico' nella costruzione di storiografia. Ciò assume un valore ancora maggiore quando le comunità vengono coinvolte nell'analisi e nella selezione delle fonti. È possibile attivare questo processo grazie a pratiche di *crowdsourcing* così come indicato dalle Citizen Humanities (Paci 2021), o sviluppando processi di metadatazione partecipativa.

Attraverso l'osservazione di casi specifici in queste pagine si proverà a far emergere i possibili benefici della rivoluzione digitale nel rapporto tra pubblico e storia, tenendo il nostro orizzonte d'analisi sulla relazione con le fonti primarie.

Il primo caso che osserveremo ci chiama a riflettere sulla serendipità del Web e sulle possibilità di condivisione e di accessibilità alle fonti in un'analisi incrociata tra Web 2.0, Sistemi Informativi e archivi digitali.

Osserveremo poi come gli archivi digitali e la presenza telepi-stemica, così come definita da Luciano Floridi in *Pensare l'infosfera* (2020, 83), possano contribuire a innovare i percorsi di didattica della storia. Infine ci confronteremo con la possibilità di costruire un archivio partecipato in cui realizzare oggetti informativi in collaborazione con una comunità di riferimento.

In tutti e tre i casi vedremo come i metadati risultino essere l'elemento decisivo per riuscire ad attuare una nuova relazione con i documenti all'interno di un sistema digitale.

2 **Da fonti a oggetti digitali**

Il potere re-ontologizzante della rivoluzione digitale (Ferraris, Pagni 2018) si manifesta in modo chiaro nella produzione, nella condivisione e nell'archiviazione di informazioni.

L'uomo è storicamente un essere informazionale che produce dati e informazioni da migliaia di anni. Secondo Luciano Floridi solo quando sono diventati disponibili sistemi per registrare eventi e, in tal modo, accumulare e trasmettere le informazioni per un futuro consumo, le lezioni apprese dalle generazioni passate hanno iniziato a evolvere esponenzialmente, in maniera debole o lamarckiana e così l'umanità ha fatto ingresso nella storia (Floridi 2017, 1).

In questo processo di produzione, conservazione e scambio di informazioni hanno un ruolo determinante le tecnologie. Il papiro, la pergamena, l'invenzione della carta, l'inchiostro e le penne, la stampa a caratteri mobili, le macchine per scrivere, sono solo alcune delle tecnologie che nel passato hanno consentito la conservazione di un dato in forma di informazione. Ciò che differenzia la produzione e la

condivisione di informazioni tra il passato e il presente, fatta salva la rilevanza dell'oralità all'interno di ogni società, è che in passato l'informazione scambiata e registrata aveva una forma tangibile e si presentava a noi come elemento materico, esisteva come oggetto fisico.

L'esistenza fisica del dato e dell'informazione ne attestava il valore in quanto documento; nulla di formale esisteva al di fuori di un artefatto documentale fisico. L'essere materico era prova dell'esistenza stessa del documento che nel suo essere tangibile diveniva parte dell'esperienza umana in società, non solo per il contenuto, ma per il fatto stesso che esistesse nel nostro mondo dell'esperienza.

Il dato per divenire informazione, essere conservato e avere un valore sociale deve passare per quello che Ferraris in *Documentalità* (2019) definisce un processo di iscrizione, che nell'era pre-digitale mutava a seconda della tecnologia con la quale si desiderava conservare e rendere fruibile l'informazione.

L'iscrizione di un dato produce un documento che interessa sempre l'esistenza di almeno due esseri umani; il documento era dunque la rappresentazione fisica di una interazione sociale, per dirla con Ferraris (2019, 176) «nulla di sociale esiste al di fuori del testo».

Il documento digitale, sia questo digitalizzato o *digital born*, si manifesta a noi come rappresentazione intellegibile di un codice binario, un linguaggio macchina elaborato da esseri umani e interpretato da macchine digitali.

Il digitale ha uniformato i processi di produzione dell'informazione. Un video digitale, una foto digitale, un suono digitale o un certificato di morte digitale, sebbene si manifestino alla nostra realtà in modo sensibilmente differente sono tutti composti di bit.

Omologando in senso computazionale i sistemi di iscrizione si ha la possibilità di generare e fruire dell'informazione in modo nuovo.

Ciò che era ontologicamente differente ad oggi è composto della stessa sostanza, bit.

L'uniformità che determina la sostanza tramite la quale viene realizzata ogni forma di documento digitale consente di sviluppare nuove modalità di fruizione, organizzazione, conservazione e condivisione di tali documenti.

All'interno delle società Iperstoriche - società in cui i processi informativi sono gestiti prevalentemente da tecnologie ICT (Floridi 2017), - in qualità di storici, storiche e public historians dobbiamo porci nuove questioni di ordine epistemologico e metodologico.

Mai come oggi i documenti, le fonti primarie (siano queste nate digitali o digitalizzate) possono essere facilmente accessibili non solo per lo storico, ma anche per chi non ha esperienza di ricerca.

Come ci mettono in guardia, tra gli altri, Valacchi (2021) e Rossi (2021), dobbiamo stare attenti a non farci ammaliare dalle semplificazioni del digitale.

La ri-mediazione (Bolter, Grusin 2002), che come scrive Pietro Montani è la rappresentazione di un medium all'interno di un altro e che può portare anche d un effetto di iper-mediazione comunemente interpretato come un incremento di immediatezza (Montani 2010, 12), è resa possibile da un processo di dematerializzazione. Sia la dematerializzazione sia la ri-mediazione delle fonti richiedono di essere trattate con cautela e soprattutto con la consapevolezza che un documento, per essere definito come oggetto digitale ha bisogno che gli vengano attribuiti dei metadata (Felicati 2009).

È dunque necessario attuare un processo di sovrascrittura per rendere possibile l'interazione tra uomo e macchina.

È in questo processo di sovrascrittura che possiamo individuare una neo-intermediazione sulla fonte, che se sviluppata in seguito a un processo ricorsivo di metadatazione può generare un oggetto informativo (Michetti 2008).

3 Metadata e serendipità, i metadata nella ricerca sul Web generalista

Come ci fa riflettere Lorenzo Pezzica, il processo trasformativo che sta coinvolgendo gli archivi e le fonti non è un processo indolore e non trova la comunità degli archivisti, e aggiungo io degli storici e delle storiche, in completo accordo su quelli che possono essere i sistemi di riorganizzazione digitale degli archivi e i reali vantaggi di questo processo.

La produzione e la circolazione dell'informazione sul Web scivolano fuori dalle maglie della funzione archivistica e il mito della ricerca globale sulle banche dati e la potenzialità totalizzante delle tecnologie di information retrieval ha fatto ritenere a qualcuno superata l'organizzazione archivistica. (Pezzica 2020, 70)

Pezzica, riprendendo il pensiero di Paola Ciandrini (Ciandrini 2018) invita gli archivisti a essere resilienti, dunque capaci di modificare il loro modo di intendere l'archivio e di non farsi limitare «dalla difesa di ciò che si fa da sempre» (Pezzica 2020, 71).

Vitali (2017, 65) nel suo saggio «La ricerca archivistica sul Web», tra le altre cose si sofferma sull'importanza del metodo che deve guidare il ricercatore nel suo approccio all'archivio, dettagliando in cinque mosse le fasi della ricerca archivistica classica:

1. individuazione dell'esigenza informativa da soddisfare e formulazione del proprio questionario di ricerca;
2. traduzione del questionario storico in un questionario archivistico;

3. identificazione dell'archivio o degli archivi che possono dare risposte positive e valutazione della loro effettiva pertinenza;
4. individuazione e localizzazione degli specifici documenti da consultare;
5. consultazione e interpretazione critica dei documenti

Aggiungendo che

Oggi alcune di queste mosse e, in taluni casi, l'intero percorso di ricerca possono essere compiuti sul web. Ciò non altera la logica profonda del modello, muta semmai le modalità operative per raggiungere i risultati prefissi e richiede non solo la conoscenza dei meccanismi di funzionamento del web ma anche la capacità di distreggiarsi fra le possibili insidie che esso può nascondere, non lasciandosi troppo sedurre dal miraggio di rapide scorciatoie e di illusorie semplificazioni e non rinunciando ad applicare alla ricerca sul web in tutti i suoi passaggi, il rigore e l'acribia filologica che lo studio critico del passato sempre richiede. (Vitali 2017, 65-6)

Grazie all'utilizzo di metadati che consentono al ricercatore o alla ricercatrice anche non esperti di dare avvio alla propria ricerca utilizzando parole chiave, è già in essere una trasformazione che modifica *de facto* i processi che Vitali individua nei punti 3 e 4.

A partire dalla necessità di avere ben chiaro oggetto e domande di ricerca, per individuare un archivio e le fonti oggi è possibile agire in modo 'semplificato' grazie a sistemi informativi a vocazione tematica come SAN (Sistema Archivistico Nazionale), SIAS (Sistema Informativo degli Archivi di Stato) e SIUSA, (Sistema Informativo Unificato per le Soprintendenze Archivistiche). Questi strumenti si sono posti innanzitutto l'obiettivo di censire, conoscere e poi descrivere di nuovo e in maniera più approfondita, per una fruizione digitale, il patrimonio documentario, diversificando, ma anche sovrapponendo la propria azione in ragione della tipologia giuridica dei soggetti produttori.

I sistemi informativi rispetto alle guide tradizionali danno la possibilità di generare sistemi di relazioni che collegano reciprocamente e dinamicamente le entità informative, riuscendo a restituire con estrema immediatezza il complesso panorama del percorso di produzione, uso e conservazione e ancorando i complessi documentari a tale contesto, che è al tempo stesso chiave primaria ai fini della reperibilità e garanzia di corretta utilizzazione di dati attendibili.

SAN, SIAS e SIUSA tra le altre cose consentono di scoprire quali risorse documentarie siano presenti all'interno di differenti archivi, velocizzando e semplificando in modo sensibile questa prima e fondamentale parte del lavoro di ogni storico e storica. Oltre a consentire al ricercatore e alla ricercatrice di sapere quali documenti sono conservati in quali archivi, questi strumenti si presentano anche

come punti di accesso unificati ad archivi federati. In questa ipotesi, se gli archivi consentono l'accesso alla riproduzione simulacrale dei documenti materici ri-mediati, è possibile proporre al fruitore un'esperienza innovativa di ricerca e fruizione di fonti storiche.

I sistemi informativi basano il proprio funzionamento su modelli di strutturazione di dati quali EAD (Encoded Archival Description)¹ e EAC-CPF (Encoded Archival Context-Corporate Bodies, Persons and Families).²

I principi che regolano questi strumenti sono:

- consentire l'accesso al materiale archivistico fornendo agli utenti informazioni relative ad esso;
- produrre strumenti di ricerca accurati, coerenti e auto-esplorativi;
- presentare il contesto e il contenuto del materiale archivistico descritto applicando le regole della descrizione multilivellare (Barbanti 2005).

Come sappiamo il Web 2.0 consente di condividere dati e informazioni in una relazione orizzontale tra gli utenti. Per quanto affascinante, questo sistema comporta alcune criticità; prima tra le quali la produzione incontrollata di informazione che contribuisce in modo sensibile a dare origine a una situazione di infodemia: la produzione e lo scambio di un'eccessiva quantità di informazione all'interno della quale è possibile trovare tutto e il contrario di tutto, rendendo così il concetto di autorità eccessivamente liquido e portando a una relativizzazione della conoscenza e della verità. Come riportato negli studi svolti da Quattrococchi e dal suo team di ricerca (Cinelli et al. 2020), nella eccessiva quantità di informazione risiede il virus della disinformazione.

Se il pericolo della *misinformation* è oggi un pericolo reale, il Web si presenta comunque come luogo di eccellenza per avere accesso alla conoscenza.

La digitalizzazione delle fonti storiche e la conseguente possibilità di condivisione possono trasformare una piattaforma come Facebook in un luogo dove è possibile confrontarsi con fonti storiche primarie anche se attraverso percorsi 'inusuali'.

1 L'origine di EAD risale al BFAP (Berkeley Finding Aids Project), avviato nel 1993, presso la Berkeley University in California, con l'intento di sviluppare uno standard non proprietario per la codifica in formato digitale di strumenti di ricerca quali inventari, elenchi sommari, indici facendo ricorso allo SGML (Standard Generalised Mark-up Language). <https://icar.cultura.gov.it/standard/standard-internazionali/ead>.

2 EAC-CPF intende costituire uno standard per la codifica in XML e l'interscambio di record di autorità basati sull'ISAAR (CPF) (International Standard Archival Authority Records for Corporate Bodies, Persons and Families). <https://icar.cultura.gov.it/standard/standard-internazionali/eac-cpf>.

Il 25 aprile 2019 un giornalista e scrittore celebrò il giorno della Liberazione condividendo sulla sua pagina Facebook il post del gruppo legato alla *Juventus Autonomia Bianconera - Juventus Supporter*.

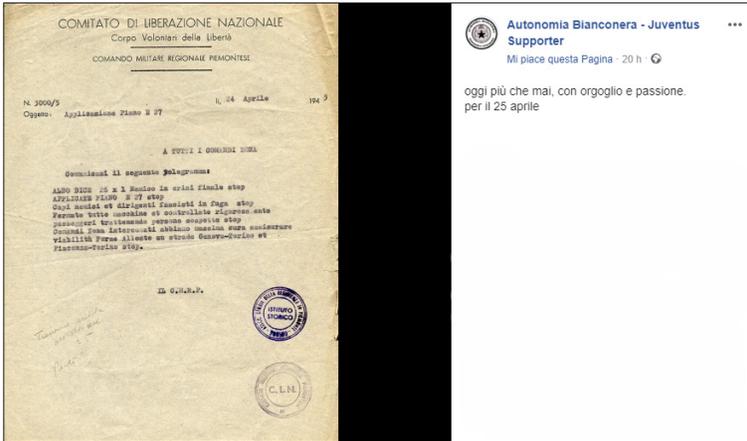


Figura 1 Post per il 25 aprile 2019 pubblicato sul gruppo Facebook Autonomia Bianconera, con documento digitalizzato del piano operativo E-27

Ciò che deve attirare la nostra attenzione è l'immagine allegata al post. La rappresentazione simulacrale di un documento materico che ha subito un processo di dematerializzazione e ri-mediazione e che diventa strumento narrativo all'interno di un social network generalista e con fini ricreativi (Scanagatta 2019).

Il documento è l'ordine dato dal CMRP (Comitato Militare Regionale Piemontese), inviato il 24 aprile del 1945 via telegramma a tutti i comandi di zona per rendere operativo il piano E27 (Emergenza 27). Un piano insurrezionale che dal 1944 veniva elaborato dal CLN (Comitato di Liberazione Nazionale), per organizzare l'insurrezione nel nord Italia contro l'occupazione nazifascista.

Possiamo immaginare che la maggior parte delle 125 persone che a venti ore dalla pubblicazione del post avevano messo un 'mi piace' non fossero storici, come non lo fossero tutti e nove gli utenti che a loro volta hanno deciso di condividere il post e dunque il documento.

Prendiamo il caso di un utente che pur non sapendo nulla sulla liberazione di Torino rimane colpito dal post perché condiviso da una pagina Facebook attinente al calcio di cui è molto tifoso.

Quello che può avvenire se l'utente dimostra un minimo di curiosità è che per capire meglio il contenuto condiviso cerchi su Google le parole che troviamo nel documento.

Conoscendo i sistemi di funzionamento di Google sappiamo che ci permettono di moltiplicare le chiavi di accesso alle informazioni e

di svincolarsi da una ricerca monodimensionale e sequenziale come quella che avviene in biblioteca.

In un sistema di *information retrieval* «l'utente specifica i propri bisogni informativi, indicando alcune parole chiave e il sistema recupera l'insieme dei documenti che si avvicinano alla domanda dell'utente» (Baeza-Yates, Ribeiro-Neto 1999) cioè i documenti che, direttamente o tramite le loro rappresentazioni, sono associati nell'indice alle parole ricercate (Vitali 2005, 83)

Google nel momento dell'indicizzazione avvia un processo automatico di estrazione di dati che poi vengono attribuiti, in qualità di metadati, alla risorsa digitale. Ciò rende possibile rintracciare il link alla risorsa informativa in seguito all'utilizzo di parole chiave corrispondenti a quelle selezionate dal sistema.

Inserendo nella barra di ricerca di Google l'oggetto del documento condiviso: «Piano E27», il primo risultato che ci fornisce il sistema di ricerca ci rimanda al sito Militarystory.org, che si presenta come un archivio inventato,³ non propone all'utente un resoconto storiografico di cosa sia stato il piano E27 ma consente la fruizione del documento originale digitalizzato, rimandando anche al documento che dal 1944 fu elaborato dal CLN.

Il secondo risultato offerto da Google in seguito alla ricerca «piano E27» ci indirizza ad accedere a www.metarchivi.it e cliccando sul link ci troviamo davanti alla videata qui di seguito riprodotta:

Archos Archivi della Resistenza e del '900
Sistema integrato dei cataloghi d'archivio

Home Metarchivi Biografie Strumenti login

esplora ricerca topografico catalogo dizionari

Sottofascicolo: "Piano E 27" [Torna all'elenco dei documenti](#)

Posizione nella struttura d'archivio

- ↳ Istituto piemontese per la storia della Resistenza e della società contemporanea "Giorgio Agostini"
- ↳ Isp. Fondi originali: Prima sezione
- ↳ Seconda sottosezione: attività militari
- ↳ Comando militare regionale piemontese
 - ↳ "Piani militari"
 - ↳ "Piano E 27"

Codice scheda: C00/00001/02/02/00004/002
Titolo: "Piano E 27"
Descrizione: Diverse redazioni e ordini relativi al piano dell'insurrezione in Piemonte.
Estremi cronologici: 30/08/1944-25/04/1945
Nota archivistica: Nel fascicolo si riscontra la presenza di carte originariamente collocate in sottofascicoli diversi.
Nota bibliografica:

Consistenze

tipologia	supporto	q.tà documenti	q.tà carte/supporti
ciclostato	carta	2	10
dattiloscritto	carta	35	93
mappa	carta	1	1
Q.tà documenti		41	Q.tà carte/supporti 104

Documenti digitalizzati (1) [mostra](#)

[Relazioni con altri documenti e biografie](#)

Figura 2 Pagina del sistema integrato dei cataloghi d'archivio ARCHOS. Area Metarchivi, sistema multilivellare navigabile

³ Sul tema degli archivi inventati vi è oggi un vivace dibattito, sia nell'ambito dell'archivistica che della storia. Per un approfondimento cf. Vitali 2017, 91-6; Paci 2021; Valacchi 2020; Noiret 2018.

ARCHOS è un sistema integrato di catalogazione, archiviazione e ricerca, che opera sulle differenti tipologie di documenti appartenenti ai fondi e alle collezioni degli enti conservatori, in questo caso prevalentemente Istituti Storici del Novecento e della Resistenza.

Il sistema è costituito da due ambienti: Biografie e Metarchivi; di interesse per l'analisi proposta è la parte relativa ai Metarchivi. Questa si presenta come un multi-catalogo in cui l'accesso alle informazioni può avvenire esplorando la struttura gerarchica degli archivi organizzata in modo multilivellare: fondi, serie, sottoserie, fascicoli, sottofascicoli, documenti. Ogni livello archivistico può essere aperto cliccando sulla relativa icona, mentre attivando il collegamento del suo titolo si accede alla relativa scheda descrittiva. Il catalogo conduce alla lista degli enti conservatori, dove scegliendo l'ente di interesse con un click si accede al fondo digitalizzato di quell'ente.

ARCHOS si presenta come un punto di accesso unificato a una confederazione di archivi digitalizzati; i documenti che è possibile osservare attraverso ARCHOS in forma digitale esistono in forma tangibile all'interno di archivi fisici e la loro trasposizione in forma digitale - come è possibile vedere anche dall'immagine - mantiene fede a specifici processi di ordinamento archivistico. Il primo e forse più importante è il vincolo archivistico, che come scrive Pigliapoco è l'insieme delle relazioni logiche e formali che esistono tra i documenti di un archivio (Pigliapoco 2018, 15).

Per dirlo con le parole di De Felice, il vincolo archivistico

non è frutto di una volontà esterna, ma è naturale, caratterizzata dalla necessità e originarietà e che determina una lunga catena di precedenti e susseguenti, catena che non può essere spezzata senza compromettere l'essenza stessa dell'archivio. (De Felice 1998, 21)

Queste parole non possono che farci riflettere rispetto al processo di indagine che stiamo descrivendo. L'utente che si è trovato di fronte al documento condiviso su Facebook è probabile che non abbia nessuna competenza in campo archivistico e interesse verso i vincoli e le classificazioni. Il documento si presenta sul social network come una complessità narrativa atomica, nel senso che basta a se stessa per veicolare l'informazione. Ciò non di meno però, l'utente che incuriosito si propone di fare una breve e semplicissima ricerca (ricordiamoci che abbiamo usato solo le parole presenti nell'oggetto del documento per compiere la ricerca su Google) potrà trovarsi di fronte a una struttura gerarchica multilivellare in cui si manifesta in modo chiaro la concatenazione di precedenti e susseguenti in cui i documenti sono in relazione. Questa concatenazione è visibile e soprattutto navigabile. Ciò comporta che l'utente solo con un click possa accedere ad altri fascicoli e dunque aumentare il contesto storico e documen-

tale all'interno del quale è inserito il documento che abbiamo visto per la prima volta su una pagina di tifosi della Juventus.

Per condurre ricerche su Google e trovare i documenti che ci interessano non c'è neanche bisogno che qualcuno li abbia preventivamente organizzati, ordinati e classificati a nostro beneficio. I meccanismi di funzionamento di Google non solo non si basano affatto su classificazioni, ordinamenti, sistemazioni preventive dei documenti, ma, in un certo senso, evitano proprio che ciò possa accadere (Vitali 2005, 85).

Per dirla con le parole dei creatori di Google, Sergey Brin e Lawrence Page,

è una estesa collezione di documenti eterogenei completamente incontrollabili, vari dal punto di vista del formato (testo, immagini, suono, ecc.), da quello della codifica (html, pdf, ecc.), delle lingue utilizzate e dei vocabolari adottati. (Brin, Page 1998)

Uno dei più grandi timori degli archivisti che si occupano dell'analisi dei processi di digitalizzazione è proprio che, come sostenuto da uno dei creatori di Google, le informazioni all'interno del Web si muovano in modo incontrollato, quando sappiamo che uno dei concetti cardine su cui si sviluppa la disciplina archivistica è quello di ordinamento. Non solo, se i documenti digitali non rimangono legati al contesto originario di produzione e conservazione c'è il rischio che nelle fasi di ricerca possano essere mal interpretati. Sono gli standard come EAD ed EAC che consentono di realizzare strutture capaci di contestualizzare i documenti digitali e di ricreare i vincoli archivistici fondamentali per assicurare una corretta fruizione dei documenti. I nuovi studi sul Web semantico e l'utilizzo di ontologie stanno cercando di superare le strutture tipiche dell'archiviazione per utilizzare paradata e metadati da incorporare sul dato originario. Così facendo ogni risorsa informativa descritta da una ontologia è pienamente autoconsistente, radunando, a grappolo intorno a sé tutta l'informazione che la descrive complessivamente (relazioni in partenza) ma risultando in connessione anche con l'informazione alla cui rappresentazione partecipa con la funzione di complemento (relazioni in arrivo) (Di Fazio 2018).

Attraverso una ricerca sul Web generalista è possibile trovare i link che ci rimandano a sistemi informativi, ad archivi digitali o federati, all'interno dei quali, prima ancora di osservare la fonte primaria, possiamo ottenere informazioni valide e utili per la ricerca o per la costruzione di un contesto di approfondimento. Standard di metadatazione, archivi digitali e sistemi informativi sono quindi strumenti di organizzazione per la conservazione e la comunicazione delle informazioni, strumenti che nel caos della rete si mostrano come potenziali e potenti argini alla *misinformation* quanto meno in ambito storico.

Perché i sistemi informativi possano davvero assumere un valore sociale diffuso e generalizzato dobbiamo in prima istanza educare noi

stessi al loro utilizzo, così che gli algoritmi di Google ce li mostrino come risultati possibili già nella prima pagina di ricerca. Se a scuola, ma anche all'università, si condividesse la conoscenza di questi strumenti e si dessero indicazioni utili e semplici per usarli, questi sarebbero davvero strumenti formidabili e dal grande impatto sociale e culturale.

In questo primo caso abbiamo osservato come la serendipità possa innescare un processo che grazie all'utilizzo di metadata, di sistemi informativi e di una corretta conservazione digitale, consenta anche a un utente non esperto di accedere a un vasto numero di fonti primarie a partire, per esempio, da un post su Facebook.

4 Metadata, accesso telepistemico e didattica

Vediamo ora come la digitalizzazione di fonti e archivi possa essere uno strumento davvero utile per la didattica della storia.

Negli ultimi anni la storia insegnata si è trasformata da impresa trasmissiva ad attività di costruzione di conoscenze, di modi per imparare a pensare la realtà storicamente e a conoscerla disponendo di concetti e categorie interattive.

L'attenzione si è progressivamente spostata dagli eventi/argomenti, oggetto della trasmissione e della memorizzazione, all'analisi di ciò che implica e richiede la costruzione di conoscenze di quegli argomenti, intesi come oggetto del discorso storico. (Girardet 2015, 11)

Perché la storia possa essere capita e non solo imparata, chi la studia - e mi riferisco a studenti e non a ricercatori - deve potersi chiedere non solo quali informazioni sia possibile ottenere leggendo un libro, ma da dove derivino le informazioni che hanno composto quel libro, chi le ha scelte e in che modo.

Mutuando sistemi di apprendimento didattico da una metodologia come il *learning by doing*, proponendo dunque a studenti e studentesse di confrontarsi con i processi della ricerca storica in archivio, si amplieranno le possibilità di conoscenza e si consentirà alla classe di comprendere che le informazioni di tipo storico sono molte e diversificate e che la storia altro non è se non ciò che rimane a noi dell'azione di altri esseri umani, che come tali commettono errori, hanno sogni e agiscono in risposta anche a emozioni e pensieri. Lavorare con le fonti vuol dire poter rintracciare ciò che Spengler definisce il volto umano della storia (Spengler 2012).

Far confrontare studenti e studentesse con la ricerca archivistica li aiuterà a comprendere come si organizzano le informazioni e li aiuterà a riflettere sul concetto di attendibilità delle fonti a partire dall'analisi del contesto, del soggetto e dell'ente produttore.

Insegnare nelle scuole il metodo storico di ricerca, analisi e critica delle fonti consente di fornire competenze utili a comprendere criticamente la complessità del passato e vorrei aggiungere ispirandomi a Carr (1966) anche del presente, soprattutto se consideriamo il costante flusso informativo in cui oggi siamo inseriti.

L'accesso ad archivi digitali richiede una riflessione sul concetto di 'presenza'.

Negli studi attuali la presenza è spesso compresa come un tipo di esperienza dell'essere qui, che coinvolge genericamente qualche mediazione tecnologica e spesso dipende da ambienti virtuali. (Floridi 2020)

Per aiutarci a meglio comprendere questi processi prenderemo in esame il laboratorio che nel 2017-18 è stato realizzato in collaborazione con l'Istituto Luce Cinecittà e con il Liceo Tacito di Roma.

A studenti e studentesse è stato chiesto di realizzare un prodotto di Public History su un argomento di storia contemporanea. Dopo una prima fase di confronto e condivisione di idee, la classe ha deciso di lavorare sui processi di fascistizzazione dell'Italia e ha realizzato una mostra multimediale in cui le fonti sono state messe in scena attraverso un processo di narrativizzazione (Scanagatta 2015).

Per realizzare il prodotto, gli studenti sono stati divisi in gruppi di lavoro, ognuno dei quali è stato assegnato a un tema specifico che avesse significato sia nel periodo storico considerato, ovvero dal 1922 al 1945, sia nel loro presente. I gruppi hanno quindi scelto di lavorare sui temi della scuola, della comunicazione e dell'urbanistica.⁴

Da un punto di vista didattico ritengo che la possibilità di risemantizzare il passato nell'oggi - e viceversa - sia un processo molto funzionale per comprendere i concetti e i fatti storici in esame.

Dopo una prima revisione e discussione della letteratura storiografica riguardo i temi scelti, studenti e studentesse hanno iniziato a interfacciarsi con il sito dell'Archivio Storico dell'Istituto Luce, tra i primi in Italia a essere digitalizzato e a consentire a chiunque acceda al portale la fruizione gratuita dei documenti, se pur in bassa definizione.

È proprio la modalità iniziale di ricerca che ci mostra da subito le potenzialità della digitalizzazione in ambito didattico. Gli studenti si sono approcciati alla ricerca in archivio utilizzando parole chiave che secondo loro avevano attinenza con i singoli argomenti da approfondire. Digitando per esempio la parola 'scuola' il sistema ha pro-

⁴ Comunicazione: <https://www.youtube.com/watch?v=1TzMFUGqibg&t=23.s>; scuola: <https://www.youtube.com/watch?v=J5zjmmUSZa0&t=17.s>; urbanistica: <https://www.youtube.com/watch?v=hVilew8I8hw&t=2s>; sport: <https://www.youtube.com/watch?v=4uW04xDJbQg>.

posto loro una serie di risultati che nella descrizione archivistica riportavano quel termine.

Interrogare l'archivio storico dell'Istituto Luce non è stato diverso dal fare una ricerca su YouTube e i risultati ottenuti hanno messo gli studenti immediatamente davanti alle fonti. Non dovevano far altro che premere play per poter iniziare a consultare delle fonti primarie.

La prima parte del lavoro di ricerca è stata svolta in completa autonomia senza che docenti e formatori li guidassero nel processo. Man mano che gli studenti andavano avanti nella navigazione dell'archivio riuscivano ad affinare sempre di più la ricerca capendo come muoversi anche grazie all'aiuto di metadati descrittivi che consentivano di utilizzare nomi propri, date e luoghi.

Fino a qui possiamo osservare come il digitale abbia consentito una nuova modalità di fruire un archivio rendendolo più accessibile e consentendo una ricerca più semplice di come si sarebbe potuta svolgere se avesse riguardato invece i documenti in forma materica.

Secondo la Società Internazionale per la ricerca sulla Presenza, organizzazione no-profit fondata nel 2002 per supportare la ricerca accademica rispetto al concetto di telepresenza (<https://ispr.info/>) se l'interazione e l'esperienza di ricerca avvengono grazie a una interazione tra umano e macchina, potrebbero mutare anche se al ricercatore potrebbe sembrare che nulla cambi. Il nodo della questione risiede nel modo in cui si è abituati a reagire alle percezioni che si hanno nell'interazione con l'ambiente nel quale si agisce. Senza il coinvolgimento delle tecnologie le percezioni dell'uomo sono veicolate unicamente attraverso i sensi, mentre quando interviene una tecnologia che ci consente un accesso in telepresenza o telepistemico all'ambiente con il quale entriamo in relazione, non sono più solo i sensi a mutuare le nostre percezioni, c'è un nuovo elemento, quello tecnologico, che non sempre cogliamo come determinante nella percezione, ma che nella realtà dei fatti non solo la rende possibile ma ne stabilisce anche le modalità (Mantovani, Riva 2001).

Se questa intromissione nelle nostre percezioni sensoriali ci può sembrare quasi distopica dobbiamo tenere di conto che oggi molte delle nostre attività sono mutate dall'utilizzo di strumenti tecnologici digitali e che soprattutto lo scambio informativo tra inforganismi avviene grazie a ICT. La generazione Z sta crescendo all'interno dell'infosfera ed è dunque a suo agio in uno stile di vita onlife (Floridi 2017).

È bene ricordare che prerogativa della tecnologia è il suo 'essere tra' un suggeritore e un utente. Spesso si sostiene che all'interno del Web non vi sia intermediazione tra il fruitore e l'informazione. Se ciò non è vero in generale - gli algoritmi sono una forma di intermediazione - lo è ancora meno quando si tratta di un archivio digitale.

Parlando di didattica non si può pensare che la docente o il formatore che decidono di utilizzare le possibilità della digitalizzazione per la didattica della storia demandino parte del loro lavoro a un

sistema che appunto viene descritto come disintermediato. Ciò che avviene è che si sviluppa un nuovo tipo di intermediazione che si ha la sensazione di non riconoscere come umana, ma, come sappiamo, alla base di qualsiasi processo tecnologico digitale vi sono indicazioni umane. Anche quando una macchina è capace di apprendere, ciò avviene perché noi esseri umani le abbiamo assegnato quel compito e fatto in modo che riuscisse a svolgerlo (Durante 2019).

Se all'interno di un archivio storico di tipo materico le fonti sono conservate mantenendo tra di loro una correlazione di natura fisica e logica, con la digitalizzazione delle fonti primarie e la costruzione di archivi digitali le fonti sono depositate in una organizzazione unicamente logica e non più fisica (Guercio 2000). Questo implica che perché una risorsa digitale possa essere correttamente archiviata e verificata necessiti dell'utilizzo di metadata.

Definire degli standard di archiviazione e gestione è fondamentale per assicurare che le fonti digitalizzate che verranno ordinate all'interno di questi sistemi di gestione mantengano tra di loro vincoli specifici, imprescindibili per la fruizione dell'archivio digitale da parte di uno storico, o in questo caso di chi storico non è ma sta provando ad applicarne il metodo di ricerca.

Grazie a *device* digitali, alla connessione a Internet e utilizzando parole chiave di ricerca, gli studenti e le studentesse del Liceo Tacito pur rimanendo in classe sono riusciti ad avere un accesso telepistemico alle fonti digitali conservate all'interno di archivi digitali. Non sono gli studenti a entrare nell'archivio, ma sono le fonti digitali a entrare nello spazio di lavoro degli studenti. Questo implica che una volta 'portata' la fonte in classe potenzialmente sarà possibile utilizzarla come meglio si crede.

Un esempio semplice per comprendere cosa accade quando si accede alle informazioni attraverso un'esperienza telepistemica è immaginarci all'interno di una stanza separata con una parete da un altro ambiente che contiene le informazioni; se eliminiamo la parete che ci divide noi non siamo entrati nella stanza in cui risiedono le informazioni, il nostro corpo non si è spostato dall'ambiente in cui si trovava, ma adesso ci è possibile fruire delle informazioni che sono entrate all'interno del nostro spazio di osservazione.

Grazie alla digitalizzazione è oggi possibile appropriarsi della fonte storica o meglio della rappresentazione simulacrale di questa. La fonte materica sarà per sempre una sola, immodificabile e rigorosamente conservata all'interno di un archivio materico, ma se dematerializzata e ri-mediata, l'oggetto digitale sarà a completa disposizione di ogni utilizzo didattico (fatto salvo per gli archivi digitali vaticani che per fruire delle fonti digitalizzate richiedono un pagamento).

Nella seconda parte del laboratorio i gruppi hanno iniziato a selezionare il materiale trovato concentrandosi sui documenti che meglio di altri riuscivano a raccontare il tema su cui stavano facendo la

ricerca. È in questa fase che studenti e studentesse si sono confrontati e confrontate con l'analisi critica delle fonti a partire dall'analisi critica dell'archivio stesso, che in questo caso come sappiamo nasce come organo di propaganda di un regime.⁵

Una volta che le fonti digitalizzate sono state scaricate nello spazio di lavoro degli studenti e delle studentesse, ogni gruppo ha iniziato un lavoro di editing delle fonti audiovisive andando a tagliare la fonte originale e selezionando le parti che ritenevano maggiormente utili per la costruzione della mostra multimediale. Obiettivo era realizzare per ogni tema un montaggio di 3-5 minuti accompagnato da una voce fuori campo montata sui suoni originali delle fonti utilizzate.

Poter lavorare direttamente con le fonti aiuta studenti e studentesse ad avvicinarsi alla storia non come qualcosa da imparare a memoria, ma come una disciplina indispensabile per la costruzione di un pensiero critico che si sviluppa a partire dall'analisi delle fonti.

L'Istituto Luce, essendo stato l'organo di propaganda del Partito Fascista e primo veicolo per la costruzione del mito della personalità di Mussolini, non ha all'interno del suo archivio nessuna fonte primaria riguardante l'opposizione al fascismo. Gli studenti e le studentesse che dovevano lavorare su quel tema si sono trovati e trovate a confrontarsi con l'assenza, che da storici sappiamo essere uno dei motori più potenti per la produzione di domande. È infatti nell'assenza di fonti riguardanti l'opposizione che si manifesta in modo assoluto la vera essenza di una dittatura.

In quell'assenza non vi poteva essere retorica, non vi era spiegazione o narrazione dello storico.

Il risultato finale del laboratorio si è composto di cinque contenuti multimediali realizzati attraverso il montaggio di fonti primarie. La presentazione dell'output ha visto la realizzazione di una piccola mostra multimediale allestita a scuola e attraversata dagli studenti dell'istituto.

In un secondo momento i video sono stati caricati su YouTube raggiungendo complessivamente più di 20.000 visualizzazioni.

5 Oggetti informativi e metadatozione partecipata

Il Progetto DCMI (Dublin Core Metadata Initiative) è stato sviluppato nell'ambito dell'OCLC (Online Computer Library Center), con l'obiettivo di creare un insieme di strumenti condivisi per l'accesso alle risorse digitali. La prima versione, rilasciata nel 1996, era composta da quindici elementi di base, ciascuno dei quali descritto con attributi in conformità con la norma ISO (<https://www.dublincore.org/>).

Il progetto nasce in seguito alla necessità di escogitare uno sche-

⁵ Per maggiori approfondimenti sul laboratorio cf. Cacciani 2018.

ma descrittivo molto agile nel quale i dati richiesti per la compilazione della descrizione possano variare a seconda delle necessità e che possano essere utilizzati direttamente da coloro che creano le risorse digitali, pur senza essere in possesso di specifiche competenze catalografiche. È con questi presupposti che nel 1995 si tenne il primo workshop sullo schema Dublin Core.

Ciò che contraddistingue lo schema conosciuto come Dublin Core Metadata Element Set è l'assenza intenzionale di una specifica destinazione d'uso. DCMI può essere utilizzato per descrivere risorse di qualunque natura.

Se noi consideriamo i metadati elementi di sovrascrittura delle fonti con i quali poter attribuire a queste anche dei significati estrapolandoli con attenzione dal contenuto o dal contesto delle fonti stesse (Weston, Sardo 2017), l'utilizzo di DCMI, con la sua semplicità e trasversalità ci consente di immaginare la possibilità di costruire all'interno di archivi digitali un nuovo rapporto tra storia e storiografia.

Per favorire la diffusione di Dublin Core e renderne possibile l'adozione come standard *de facto*, sono stati stabiliti un numero minimo di elementi dal significato immediatamente comprensibile e con una struttura talmente flessibile da consentire la descrizione di risorse di qualunque natura e relative a un qualsiasi ambito disciplinare. L'insieme degli elementi che costituiscono lo schema Dublin Core è il risultato di un consenso ampio e interdisciplinare sui dati che sono necessari per l'effettuazione della ricerca di una risorsa.

Nell'elaborare lo schema i promotori si sono attenuti ai seguenti criteri:

- l'intrinsecità, cioè il fatto di descrivere le proprietà intrinseche della risorsa, come il contenuto intellettuale o la forma fisica, piuttosto che quelle estrinseche, che descrivono il contesto nel quale la risorsa viene utilizzata;
- l'estensibilità, cioè la possibilità di introdurre un limitato numero di personalizzazioni per adeguare la descrizione a esigenze specifiche;
- l'indipendenza dalla sintassi, per non vincolare gli utilizzatori al rispetto di definizioni formali;
- le opzionalità degli elementi, per consentire agli utilizzatori di redigere descrizioni di qualunque estensione, anche minima;
- la ripetibilità degli elementi, per non imporre limiti al numero di occorrenze del medesimo elemento: autori, date, soggetti e così via;
- la modificabilità degli elementi, cioè la possibilità di introdurre dei qualificatori per rendere più specifico il campo di applicazione di un determinato elemento.

La flessibilità di DCMI consente di utilizzare questo schema per qualsiasi risorsa digitale così da ottenere un'interoperabilità semantica che

potenzialmente rende possibile un'aggregazione di diverse fonti attraverso l'utilizzo di semplici parole chiave. Essendo più semplice da gestire rispetto ad altri standard richiede un livello di specializzazione inferiore nella codifica dei metadati e nell'ampliamento dei set descrittivi.

È in relazione a questa apertura e modificabilità che credo ci si possa chiedere se i principi di DCMI non ci debbano far riflettere sul lavoro dello storico e delle storiche e su quale possa essere il loro ruolo oggi all'interno di una società in cui ogni tipo di informazione viene cercata e trovata prevalentemente nel Web grazie a diversi motori di ricerca.

La nuova accessibilità alle fonti primarie che si ha grazie al Web può modificare in modo sensibile il rapporto tra storia e storiografia. La storia, di cui l'unica reale rappresentazione sono le fonti primarie, non emerge più solo grazie al lavoro delle storiche ma anche grazie allo sviluppo di articolati processi di ordinamento, sovrascrittura e comunicazione delle fonti digitali attraverso l'utilizzo di metadati.

Un lavoro storiografico per prendere forma ha bisogno oltre che dell'analisi critica e comparativa delle fonti anche della spiegazione e della narrazione, che come ci spiega Topolski (1997) è inevitabilmente guidata anche dal *savoir* dello storico o della storica. Ma se noi immaginiamo che storici e storiche in seguito all'analisi delle fonti ne estrapolassero dei concetti da trasformare in metadati, questi se pur in forma limitata potrebbero avere il ruolo di spiegazione, mentre l'interoperabilità delle fonti potrebbe assumere il ruolo di narrazione.

La conoscenza storica si manifesterebbe al fruitore attraverso una nuova tipologia di intermediazione, che tiene al centro della narrazione la fonte stessa che a sua volta avrà un ruolo differente. Non più solo solido appiglio scientifico per la costruzione storiografica da parte di un autore, ma vero e proprio elemento narrativo che si manifesta attraverso dei collegamenti carichi di significato da chi in una fase antecedente a quella della fruizione ha composto la metadattazione.

Questa ipotesi apre ad alcune domande. Chi lavorerà sull'analisi dei documenti per individuare i significati rilevanti? Come estrapolare i concetti? Quali parole chiave usare? Come sapere che le parole scelte hanno effettiva rilevanza per i fruitori che si avventurano in un archivio digitale?

Alcune di queste domande credo possano trovare una risposta o quanto meno una proposta nelle pratiche della Public History, se si osserva questo metodo non solo per la diffusione o divulgazione del contenuto storiografico, ma anche - o soprattutto - per la costruzione dal basso di processi di analisi critica delle fonti e di elaborazione di prodotti storiografici.

Nel 2022 durante la Summer School internazionale in Digital Humanities organizzata dall'Università di Modena e Reggio Emilia, con il dott. Matteo De Cristofaro, linguista che sta studiando le potenzialità dei metadati nell'analisi di corpora, si è tenuto un workshop dal titolo *Metadata* che veniva presentato ai frequentati con queste parole

With many misconceptions surrounding it, metadata has come to play a central role in the interaction between humans and computers, making exchanges between human knowledge and computer information possible. From search engines to semantic web, metadata allows the retrieval of information from digital data, as well as allowing machines to represent 'traces' of meaning across pieces of information stored as strings of 0s and 1s, de-facto translating human knowledge into digital objects that can be interpreted by computers. Consequently, how to correctly translate and represent knowledge into digital format becomes central to the possibility of accessing and querying digital data, and to ensure that this possibility is not hindered in the future due to technological or conceptual obsolescence. Similarly metadata should provide ways to interact with the data that are not limited to one single perspective or purpose, but rather that allow for multidisciplinary (re-)use of its contents. The workshop presents an experimental approach to the creation of metadata, combining theories and practices from historiography and linguistics in an effort to enrich digital objects with metadata derived from the analysis of both the context and the contents of digital sources.⁶

Durante il workshop si è deciso di sperimentare la realizzazione/prototipazione di un 'archivio corpus partecipato' andando a lavorare sulla metadattazione di documenti storici nati materici e digitalizzati relativi al conflitto jugoslavo del 1991, e documenti nati digitali relativi al conflitto ora in atto in Ucraina. Per questo secondo set di fonti abbiamo utilizzato i documenti prodotti dal Parlamento Europeo.⁷

Si è deciso di lavorare su questi due eventi in quanto si tratta della prima e dell'ultima guerra in Europa dopo la fine della seconda guerra mondiale.

I contenuti delle fonti sono stati individuati attraverso due processi: si è lavorato sia attraverso l'analisi semantica, umana, delle fonti, cercando nei contenuti parole chiave e temi che i gruppi di lavoro ritenessero interessanti per la descrizione; sia sintattica, utilizzando LancsBox⁸ per individuare all'interno del corpus le ricorren-

⁶ Cf. <https://www.summerschooldigitalhumanities.unimore.it/2022-edition/scientific-programme/>.

⁷ Il portale del Parlamento Europeo si mostra come uno strumento di incredibile interesse. Oltre a contenere documenti tradotti in tutte le lingue, ciò a cui si ha accesso è niente di diverso da un archivio istituzionale corrente o in certi casi di deposito. Questo ci dovrebbe far riflettere ulteriormente sui principi dell'accessibilità in epoca digitale e sulla necessità di definire sistemi di ordinamento e metadattazione già nelle fasi correnti e di deposito di un archivio.

⁸ Lancaster University corpus toolbox un software di analisi computazionale (<http://corpora.lancs.ac.uk/lancsbox/>).

ze e le relazioni sintattiche delle parole. Agendo su questo doppio livello si sono ottenuti risultati sia di ordine qualitativo sia di ordine quantitativo.

Questa fase è stata seguita da un lavoro di comparazione e sintesi grazie al quale sono state estrapolate le *keywords* da utilizzare per la creazione di metadati a partire dai contenuti degli oggetti digitali.

Per sviluppare il workshop si è deciso di utilizzare Omeka S, un programma open source di web-publishing per la creazione e la fruizione di archivi digitali (<https://omeka.org/s/>).

La scelta di utilizzare Omeka S è stata presa perché, oltre alla semplicità di utilizzo, questo strumento consente di interagire con le fonti attraverso diversi vocabolari e standard di metadattazione per costruire ontologie.

Nella voce *Vocabularies* il sistema permette di scegliere tra Bibliographic Ontology, Dublin Core, Dublin Core Type e Friends of a Friend. La nostra scelta è ricaduta sull'utilizzo di Dublin Core, che su Omeka S si presenta con uno schema composto da 55 elementi.

Durante il workshop si è deciso di lasciare libera scelta ai frequentanti sia sui documenti da analizzare sia sui campi di metadati da importare, sottolineando però la necessità di utilizzare metadati descrittivi relativi a: contesto archivistico di origine, titolo del documento, anno di produzione, soggetto produttore, oggetto e tipologia di documento. Indicazioni necessarie per una corretta conservazione all'interno di un archivio e per validare l'esistenza, la ricerca e la tracciabilità dell'oggetto digitale.

La parte più interessante è emersa nel momento in cui i partecipanti hanno iniziato a chiedersi come poter utilizzare il campo *Description* nello schema di Dublin Core. Questo elemento consente di utilizzare elementi descrittivi estrapolati dal contenuto del documento stesso e capaci di andare a costruire significazione sull'oggetto digitale così da renderlo un oggetto informativo.

Un oggetto digitale può essere definito oggetto informativo nel momento in cui le informazioni di rappresentazione contengono indicazioni sia di natura sintattica che semantica e svolgono un ruolo analogo a quello che il dizionario della lingua italiana e le regole grammaticali assumono per la comprensione di un testo scritto in italiano (Pigliapoco 2018).

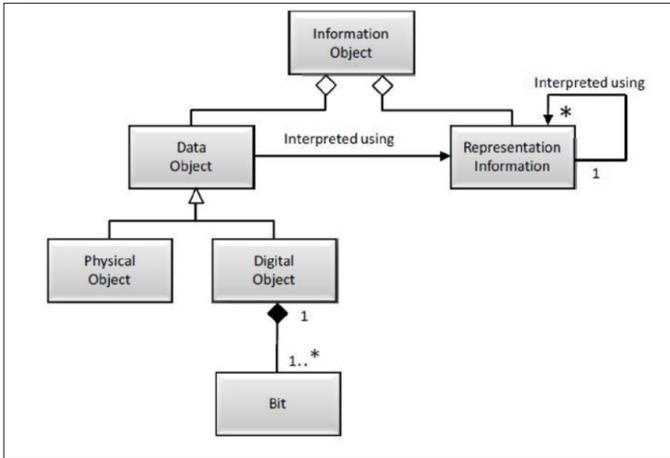


Figura 3 Schema logico di funzionamento per la realizzazione di un oggetto informativo all'interno del modello di riferimento OAIS (Open Archival Information System)

Il modello, che possiamo osservare nell'immagine, è stato sviluppato per il funzionamento di OAIS e muove dai dati per costruire su di essi una struttura concentrica di metadati tramite il meccanismo ricorsivo: i dati richiedono altri dati per poter essere correttamente interpretati; questi a loro volta possono richiedere altri dati interpretativi; e così via, finché non si raggiunge un livello autosufficiente, determinato dalla base di conoscenza della comunità di riferimento. Qui si arresta un'iterazione che altrimenti condurrebbe paradossalmente alla conoscenza del mondo (Michetti 2008, 42).

Uno dei gruppi di lavoro che ha partecipato al workshop ha deciso di metadattare la Risoluzione 721 del Security Council delle Nazioni Unite rilasciata nel novembre del 1991, utilizzando come descrizione anche i concetti di *peace keeping operation* e *cease fire agreement*. Questi termini non sono esplicitamente espressi nel titolo del documento e potrebbero non essere formalmente necessari all'indicizzazione del documento stesso. In seguito all'analisi computazionale e semantica del testo sono emersi come elementi significativi per la comprensione e per la ricerca del documento all'interno di un sistema di archiviazione digitale. La comunità di riferimento ha dunque deciso di utilizzare questi termini come rappresentazione dell'informazione.

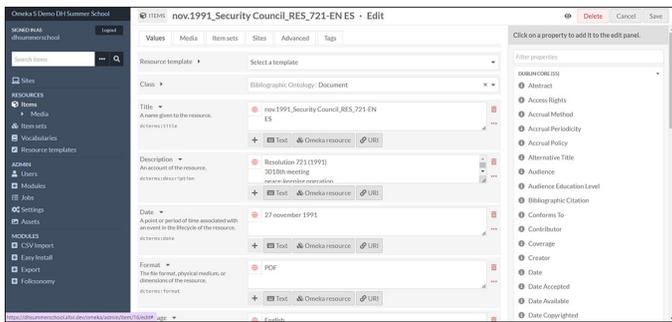


Figura 4 Screenshot del backend dell'istanza in Omeka S creata per la Summer School in Digital Humanities dell'Università degli Studi di Modena e Reggio Emilia. Metadattazione ricorsiva nel campo 'description'

L'utilizzo di questo tipo di metadattazione non è utile solo a dare chiavi di ricerca funzionali per fruitori che non necessariamente conoscono l'oggetto o il titolo dei documenti di cui hanno bisogno, è uno strumento utile anche per sviluppare un'interoperabilità tra documenti (testo, video, audio), attraverso chiavi di lettura tematiche e concettuali.

Se nella barra di ricerca del sistema di Omeka S si scrivono le parole *peace* o *cease fire* il sistema restituirà come risultato tutti i documenti che sono stati metadattati con questi termini. Questo consentirà al fruitore di trovarsi davanti a un'eterogeneità di elementi informativi accomunati da termini o concetti specifici.

In questo caso possiamo vedere che utilizzando il concetto *cease fire* il sistema risponde fornendoci l'accesso a tre documenti differenti caricati e metadattati da due diversi gruppi di lavoro: un verbale previsionale, una risoluzione del Consiglio di Sicurezza e un articolo di giornale.

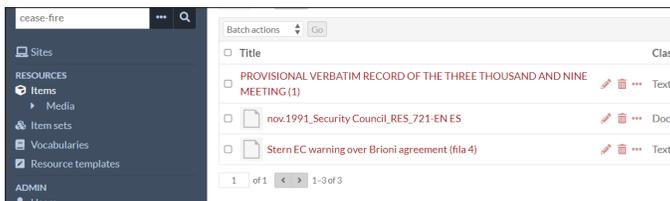


Figura 5 Screenshot del backend dell'istanza in Omeka S creata per la Summer School in Digital Humanities dell'Università degli Studi di Modena e Reggio Emilia. Restituzione degli oggetti informativi in seguito ad una ricerca utilizzando dei concetti come parole chiave

L'utilizzo del concetto di *cease fire* come metadati per descrivere tre documenti differenti consente anche di poter meglio comprendere il concetto di *cease fire* ponendolo in relazione con i contenuti informativi dei diversi documenti.

Il campo *Media* consente di inserire all'interno di Omeka documenti di ogni tipo. Come abbiamo visto, video, testi, immagini o audio, una volta subito il processo di ri-mediazione o se nati digitali, sono tutti composti di bit. Per il sistema digitale sarà possibile accogliere al proprio interno e rendere interoperabili tipologie di documenti molto differenti tra loro e una volta metadati - dopo aver subito un processo di sovrascrittura - ogni documento risponderà alle chiavi di ricerca espresse dal ricercatore o dalla ricercatrice.

Il workshop è stato un primo esperimento molto utile per comprendere meglio il funzionamento dei metadati e i possibili processi di metadazione in relazione a una comunità di riferimento. Senza nessuna pretesa di essere una sperimentazione sufficientemente valida grazie a questo workshop si è comunque potuto testare un processo d'analisi dei documenti, sia in senso qualitativo che quantitativo. Il coinvolgimento di una comunità di riferimento molto eterogenea ci ha mostrato le possibilità e le complessità di una metadazione partecipata durante la quale siamo riusciti a estrapolare diversi significati dalle fonti facendone dunque emergere la complessità informativa. L'interoperabilità delle fonti ci ha mostrato come sia possibile costruire delle narrazioni all'interno di un archivio digitale utilizzando i metadati come elemento portante della 'spiegazione'. In una prospettiva di Public History un processo come quello descritto ha consentito a un pubblico non specialistico di confrontarsi con le fonti primarie e di interrogarsi sui processi di analisi critica tipici del lavoro di storiche e storici. Una volta conclusa la fase di metadazione i frequentanti hanno potuto osservare e sperimentare un nuovo processo di narrazione della storia.

Con la Cooperativa Oz-Officine Zero abbiamo presentato un progetto all'interno dell'European Urban Initiative, che prevede la realizzazione di un archivio digitale partecipato relativo al territorio di Vigne Nuove a Roma. Con la Sovrintendenza ai Beni Archeologici, il III Municipio di Roma e il dipartimento di Architettura e Urbanistica di Roma 3 si è pensato che il processo di metadazione e significazione dal basso di luoghi, documenti e memorie poteva essere il giusto punto di partenza per iniziare una più ampia azione di rigenerazione urbana dell'area. Come ci indica Giannachi (2021) le memorie individuali grazie a processi performativi possono produrre memorie collettive capaci di attivare azioni di innovazione sociale.

6 Conclusioni

In queste pagine si è provato a far emergere il ruolo dei metadata all'interno dei processi di digitalizzazione delle informazioni con un'attenzione specifica riguardo al tema dell'accessibilità e del possibile utilizzo di fonti storiche da parte di fruitori non necessariamente esperti nella ricerca storiografica.

Avvalendoci dei più recenti studi svolti nell'ambito dell'archivistica informatica, delle Digital Humanities, della Public History e delle scienze dell'informazione, attraverso l'analisi di alcuni esempi pratici abbiamo potuto osservare come i processi di digitalizzazione favoriscano un rinnovato rapporto tra fonti e pubblico.

I sistemi informativi e gli archivi digitali si sono mostrati a noi non solo come strumenti di conservazione, ma come potenti strumenti di comunicazione capaci di intervenire all'interno dei serendipitosi processi di ricerca nel Web generalista.

Hilda Girardet (2015) con il suo testo sull'insegnamento della storia con le fonti fa emergere le positive implicazioni di una didattica sviluppata non solo con i manuali, ma anche attraverso le risorse dello storico. L'analisi del laboratorio svolto al Liceo Tacito di Roma in collaborazione con l'Istituto Luce ha messo in evidenza quanto le pratiche di Public History unite a un accesso telepistemico alle fonti digitali consentano di implementare ulteriormente metodologie didattiche, coinvolgendo studenti e studentesse nelle azioni di ricerca d'archivio, di critica delle fonti e di produzione di contenuti storiografici multimediali.

Infine grazie a sistemi logici complessi come quelli che sottendono al funzionamento di OAIS, alla flessibilità di standard di metadatazione come DCMI e alla facilità di utilizzo di un software come Omeka S, abbiamo potuto osservare l'impatto che può avere la costruzione di un oggetto informativo in relazione a una comunità di riferimento coinvolta in una metadatazione partecipata.

I tre casi studio riportati ci aiutano anche a osservare le possibili insidie legate ai processi di digitalizzazione dell'informazione. L'insidia maggiore credo la si possa riscontrare nel rischio che i processi di digitalizzazione rendano l'informazione eccessivamente liquida e dunque svincolata dai corretti contesti informativi di produzione e conservazione. Il reperimento di oggetti digitali all'interno di contesti informativi differenti da quelli entro cui sono stati originariamente generati rischia di mostrarsi come una importante limitazione nei processi di ricerca e analisi delle fonti. Anche per lavorare in modo corretto a una metadatazione collaborativa è importante condividere con la comunità di riferimento una riflessione attenta sui quadri contestuali informativi di origine e destinazione del documento, onde evitare che l'analisi interpretativa si sostituisca completamente alle relazioni di contesto del documento.

Pur senza soffermarsi eccessivamente sulle analisi teoriche e i risvolti teoretici dei processi di digitalizzazione all'interno della storia come disciplina, tutti e tre i casi credo ci mostrino come l'idea di metacatalogazione possa coniugare evoluzioni tecniche e tecnologiche con evidenti persistenze metodologiche, ponendosi l'obiettivo di concettualizzare la conoscenza ancora prima di classificarla per difonderla (Valacchi 2022).

Bibliografia

- Baeza-Yates, R.; Ribeiro-Neto, B. (1999). *Modern Infomation Retrieval*. Boston: Addison Wesley.
- Barbanti, I. (2005). «La codifica informatica dei dati d'archivio: EAD ed EAC», *Regesta.com. New Media and Historical heritage*, 2(3), 10-25. <https://www.regesta.com/2005/12/31/La-codifica-informatica-dei-dati-di-archivio-ead-ed-eac/>.
- Barbanti, I. (2021). «Il motore di una evoluzione possibile. L'informatica spinge gli archivi». Pezzica, L.; Valacchi, F. (a cura di), *Dimensioni archivistiche. Una piramide rovesciata*. Milano: Bibliografica, 103-22.
- Bolter, J.D.; Grusin, R. (2002). *Remediation. Competizione e integrazione tra media vecchi e nuovi*. Milano: Guerini e Associati.
- Brin, S.; Page, L. (1998). "The Anatomy of a Large-Scale Hypertextual Web Search Engine". *Proceedings of the 7th International World Wide Web Conference (WWW7)*. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X).
- Cacciani, P. (2018). *Narrare la storia. Ricerca e creatività. A cura di Manfredi Scanagatta con gli studenti della VB Liceo classico Cornelio Tacito di Roma*. <https://shorturl.at/fpv08>.
- Carr, E. (1966). *Sei lezioni sulla storia*. Torino: Einaudi.
- Ciandrini, P. (2018). «Lo zen e l'arte della manutenzione archivistica. Ibridi e resilienti». *JLIS.it*, 9(3), 171-99. <https://doi.org/10.4403/jlis.it-12485>.
- Cinelli, M. et al. (2020). «The COVID-19 Social Media Infodemic». *Sci Rep.*, 10(1). <http://doi.org/10.1038/s41598-020-73510-5>.
- De Felice, R. (1998). *L'archivio contemporaneo. Titolare e classificazione sistematica di competenza nei moderni archivi*. Roma: Carocci.
- Di Fazio, S. (2018). «Le ontologie». Di Marcantonio, G.; Valacchi, F. (a cura di), *Descrivere gli archivi al tempo di RIC-CM*. Macerata: EUM, 81-116.
- Durante, M. (2019). *Potere computazionale. L'impatto delle ICT su diritto, società, sapere*. Milano: Meltemi.
- Feliciati, P. (2009). *Gestione e conservazione di dati e metadati per gli archivi: quali standard?*. Macerata: Unimc, Pubblicazioni Aperte Digitali. <http://eprints.rclis.org/13657/1/atti-testoFeliciati09.pdf>.
- Ferraris, M. (2019). *Documentalità. Perché è necessario lasciare tracce*. Roma-Bari: Laterza.
- Ferraris, M.; Pains, G. (2018). *Scienza nuova. Ontologia della trasformazione digitale*. Torino: Rosenberg e Sellier.
- Floridi, L. (2017). *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*. Milano: Raffaello Cortina Editore.
- Floridi, L. (2020). *Pensare l'infosfera. La filosofia come design concettuale*. Milano: Raffaello Cortina Editore.

- Giannachi, G. (2021). *Archiviare tutto. Una mappatura del quotidiano*. Roma: Treccani.
- Girardet, H. (2015). *Vedere, toccare, ascoltare. L'insegnamento della storia attraverso le fonti*. Roma: Carocci.
- Guercio, M. (2000). *Archivistica informatica. I documenti in ambiente digitale*. Roma: Carocci.
- Mantovani, G.; Riva, G. (2001). «Building a Bridge Between Different Scientific Communities: On Sheridan's Eclectic Ontology of Presence». *Presence: Teleoperators and Virtual Environments*, 10(5), 537-43. <https://doi.org/10.1162/105474601753132704>.
- Michetti, G. (2008). «Il modello OAI». *Digitalia*, 3(1), 32-49. <https://digitalia.cultura.gov.it/article/view/441>.
- Montani, P. (2010). *L'immaginazione intermediale. Perlustrare, rifigurare, testimoniare il mondo visibile*. Roma-Bari: Laterza.
- Noiret, S. (2018). *Citizen History. Storia partecipata*. <https://www.raiplay.it/video/2018/10/339-punt20-5e4fabeb-88fa-4fb0-8cba-533f98d623c0.html>.
- Paci, D. (2021). «Conoscere è partecipare: digital public history, wiki e citizen humanities». *Umanistica Digitale*, 10, 235-49. <https://umanisticadigitale.unibo.it/article/view/12555/13085>.
- Pezzica, L. (2020). *L'archivio liberato. Guida teorico-pratica ai fondi storici del Novecento*. Milano: Bibliografica.
- Pigliapoco, S. (2018). *Progetto archivio digitale. Metodologia, sistemi, professionalità*. Lucca: Civita Editoriale.
- Rossi, A. (2021). «La palla di vetro. Immaginare la tutela e la comunicazione archivistica diffusa». Pezzica L.; Valacchi F. (a cura di), *Dimensioni archivistiche. Una piramide rovesciata*. Milano: Bibliografica, 37-56.
- Scanagatta, M. (2015). «Public historian, ricerca e azione creativa». Bertucelli, L.; Bertella Farnetti, P.; Botti, A. (a cura di), *Public History. Discussioni e Pratiche*. Milano; Udine: Mimesis, 315-31.
- Scanagatta, M. (2019). «Storia e narrativizzazione digitale». Paci, D. (a cura di), *La storia in digitale. Teorie e metodologie*. Milano: Unicopli, 320-2.
- Spengler, O. (2012). *Il tramonto dell'Occidente*. Milano: Longanesi.
- Topolski, J. (1997). *Narrare la storia. Nuovi principi di metodologia storica*. Milano: Bruno Mondadori.
- Valacchi, F. (2020). *Gli archivi tra storia e uso futuro*. Milano: Bibliografica.
- Valacchi, F. (2021). «La memoria dell'acqua». Pezzica, L.; Valacchi, F. (a cura di), *Dimensioni archivistiche. Una piramide rovesciata*. Milano: Bibliografica, 173-87.
- Valacchi, F. (2022). «Le parti e il tutto: integrare la conoscenza». *Jlis.it*, 13(3), 6-15. <http://doi.org/10.36253/jlis.it-477>.
- Vitali, S. (2005). «Ordine e caos: Google e l'arte della memoria». Spaziante, A. (a cura di), *Il futuro della memoria: la trasmissione del patrimonio culturale nell'era digitale*. Torino: CSI Piemonte, 71-96β.
- Vitali, S. (2017). «La ricerca archivistica sul Web». Minuti, R. (a cura di), *Il Web e gli studi storici*. Roma: Carocci, 61-96.
- Weston, P.G.; Sardo, L. (2017). *Metadati*. Milano: Associazione Italiana Biblioteche.

Riferimenti aggiuntivi

- DCMI – Dublin Core Metadata Initiative, <https://www.dublincore.org/>.
- EAC-CPF – Encoded Archival Context-Corporate Bodies, Persons and Families, <https://icar.cultura.gov.it/standard/standard-internazionali/eac-cpf>.
- EAD – Encoded Archival Description, <https://www.loc.gov/ead/>.
- ISAAR-CPF, <https://icar.cultura.gov.it/standard/standard-internazionali/isaar-cpf>.
- ISAD (G), <https://icar.cultura.gov.it/standard/standard-internazionali/isad-g>.
- Lancsbox, <http://corpora.lancs.ac.uk/lancsbox/>.
- OAIS – Open Archival Information System, <http://www.oais.info/>.
- Omeka S, <https://omeka.org/s/>.
- SAN – Sistema Archivistico Nazionale, <http://san.beniculturali.it/web/san/dati-san-lod>.
- SIAS – Sistema Informativo per gli Archivi di Stato, <https://sias.archivi.beniculturali.it/cgi-bin/pagina.pl>.
- SIUSA – Sistema Informativo Unificato per le Sovrintendenze Archivistiche, <https://siusa.archivi.beniculturali.it/>.

La ‘Galassia MQDQ’: un concetto di filologia tradizionale, digitale, sostenibile

Martina Venuti, Angelo Mario Del Grosso, Federico Boschetti,
Luigi Tassarolo, Alessia Prontera, Dylan Bovet,
Gianmario Cattaneo, Valeria Melis

Abstract The investigation of intertextuality within a corpus of Latin poetry is the main objective of the research functions of the so-called ‘Musisque Deoque Galaxy’: formular recurrences, and lexical and metric-verbal co-occurrences draw a dense network of relationships between texts, where poetic memory presents itself in various modalities. This contribution aims to analyse the latest developments in the “Musisque Deoque Galaxy” both from the point of view of the tools created to enable new textual acquisitions and new analyses, and from the technological point of view, with reference to the work carried out to ensure a long-term sustainability.

Keywords Digital textual corpora. Latin poetry. MQDQ Galaxy. Carmina Latina Epigraphica. Domain Specific Languages (DSL). Sustainability.

Sommario 1 Introduzione alla nuova Galassia MQDQ. – 2 I testimoni manoscritti: revisione e ristrutturazione. – 3 Wiki-MQDQ e le nuove acquisizioni tra i testi letterari. – 4 I *Carmina Latina Epigraphica*. – 5 Ridisegnare l’architettura della Galassia MQDQ: workflow collaborativo e sostenibile in MQDQ.

Benché il lavoro sia il frutto di una stretta collaborazione tra i co-autori, si specifica che responsabile generale è Martina Venuti (primo autore, Università Ca’ Foscari Venezia), cui si deve la stesura dei §§ 1 e 3; Alessia Prontera (Università Ca’ Foscari Venezia) è autrice dei §§ 2 e 3.1, 4.1-3; Valeria Melis (Università Ca’ Foscari Venezia) del § 3.2; Gianmario Cattaneo (Università del Piemonte Orientale) del § 3.3; Dylan Bovet (UNIL, Université de Lausanne) dell’introduzione al § 4 e dei §§ 4.4-5; Angelo Mario Del Grosso (CNR-ILC) dei §§ 5.1-9 (il § 5.7 con la consulenza di Luigi Tassarolo, Università Ca’ Foscari Venezia) e Federico Boschetti (CNR-ILC) del § 5.10. Si ringrazia Paolo Mastandrea, insieme ai revisori anonimi, per i preziosi consigli.



Peer review

Submitted 2023-03-31
Accepted 2023-06-15
Published 2023-08-02

Open access

© 2023 Venuti et al. | © 4.0



Citation Venuti, M. et al. (2023). “La ‘Galassia MQDQ’: un concetto di filologia tradizionale, digitale, sostenibile”. *magazén*, 4(1), 71-120.

1 Introduzione alla nuova Galassia MQDQ

Gli archivi che compongono quella che nel tempo è andata configurandosi come la 'Galassia *Musisque Deoque*' sono ormai annosi: dopo alcune esperienze di pubblicazione in CD-Rom di testi antichi e medioevali, nel 1999 entravano in rete i pionieristici *Poeti d'Italia in lingua latina* (d'ora in avanti *Poeti d'Italia*),¹ nati con lo scopo di raccogliere e digitalizzare una messe imponente di versificazione in latino prodotta da autori di area geografica e culturale italiana fra pre-Umanesimo e pieno-Rinascimento, in larga parte negletti dagli studiosi. Oggi la biblioteca conta 316 autori con 840 opere, per molte delle quali costituisce l'unica fonte digitalizzata, ricercabile e *open access*.² Pochi anni dopo, nel 2005, un altro grande progetto collaborativo guidato da Paolo Mastandrea dava origine a *Musisque Deoque* (d'ora in avanti *MQDQ*),³ un archivio che comprende l'intero corpus della poesia latina antica (inclusi i *Carmina Latina Epigraphica*)⁴ fino all'epoca precarolingia (276 autori, 642 opere),⁵ corredato da una serie di strumenti specificamente destinati all'indagine scientifica: apparati critici ricercabili e selettivi (ad oggi, circa la metà dei testi dell'archivio è dotata di apparato critico) in funzione di un'idea di intertestualità da applicarsi anche alle varianti di tradizione, onde disegnare una storia stratificata dei testi; funzioni di ricerca avanzate (dalle semplici ricerche di tipo lessicale a quelle sulle co-occorrenze metrico-verbali o foniche); strumenti di analisi metrica come *Pedecerto*,⁶ che permette la scansione automatica dei versi dattilici contenuti nel corpus (circa 250.000).⁷ A questi archivi principali si aggiungono alcuni corpora minori, come quello di *Hellenica*,⁸ una per ora piccola raccolta di testi poetici greci, pensata come prototipo di un archivio *open access*, auspicabilmente da ampliare in futuro.

1 <https://www.poetiditalia.it/public/>.

2 Il sito è frutto di due finanziamenti PRIN (1991 e 2001), coordinati da Paolo Mastandrea, con progetto informatico di Luigi Tessarolo. All'archivio principale si è aggiunta successivamente la piccola raccolta dei *Carmina Libraria*, accessibile da una voce specifica dell'indice: cf. Mastandrea 2004.

3 <https://www.mqdq.it/public/>.

4 <https://www.mqdq.it/public/ce/presentazione>.

5 Il progetto è frutto di due finanziamenti PRIN (2005 e 2007), che hanno visto la collaborazione, negli anni, di diverse Università italiane e di decine di studiosi. Per una breve storia dei progetti della Galassia *MQDQ* e dei diversi partner coinvolti si vedano, oltre a Mastandrea 2009; Mastandrea, Spinazzé 2011; Boschetti, Del Grosso, Spinazzé 2021.

6 <https://www.pedecerto.eu/public/>.

7 Lo strumento nasce nel 2011 grazie a un finanziamento FIRB dell'Università di Udine (P.I. Emanuela Colombi; progetto scientifico di Luca Mondin; progetto informatico di Luigi Tessarolo).

8 <https://mizar.unive.it/hellenica/public/indici/autori>.

Oggi la comunità scientifica internazionale, con particolare riguardo agli studiosi di letteratura latina, utilizza gli strumenti offerti dalla Galassia MQDQ come base di lavoro quotidiano.⁹

Alla luce dell'età e della consistenza dell'archivio, esso ha di recente dovuto affrontare una serie di sfide, sempre più dure, riguardanti il futuro: quali accorgimenti assumere per rendere sostenibile nel tempo il sistema? Quali soluzioni adottare, sia dal punto di vista dello sviluppo e dell'ampliamento dei corpora e delle funzioni ad essi collegate, sia - per certi versi soprattutto - dal punto di vista tecnologico? Come contrastare l'obsolescenza e garantire la necessaria manutenzione? Comuni a tutti i progetti DH, i problemi sono di ordine economico e tecnico: da una parte, la mancanza di fonti di finanziamento durature, costanti, che permettano la gestione efficiente di un organismo 'vivo' come una biblioteca digitale - peraltro nata grazie a fondi pubblici e dunque da preservare come bene comune; dall'altra, la necessità di migrare verso architetture di sistema più 'moderne', sia quanto a standard formali sia quanto a sostenibilità dei processi di aggiornamento e più in generale dei flussi di lavoro.¹⁰ La cornice del VeDPH¹¹ - sorto nel 2018 e attivo dal 2019 come esito del Progetto di Eccellenza del Dipartimento di Studi Umanistici dell'Università Ca' Foscari Venezia -¹² e la collaborazione con l'Istituto di Linguistica Computazionale A. Zampolli di Pisa¹³ hanno reso possibile affrontare alcune di queste sfide, sicché la Galassia MQDQ ci sembra divenuta un interessante caso di studio. In questa sede ci proponiamo di presentare le soluzioni adottate e gli avanzamenti

9 Impossibile dare conto di tutti i contributi per i quali gli archivi di *Musisque Deoque* e i suoi strumenti sono stati fonti imprescindibili: si rimanda alle numerose pubblicazioni che hanno illustrato le diverse fasi del progetto e della riflessione filologico-critica ad esso connessa: Zurli, Mastandrea 2009; Manca 2010; Mastandrea, Spinazzè 2011; Mastandrea 2012; Mastandrea 2017.

10 Il tema della sostenibilità per i progetti DH è stato largamente dibattuto in letteratura. Tra gli altri si vedano i contributi di Barats, Schafer, Fickers 2020; Drucker 2021; Tucker 2022; Ciula, Smithies 2023.

11 <https://www.unive.it/pag/39287>.

12 Il VeDPH, diretto da Franz Fischer, è un centro fortemente interdisciplinare che si basa su quattro pilastri: Digital Textual Scholarship, Digital Public History, Digital Art History e Digital Archaeology. Come espresso nella sua denominazione, il VeDPH affianca all'aspetto digitale (metodi, risorse e strumenti per l'acquisizione e la fruizione di oggetti digitali o digitalizzati di interesse culturale) l'aspetto pubblico (coinvolgimento del pubblico o *public engagement*, partecipazione di soggetti non accademici alla produzione di contenuti di supporto alla ricerca o *citizen science*).

13 La collaborazione è stata resa possibile da una convenzione tra il Dipartimento di Studi Umanistici dell'Università Ca' Foscari Venezia (responsabile Martina Venuti) e il CNR-ILC di Pisa (responsabili la direttrice Simonetta Montemagni e Monica Monachini, coordinatrice di CLARIN-IT e membro del comitato di gestione). È stata costituita una specifica Unità di Ricerca presso Terzi (URT), di cui è coordinatore Federico Boschetti, focalizzata su temi e progetti di filologia digitale e computazionale, da affrontare in stretta collaborazione con il VeDPH.

nei due campi di intervento, quello più propriamente filologico-letterario e quello prettamente tecnologico: saranno così mostrate da una parte alcune nuove acquisizioni nei contenuti degli archivi (§§ 2-4), dall'altra le modifiche all'architettura di MQDQ, che mira a costituire - e in parte già costituisce - un modello di riferimento (§ 5).¹⁴

Il primo importante passo in questo senso è stato quello di sviluppare un complesso di funzionalità per l'incremento e insieme la manutenzione di tutti gli archivi. Dopo alcune sperimentazioni portate avanti con *Cadmus*, editor modulare di contenuti testuali e multimediali messo a punto da Daniele Fusi e che ad oggi serve diversi importanti progetti DH,¹⁵ il nuovo back-end è stato sviluppato da Luigi Tassarolo entro il progetto *MQDQ Galaxy* coordinato da Martina Venuti (www.unive.it/mqdqgalaxy). Attraverso una modulazione nelle autorizzazioni concesse ai diversi tipi di utente (utente 'super-amministratore', utente 'amministratore', utente 'editore', utente 'operatore' e così via) i collaboratori operano entro tale piattaforma a livelli diversi e con mansioni diverse: gli utenti editori, ad esempio, sono autorizzati solo ad aggiungere dati (un nuovo apparato critico, la descrizione di un manoscritto), generando una versione in bozza che deve poi essere approvata dal comitato scientifico. Solo a un gruppo ristretto di persone, gli amministratori, è invece possibile compiere tutte le operazioni, dall'inserimento di nuovi testi all'aggiornamento generale degli archivi. Una 'guida' per i collaboratori e una descrizione di tutti gli strumenti di lavoro e delle procedure relative sono rese disponibili nella pagina di accesso [fig. 1].

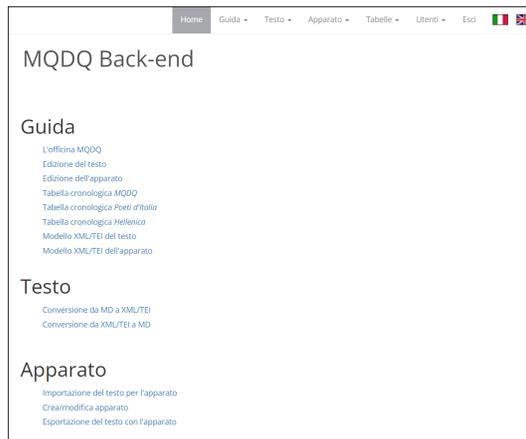


Figura 1
Pagina di accesso
del back-end della Galassia
MQDQ: www.mqdqgalaxy.it

14 Il metodo di lavoro e parte degli strumenti della Galassia MQDQ sono stati recentemente adottati anche dal progetto ERC *Florilegia Syriaca. The Intercultural Dissemination of Greek Christian Thought in Syriac and Arabic in the First Millennium CE* (Grant Agreement ID 758732-FLOS).

15 Una recente presentazione del sistema qui: <https://youtu.be/LYykjz26TCg>.

Il progetto *MQDQ Galaxy* ha dunque reso disponibile uno strumento collaborativo facile da usare ed efficiente con cui procedere nella direzione dell'aggiornamento e della futura sostenibilità.

Un primo ambito di intervento, da un punto di vista filologico, ha riguardato la descrizione dei manoscritti che compaiono come fonti per gli apparati critici: come è noto, la presenza degli apparati è elemento distintivo dell'archivio di *MQDQ* rispetto ad altre risorse digitali che offrono corpora di testi latini. Data la natura collaborativa e l'estensione diacronica del progetto, quasi trentennale, tale descrizione presentava un notevole grado di disomogeneità, essendo frutto di volta in volta della scelta autonoma dei singoli studiosi, che si erano rifatti a linee guida di catalogazione non sempre armonizzate tra loro: va ricordato che ciascun testo attualmente dotato di apparato è stato trattato da un editore esperto, che, sulla base della propria competenza e del proprio giudizio, ha scelto di volta in volta quali testimoni manoscritti e/o a stampa utilizzare per il lavoro critico, di cui risulta pubblicamente responsabile. L'acquisizione di tali testimoni è avvenuta tramite inserimento e descrizione manuali. Si è proceduto dunque a creare una 'grammatica normalizzata' per la compilazione, che si è applicata a tutti i testimoni presenti nell'archivio e che verrà automaticamente applicata in futuro.

2 I testimoni manoscritti: revisione e ristrutturazione

2.1 L'organizzazione nel back-end

Il numero complessivo dei mss citati negli archivi *MQDQ* è attualmente di 4845, comprese le *secundae manus* e i testimoni perduti. Come accennato, il processo di omogeneizzazione delle fonti è avvenuto attraverso l'adozione di una 'descrizione strutturata' di tutti i manoscritti, ovvero la scomposizione della descrizione in un numero progettato di segmenti obbligati, il cui valore è scelto entro una limitata gamma di elementi predefiniti, che in questo caso sono voci di tabelle condivise. Sono stati esclusi da questa operazione i testimoni perduti e le seconde mani.

Si elencano qui di seguito i campi della strutturazione, individuati dal comitato scientifico, sulla base degli standard correnti di catalogazione e secondo un principio di bilanciamento tra quantità di informazioni e utilità per l'utente:

- **Città (Settlement)**: compare sempre nella lingua nazionale.
- **Biblioteca (Repository)**: citata nella lingua nazionale secondo la denominazione attuale.
- **Fondo (Collection)**: indicato solo laddove sia prevista dalla singola organizzazione.

- **Segnatura (Shelfmark):** quella attuale, affiancata da quella antica solo se previsto dal catalogo bibliografico.
- **Datazione (Date):** espressa tramite il secolo in numero romano preceduto dall’abbreviazione *saec.*
- **Nome (Name):** è l’elemento maggiormente soggetto a variabilità poiché rispecchia lo stato precedente all’intervento di revisione e dove perlopiù si trova la denominazione filologica in latino.
- **Link (URL):** la stringa contiene il permalink della scheda catalografica digitale del singolo manoscritto offerta dai *websites* ufficiali delle maggiori biblioteche europee.

2.2 La consultazione nel front-end

La consultazione della descrizione dei testimoni da parte degli utenti del front-end può avvenire in due modi: la lista completa è visualizzabile, per ciascun testo tra quelli dotati di apparato, nella relativa sezione ‘Testimoni’, in alto a destra (es. *Verg. ecl. 1*).¹⁶ Qui, cliccando sulla descrizione dei singoli manoscritti, si apre un menù a tendina che contiene, laddove possibile, il permalink alla biblioteca di appartenenza, con relativa scheda catalografica, e quello alla riproduzione digitale del codice. La strutturazione ha poi reso possibile la visualizzazione dell’informazione forse più rilevante, vale a dire i testi del corpus di *MQDQ* per i quali quel manoscritto è stato usato come testimone (‘Testi’). Per chiarezza si veda l’esempio del *Vergilius Romanus*, il *Vaticanus Latinus 3867*,¹⁷ celebre codice tardoantico di Virgilio, portatore anche degli *argumenta Aeneidos* pseudo-ovidiani editi nell’*Anthologia Latina* [fig. 2].

Figura 2
Il *Vaticanus Latinus 3867*
nei ‘Testimoni’



I singoli codici sono visualizzabili poi, ovviamente, in corrispondenza delle voci di apparato in cui compaiono: al passaggio del cursore

¹⁶ <https://www.mqdq.it/public/testo/testo/codice/VERG%7ceclo%7c001>.

¹⁷ https://digi.vatlib.it/view/MSS_Vat.lat.3867.

sulla sigla del ms si aprirà un *pop-up* che contiene sia la descrizione strutturata (in rosso) sia la denominazione latina corrispondente al campo *Name* della maschera contenuta nel back-end (in nero)¹⁸ [fig. 3].

The screenshot shows the MQDQ interface. At the top, there's a navigation bar with 'Info', 'Testimoni', 'Metri', and 'Scansione metrica'. Below that, the text 'Vergilius Aeneis 6' is visible. The main content area displays a Latin text snippet with line numbers 225, 230, and 235. A pop-up window is overlaid on the text, showing the following information:

231
uiros F M P² e y
 domos P a.c. **R a b o r**
 Città del Vaticano, Biblioteca Apostolica
 Vaticana, Vat. lat. 3867 «Vergilius
 Romanus» (saec. V-VI)
 codex Vaticanus 3867 Vergilius

Figura 3 Doppia denominazione del *Vaticanus Latinus* 3867

3 Wiki-MQDQ e le nuove acquisizioni tra i testi letterari

Un secondo intervento di particolare urgenza è stato quello relativo al rinnovamento della piattaforma di inserimento dei dati nell'archivio. Si era infatti reso necessario sostituire il vecchio procedimento - ormai obsoleto perché basato sul download in locale di un software proprietario a uso interno del gruppo di lavoro, predisposto ad hoc dal Luigi Tassarolo e non rilasciato esternamente - con uno strumento, moderno e *web-based*. Scopo primario era agevolare il lavoro degli editori, filologi puri spesso digiuni di competenze informatiche avanzate, e in generale degli operatori, di tutti i livelli.¹⁹ Il

¹⁸ Per la doppia visualizzazione della denominazione delle fonti in MQDQ, cf. Spinazzé 2011.

¹⁹ Un punto a parte, ma di grande rilevanza e che meriterebbe un approfondimento entro la discussione sulla sostenibilità del progetto, riguarda la composizione del gruppo di lavoro di MQDQ Galaxy. Come già accennato, il progetto ha carattere collaborativo e diffuso e gli editori dei testi sono rappresentati indicativamente da due diverse tipologie di collaboratori: studiosi affermati e già strutturati nel sistema universitario, italiano o straniero, che si propongono per la cura dell'edizione di un testo di loro specifico interesse, oppure collaboratori selezionati tramite bando di concorso (ad esempio destinatari di assegni di ricerca o di collaborazioni occasionali) cui viene affidato un incarico preciso (è il caso, ad esempio, dell'attuale lavoro in corso su Ausonio). Numerosi poi sono i contributi volontari di quanti collaborano per interesse di ricerca (è il caso di Dylan Bovet: vedi *infra*, il § 4 sui CLE), per impegno formativo (ad esempio, gli studenti magistrali impegnati nei *workshops* organizzati entro l'offerta didattica di Ca' Foscari dal gruppo di lavoro MQDQ) o per spirito di servizio nei

principale problema da affrontare riguardava la specificità dei dati da inserire. Ciascun testo poetico latino ha infatti naturalmente la propria storia e la propria tradizione; gli editori devono dunque fronteggiare di volta in volta problematiche differenti, che spaziano dalla scelta dei testimoni di un determinato testo alle criticità consuete per qualsiasi testo antico (lacune, varianti, interpolazioni, ecc.). Nella filosofia del progetto MQDQ, l’edizione digitale proposta – è bene ricordarlo – non replica un’edizione cartacea, ma vuole restituire all’utente la fluidità del testo sul piano diacronico, selezionando le informazioni e le varianti non tanto per costituire il testo ‘migliore’ (e infatti spesso le congetture sono escluse dagli apparati di MQDQ), quanto per restituire quelle forme che, oggi scartate o non più accettate, in passato possono aver avuto fortuna e generato tradizione, spiegando così fenomeni di memoria intertestuale. Per supportare gli editori nel lavoro di selezione e di successiva resa dei dati è stata creata la nuova piattaforma *Wiki-MQDQ*, che ha permesso di lavorare direttamente sugli apparati critici.

3.1 Columella, *De re rustica* 10: la prima sperimentazione di *Wiki-MQDQ*

Alla piattaforma si accede inserendo le proprie credenziali (username e password, fornite dall’amministratore di sistema) nel back-end (www.mqdqgalaxy.it). Una volta effettuato l’accesso, per iniziare a operare sul testo è sufficiente aprire il menù a tendina in corrispondenza della voce *Apparatus* [fig. 4].



Figura 4 Menù a tendina della voce *Apparatus*

confronti della comunità scientifica (ad esempio, Marco Sartor, neo-dottore di ricerca presso l’Università degli Studi di Parma, che contribuisce regolarmente a *Poeti d’Italia*). Infine, l’apporto più consistente è ovviamente fornito dal comitato scientifico e dal comitato tecnico, composto dal team di Ca’ Foscari (Paolo Mastandrea, Luca Mondin, Luigi Tassarolo, Martina Venuti) e dell’Unità di ricerca CNR-ILC (Federico Boschetti, Angelo Mario Del Grosso).

Giacché l’intero corpus della poesia latina antica è stato acquisito nelle prime fasi del progetto *MQDQ*, tutti i testi sono già presenti nell’archivio: è dunque sufficiente digitare la voce *Create/Edit* per poter selezionare l’autore e l’opera su cui lavorare [fig. 5].

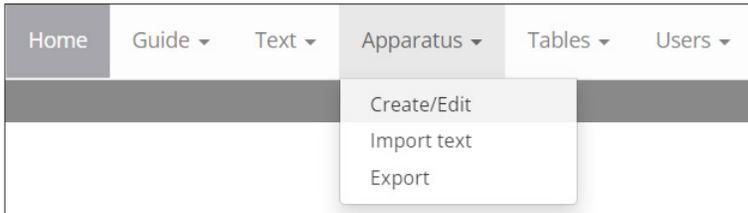


Figura 5 Opzione *Create/Edit*

Nella nuova schermata la selezione di autore (*Author*) e opera (*Work*) avviene in ordine alfabetico e, qualora l’opera sia suddivisa in sezioni (*Sections*), che possono essere libri (per esempio *l’Eneide*), carmi (nel caso delle *Eclogae* di Nemesiano) o epigrammi (per gli *Epigrammata* di Marziale), si dovrà procedere scegliendo, dai relativi menù a tendina, i tre campi in progressione [fig. 6].

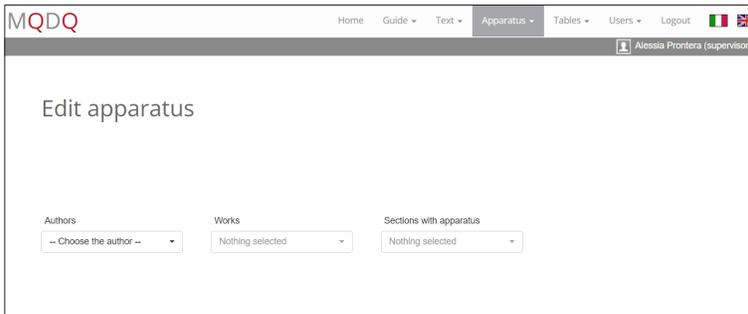


Figura 6 Selezione autore, opera, sezione

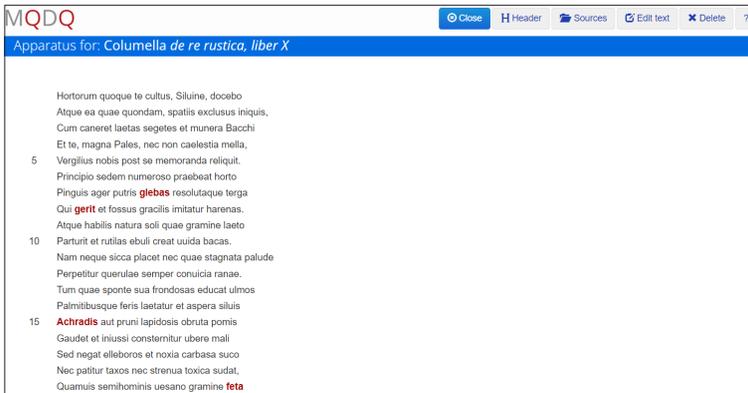
Andrà per completezza rilevato che, nel caso in cui l’editore ritenesse necessari interventi consistenti sul testo disponibile attualmente nell’archivio come testo base (per esempio, una modifica nell’ordine dei versi o una diversa suddivisione delle sezioni di un componente rispetto all’edizione acquisita nel corpus), dovrà operare preventivamente su un file di testo e successivamente importarlo nel back-end tramite il comando *Import text* [fig. 7]. Il file da modificare si trova nel repository privato del progetto sulla piattaforma *github* (vedi *infra*) e può essere reso disponibile all’operatore o per intervento dell’amministratore o per diretta condivisione del repository.

Il file è in formato TEI, ma il back-end di *MQDQ Galaxy* fornisce delle *utilities* che trasformano il file XML nel formato MD, molto più semplice e intuitivo dell'XML, e viceversa.



Figura 7 Pagina di *Import Text*

Una volta verificato il testo base dell'opera, si procederà con l'inserimento dell'apparato digitale, a cominciare da una serie di dati: l'edizione di riferimento utilizzata come base di collazione (normalmente quella considerata più autorevole o più diffusa), il nome del curatore dell'edizione digitale (cioè di colui che è responsabile dell'edizione *MQDQ*) e l'indicazione della forma metrica del testo. Questi dati andranno inseriti nello spazio *Header*; nel campo *Sources* compariranno i testimoni, manoscritti e a stampa (i codici con *sigla* in grassetto e le edizioni con abbreviazione in corsivo); l'ultimo comando (*Edit text*) permette, direttamente dall'interfaccia *Wiki-MQDQ*, di apportare modifiche al testo, operando su singole lezioni o sull'interpunzione [fig. 8].

Figura 8 Columella nel *back-end*

Per il testo di Columella 10, l'edizione di riferimento è quella di Boldrer 1996, mentre responsabile dell'edizione MQDQ è Alessia Prontera. L'edizione MQDQ, secondo le linee-guida del progetto, non segnala tutte le varianti grafiche, poco utili per la ricerca intertestuale, ma indica ad esempio le lezioni corrotte nella tradizione. È il caso di alcuni termini tecnici, poco poetici, come le *trichilae* (pergolati dedicati alla coltivazione del cetriolo) al v. 378: *tum modo dependens trichilis, modo more chelydri*. Laddove i mss *Petropolitanus* F.v.1 e *Ambrosianus* L 85 sup. riportano la lezione *triplicis* e i mss *recentiores* la lezione *richili*, è stato solo l'intervento congetturale di Gesner (1795, 532) a restituire la forma corretta.

Un esempio interessante riguarda inoltre il passo in cui Columella descrive la statuetta lignea di Priapo:

neu tibi Daedaliae quaerantur munera dextrae
 nec Polyclitea nec Phradmonis aut Ageladae 30
 arte laboretur, sed truncum forte dolatum
 arboris antiquae numen uenerare **Priapi**
 terribilis membri, medio qui semper in horto
 inguinibus puero, praedoni falce minetur.
 (*De re rustica* 10.29-34)

32 Priapi **SAR^p** | ithyphalli *in marg.* **k²t¹** triphalli **qsz**

né da te si ricerchino opere degne della mano di Dedalo,
 e non di Policleteo o di Fradmone o di Agelade
 l'arte si impieghi ma, in forma di tronco sgrossato alla meglio di
 un albero vecchio, tu venera il nume di Priapo
 dal terribile membro, che sempre al centro dell'orto

minacci con il fallo il ragazzo, con la falce il ladrone.
(Trad. Boldrer 1996, 59)

Il genitivo al v. 32 presenta nella tradizione la variante *ithyphalli* che si trova in margine a **k**² = *Laurentianus plut.* 53.24 (*saec.* XV) e a **t** = *Bononiensis* 2523 (*saec.* XV), mentre *triphalli* compare a testo in altri tre testimoni (**q** = *Laurentianus plut.* 91 inf. 6, *saec.* XV; **s** = *Laurentianus Strozzi* 69, *saec.* XV; **z** = *Vindobonensis* 33, *saec.* XV). Queste varianti sono considerate da Boldrer 1996, 131 glosse marginali al testo, dunque da scartare. In MQDQ tali varianti sono invece segnalate per l'intrinseco valore storico e culturale che le rendono 'varianti significative' (cf. Manca 2010): *triphallus* è un epiteto di Priapo molto raro, che si ritrova nel *Priapeum* 83.9 Bücheler – *quid hoc novi est dell'Appendix Vergiliana (placet, Triphalle, qui sub arboris coma)* – dove costituisce un'emendazione, proposta da Scaligero nella sua edizione del 1573, alla lezione *Tripalle*, dalla grafia incerta, contenuta in un testimone oggi perduto, appartenuto al Cujas. La terza attestazione, di altrettanto interesse, è contenuta al v. w26, verso che fa parte di un brano (6.w1-34) della *Satira* 6 di Giovenale attestato in un solo ms conservato alla Bodleian Library di Oxford, il *Canon. Class. Lat.* 41 (*saec.* XI-XII), e identificato per la prima volta da Winstedt 1899. L'eccezionalità della scoperta ha suscitato dubbi sull'autenticità del passo (cf. da ultimi Reeve 1973 e Willis 1989), che tuttavia è stata riaffermata da coloro che nei versi riportati dal *codex unicus* vedono una prova della doppia recensione operata da Giovenale stesso: secondo Leo 1909 e Luck 1972 si tratterebbe di un passo autentico, poi cassato dal poeta stesso nella revisione dell'opera.

Inoltre, la rara epiclesi è attestata anche per tradizione indiretta, sotto la voce 'LONGURIO' nel libro 2 del *De compendiosa doctrina* di Nonio Marcello: essa sarebbe stata il titolo di una delle *Satire Menippeae* di Varrone (cf. Non. p. 191.27-29 Lindsay: *LONGURIO, id est longus. Varro Triphallo, περί ἀρρενότητος: 'ego nihil Varro video: ita hic obscurat, qui ante me est, nescio qui longurio'*). La possibilità di estendere la ricerca alle varianti d'apparato permette di aggiungere alle fonti del raro epiteto il passo di Columella [fig. 9].

Source	Text
VERG. app. quid hoc 9	▶ At, o <i>Triphalle</i> , saepe floribus nous
COLVM. rust. 32	▶ Arboris antiquae numen uenerare <i>Priapi</i>
IVV. sat. 6, W 26	▶ Personam docili Thais saltata <i>Triphallo</i> .

Figura 9 Ricerca della voce *triph**

3.2 Lucrezio, *De rerum natura* 1: il caso di *Lucr.* 1.66

Tra le nuove acquisizioni di *Musisque Deoque* va annoverata l'edizione digitale del primo libro del *De rerum natura* di Lucrezio, a cura di Paolo Mastandrea, Valeria Melis e Alessia Prontera.²⁰ Era da tempo che l'opera attendeva il suo completamento sulla piattaforma digitale: il libro sesto, curato da Fabrizio Feraco, già vi compare dal 2009, mentre al 2010 risale la pubblicazione dei libri secondo e terzo, a cura di Monica Berti, e dei libri quarto e quinto, curati da Francesco Mambrini. Diverse sono state le scelte editoriali compiute da ciascun editore: se Mambrini (2010) assume quale base di partenza una versione del testo da lui stesso elaborata *ad hoc*, Feraco (2009) e Berti (2010) partono dalla datata, ma sempre autorevole, edizione di Cyril Bailey (1921). Poiché nel 2019 ha visto la luce una nuova, importante edizione del *De rerum natura*, curata da Marcus Deufert per i tipi di Walter de Gruyter, ai curatori è parso opportuno prendere le mosse da quella per costruire l'edizione digitale del *liber primus*. E tuttavia continuo è stato il dialogo con le edizioni del passato, con quella di Bailey (1921) innanzi tutto, ma anche, ad esempio, con quelle di Hugh Andrew Johnstone Munro (1978), Alfred Ernout (1978), Joseph Martin (1969) e di Enrico Flores (2002), che di volta in volta hanno fornito dati interessanti. Andrà peraltro sottolineato che le edizioni MQDQ, proprio per il tipo di intervento che prevedono (una rielaborazione originale, che utilizza con uno scopo specifico i dati provenienti dalla tradizione manoscritta e dalle edizioni a stampa esistenti), non presentano problemi di copyright, dal momento che configurano a tutti gli effetti un nuovo prodotto editoriale, frutto del lavoro di uno specialista del testo, che ne ha dichiaratamente la responsabilità scientifica.

La prima operazione necessaria è consistita nel riordino della sezione relativa alle fonti, dirette e indirette, del poema lucreziano: ovviamente, per il primo libro l'elenco dei testimoni, delle edizioni e dei commenti, consultabile tramite il tasto 'Testimoni', ricavato da quello dei libri già editati (libri 2-6), non coincideva con l'*Index editionum et commentationum* e con i testimoni elencati nel *Conspectus siglorum* dell'edizione Deufert (2019, xxxv-xlv, xlvii-xlix). Il riordino dei testimoni e il completamento della relativa metadattazione hanno consentito, nel back-end, di procedere all'inserimento dei dati evitando confusioni e incoerenze, e, nel front-end, non solo di reperire le informazioni sul dato, ma anche di ricostruire facilmente la 'storia' dello stesso processo di metadattazione.²¹ Secondo quanto già det-

²⁰ <https://www.mqdq.it/public/testo/testo/codice/LVCR%7crena%7c001>.

²¹ Sull'importanza dei metadati in quanto potenziali fonti di memoria del loro stesso ciclo di vita e dei processi di creazione delle entità digitali, cf. Barbuti 2019. Sulle

to, nella preparazione dell’apparato si sono escluse le *uoces nihili*, che pure, per ragioni filologiche e codicologiche, sono presenti in una certa quantità nell’edizione di Deufert (2019). Ad esempio, non si dà conto del fatto che, al verso 8, il manoscritto O²² trasmette la lezione *summittit*, mentre la *uox nihili* testimoniata da Γ²³ è *summit*, o che, al verso 85, il manoscritto O testimonia *sanguine* e Γ l’impossibile *sanuine*. Tuttavia, data la loro rilevanza, sono state segnalate tutte le *uoces nihili* che denunciano un guasto d’archetipo. Nell’apparato relativo al verso 659, ad esempio, è parso significativo mostrare che, mentre l’archetipo (Ω) trasmette l’impossibile *ueraula*, si deve alla seconda mano di O (O²) la lezione corretta *uera uiui* [fig. 10].²⁴



Figura 10 Esempio di indicazione di guasto d’archetipo nell’apparato di MQDQ

Nelle note a margine, rappresentate da un’icona gialla, sono segnalati fenomeni diversi dalla trasmissione di varianti, come cambi nell’ordine dei versi, espunzioni, omissioni, aggiunte,²⁵ particolarità della tradizione manoscritta. In quest’ultima categoria rientrano, ad esempio, i vv. 1068-1075, trasmessi da O mutili nella parte finale e omessi del tutto da Γ [fig. 11].



Figura 11 Esempio di note a margine (Lucr. l. 1.1068-1075)

difficoltà legate alla metadateazione, cf. Currier et al. 2004.

²² *Leidensis Voss. Lat. F 30* (manoscritto meglio noto come *Oblongus*; saec. IXⁱⁿ).

²³ Cioè, la fonte perduta del *Leidensis Voss. Lat. Q 94* (*siglum Q*; saec. IX^{med}) e delle *schedae G, V e U* (per le quali rimando a Deufert 2019, ix-x e xlvii).

²⁴ Altri casi si trovano, ad esempio, nei vv. 27 (ornatum O² *Prisc.*: oratum Ω), 542 (nihiloque renata *Lambinus*: nihiloque creata Q²: nihilo quaeranta Ω) e 665 (*alia Lachmann*: mia Ω).

²⁵ Rispettivamente, vv. 15, 155, 205-7, 434, 984-7, 1093; vv. 44-9, 873-4; vv. 873-4, 1068-75, 1093; vv. 1022-3.

Gli editori non hanno accolto la soluzione adottata dal Deufert (2019, 42) – lasciare in bianco la parte di testo mutila riferendo nell'apparato tutte le integrazioni proposte – e hanno optato per porre a testo l'integrazione più autorevole riportando in apparato il nome dell'editore che l'ha proposta.

Uno dei casi di variantistica più interessanti si trova a *Lucrez. 1.66*. Lucrezio intesse con epica solennità l'elogio di Epicuro, il primo dei mortali a opporsi alla *religio*, la 'superstizione religiosa' che allontana gli uomini dalla verità: *primum Graius homo mortalis tendere* (*Non. 411.3 M. [= 662.3 L.] s.v. tendere: tollere Ω*) *contra / est oculos ausus primusque obsistere contra*. Una buona parte degli editori pone a testo la lezione *tollere*,²⁶ concordemente trasmessa dai manoscritti, e scarta *tendere* (accolta invece dal Deufert 2019),²⁷ adducendo varie motivazioni: essa sarebbe una lettura scorretta di Nonio, secondo Munro (1978, 41), il quale taccia il grammatico di essere un inaffidabile «careless writer»; Flores (2002, 46) ipotizza che *tendere* sia una «uarìa auctoris lectio quae postea repulsa est».²⁸ Tuttavia, poco più di vent'anni prima, Flores (1980, 87 nota 8) adduceva una valida ragione per sostenere la bontà di questa lezione: «l'espressione *tollere contra oculos* non trova riscontri nell'*usus scribendi* di Lucrezio»,²⁹ mentre *tendere contra* è impiegato in 4.299-300, ove il poeta ammonisce: *splendida porro oculi fugitant uitantque tueri. / Sol etiam caecat, contra si tendere pergas*.³⁰ Quindi, Lucrezio impiega l'espressione *tendere contra oculos* per designare l'atto di 'fissare lo sguardo dritto' su qualcosa. In 1.66, tale immagine assume una connotazione di sfida guerresca adatta al contesto epico e 'militaresco' della scena: Epicuro ha sfidato la *religio* rimanendo saldo nella sua posizione.³¹

Solitamente le citazioni noniane del *De rerum natura* concordano con il testo della tradizione diretta. In un caso come quello di 1.66,

²⁶ Ad esempio, cf. Bailey (1921, v. 66) e Bailey (1947, 611 nota 66); Ernout (1978, 4); Martin (1969, 4).

²⁷ Ma già prima *tendere* aveva avuto credito presso autorevoli editori. Cf. per esempio Lambin (1570, 12): «tendere oculos autem dixit, pro contendere»; Lachmann (1853, 5: 21-2 nota 66): «tendere ut a teste nostris antiquiore editum praefendum duco».

²⁸ Contro questa ipotesi Mastandrea 2009, 66 e nota 52.

²⁹ Anche Mazzacane (1984, 162) osserva che «*tendere* sembra rientrare nell'*usus scribendi* del poeta».

³⁰ Cf. anche *Lucrez. 1.324: nulla potest oculorum acies contenta tueri*. Qui il poeta afferma che nemmeno lo sguardo più acuto è in grado di cogliere quel che provoca l'accrescersi graduale dei corpi.

³¹ Canali (1990, 76 ad 62-79) osserva come la «eroicizzazione della *ratio* di Epicuro» sia realizzata qui mediante l'adozione della «terminologia militare romana (cf. *uictor, refert*)». Cf. anche Bailey 1947, 612 nota 75. Per il valore militare di *obsistere contra*, cf. Stat. *Theb.* 10.753 (nessun nemico osa *obsistere contra* il *turbidus* Capaneo). Mantenere la posizione in battaglia è una delle maggiori virtù dei guerrieri dall'*epos* omerico in poi (cf. per esempio *Il. 5.522-7; Tyrt. fr. 12 W.² = 9 G.-P., vv. 15-20; Hdt. 7.104.5*).

la divergenza è attribuibile alla circolazione di un testo dell’opera diverso da quello della tradizione diretta e,³² comunque, a una «precoce convivenza fra le due lezioni» (Mastandrea 2009, 66).³³ La motivazione principale per cui *tendere* può essere ritenuta la lezione corretta consiste nel fatto che Nonio la adotta come lemma: essa non rientra in una delle molte citazioni del testo lucreziano, ma è la parola che il grammatico spiega traendola dal testo del *De rerum natura* a sua disposizione. Dunque, non pare sussistere alcun valido motivo per respingere la lezione trasmessa da un testimone antico come Nonio, il quale, *pace* Munro (1978), in questo caso assai difficilmente può aver commesso un errore mnemonico:

il lessicografo evoca questo luogo lucreziano all’esplicito scopo di definire e documentare una particolare accezione del verbo (*TENDERE est dirigere*), per cui sembra arduo supporre che egli allentasse i controlli in tali circostanze e cadesse vittima di un banale *lapsus*. (Mastandrea 2009, 65)

Inserire entrambe le lezioni entro l’archivio digitale MQDQ rende possibile sottoporle (insieme al loro contesto poetico) alle funzionalità di ricerca raffinata offerte dal sistema.³⁴

3.3 ‘Completare Virgilio’: lavori in corso sulle Georgiche

Chiunque consulti attualmente la pagina MQDQ dedicata a Virgilio³⁵ noterà che tutti gli apparati critici caricati finora (*ecl.* e *Aen.*) risalgono al 2009: l’anno successivo a quello della ristampa corretta dell’edizione di Virgilio curata da Geymonat (2008). La pubblicazione di questo volume non poté non influire sia sulla scelta dell’edizione di riferimento per il testo base, sia, soprattutto, sulla selezione dei dati da inserire nell’apparato critico di *ecl.* e *Aen.* Infatti, rispetto alle edizioni critiche di Sabbadini (1930), Sabbadini-Castiglioni (1944-1945)³⁶ e Mynors (1969), già la prima edizione di Geymonat (1973) spiccava per la notevole ampiezza di informazioni presenti in

³² Mazzacane (1984, 151, 175) ipotizza che Nonio si sia servito di un testo circolante in ambiente africano nel III o nel IV sec. d.C.

³³ Mastandrea (2009, 66) mostra come l’espressione *tendere contra* di Lucr. 1.66 sia stata ripresa da Paolino di Petricordia (un colto vescovo gallo-romano della seconda metà del V sec. d.C., non digiuno, a quanto pare, della conoscenza diretta del testo di Lucrezio) nella *Vita metrica* di San Martino (*Mart.* 1.243).

³⁴ <https://www.mqdq.it/public/ricerca/avanzata>.

³⁵ <https://www.mqdq.it/public/indici/autori/tipo/alpha/lettera/v/idAuthors/78>.

³⁶ Sull’edizione di Sabbadini e la revisione di Castiglioni, cf. Stok 2016.

apparato, dovuta sia alla ricollazione sistematica dei testimoni manoscritti sia a una minuziosa indagine della tradizione indiretta.

Il lavoro di Geymonat (2008) poté quindi offrire agli editori di MQDQ (M. Gioseffi e M. Venuti per *ecl.*, M. Gioseffi e I. Canetta per *Aen.*) abbondante materiale da includere nell’apparato critico, in cui trovarono spazio brani significativi (per esempio, quelli forieri di *variae lectiones*) provenienti dai commentatori antichi di Virgilio. In quell’occasione, e proprio per cercare di includere, oltre ai testimoni, anche il commento degli scoliasti, venne creata la possibilità di inserire all’interno della piattaforma MQDQ due differenti tipologie di annotazioni: ‘apparato’ e ‘nota a margine’. I passi in questione sono perlopiù ricavati dall’edizione Geymonat (2008), ma non è da trascurare il contributo degli editori all’arricchimento di un già cospicuo nucleo di testimonianze indirette, con particolare attenzione all’immediata accessibilità, per gli utenti, dei brani notevoli, spesso inseriti entro rimandi generici o di non immediata reperibilità.³⁷ Ad esempio, per *Aen.* 1.2 (*Italiam fato profugus Lauiniaque uenit*) è riportato non solo Serv. *Aen.* 1.2, citato da Geymonat (2008, 174) (*Lauina legendum est, non Lauinia, quia post aduentum Aeneae Lauinium nomen accepit*), ma anche Gell. 10.16.6 (*neque simile, inquit, illud uideri debet, quod est in primo carmine: “Italiam fato profugus Lauinaque uenit/litora”*), che Geymonat cita (174), ma non riporta per esteso (cf. Canetta 2011).

Chi ora si accinge a completare il lavoro virgiliano entro l’archivio MQDQ, approntando l’apparato delle *Georgiche*, si trova di fronte, nell’ambito degli studi, a nuove importanti acquisizioni di cui tenere necessariamente conto, prima fra tutte la nuova edizione curata da Conte nel 2013, affiancata da un volume di note critiche in cui l’editore dà conto delle sue scelte testuali (Conte 2016). Pur mantenendo come testo base quello fissato da Geymonat (2008) già presente nell’archivio, il nuovo testo digitale delle *Georgiche* ospiterà dunque un buon numero di variazioni che dipendono dall’edizione di Conte (2013) e riguardano non solo le varianti da porre a testo, ma anche l’ordine dei versi e la punteggiatura.

Operativamente, siccome si è reso necessario ‘completare’ il lavoro su un autore per il quale esistevano già apparati critici, è stata creata la possibilità di esportare in *georg.* la lista di testimoni di *ecl.* e *Aen.*, già sottoposti a revisione formale (si veda *supra*), evitando così la creazione di doppioni all’interno del database, nonché possibili errori prodotti dall’inserimento di nuove entrate. Naturalmente, nella lista troveranno spazio non solo i codici impiegati da Geymonat (2008), ma anche manoscritti utilizzati per la prima volta da Conte

³⁷ Un lavoro cospicuo è stato fatto, ad esempio, per rendere immediatamente comprensibili e visualizzabili i rimandi ai diversi volumi dei *Grammatici Latini* (Keil 1857-80). Si veda Venuti 2011.

(2013) per la *constitutio textus*, con particolare riguardo a un nucleo consistente di testimoni di X-XI secolo vergati in scrittura beneventana e indicati con le lettere dell'alfabeto greco (ε, δ, ζ, η).³⁸ Per quanto riguarda il lavoro sul testo si consideri, a titolo esemplificativo, il solo caso di *georg.* 2.69-72:

inseritur uero et fetu nucis arbutus horrida,
et sterilis platani malos gessere ualentis,
castaneae fagos; ornusque incanuit albo
flore piri glandemque sues fregere sub ulmis.
(Ed. Mynors 1969 e Geymonat 2008)

La revisione della punteggiatura, con l'accoglimento della variante *fagus* anziché *fagos*, rende la sintassi più scorrevole e migliora notevolmente la comprensione della frase:

inseritur uero et fetu nucis arbutus horrida,
et steriles platani malos gessere ualentis;
castaneae fagus ornusque incanuit albo
flore piri [...].
(Ed. Conte 2013, 147; 2016, 17-19)³⁹

Ovviamente, il testo *MQDQ* delle *Georgiche* non sarà da intendersi come una nuova edizione critica, del tutto superflua dopo le ultime autorevoli acquisizioni, ma si propone come uno strumento filologicamente raffinato per lo studio dell'intertestualità poetica, e come tale andrà citato da quanti utilizzano l'archivio per le proprie ricerche.⁴⁰

38 Questi codici sono descritti da S. Ottaviano nella sua edizione delle *Bucoliche*, contenuta in Conte 2013, 18-19.

39 Cf. Conte 2013, 147: «at *castaneae* genetiuus est qui genetiuo *piri* respondet (sc. *flore*); haud secus quam *ornus*, *fagus* nominatiuus singularis est»; Conte 2016, 18-19: «ma collegare asinteticamente *castaneae fagos* al verso precedente, in cui i meli erano innestati sui platani, comportava che una pianta sterile (il faggio) fosse innestata su una fruttifera (il castagno), contro ogni buon senso [...]. Ma la soluzione corretta è un'altra: ce l'offre la lezione *fagus* tramandata da Prisciano (e da due manoscritti carolingi). Cambiata l'interpunzione, si fa terminare il periodo a *gessere ualentes*; ha così inizio un nuovo periodo in cui *fagus* è coordinato a *ornus* ed è anch'esso soggetto di *incanuit*, mentre *castaneae* è genitivo retto da *flore* (in parallelo con *piri*)».

40 La questione dei corretti criteri di citazione di *MQDQ*, di tutti i corpora e delle funzionalità che lo compongono è un problema aperto, al quale si sta lavorando: a breve le pagine del progetto subiranno una revisione e verranno inserite indicazioni su questo punto.

4 I *Carmina Latina Epigraphica*

I *Carmina Latina Epigraphica* sono un corpus di iscrizioni latine in versi, compilato nel corso dei secoli e pubblicato nella sua forma complessiva da Bücheler e Lommatzsch tra il 1895 (voll. 1 e 2) e il 1926 (vol. 3). I tre volumi costituiscono la seconda parte dell'*Anthologia Latina* curata da Riese. A questa autorevole edizione di più di 2.500 carmi epigrafici è stato aggiunto un *supplementum* di più di 1.200 iscrizioni rinvenute in epoca successiva, tuttora prive di collocazione definitiva in un qualsiasi corpus organico.⁴¹ Inoltre, nel corso degli ultimi anni, si sono naturalmente sviluppati diversi database epigrafici online,⁴² facendosi sempre più rilevante la necessità, per questo tipo di testi, di un approccio digitale e ad ampio raggio (cf. Fernández Martínez 2021). È stata così creata la piattaforma CLEO (*Carmina Latina Epigraphica Online*), resa disponibile sin dal 2020, che consente di svolgere ricerche su circa 390 carmi dell'*Hispania*, delle Gallie e della Britannia.⁴³

Come detto, l'archivio di *Musisque Deoque* comprende anche i CLE:⁴⁴ nel loro insieme i testi riportati sono datati tra la metà del III secolo a.C. e il VI secolo d.C., ma la maggior parte risale al II e III secolo d.C. I carmi provengono da tutte le regioni dell'Impero Romano, in particolare quelle occidentali (Italia, Spagna, Gallia, Germania, Nord Africa).

⁴¹ Questo *supplementum* è reperibile, oltre che nelle pubblicazioni nell'*Année épigraphique*, anche nelle molteplici edizioni cartacee di corpora locali pubblicati nell'arco degli ultimi trent'anni: si vedano ad esempio Pikhau 1994 e Hamdoune 2016 per l'Africa; Buonocore 1997, Masaro 2017, Morelli 2020 per l'Italia; Cugusi, Sblendorio Cugusi 2016 per varie province e, infine, un gruppo di studiosi spagnoli che ha lavorato ad esempio sulle iscrizioni della Baetica (Fernández Martínez 2007). Da tempo si attende l'edizione del *CIL XVIII* (cf. Krummrey 1964), la quale sarà articolata in almeno due volumi: *CIL XVIII/1 (Carmina Latina epigraphica urbis Romae*, dir.: B.E. Thomasson) e *CIL XVIII/2 (Carmina Latina epigraphica provinciarum Hispanarum*, dir.: C. Fernández Martínez). Il primo volume prevede di raggruppare tutti i carmi provenienti da Roma, il secondo i carmi dell'*Hispania*.

⁴² Occorre menzionare: *Epigraphic Database Roma* per le iscrizioni italiane (<http://www.edr-edr.it/default/index.php>); *Epigraphic Database Heidelberg* per le iscrizioni delle province (<https://edh.ub.uni-heidelberg.de/>); *Epigraphic Database Bari* per le iscrizioni cristiane (<https://www.edb.uniba.it/>); l'esautiva *Epigraphic Database Clauss-Slaby* (<http://www.manfredclauss.de/it/index.html>). Per una discussione, cf. Feraudi-Gruénais 2010.

⁴³ <https://institucional.us.es/cleo/?lang=es>. La questione è discussa da Fernández Martínez, Limón Belén 2014.

⁴⁴ <https://www.mqdq.it/public/ce/presentazione>.

4.1 Aggiornamento e integrazione in MQDQ

Il recente aggiornamento dei *CLE* ha portato a un ampliamento dell’archivio epigrafico di *MQDQ*.⁴⁵ A lavoro concluso (novembre 2022) il numero delle iscrizioni metriche è risultato incrementato di poco più di 410 unità. I nuovi carmi appartengono a epoche differenti: all’età repubblicana risale *CIL* I 1209, l’epigramma sepolcrale di L. Mecio Filotimo che ricalca quello di Pacuvio citato da Aulo Gellio (1.24.4); di epoca imperiale è il carme funerario per Valeria, scoperto di recente presso l’odierna Cadice, in cui si rievoca il mito di Paride ed Elena (Martínez 2016, 54); più ampio è il numero delle iscrizioni tarde, prevalentemente funerarie cristiane, come ad esempio i due distici dedicati ad Albino, rinvenuti nel 2019 presso Vico Equense (Pagano 2019, 692). Una significativa percentuale della nuova implementazione è costituita dai carmi tramandati per via indiretta. È il caso delle iscrizioni metriche ravennati risalenti all’età tardoantica che Agnello (VIII-IX secolo) riporta nel suo *Liber pontificalis ecclesiae Ravennatis*.

Ciascun carme presenta un apparato con elementi precisi. La pagina web, dotata di un permalink copiabile, offre in alto a sinistra, al di sopra del testo, le concordanze dei corpora in cui è presente il carme. È possibile conoscere lo scioglimento delle abbreviazioni passandovi sopra il cursore, come indicato nella [fig. 12].

The screenshot shows the MQDQ website interface. At the top, the logo 'MQDQ' is displayed. Below it, a blue bar contains the text 'Permalink: https://www.mqdq.it/texts/CE|ce|1521' followed by a 'Copia' button. The main content area lists several abbreviations and their corresponding corpora: 'CIL 08, 00241 = CIL 08, 11347 = ILTun 00355 = CLE 01521 = Cholodniak 00990 = D 07801 = CLEAfrigue 00008'. Below this list, there is a section for a specific inscription, 'Marcellus hic quiescit', followed by a list of lines: 'Medica nobilis arte, Annis qui fere uixit Triginta et duobus, 5 Sed eum cuncta parasset Edendo placiturus, Tertium muneris ante Valida febre crematus Diem defunctus obiit.' A tooltip is visible over the list, containing the text: 'C. Hamdoune (Hg.), Vie, mort et poésie dans l'Afrique romaine d'après un choix de Carmina Latina Epigraphica, Brüssel 2011'.

Figura 12 Esempio di scioglimento delle abbreviazioni *CLE*

Nella sezione inferiore sono riportate la provincia romana (o *regio* augustea) di rinvenimento e la località indicata secondo la

⁴⁵ https://www.mqdq.it/public/ce/corpora#HI_anchor.

toponomastica antica e odierna. Segue la datazione, indicata in un arco di decenni o di secoli, più raramente *ad diem*. Laddove presenti, vengono indicati anche il *praescriptum* e il *postscriptum* in prosa. In sequenza compaiono poi i permalink ai maggiori database epigrafici a partire da EDCS. Infine, si indicano l’edizione di riferimento e il nome del/-la curatore/-rice dell’edizione digitale seguito dall’anno. Laddove il carme disponga anche di riferimenti bibliografici citati nei lemmi di apparato, vengono riportate le abbreviazioni corredate dei rispettivi scioglimenti. La bibliografia può essere composta da edizioni,⁴⁶ corpora epigrafici (cf. *CIL*; Pais 1884) e studi su specifiche iscrizioni (cf. Puglia 2021).

L’ultimo metadato è relativo al metro. L’eventuale polimetria di un’iscrizione viene specificata per ogni verso o per gruppi di versi. Nella [fig. 13] un esempio della visualizzazione dei metadati tratto da *CLE* 1191: carme dedicato al *faber naualis* Cecilio Nigro da parte dei colleghi ad Arelate/Arles.

Provincia: Gallia Narbonensis / Località: Arelate / Arles Datazione: 1-200 d.C. Praescriptum: [Cae]c[illo] / Nigro fa[bro navali] EDCS CLEo TM Testo base di riferimento: CLE 01191 Cura dell’edizione digitale: A. Prontera (2022) Bücheler = F. Bücheler, <i>Carmina Latina Epigraphica</i> , I-II, Lipsiae 1895-1897 Bormann = E. Bormann ap. O. Hirschfeld, <i>Corpus Inscriptionum Latinarum (CIL) XII: Inscriptiones Galliae Narbonensis Latinae</i> , Berolini 1888 De Lauriere = J. De Lauriere, <i>Séance du 5 Septembre</i> , “Bulletin de la Société nationale des antiquaires de France” 1888, pp. 262-267 Espérandieu = E. Espérandieu, <i>Inscriptions latines de Gaule: Narbonnaise</i> , Paris 1929 Hirschfeld = O. Hirschfeld, <i>Corpus Inscriptionum Latinarum (CIL) XII: Inscriptiones Galliae Narbonensis Latinae</i> , Berolini 1888 Mommsen = Th. Mommsen ap. O. Hirschfeld, <i>Corpus Inscriptionum Latinarum (CIL) XII: Inscriptiones Galliae Narbonensis Latinae</i> , Berolini 1888 Distici elegiaci

Figura 13 I metadati di *CLE* 1191

4.2 La nota a margine di verso

L’apparato dei *CLE* si compone prevalentemente di congetture in cui si rende conto delle varie proposte di integrazione in corrispondenza delle parti lacunose. Ma le varianti possono insistere anche sulla parte di testo tramandata per via codicologica. Spesso si ricorre anche ad altre funzionalità volte a spiegare particolarità legate all’epigrafia in versi, che non è possibile inquadrare all’interno di un lemma. A queste informazioni è dedicato l’uso della ‘nota a margine di verso’, spesso il primo. Composta in lingua latina, secondo le norme dell’edizione cartacea, è dedicata a fornire i seguenti dati:

⁴⁶ Cf. Bücheler 1895-1926; Hamdoune 2011; Cugusi, Sblendorio Cugusi 2008.

- **Acrostico:** la sequenza verticale delle prime lettere dei versi disposte in modo tale da creare un nome, spesso quello del defunto, viene indicata da una nota a margine del primo verso del carme secondo la forma *Achrosticon [...] legitur*.⁴⁷ Notevole è il caso del quadruplicato acrostico dedicato a Licinia Leontia Ampelia Flavia dalla poetessa Taurina nel V sec. d.C. (CLE 748).
- **Particolarità epigrafiche:** alcune peculiarità meritano una segnalazione. Nell’iscrizione dell’*Africa Proconsularis* dedicata a un bambino morto dopo pochi anni di vita (CLE *Afrique* 105), al v. 1 l’insolita espressione *pueri nominandi* presenta una nota che fornisce la spiegazione della formula:⁴⁸ questa sarebbe stata sostituita con il nome del defunto, una volta che la lastra fosse stata venduta e impiegata. Una nota segnala anche la presenza, all’interno della sezione versificata, di una parte in prosa contenente il nome del defunto o la dedica a quest’ultimo (CLE 1012 e CLE 1215).
- **Riferimenti a personaggi storici:** talvolta i versi epigrafici veicolano testimonianze di personalità storicamente rilevanti. Con la nota a margine si segnala l’identità di Manlia Daedalia, sorella di Manlio Teodoro, console del 399 d.C. (CLE 1434) e di alcuni consanguinei di Magno Felice Ennodio: Probo (PLRE II, s.v. *Fl. Probus* 9, 913) in CLE 1408 e l’omonimo parente del poeta (PLRE IIIA, s.v. *Felix Ennodius* 1, 442).

4.3 Le particolarità linguistiche

Per poter rendere facilmente consultabili i CLE linguisticamente devianti dal latino classico si è scelto di applicare al testo nella versione originale un lemma che segnalasse la forma corretta del vocabolo. Le occorrenze di *orna* è normalizzato in *urna* nel lemma corrispondente (CLE 791.2, AE 1931, 39.2); la forma scempia *cipum* viene corretta in *cippum* in CLE 1193.1; l’apertura della vocale ‘i’ in ‘e’ e l’oscillazione tra ‘b’ e ‘u’ è normalizzata in tutte le occorrenze in CLE 1384: *semplicitate* in *simplicitate*, *gemeno* in *gemino*, *noueletate* in *nobilitate* e lo stesso vale per CLE 1339, laddove *gracelis* diviene *gracilis*, *nemis* diviene *nimis* e *uibat* diviene *uiuat*.

⁴⁷ Sull’acrostico nell’epigramma antico si veda almeno il recente lavoro di Kwapisz 2022.

⁴⁸ Su questa particolarità epigrafica, cf. Cooley 2012, 291; Ireland 1983, 221; Cugusi, Sblendorio Cugusi 2014, 248.

4.4 ‘Taggatura’ su *Pedecerto*

I carmi epigrafici creano inevitabilmente alcuni problemi a uno strumento di analisi metrica automatica come *Pedecerto*. Molte difficoltà sono legate alla prassi epigrafica: circa l’80% dei carmi epigrafici sono epitaffi in cui devono comparire dati quali l’età e i nomi dei defunti, non sempre inquadrabili o conteggiabili in una sequenza dattilica corretta (Schmidt 2014, 772-3). Si veda ad esempio un carme di Cirta (CLE 1327, v. 14) dove il nome sta fuori metro; esso viene evidenziato in verde e non contabilizzato nel calcolo dello schema metrico [fig. 14]:

Tér dēnós ēt bīs quīn[ós uīx] ; sumpserāt áno[s]	SSSS
Ātāté, mý[s]ēró quē ; mīhī fūit ; únīcā ná[t]a.	SDDD
Quót dēdit, ít rēpētít nātúra, ; nōn [ā]uīá pēccat.	DDSS
Dícērē né pīgēāt: P. Sitti Optati móllitēr óssā cūbént.	DD—DD—

Figura 14 Scansione di CLE 1327

Allo stesso modo, data la dimensione commemorativa e funeraria, ci sono a volte parole aggiunte in maniera enfatica o affettiva che ci portano al limite tra poesia metrica e prosa.⁴⁹ Questi elementi conducono spesso a compromessi con la metrica classica, nonché a strutture non previste da essa: versi a volte troppo lunghi per via dell’inserimento di informazioni aggiuntive che cercano però di rispettare le cesure e le clausole.⁵⁰

Come si capisce, di fronte a questa situazione non è possibile una scansione automatica: *Pedecerto* è in grado di scandire automaticamente circa il 98% dei versi sottoposti all’analisi; per il restante 2% non riesce a proporre una scansione e, in questo gruppo, i versi epigrafici costituiscono ancora una parte non trascurabile. In tali casi, quindi, diviene necessaria una marcatura ‘a mano’ per indicare al programma come gestire le anomalie. Inoltre, certe iscrizioni sono frammentarie e la marcatura diventa indispensabile anche laddove si riscontrino lacune che *Pedecerto*, per motivi ovvi, non è in grado di prevedere e di risolvere. Per salvaguardare la forma originale dei versi quando nessuna ipotesi di restituzione filologica prevale e si constata dunque che il testo appare irrimediabilmente perduto, si

⁴⁹ La nozione di ‘prosa affettiva’ è stata proposta da Massaro 2012, ma composizioni epigrafiche tra prosa e verso sono state identificate già da Bücheler, che le chiamò, ambigualmente, «commatica» (ad CLE 1851). Sul tema si vedano Kruschwitz 2002; Carande Herrero 2010.

⁵⁰ Pochi sono, oltre le osservazioni puntuali sul singolo carme, gli studi di fondo sulla metrica dei CLE. Si veda ad esempio il commento di Courtney 1995, 22-31.

è scelto di inserire una stringa di ‘sillabe teoriche’, senza così risolversi arbitrariamente tra dattilo o spondeo. Tale caso si riscontra ad esempio nel numidico *CLE* 1246, v. 5 [fig. 15]:

Scansione eseguita da **PEDECERTO** © 2011— www.pedecerto.eu 

[Éssé]t sí, gēnītōr, nōbī[s sūb]stāntiā lārga,	SDSS
[Ín]sitā iām pīētās rédd[érēt] óbsēquíum	DD— DD—
[Ét t]ībī mágñificá strūēr[ét, uē]nērándē, sēpúlcrha	DDDD
Éx āuró fūlud_ ét sólīd[o]_éx ēbōré	SS— DD—
[Súblīm]ésquē tībī st[ātū]á[s 	SDDS

Figura 15 Scansione di *CLE* 1246

Il lavoro è stato condotto a partire dal 2021: prima dell’intervento risultavano 8.400 versi epigrafici non scanditi e/o da integrare. Una prima ricognizione, effettuata da Luigi Tessarolo a partire dai distici elegiaci dell’edizione di Bücheler, Lommatzsch (1895-26),⁵¹ ha rilevato 306 versi non scansionabili automaticamente da *Pedecerto* e 224 versi frammentari, cioè in totale 530 versi (6,3%) su cui effettuare una marcatura manuale. Essa ha necessitato di oltre 729 *tag* metrici specifici sul file XML dei testi che sta alla base della piattaforma.⁵² Questo intervento ha permesso a *Pedecerto* di raggiungere una percentuale totale del 97,5% di versi dattilici scanditi in tutto l’archivio *MQDQ*; rimangono da trattare nei *CLE* solamente 62 versi. Infine, andrà rilevato che permangono alcuni minimi errori sul materiale trattato, ad esempio laddove il programma non riconosce la prosodia di vocaboli specifici, che richiedono dunque un’integrazione del suo dizionario prosodico (vedi *infra* § 5.7).

51 Considerando le particolarità metriche summenzionate, si impiega qui il termine ‘distici’ per indicare non solo i carmi strutturati in una coppia di versi, ma anche quelli che contengono almeno un pentametro. Quest’ultimo, infatti, identifica il componimento come non esclusivamente esametrico.

52 Le regole della taggatura specifica impiegata da *Pedecerto* compaiono pubblicamente nella funzione ‘scansioni libere’ (<https://www.pedecerto.eu/public/scansioni-libere>). Questa funzione consente all’utente di inserire qualsiasi verso dattilico latino per eseguirne la scansione scegliendo tra più possibilità di versi ed è stata continuamente utilizzata nel corso della marcatura come strumento di verifica.

4.5 Criticità, prospettive di sviluppo e riflessioni

I lavori in corso, con particolare riguardo all'applicazione di uno strumento di analisi metrica automatico al materiale epigrafico, fanno emergere nuove riflessioni per lo studio dei *CLE*. Il traguardo più importante è la possibilità di ricavare dati statistici paragonabili a quelli a disposizione per gli autori letterari e poter sfruttare le opportunità offerte da un motore di ricerca appositamente sviluppato per interrogare su base metrica il corpus delle iscrizioni. Per molto tempo, infatti, l'analisi dei dati statistici non è stata possibile, semplicemente per mancanza di dati, sia perché essi erano sparsi in vari corpora, sia perché erano profondamente disomogenei tra loro. Del resto, il confronto con i testi letterari rimane una sfida più generale:

these [*scil. carmina epigraphica*] are texts which generally cannot be ascribed to identifiable authorial personalities, are of rather limited extent and of dates that cannot always be established; they therefore demand a different type of treatment from that of the literary distichs. A study of the epigraphic distichs, which would certainly not be without interest, would therefore need to be conducted along very different lines from those adopted for the literary distichs. (Ceccarelli 2018, 10)

Alla luce di queste considerazioni, sarà opportuno ripensare in *Pedecerto* una modalità adatta di organizzare e presentare il materiale statistico per i *CLE*, sfruttando le potenzialità del mezzo digitale: pare poco significativo lo stato attuale, nel quale i carmi sono divisi secondo le raccolte e le edizioni tradizionali. Così, l'analisi del *carmen* funerario degli Scipioni (metà del III secolo a.C.) si trova accanto alle iscrizioni dedicatorie cristiane composte da Venanzio Fortunato (VI secolo d.C.). Un criterio più efficace potrebbe essere quello di raggruppare i carmi epigrafici secondo elementi di datazione, in periodi specifici, cosicché formino insieme di testi con caratteristiche analoghe alle opere di un autore, la cui produzione è limitata nel tempo. In questo modo, sarebbero collegabili non soltanto a un periodo ma anche a un ambito o ambiente letterario, rispetto alla letteratura a noi pervenuta. Inoltre, si potrebbe pensare a distinzioni 'generiche',⁵³ cioè almeno tra *carmina sepulcralia*, carmi dedicatori e graffiti, ad oggi indistinti. Inoltre, sfruttando l'efficienza degli strumenti di ricerca intertestuale e metrica offerti da *Musisque*

⁵³ La questione di sapere se esistono davvero uno o più generi letterari epigrafici - cioè con caratteristiche al di là di quelle dei generi letterari a cui i *carmina* si rifanno - è discussa da Fernández Martínez 1999. Sanders (1991, 219) preferisce non parlare di genere o generi lapidari, ma piuttosto di pratiche epigrafiche distinte. La mia proposta, minimale, sta all'incrocio delle due posizioni.

Deoque e *Pedecerto*, si potrebbero sviluppare più opzioni specifiche per i *CLE*: ricerca dei casi di alternanza irregolare tra esametro e pentametro – particolarità propria alla prassi epigrafica – o dei casi di mancanza o di eccedenza di sillabe nel verso. Queste andrebbero correlate, nell'archivio di *Musisque Deoque*, con la costituzione di un apparato metrico-prosodico parallelo a quello filologico per registrare in che modo la prosodia sia influenzata dalla metrica e viceversa.

5 **Ridisegnare l'architettura della Galassia MQDQ: workflow collaborativo e sostenibile in MQDQ**

5.1 **La sfida della sostenibilità**

Come detto in principio, il gruppo che opera in seno alla Galassia *Musisque Deoque* è stato chiamato, tra le altre cose, ad affrontare le delicate sfide relative allo sviluppo sostenibile dei prodotti digitali.⁵⁴ Con questo intento sono state curate le risorse e gli strumenti della Galassia come oggetti di sostenibilità (Schneider et al. 2019), stabilendo che la sostenibilità digitale è «la capacità di soddisfare i bisogni di un sistema digitale nel presente e nel futuro a breve, a medio e a lungo termine» (Fischer et al. 2023, 118-19). Conseguentemente, è stata ridisegnata l'architettura dell'ambiente *MQDQ* al fine di reingegnerizzarne l'esistente in una forma il più possibile vicina all'idea di prodotto di ricerca sostenibile (Boschetti et al. 2021, 414, 418).

Per quel che concerne la dimensione tecnico-infrastrutturale, le attività di reingegnerizzazione si sono sviluppate seguendo molteplici direttive che contemplassero, contemporaneamente, sia i requisiti funzionali sia i requisiti non funzionali del dominio d'interesse. Vale a dire, rispettivamente, tanto soddisfare le necessità degli utenti del sistema *MQDQ* – a partire dalla preparazione delle risorse e dalla fruizione delle stesse, fino alla fase di ricerca e di approfondimento – quanto implementare soluzioni robuste, riusabili ed estensibili – a partire dall'adozione di architetture flessibili, di tecnologie allo stato dell'arte e di infrastrutture digitali in grado di garantire non solo una conservazione a lungo termine e un accesso universale agli archivi testuali, ma anche una batteria di strumenti per le successive attività di ricerca. Il risultato si è configurato nei termini di una definizione di un preciso metodo di lavoro collaborativo e di un modello generico per la produzione, la gestione e la fruizione di corpora testuali e l'implementazione di un'architettura distribuita

⁵⁴ Uno dei primi ad aprire il dibattito sul tema è stato Bradley nel 2007, i cui rilievi hanno ispirato anche le attività finalizzate alla sostenibilità della Galassia *MQDQ*.

finalizzata allo sviluppo di un *workflow* cooperativo (Boschetti, Del Gratta, Del Grosso 2017).

5.2 Architettura distribuita per lavoro cooperativo

Le attività per lo sviluppo della nuova architettura sono state precedute da un lavoro di analisi dell’esistente che ha portato alla selezione delle risorse *legacy* da recuperare nonché all’individuazione delle criticità da sanare. Dopo una fase di descrizione delle modalità e dei canali del processo editoriale preesistente, assai frammentato, si sono individuate le caratteristiche essenziali per il quadro operativo della Galassia MQDQ. Esse prevedono:

1. l’implementazione di una architettura distribuita;
2. la riorganizzazione del flusso di lavoro editoriale;
3. la produzione di dati conformi ai principi FAIR (Wilkinson et al. 2016) al fine di accogliere le politiche della *scienza aperta* (European Commission, Directorate-General for Research, and Innovation 2016);
4. l’utilizzo di infrastrutture conformi ai principi TRUST (Lin et al. 2020) come l’infrastruttura CLARIN (De Jong et al. 2018) per la conservazione delle risorse digitali;
5. l’implementazione di Domain-Specific Languages (DSL) per l’editing scientifico dei testi e delle annotazioni (Boschetti, Del Grosso 2020);
6. l’adozione di formati di interscambio standard e riconosciuti dalla comunità di ambito DH quali XML/TEI per la distribuzione delle risorse (Driscoll, Pierazzo 2016);
7. la progettazione e implementazione di Web API conformi alle specifiche RESTful (Fischer et al. 2019).

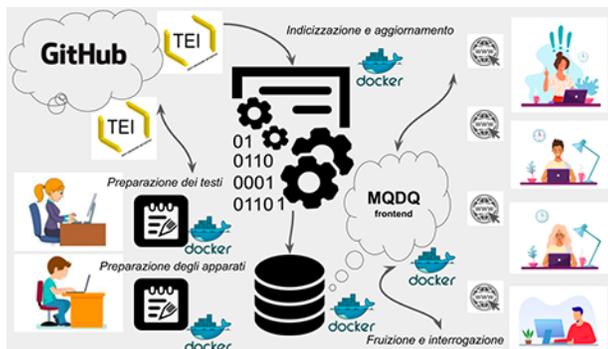


Figura 16 Schema dell'architettura del *workflow* distribuito implementato per la Galassia MQDQ

La figura 15 schematizza il nuovo flusso di lavoro volto alla produzione, all’elaborazione e alla fruizione delle risorse gestite dalla Galassia MQDQ (principalmente testo e apparato) [fig. 16]. Il flusso è implementato mediante un’architettura di tipo distribuito e si compone delle seguenti fasi:

1. registrazione dei metadati relativi all’edizione di riferimento;
2. trascrizione del testo accolto dall’edizione di riferimento;
3. preparazione degli apparati critici;
4. creazione di indici multidimensionali al fine di ottimizzare le ricerche intertestuali sia in testo sia in apparato;
5. realizzazione di un’interfaccia grafica per l’interazione con gli studiosi;
6. progettazione e implementazione di Web API conformi alle specifiche RESTful.

L’architettura integra differenti soluzioni tecnologiche che garantiscono flessibilità e robustezza al sistema. In dettaglio, il componente di produzione delle risorse (cioè metadati, testi e apparato) fa uso, dove conveniente, di ‘Linguaggi Specifici di Dominio’ (DSL) per ricalcare il più possibile le pratiche editoriali degli umanisti ‘tradizionali’ senza rinunciare alle prassi collaborative proprie degli umanisti ‘digitali’. Laddove l’uso dell’approccio orientato ai DSL venisse percepito meno adatto dagli utenti-editori, l’ambiente prevede l’uso di interfacce grafiche per l’inserimento dei dati. Per quanto riguarda l’editing del testo, infatti, è stato implementato l’approccio DSL basato sul linguaggio *markdown*; per la gestione delle entrate d’apparato, invece, è stato preferito l’approccio con interfaccia grafica; infine, per i metadati è disponibile un sistema misto DSL-GUI. Indipendentemente dalla modalità e/o dal formalismo per la codifica dei dati, il formato di rappresentazione e di interscambio delle risorse testuali prescinde da come questi vengono prodotti: le risorse rispettano sempre lo schema di codifica MQDQ conforme alle linee guida della *Text Encoding Initiative* (TEI).⁵⁵ All’atto pratico, ciò comporta che i dati memorizzati, versionati, pubblicati ed elaborati sono rappresentati in formato standard e riconosciuti dalla comunità degli umanisti digitali (Pierazzo 2015).

Il componente di produzione delle risorse si interfaccia con la tecnologia *git* (Chacon, Straub 2014) e con la piattaforma *github*;⁵⁶ tutta-

⁵⁵ <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.

⁵⁶ <https://docs.github.com/en>. *Git* è un sistema distribuito per il tracciamento delle modifiche di risorse testuali. Sviluppato con lo scopo di supportare l’evoluzione di grandi progetti software *open source*, è adatto anche per tenere traccia delle modifiche di risorse testuali, come nel caso dei corpora della Galassia MQDQ. *GitHub*, dal canto suo, è una piattaforma web che, tra le altre innumerevoli funzionalità, permette di ospitare i dati assieme a tutta la storia delle versioni (*commit*).

via, il sistema *MQDQ* non fa uso di *github* per la pubblicazione *open* delle risorse alla comunità, ma le *releases* ufficiali dei corpora vengono depositate su repositories CLARIN.⁵⁷ L'infrastruttura rilascia un *PID handle* e garantisce la conservazione e l'accesso a lungo termine dei dati depositati nelle varie *releases*.⁵⁸ Le soluzioni così adottate garantiscono da una parte l'aderenza della Galassia *MQDQ* ai principi FAIR e ai principi TRUST, dall'altra la collocazione degli archivi gestiti dalla Galassia in una posizione favorevole rispetto alle direttive della scienza aperta e rispetto a un quadro di sostenibilità del progetto nella sua interezza.

Prodotte le risorse, il flusso di lavoro prevede poi una fase di elaborazione per l'indicizzazione e l'analisi dei dati testuali. Quest'ultima produce una collezione di indici strutturati che abilitano la fruizione puntuale e la ricerca avanzata nei vari corpora. La fruizione in *MQDQ* può avvenire mediante interfaccia grafica per gli utenti della piattaforma (GUI) oppure mediante interfaccia di programmazione per agenti software (API).⁵⁹ Infine, la tecnologia usata per il *deployment* e la messa in produzione delle varie componenti è la tecnologia *docker*: una piattaforma per lo sviluppo e il rilascio di applicazioni 'pronte all'uso' (<https://docs.docker.com/>). Grazie all'impiego di questa tecnologia è possibile costruire pacchetti software che isolano il contesto applicativo dal contesto infrastrutturale, ottimizzando l'installazione delle applicazioni per la rapida messa in produzione. I pacchetti così realizzati prendono il nome di 'immagini'; a loro volta, le immagini vengono utilizzate dall'ambiente *docker* per eseguire l'applicazione nella sua forma preconstituita. La piattaforma *docker* costruisce in questo modo quello che viene chiamato *container*, istanza cioè dell'immagine precedentemente costruita.

⁵⁷ <https://dspace-clarin-it.ilc.cnr.it/repository/XMLui/page/deposit>.

⁵⁸ Per esempio, l'attuale identificativo persistente rilasciato dall'infrastruttura per il corpus *MQDQ* è <http://hdl.handle.net/20.500.11752/OPEN-555>.

⁵⁹ Il sistema citazionale sviluppato nel corso degli anni dalla piattaforma *MQDQ* è stato adottato anche nel recupero dei dati mediante Web API per garantire retrocompatibilità degli indirizzi URL in conformità ai permalink per l'accesso alle risorse (per esempio in *MQDQ* il primo libro dell'Eneide è citato con l'URL <https://www.mqdq.it/public/testo/testo/codice/VERG%7caene%7c001>). L'architettura sviluppata dai *Canonical Text Services* (Berti 2019) - adottata anche dal progetto *Perseus* - è stata considerata come ulteriore tecnologia di accesso canonico che in futuro potrà accompagnare il sistema di recupero dei dati testuali.

5.3 Inserimento metadati, dati ed edizione

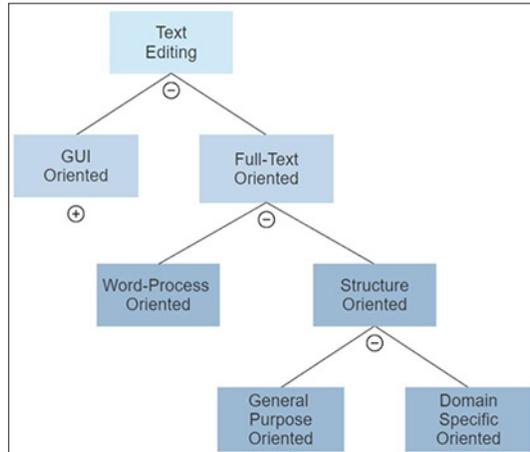


Figura 17 Modalità di inserimento dati testuali strutturati

All’interno dell’ambiente per la produzione delle risorse testuali (cioè i metadati, i testi e gli apparati critici) sono state implementate le componenti per l’editing dei dati. Le soluzioni adottate per la creazione delle risorse si basano su diversi approcci (Zenzaro et al. 2022) e seguono sostanzialmente due principi: 1) soluzioni orientate alle interfacce grafiche, dette *GUI*; 2) soluzioni orientate al testo, dette *full-text*. Come mostra la [fig. 17], l’approccio alla codifica con i linguaggi specifici di dominio (DSL) appartiene al ramo originato dalla soluzione full-text; le altre ramificazioni rappresentano: 1) la soluzione mediante l’uso di programmi di video scrittura quali *word processor* (come il pacchetto Office e simili); 2) la soluzione orientata all’uso di linguaggi descrittivi generici (quali XML, JSON, ecc). La prima soluzione – comoda per l’editing testuale – è stata adottata dai consorzi accademici francesi afferenti all’infrastruttura di ricerca HumaNum (<https://www.huma-num.fr/>) per semplificare il più possibile la produzione di edizioni digitali. Ad ogni modo, tale soluzione risulta non solo meno flessibile ed efficiente rispetto alla seconda dal punto di vista computazionale, ma anche del tutto inadatta per la conservazione a lungo termine delle risorse testuali e per il riuso in sistemi di analisi *general-purpose*. La seconda soluzione, ad oggi considerata soluzione standard per la rappresentazione dei testi digitali, risulta adatta per la conservazione e lo scambio dati, ma non sempre familiare allo studioso che ne deve produrre una corretta rappresentazione digitale, poiché spesso ridondante, macchinosa e lontana dalle pratiche tradizionali. Tali considerazioni portano a

preferire l’adozione di linguaggi specifici di dominio quando le esigenze di editing e la complessità del contesto lo consentono: tale circostanza si è verificata per la codifica dei metadati e dei testi per la Galassia MQDQ, mentre si è preferito registrare i dati di apparato attraverso interfaccia grafica oppure mediante l’uso diretto del formato di rappresentazione del testo XML/TEI. Utile ribadire che per i corpora della Galassia MQDQ il formato di conservazione e di intercambio in ogni caso rispetta le norme dello schema XML/TEI, indipendentemente dalla modalità con cui sono stati inseriti i dati (cioè GUI oppure DSL). Ciò rappresenta una netta distinzione tra il linguaggio utilizzato per ‘dare forma’ digitale ai dati testuali e il formato di ‘rappresentazione’ digitale impiegato per la pubblicazione, memorizzazione e serializzazione dei dati.

<pre>- corpus > `name` mdaq - work > `main-title` carmina > `alt-title` carmina > `short-title` carm. > `form` poetry > `meter` various > `source` literary > `number` 1.00 - author > `main-name` Optatianus Porfyrius > `complex-name` Publilius Optatianus Porfyrius > `short-name` OPT. PORF. > `date` aetate Constantini > `age` 5 > `numeber` 164.000 - bibl > `paper` G. Polara, 1973</pre>	<pre>• corpus name mdaq • work main-title carmina alt-title carmina short-title carm. form poetry meter various source literary number 14.00 • author main-name Optatianus Porfyrius complex-name Publilius Optatianus Porfyrius short-name OPT. PORF. date aetate Constantini age 5 numeber 130.000 • bibl paper G. Polara, 1973</pre>
---	---

Figura 18 Indicazione dei metadati con metodo DSL-based tramite linguaggio markdown

In [fig. 18] è riportato un esempio di inserimento dei metadati relativi a un’opera oggetto di recente intervento nel corpus MQDQ: i *carmina* di Optaziano Porfirio. Si nota sulla sinistra l’indicazione dei campi e dei valori in formato *markdown* (.md)⁶⁰ e sulla destra la preview in formato HTML. La scelta di costruire il linguaggio specifico per la descrizione dei metadati a partire dalla sintassi del linguaggio

⁶⁰ <https://daringfireball.net/projects/markdown/>.

markdown è stata operata considerando la capacità intrinseca di visualizzazione web dei dati codificati.

Successivamente, i metadati inseriti sono automaticamente trasformati in formato XML/TEI convertendo i dati editati con la sintassi *markdown* nei rispettivi campi dell'intestazione del documento TEI. Nello specifico si notano nella [fig. 19] i metadati descrittivi del *corpus* e del titolo (registrati nel descrittore <titleStm<), i termini che connotano il profilo dell'opera (registrati nel descrittore <textClass> del blocco <profileDesc>) e l'indicazione dell'edizione a stampa da cui il testo è stato digitalizzato (Polara 1973).

```

<fileDesc>
  ... <titleStm>
  ...   ... <title type="main" xml:id="OPT_PORF-carm">carmina</title>
  ...   ... <title type="alt">carmina</title>
  ...   ... <title type="short">carm.</title>
  ...   ... <author xml:id="OPT_PORF" n="164.000"> ...
  ...   ... </author>
  ... </titleStm>
  ... <extent> ...
  ... </extent>
  ... <publicationStm> ...
  ... </publicationStm>
  ... <seriesStm> ...
  ... </seriesStm>
  ... <sourceDesc>
  ...   ... <bibl>G. Polara, 1973</bibl>
  ... </sourceDesc>
</fileDesc>
<encodingDesc> ...
</encodingDesc>
<encodingDesc> ...
</encodingDesc>
<profileDesc>
  ... <langUsage>
  ...   ... <language ident="lat">Latin</language>
  ... </langUsage>
  ... <textClass>
  ...   ... <keywords>
  ...     ... <term type="form">poetry</term>
  ...     ... <term type="source">literary</term>
  ...     ... <term type="meter">various</term>
  ...   ... </keywords>
  ... </textClass>
  ... <creation>
  ...   ... <idno>1.00</idno>
  ...   ... <date></date>
  ... </creation>
</profileDesc>

```

Figura 19 Formato XML/TEI Compliant dei metadati inseriti in *markdown* usato come formalismo

Allo stesso modo, anche il contenuto delle opere è codificato ricorrendo al linguaggio *markdown*. In [fig. 20] e in [fig. 21] è presente una parte della trascrizione del testo di Optaziano Porfirio. Nel dettaglio, la struttura del testo in MQDQ registra più livelli per identificare le varie divisioni e i blocchi di testo. La struttura viene rappresentata dal linguaggio mediante il segno di 'cancelletto' (# rappresenta il livello 1 della struttura) oppure mediante il segno di 'doppio cancelletto' (## rappresenta il livello 2 della struttura). Il titolo della divisione è indicato semplicemente dai caratteri presenti sulla riga del livello: per esempio in [fig. 21] il titolo *Versus intexti*. Il testo, verso per verso, è inserito indicando la sequenza di parole in formato *plain text*, con l'aggiunta del ritorno a capo a fine riga per registrare la fine del verso e l'inizio del verso successivo. Le aggiunte di informazioni metatestuali, che possono riguardare la tipologia, la funzione e la semantica delle divisioni, si apportano mediante un meccanismo di decorazione dei segmenti di testo: si seleziona il testo da annotare per mezzo di parentesi quadre e si fa seguire la selezione dall'indicazione di un'etichetta testuale definita in calce al testo in editing.⁶¹ Per maggiore chiarezza, in [fig. 20] si evidenziano tre annotazioni di questo tipo:

1. l'indicazione del tipo di metro (# [1][E], indica che il primo blocco di testo appartiene ai distici elegiaci);
2. un'ulteriore segnalazione metrica (# [2][H], indica che il secondo blocco di testo è costituito da esametri);
3. un'annotazione in testo che dettaglia un dato di tipo analitico relativo alla lunghezza della prima 'e' di *sedent* ([sedent] [desc], riporta l'etichetta *desc* in testo, definita poi con la stringa formattata [desc]: # 'SE+DENT' in calce).

Nella visualizzazione della preview è possibile verificare visivamente il contenuto delle annotazioni aggiunte; di particolare rilievo è la possibilità di reimpiegare etichette precedentemente definite: si faccia riferimento, ad esempio, ai dati sui metri in [fig. 20] e [fig. 21] per i distici elegiaci e per gli esametri. Tuttavia, non è esclusa la possibilità di inserire *inline* le annotazioni metatestuali mediante l'uso di coppie di parentesi tonde (si veda in [fig. 20] l'indicazione della stringa [metri](# 'METRI_') che registra un'annotazione in modalità *inline*).

⁶¹ La sintassi del linguaggio specifico di dominio è definita assieme ai collaboratori ed è il più familiare possibile all'opera del filologo. Gli errori di sintassi vengono automaticamente riconosciuti dal sistema e segnalati all'operatore durante la preparazione dei testi. Vengono quindi concordate poche regole sintattiche che tutti i collaboratori possono seguire in modo intuitivo.

<pre># [1][E] Quae quondam sueras pulchro decorata libello Carmen in Augusti ferre Thalia manus, Ostro tota nitens, argento auroque coruscis [...] Purpureo fulgens habitu, radantibus intus, Vt quondam, scriptis ambitiosa tuis. # [2][H] Sancte, tui uatis, Caesar, miserere serenus. Auguste omnipotens, almo mortalia cuncta Numine laetificans, nobis ad gaudia nomen, Constantine, tuum fecundi carminis ex hoc Te duce det Musas; nam tristis cura recusat > Egregios actus: iam [sedent][desc] crimina Parcae. Tunc melius dominum te uox secura sonabit, Virtutum rector. potuit uix pangere uersu [...] Sancte, tui uatis, Caesar, miserere serenus. [E]: # 'disticti elegiaci / elegiac couplets' [H]: # 'esametri / hexameters' [desc]: # 'SE+DENT'</pre>	<p><u>1</u></p> <p>Quae quondam sueras pulchro decorata libello Carmen in Augusti ferre Thalia manus, Ostro tota nitens, argento auroque coruscis [...] Purpureo fulgens habitu, radiantibus intus, Vt quondam, scriptis ambitiosa tuis.</p> <p><u>2</u></p> <p>Sancte, tui uatis, Caesar, miserere serenus. Auguste omnipotens, almo mortalia cuncta Numine laetificans, nobis ad gaudia nomen, Constantine, tuum fecundi carminis ex hoc Te duce det Musas; nam tristis cura recusat</p> <p><i>Egregios actus: iam <u>sedent</u> crimina Parcae.</i></p> <p>Tunc melius dominum te uox secura sonabit, Virtutum rector. potuit uix pangere uersu [...] Sancte, tui uatis, Caesar, miserere serenus.</p>
---	---

Figura 20 Codifica del testo mediante approccio DSL-based adottando il linguaggio markdown

<pre># [31][n] Versus intexti ## [2][H][.][c] Sancte, tui uatis, Caesar, miserere serenus; Aurea sic mundo disponas saecula toto. ## [3][H][.][c] Fingere Musa queat tali si carmine uultus > Augusti, et [metri](# 'METRI_') et uersus lege manente, Picta elementorum uario per musica textu Vincere Apelleas ardebit pagina ceras. Grandia quaerentur, si uatis laeta Camena Orsa iuuat, uersu consignans aurea saecula. [H]: # 'esametri / hexameters' [c]: # 'carmen' [E]: # 'disticti elegiaci / elegiac couplets' [desc]: # 'SE+DENT' [n]: # 'versus intexti'</pre>	<h2><u>31</u> Versus intexti</h2> <p><u>2.</u></p> <p>Sancte, tui uatis, Caesar, miserere serenus; Aurea sic mundo disponas saecula toto.</p> <p><u>3.</u></p> <p>Fingere Musa queat tali si carmine uultus</p> <p><i>Augusti, et <u>metri</u> et uersus lege manente,</i></p> <p>Picta elementorum uario per musica textu Vincere Apelleas ardebit pagina ceras. Grandia quaerentur, si uatis laeta Camena Orsa iuuat, uersu consignans aurea saecula.</p>
--	---

Figura 21 Codifica doppio livello con approccio DSL-based in linguaggio markdown

La [fig. 22] mostra la codifica XML/TEI delle tre tipologie di entrata d’apparato illustrate più sopra (§ 3.2) per Lucrezio, così come i dati dei testimoni e della tradizione a stampa. Nello specifico, a sinistra è riportata sia la lezione guasta dell’archetipo sia il lemma accolto a testo; si nota l’uso dei tag <app> (entrata d’apparato), <lem> (lemma della lezione accolta a testo) e <rdg> (lezione variante registrata in apparato ma non accolta a testo) con gli attributi @from (riferimento al luogo del testo dove inizia la lezione), @to (riferimento al luogo del testo dove termina la lezione) e @wit (riferimento al testimone che riporta la lezione oggetto di codifica). Appena sotto, è riportata la descrizione dell’archetipo Ω : si nota l’uso del tag <witness> (descrizione del testimone), il codice identificativo del testimone nonché la sigla del testimone registrato nell’attributo @n (nome).

Similmente, per la notizia d’apparato relativa alle integrazioni del poema lucreziano, è evidente la presenza del collegamento con l’editore mediante l’indicazione della fonte a stampa: questa è registrata nel sistema per mezzo del tag <bibl> (entrata bibliografica), elemento della lista bibliografica <listBibl> (lista di entrate bibliografiche), dove è possibile notare l’attributo @n che definisce il nome dell’editore (nella fattispecie *Bailey*). Ultima tipologia di entrata d’apparato è quella relativa alle note a margine indicata con attributo @type (tipologia del dato) dell’elemento <app> valorizzato con la stringa *margin-note*.

5.5 Il sistema di gestione delle modifiche GIT

Il sistema di editing della Galassia MQDQ si interfaccia con l’ambiente di tracciamento delle risorse digitali *git*. La [fig. 23] mostra il confronto tra due versioni del documento XML/TEI della risorsa *OPT_PORF-carm.XML* che riporta il testo dei *carmina* di Optaziano Porfirio. Si notano due tipologie di interventi: extratestuale e testuale. Nel primo caso è stata aggiunta un’indicazione per l’analisi dei versi (che andrà in ultima istanza codificata in XML/TEI con tag <w> assieme agli opportuni attributi); il secondo caso, invece, emenda la sequenza di caratteri latini e greci *uictoreꝑ* con la sequenza di caratteri latini *uictorem*.

1041	<l xml:id="d03115" n="1">Fingere Musa queat tali si carmine uultus</l>	1041	<l xml:id="d03115" n="1">Fingere Musa queat tali si carmine uultus</l>
1042 -	<l xml:id="d03116" n="2">Augusti, et metri et uersus lege manente,</l>	1042 +	<l xml:id="d03116" n="2">Augusti, et metri(=METRI) et uersus lege manente,</l>
1043	<l xml:id="d03117" n="3">Picta elementorum uario per musica textu</l>	1043	<l xml:id="d03117" n="3">Picta elementorum uario per musica textu</l>
1044	<l xml:id="d03118" n="4">Vincere Apelleas ardebit pagina ceras.</l>	1044	<l xml:id="d03118" n="4">Vincere Apelleas ardebit pagina ceras.</l>
1045	<l xml:id="d03119" n="5">Grandia quaerentur, si uatis laeta Camena</l>	1045	<l xml:id="d03119" n="5">Grandia quaerentur, si uatis laeta Camena</l>
+	@@ -1138,7 +1138,7 @@		
1138	<l xml:id="d031166" n="1">Alme tuas laurus aetas sustollet in astra.</l>	1138	<l xml:id="d031166" n="1">Alme tuas laurus aetas sustollet in astra.</l>
1139	<l xml:id="d031167" n="2">Aurea lux uatum, siluae mihi praemia serua.</l>	1139	<l xml:id="d031167" n="2">Aurea lux uatum, siluae mihi praemia serua.</l>
1140	<l xml:id="d031168" n="3">Aucta Deo uirtus Musas magis ornat aperta.</l>	1140	<l xml:id="d031168" n="3">Aucta Deo uirtus Musas magis ornat aperta.</l>
1141 -	<l xml:id="d031169" n="4">Aurea uictorey pietas sonat ubere lingua.</l>	1141 +	<l xml:id="d031169" n="4">Aurea uictorem pietas sonat ubere lingua.</l>
1142	<l xml:id="d031170" n="5">Aonios latices pietas iuuat armaque diua.</l>	1142	<l xml:id="d031170" n="5">Aonios latices pietas iuuat armaque diua.</l>

Figura 23 Visualizzazione tramite piattaforma *github* per il confronto alla storia delle modifiche della risorsa

Le risorse in editing sono collegate alla piattaforma *github* in modo da sincronizzare le risorse con gli aggiornamenti di carattere filologico e testuale provenienti dal gruppo di lavoro. Le procedure per l’elaborazione del testo e l’indicizzazione faranno ricorso ai soli dati più freschi presenti sul repository *github* di progetto. Se da una parte i repositories *github* dove risiedono le risorse in lavorazione sono evidentemente tenuti privati e ad accesso limitato, le versioni pubbliche dei corpora sono ospitate e rese conformi ai principi *FAIR* tramite il deposito su repository *CLARIN* (vedi §§ 5.8-10).

5.6 Le Application Programming Interfaces

La nuova architettura *MQDQ* prevede non solo la consultazione delle risorse per mezzo di interfacce grafiche, amichevoli per i non addetti ai lavori, ma anche la costruzione di interfacce di programmazione *API REST* per l’accesso e l’interrogazione degli archivi testuali in modalità programmatica. Questo consente nuovi scenari d’uso e nuove prospettive di ricerca grazie alla possibilità di elaborare i dati indipendentemente dall’ambiente grafico. Terze parti sono così in grado di usare i servizi di analisi e di accesso ai contenuti editoriali. Il protocollo utilizzato per l’implementazione delle prime API è quello *RESTful* (Pautasso, Wilde, Alarcon 2013). Il protocollo è definito formalmente dalle specifiche della tecnologia *OpenAPI* e le API sono implementate mediante il framework *Java SpringBoot* (<https://spring.io>).

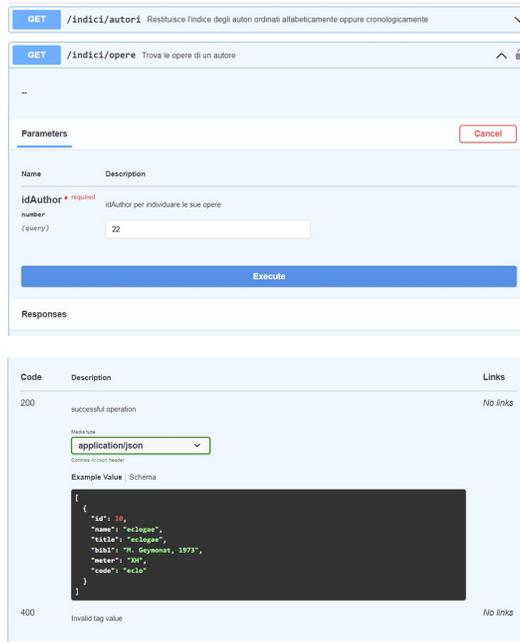


Figura 24 Vista delle API in sviluppo con sistema Swagger

Le API consentono di interagire con il sistema MQDQ per ottenere i dati degli archivi via software con la stessa efficacia offerta dall'interfaccia grafica. La documentazione è pubblicata tramite l'ambiente web Swagger (<https://swagger.io>). La [fig. 24] mostra due servizi di accesso ai dati (*end-point*) per due liste esposte da MQDQ: quella degli autori e quella delle opere per singolo autore. Grazie all'uso dell'ambiente Swagger è possibile dettagliare il funzionamento dell'API e le specifiche per il loro uso. Oltre alle suddette liste, mediante API RESTful è possibile accedere a:

1. numero degli autori gestiti;
2. lista delle opere con metro specificato;
3. sezioni testuali di un'opera specifica;
4. lista di versi di una specifica sezione;
5. lista dei versi con metro specificato;
6. lista di occorrenze specificate da una interrogazione full-text;
7. autore e opera di una specifica occorrenza testuale;
8. citazione completa di una singola occorrenza testuale;
9. lista di occorrenze in *multiword*;
10. lista di occorrenze con lezioni varianti;
11. informazioni di apparato complete per una specifica lezione variante.

5.7 Scansione metrica e *Pedecerto*

Il componente di elaborazione del testo è funzionale a due precise attività: 1) costruire indici che possano sostenere nel miglior modo possibile l'espressività delle ricerche della piattaforma e la rapidità della restituzione dei risultati; 2) l'analisi metrica dei versi codificati nei corpora testuali.

Pedecerto nella sua scansione deve essenzialmente risolvere due aspetti del verso: 1) definire la sillabazione; questo implica soprattutto che si riconosca il valore di vocale o consonante delle lettere 'u' e 'i', e il valore mono- o plurisillabico degli incontri di vocali; questa fase può anche concludersi con la proposta di più di una sillabazione (ad esempio una forma di perfetto come *uoluit* potrebbe essere tanto bisillabica, lemma *uoluo*, quanto trisillabica, lemma *uolo*); 2) riconoscimento delle quantità delle sillabe, basata sia sul dato morfologico (dittonghi, sillabe chiuse) sia sulla consultazione di una tabella di circa 76.900 prosodie (in parte autocostruita da *Pedecerto* stesso).

Come risultato della scansione è possibile ottenere: 1) una sola soluzione; 2) più ipotesi di soluzione, per la presenza di ambiguità non risolte nella sillabazione, nell'assegnazione delle quantità, nella posizione di uno iato ecc.; 3) nessuna ipotesi di soluzione: le cause possono essere varie, ad esempio usi poetici poco frequenti di cui *Pedecerto* non tiene conto (*omnia* bisillabo, *Launia* trisillabo) o quantità non rispettate in autori tardi. È prevista una speciale codifica manuale che fa accettare a *Pedecerto* anche i versi del tipo 2 o 3. Va poi sottolineato che *Pedecerto* entra in relazione con i due corpora latini in maniera diversa:

1. *Musisque Deoque*: i versi non risolvibili automaticamente (4671) sono stati marcati manualmente, per cui la scansione viene eseguita preliminarmente dalla versione Java di *Pedecerto*, integrata nelle *routines* di indicizzazione. Ne consegue che: a) nel front-end è possibile visualizzare la scansione della totalità dei versi dattilici; b) le scansioni memorizzate sono sfruttate per la ricerca delle co-occorrenze metrico-verbali (<https://www.mqdq.it/public/cooccorrenze/cometri>); c) è disponibile un'interfaccia web specializzata (<https://www.pedecerto.eu>), che consente di esplorare il corpus di *MQDQ* attraverso svariate ricerche di natura metrico-lessicale.
2. *Poeti d'Italia*: non è stata eseguita la marcatura manuale dei versi non risolti, per cui il front-end propone solo una scansione estemporanea ad opera della versione PHP di *Pedecerto*, in cui alcuni versi risultano ignorati.

5.8 Verso la CLARIN federated content search

L’implementazione delle API programmatiche è il pre-requisito obbligatorio per il collegamento tra il motore di ricerca MQDQ e i servizi offerti dalla ricerca federata di CLARIN (Federated Content Search, FCS).

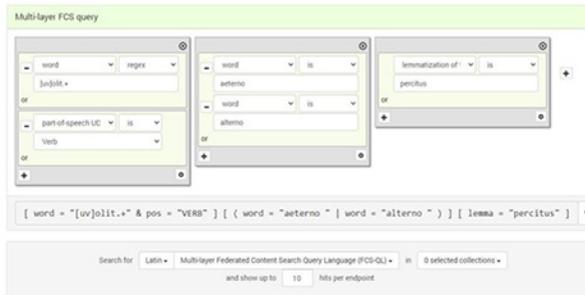


Figura 25 Interfaccia per l’interrogazione dei corpora nella ricerca federata CLARIN

Grazie alla FCS sarà possibile dar valore ai corpora del progetto e alle funzionalità ad esso collegate, mettendole a sistema con una delle infrastrutture di ricerca di ambito umanistico più grandi nel panorama europeo.⁶⁵ L’espressività dell’interrogazione della ricerca federata è basata sul linguaggio CQL (Corpus Query Language).⁶⁶ La [fig. 25] esemplifica un’espressione di ricerca costruita articolando caratteristiche paradigmatiche e sintagmatiche. In particolare, è formulata una ricerca per la stringa presente in Lucrezio, 4.47: *uolitent aeterno percita*. La stringa assembla graficamente le componenti dell’espressione d’interrogazione CQL: `[word = "[uv]olit.+ " & pos = "VERB"] [(word = "aeterno " | word = "alterno ")] [lemma = "percitus"]`. L’espressione ottenuta ricerca all’interno dei corpora latini collegati alla FCS tutti i passi che verificano la sequenza di tre parole: la prima parola deve rispondere a una espressione regolare (*[uv]olit.+*) con anche l’indicazione linguistica di categoria grammaticale verbo (*VERB*); la seconda parola è un’alternativa rispetto alla sequenza di caratteri *aeterno* oppure *alterno* (per tenere traccia delle informazioni d’apparato); infine,

⁶⁵ L’integrazione tra MQDQ e la ricerca federata sviluppata da CLARIN prevede l’implementazione di uno strato software di raccordo tra i formati e i servizi dell’infrastruttura con i formati e i servizi del nodo che espone le risorse. I dettagli sono descritti al seguente link: <https://www.clarin.eu/content/federated-content-search-clarin-fcs-technical-details>.

⁶⁶ <https://www.sketchengine.eu/documentation/corpus-querying/>.

la terza parola deve avere come lemma *percitus*.⁶⁷

5.9 Analisi dei dati con *Voyant Tools*

Un’ulteriore integrazione che si auspica nel prossimo futuro per la Galassia MQDQ è quella relativa alle funzionalità disponibili nell’ambiente web-based *Voyant Tools*,⁶⁸ sviluppato da Stéfán Sinclair e Geoffrey Rockwell. Integrazione resa ancor più semplice dalle possibilità offerte dal Tool Inventory del Language Resource Switchboard di CLARIN, che presenta l’ambiente *Voyant-Tools* tra le varie opzioni, nella categoria degli strumenti per il Distant Reading.⁶⁹

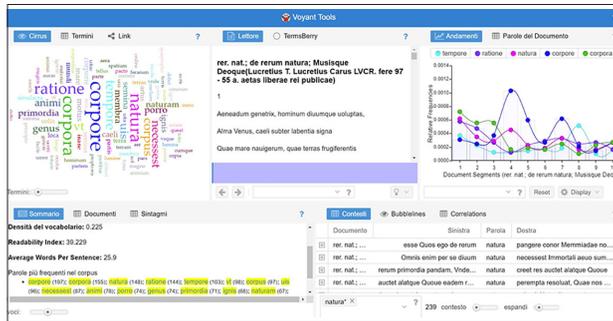


Figura 26 Analisi dei corpora via *Voyant Tools*

La [fig. 26] mostra la schermata dell’ambiente *Voyant Tools* per l’analisi dell’opera di Lucrezio. Nella fattispecie, si colgono diversi strumenti grafici per lo studio del testo: in alto a sinistra, il componente per la visualizzazione del *word cloud*, chiamato Cirrus; in alto, al centro, lo strumento per la visualizzazione dei collocati di una parola frequente; questo strumento è chiamato TermsBerry; in alto a destra, invece, il componente relativo alla distribuzione delle occorrenze di una parola all’interno del testo selezionato. Interessante in

⁶⁷ Il lavoro di integrazione è in fase di sviluppo e stiamo lavorando alla realizzazione di un prototipo che mostri i vantaggi della ricerca federata. Le potenzialità di indagine avanzate permesse dal linguaggio CQL sono esemplificate nella stringa d’interrogazione che al momento è del tutto funzionale a dimostrare le potenzialità dell’infrastruttura. Grazie all’uso degli strumenti di analisi linguistica e dei modelli per la lingua latina, sarà possibile interrogare i testi della Galassia MQDQ combinando anche i dati di annotazione linguistica. L’attuale sistema di ricerca con lemma prevede un approccio simile a quello implementato dal motore di ricerca testuale Diogenes (<https://d.iogen.es/>).

⁶⁸ <https://voyant-tools.org/docs/#!/guide/start>.

⁶⁹ <https://switchboard.clarin.eu/tools>.

questa visualizzazione è la possibilità di selezionare più parole e verificare la distribuzione, comparando le occorrenze nei diversi segmenti del testo. È evidente, per esempio, come la parola *natura* trovi un massimo relativo nella sezione 4, mentre in questa stessa sezione la parola *ratio* trovi il suo minimo relativo. In basso a sinistra sono presenti le statistiche relative al testo (totale parole, forme uniche, densità del vocabolario, indice di leggibilità, media delle parole per frase, ecc.). Lo *skin*, infine, mostra il classico strumento del Keyword in Context (KWIC) per la parola *natura*.⁷⁰

5.10 Le risorse digitali per lo studio del latino nel contesto delle infrastrutture di ricerca europee: prospettive future

Le principali infrastrutture di ricerca europee che operano nel campo delle scienze umane sono:⁷¹ CLARIN,⁷² ERIC (Research Infrastructure for Language as Social and Cultural Data),⁷³ DARIAH,⁷⁴ ERIC (Digital Research Infrastructure for Arts and Humanities), E-RIHS (European Research Infrastructure for Heritage Science),⁷⁵ e OPERAS (Open Scholarly Communication in the European Research Area for Social Sciences and Humanities).⁷⁶

L'ambito originario di CLARIN, come indicato dalla dicitura in inglese che ne determina l'acronimo (Common Language Resource Infrastructure), era prevalentemente costituito dalla catalogazione e dalla interconnessione di risorse linguistiche, come dizionari, *wordnet*,⁷⁷ o *treebank*.⁷⁸ Tuttavia, l'interesse si è oggi notevolmente ampliato verso le risorse testuali in sé, come espressioni culturali dotate di valore intrinseco, oltre che come campioni casuali o

⁷⁰ Per una lista completa degli strumenti disponibili in Voyant Tools è possibile consultare la documentazione al seguente link: <https://voyant-tools.org/docs/#!/guide/tools>.

⁷¹ Le infrastrutture di ricerca europee sono «strutture, risorse e servizi collegati, utilizzati dalla comunità scientifica per condurre ricerche di alta qualità nei rispettivi campi, senza vincolo di appartenenza istituzionale o nazionale» (D.M. 1082 del 10/09/2021 - PNIR 2021-27, <https://bit.ly/3LTXHgo>).

⁷² <https://www.clarin.eu>.

⁷³ La sigla ERIC indica un European Research Infrastructure Consortium.

⁷⁴ <https://www.dariah.eu>.

⁷⁵ <https://www.e-rihs.eu>.

⁷⁶ <https://operas-eu.org>.

⁷⁷ Una *wordnet* è una risorsa lessico-semantica a doppio accesso: sia semasiologico (dal vocabolo ai concetti che esprime), sia onomasiologico (dal nodo concettuale ai vocaboli che lo esprimono).

⁷⁸ Una *treebank* è un corpus di testi annotati sintatticamente, dove le relazioni sintattiche di dipendenza sono rappresentate da strutture ad albero.

selezionati, funzionali allo studio della lingua. DARIAH opera invece nell’ambito delle DH in senso lato: a questa infrastruttura ricorrono sia università ed enti di ricerca, sia musei, archivi e biblioteche. L’ambito di E-RIHS è la scienza del patrimonio culturale, inteso soprattutto come insieme dei beni materiali di rilevanza culturale, specialmente in campo artistico e archeologico. Infine, OPERAS si focalizza sulla comunicazione e sulla pubblicazione dei risultati della ricerca in campo umanistico (e sociale) ad accesso aperto. Mentre CLARIN e DARIAH si occupano principalmente di dati digitalizzati o nativamente digitali e di strumenti computazionali, E-RIHS prende in considerazione anche metodi e strumenti rivolti allo studio degli oggetti fisici, come manufatti dotati di valore artistico e/o culturale. OPERAS, occupandosi della destinazione dei prodotti dell’editoria scientifica in ambito umanistico, sta a valle del processo di produzione (di dati digitali), elaborazione (tramite strumenti computazionali) ed esposizione (dei risultati in pubblicazione scientifica digitale).

I quattro consorzi italiani afferenti a queste infrastrutture (CLARIN-IT, DARIAH-IT, E-RIHS.it e OPERAS.it), che vedono un forte coinvolgimento del CNR a fianco delle università, si sono recentemente federati in H2IOSC (Humanities and Heritage Italian Open Science Cloud)⁷⁹ per armonizzare gli sforzi verso alcuni obiettivi comuni, fra cui la digitalizzazione e la fruizione online del patrimonio culturale italiano, materiale e immateriale. È in fase di attivazione presso il CNR-ILC un nuovo *data center* che ha lo scopo di supportare la sostenibilità dei progetti di interesse per le scienze umane, mettendo a disposizione spazio disco e potenza di calcolo.

Con l’obiettivo di rendersi sempre più conformi ai principi FAIR all’interno di un’infrastruttura di ricerca, le grandi biblioteche digitali italiane di testi latini hanno provveduto alla metadattazione delle proprie collezioni sul VLO (Virtual Language Observatory)⁸⁰ di CLARIN.⁸¹ Oltre a MQDQ, tali biblioteche digitali sono costituite dal corpus giuridico BIA-Net FONTES (<https://bia.igsg.cnr.it>), dai testi del tardo antico digilib-LT (<https://digiliblt.uniupo.it>) e dai testi medievali del progetto ALIM (<http://alim.unisi.it>).⁸²

BIA-Net FONTES, digilib-LT e MQDQ hanno scelto di depositare

⁷⁹ <https://www.h2iosc.cnr.it>.

⁸⁰ <https://vlo.clarin.eu>.

⁸¹ La metadattazione consiste nella creazione di schede contenenti, fra le altre informazioni, anche le seguenti: a) la descrizione di un corpus o dei singoli documenti che vi appartengono; b) il tipo di licenza d’uso (prevalentemente CC-BY-SA [<https://creativecommons.org/licenses/by-sa/4.0/>] 4.0, ma in alcuni casi CC-BY-NC-SA [<https://creativecommons.org/licenses/by-nc-sa/4.0/>] 4.0 oppure CC-BY-NC-ND [<https://creativecommons.org/licenses/by-nc-nd/4.0/>] 4.0); c) *l’handler*: un URL che identifica in modo permanente la scheda e ne garantisce l’accesso a lungo termine.

⁸² Boschetti et al. (2021) discutono l’incontro di ALIM con il mondo di CLARIN-IT.

presso il repository ILC4CLARIN anche copia delle loro collezioni in formato XML/TEI ai fini della preservazione a lungo termine, per tutta la comunità, non solo dei metadati ma anche dei dati stessi.⁸³

Il VLO permette di filtrare la ricerca per lingua, collezione, tipo di risorsa, modalità, formato, parola chiave, copertura temporale (dei corpora) e tipo di licenza. Poiché i metadati sono inseriti dalle singole istituzioni afferenti ai nodi nazionali di CLARIN, nonostante le indicazioni delle linee guida, sono possibili disomogeneità nella descrizione delle risorse. Per quanto riguarda il latino, le risorse presenti nel VLO attualmente sono poco meno di 4500: la varietà è molto elevata, perché si trovano corpora testuali in formato *plain text* o in formato XML/TEI; corpora annotati linguisticamente (fra gli altri spiccano l'*Index Thomisticus Treebank*, con l'annotazione sintattica di Tommaso d'Aquino e la *PROIEL collection*, con testi poliglotti in parallelo del Nuovo Testamento con annotazione sintattica e pragmatica); opere di singoli autori; risorse lessico-semantiche come *Latin WordNet* all'interno di *MultiWordNet* e le molteplici risorse linguistiche depositate dal CIRCSE (cf. Passarotti, Sprugnoli 2021). Oltre ai dati, il VLO cataloga anche gli strumenti computazionali per l'analisi linguistica semiautomatica. Fra tutti va senz'altro menzionato UDPipe 2,⁸⁴ costituito da una catena di servizi web per il riconoscimento delle parti del discorso (*PoS tagger*), la lemmatizzazione (*lemmatizer*) e l'analisi sintattica secondo l'impianto teorico della grammatica di dipendenza (*dependency parser*).

Per quanto riguarda il latino, diverse sono le potenzialità ancora da esplorare e gli interventi auspicabili: ad esempio, l'accuratezza delle analisi automatiche è ancora notevolmente inferiore rispetto alle lingue moderne, che possono avvalersi di enormi quantità di dati per l'addestramento; inoltre, allo stato attuale non è ancora confluita in CLARIN una batteria di strumenti pensati per venire incontro alle esigenze specifiche del filologo e dello storico della letteratura. Se al linguista una concordanza serve soprattutto per verificare la presenza e la dimensione, in termini di quantità di occorrenze, di un fenomeno attestato nei corpora, al filologo – per menzionare alcune esigenze che caratterizzano gli studi filologici e letterari rispetto agli studi linguistici – serve poter identificare e citare in modo canonico il passo trovato; insieme all'analisi morfosintattica serve l'analisi metrica; serve la possibilità di interrogare la variantistica; servono metadati arricchiti con informazioni sulle fonti primarie, sulle datazioni incerte, sui generi letterari. La sinergia fra MQDQ, BIA-Net

⁸³ I corpora attualmente depositati su CLARIN sono *MQDQ* (<http://hdl.handle.net/20.500.11752/OPEN-555>) e *Poeti d'Italia in lingua latina* (<http://hdl.handle.net/20.500.11752/OPEN-556>).

⁸⁴ <https://hdl.handle.net/11234/1-4816>.

FONTES, ALIM e digilib-LT con CLARIN-IT e la sinergia fra le quattro infrastrutture di ricerca italiane di ambito umanistico all'interno di H2IOSC possono fornire grandi opportunità per studiare il modo migliore di estendere gli strumenti esistenti con nuove funzionalità necessarie per gli studi filologici e letterari.

Abbreviazioni

- CIL = Mommsen, Th. et al. (edd.) (1893-). *Corpus Inscriptionum Latinarum*. Berlin: Apud Georgium Reimerus; De Gruyter.
- CLE = Bücheler, F.; Lommatzsch, E. (edd.) (1895-1926). *Carmina Latina Epigraphica*. 3 vols. Lipsiae: Teubner.
- MQDQ = *Musisque Deoque*. Un archivio digitale di poesia latina (<https://www.mqdq.it/>).
- PLRE = Jones, A.H.M. (1971-92). *Prosopography of the Later Roman Empire*. 3 vols. Cambridge: Cambridge University Press.

Bibliografia

- Bailey, C. (ed.) [1898] (1921). *Lucreti De rerum natura*. Oxonii: e typographeo Clarendoniano.
- Bailey, C. (ed.) (1947). *Titi Lucreti Cari De rerum natura libri sex*. Vol. 2, *Commentary, books I-III*. Oxford: Clarendon Press.
- Barats, C.; Schafer, V.; Fickers, A. (2020). «Fading Away... The Challenge of Sustainability in Digital Studies». *Digital Humanities Quarterly*, 14(3), nr. 484.
- Barbuti, N. (2019). «Ripensare i formati, ripensare i metadati: prove 'tecniche' di conservazione digitale». *Umanistica digitale*, 5, 121-38. <https://doi.org/10.6092/issn.2532-8816/9055>.
- Barchiesi, A. (a cura di) [1980] (2019). *Virgilio: Georgiche*. Milano: Mondadori.
- Berti, M. (ed.) (2019). *Ancient Greek and Latin in the Digital Revolution*. Berlin; Boston: De Gruyter Saur. <https://doi.org/10.1515/9783110599572>.
- Boldrer, F. (ed.) (1996). *L. Junii Moderati Columellae rei rusticae liber decimus: carmen de cultu hortorum*. Pisa: Edizioni ETS.
- Boschetti, F.; Del Gratta, R.; Del Grosso, A.M. (2017). «The Role of Digital Scholarly Editors in the Design of Components for Cooperative Philology». Boot, P.; Cappelletto, A. (eds), *Advances in Digital Scholarly Editing Papers Presented at the DIXIT Conferences in The Hague, Cologne, and Antwerp*. Leiden: Sidestone Press, 249-53.
- Boschetti, F.; Del Grosso, A.M. (2020). «L'annotazione di testi storico-letterari al tempo dei social media». *Italica Wratislaviensia*, 11(1), 65-99. <http://dx.doi.org/10.15804/IW.2020.11.1.03>.
- Boschetti, F.; Del Grosso, A.M.; Spinazzè, L. (2021). «La galassia *Musisque Deoque*: storia e prospettive». Venuti, M.; Manca, M. (a cura di), «*Paulo maiora canamus*». *Raccolta di studi per Paolo Mastandrea*. Venezia: Edizioni Ca' Foscari, 404-19. <https://doi.org/10.30687/978-88-6969-557-5>.
- Boschetti, F.; Del Gratta, R.; Monachini, M.; Buzzoni, M.; Monella, P.; Rosselli Del Turco, R. (2021). «'Tea for Two': The Archive of the Italian Latinity of the

- Middle Ages Meets the CLARIN Infrastructure». Navarretta, C.; Eskevich, M. (eds), *Selected Papers from the Clarin Annual Conference 2020* (Virtual Event, 5-7 October 2020). Linköping: Linköping University Electronic Press, 37-46. <https://doi.org/10.3384/ecp180>.
- Bradley, K. (2007). «Defining Digital Sustainability». *Library Trends*, 56(1), 148-63.
- Bücheler, F. (ed.) (1895-97). *Carmina Latina Epigraphica*, vol. 2(1-2). Lipsiae: Teubner.
- Buonocore, M. (1997). «*Carmina Latina Epigraphica Regionis IV Augusteae*. Avvio ad un censimento». *Giornale Italiano di Filologia*, 49, 21-50.
- Canali, L. (ed.) (1990). *Tito Lucrezio Caro. La natura delle cose*. Milano: Rizzoli.
- Canetta, I. (2011). «*Lauinia/Lauina* (Aen. 1, 2): variante testuale e testimonianze indirette». Mastandrea, P.; Spinazzé, L. (a cura di), *Nuovi archivi e mezzetti d'analisi per testi poetici. I lavori del progetto Musisque Deoque*. Amsterdam: Hakkert, 35-41.
- Carande Herrero, R. (2010). «Carmen o no carmen: problemas de catalogación en *CIL II 2/ 5* y *CIL II 2/7*». *Habis*, 41, 219-39.
- Ceccarelli, L. (2018). *Contributions to the History of Latin Elegiac Distich*. Turnhout: Brepols publishers.
- Chacon, S.; Straub, B. (2014). *Pro Git*. Apress.
- Ciula, A.; Smithies, J. (2023). «Sustainability and Modelling at King's Digital Lab: Between Tradition and Innovation». Nyhan, J. et al. (eds), *On Making in the Digital Humanities*. London: University College London Press, 78-104. <https://doi.org/10.14324/111.9781800084209>.
- Conte, G.B. (ed.) (2009). *P. Vergilius Maro: Aeneis*. Berlin; Boston: De Gruyter.
- Conte, G.B. (ed.) (2013). *P. Vergilius Maro: Georgica*. Berlin; Boston: De Gruyter.
- Conte, G.B. (2016). *Marginalia. Note critiche all'edizione teubneriana di Virgilio*. Pisa: Edizioni della Normale.
- Cooley, A.E. (2012). *The Cambridge Manual of Latin Epigraphy*. Cambridge: Cambridge University Press.
- Courtney, E. (1995). *Musa Lapidaria: A Selection of Latin Verse Inscriptions*. Atlanta: Scholars Press.
- Cugusi, P.; Sblendorio Cugusi, M.T. (edd.) (2008). *Carmina latina epigraphica Moesica (CLEMoes); Carmina latina epigraphica Thraciae (CLEThr)*. Bologna: Pàtron.
- Cugusi, P.; Sblendorio Cugusi, M.T. (edd.) (2014). *Carmina latina epigraphica Africarum provinciarum post Buechelerianam collectionem editam reperta cognita (CLEAfr)*. Faenza: Fratelli Lega.
- Cugusi, P.; Sblendorio Cugusi, M.T. (a cura di) (2016). *Versi su pietra. Studi sui "Carmina Latina Epigraphica": metodologia, problemi, tematiche, rapporti con gli auctores, aspetti filologici e linguistici, edizione di testi. Quaranta anni di ricerche*. Faenza: Fratelli Lega.
- Currier, S.; Barton, J.; O'Beirne, R.; Ryan, B. (2004). «Quality Assurance for Digital Learning Object Repositories: Issues for the Metadata Creation Process». *ALT-J. Research in Learning Technology*, 12(1), 5-20. <https://doi.org/10.1080/0968776042000211494>.
- De Jong, F.; Maegaard, B.; De Smedt, K.; Fišer, D.; Van Uytvanck, D. (2018). «CLARIN: Towards FAIR and Responsible Data Science Using Language Resources». Calzolari, N. et al. (eds), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (Miyakazi, Japan,

- May 7-12, 2018). Japan: European Language Resources Association (ELRA), 3259-64.
- Deufert, M. (ed.) (2019). *Titus Lucretius Carus. De rerum natura libri VI*. Berlin; Boston: De Gruyter.
- Driscoll, M.J.; Pierazzo, E. (eds) (2016). *Digital Scholarly Editing: Theories and Practices*. Cambridge: Open Book Publishers. Digital Humanities Series 4.
- Drucker, J. (2021). «Sustainability and Complexity: Knowledge and Authority in the Digital Humanities», in «Digital Humanities 2019: Complexities», suppl. 2, *Digital Scholarship in the Humanities*, 36, 86-94. <https://doi.org/10.1093/llc/fqab025>.
- Ernout, A. (éd.) [1920] (1978). *Lucrèce. De la nature*, vol. 1. Paris: Les Belles Lettres.
- European Commission, Directorate-General for Research, and Innovation (2016). *Open Innovation, Open Science, Open to the World: A Vision for Europe*. Luxembourg: Publications Office of the European Union. <https://doi.org/doi/10.2777/061652>.
- Feraudi-Gruénais, F. (2010). «An Inventory of the Main Electronic Archives of Latin Inscriptions». Feraudi-Gruénais, F. (ed.), *Latin on Stone. Epigraphic Research and Electronic Archives*. Plymouth: Lexington Books, 157-60.
- Fernández Martínez, C. (1999). «Acercamiento a los CLE: ¿existe un género lapidario?». Aldama Roy, A.M.; Barrio Vega, M.F.; Conde Salazar, M.; Espigares Pinilla, A.; López de Ayala y Genovés, M.J. (eds), *La filología latina hoy. Actualización y perspectivas*. Madrid: Sociedad de Estudios Latinos, 119-26.
- Fernández Martínez, C. (2007). «*Carmina Latina Epigraphica de la Bética Romana: las primeras piedras de nuestra poesía*. Séville: Universidad de Sevilla.
- Fernández Martínez, C. (2021). «The Need for an Innovative Approach to the Study of Latin Epigraphic Poetry». Velázquez Soriano, I.; Espinosa Espinosa, D. (eds), *Epigraphy in the Digital Age, Opportunities and Challenges in the Recording, Analysis and Dissemination of Inscriptions*. Oxford: Archeopress, 9-14.
- Fernández Martínez, C.; Limón Belén, M. (2014). «Latin Epigraphic Poetry Database Project (Hispania & Galliae)». Orlandi, S.; Santucci, R.; Casarosa, V.; Luizzo, P.M. (eds), *Information Technologies for Epigraphy and Cultural Heritage = Proceedings of the First EAGLE International Conference (Paris, September 29-30, October 1, 2014)*. Roma: Sapienza Università Editrice, 369-79.
- Fischer, F.; Börner, I.; Göbel, M.; Hechtel, A.; Kittel, C.; Milling, C.; Trilcke P. (2019). «Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama». *OpenMethods: Highlighting Digital Humanities Methods and Tools*. <https://openmethods.dariah.eu/2021/02/08/programmable-corpora-introducing-dracor-an-infrastructure-for-the-research-on-european-drama/>.
- Fischer, F.; Boschetti, F.; Del Grosso, A.M.; Montefusco, A.; Mancinelli, T.; Macchiarelli A. (2023). «Sinergie fra VeDPH e CNR-ILC in termini di condivisione della conoscenza e sostenibilità dei progetti digitali». Di Maro, M.; Merola, V.; Nocita, N. (a cura di), *DH.22 – Digital Humanities. Per un confronto interdisciplinare tra saperi umanistici a 30 anni dalla nascita del World Wide Web.22*. Roma: L'Erma di Bretschneider, 113-33.
- Flores, E. (1980). *Le scoperte di Poggio e il testo di Lucrezio*. Napoli: Liguori Editore.
- Flores, E. (ed.) (2002). *Titus Lucretius Carus. De rerum natura*. Vol. 1, *Libri 1-3*. Napoli: Bibliopolis.

- Gesner, J.M. (ed.) (1795). *L. Iunii Moderatae Columellae De Re Rustica libri XII curante Io. Matthia Gesnero, Tomus Primus [...]*. Flensburgi: Sumptibus Korteanis.
- Geymonat, M. (ed.) [1973] (2008). *P. Vergili Maronis opera*. Roma: Storia e Letteratura.
- Hamdoune, C. (2011). *Vie, mort et poesie dans l'Afrique romaine: d'après un choix de Carmina Latina Epigraphica*. Bruxelles: Editions Latomus.
- Hamdoune, C. (2016). *Parure monumentale et paysage dans la poésie épigraphique de l'Afrique romaine: recueil de Carmina Latina Epigraphica*. Bordeaux: Ausonius.
- Ireland, R. (1983). «Epigraphy». Henig, M. (ed.), *A Handbook of Roman Art: A Survey of All the Arts of the Roman World*. Oxford: Paidon, 220-33.
- Keil, H. (ed.) (1857-80). *Grammatici Latini*. 7 vols. Lipsiae: in aedibus B.G. Teubneri.
- Krummrey, H. (1964). «Zum Plan einer neuen Sammlung der *Carmina Latina Epigraphica*». *Philologus*, 108, 304-10.
- Kruschwitz, P. (2002). «Überlegungen zum Begriff *Commaticum*. Theorie und Praxis am Beispiel von *CLE Engström 410*». Del Hoyo Calleja, J.; Gómez Pallarès, J. (eds), «*Asta ac pellege*». 50 años de la publicación de las «*Inscriptiones hispanas en verso*», de S. Mariner. Madrid: Signifer Libros, 39-45.
- Kwapisz, J. (2022). «Acrostiche». Urlacher-Becht, C. (éd.), *Dictionnaire de l'Épigramme littéraire dans l'Antiquité grecque et romaine*. Turnhout: Brepols Publishers, 2-4.
- Lachmann, K. (ed.) (1853). *T. Lucreti Cari De rerum natura libri VI*. 2a ed. Berolini: Impensis Georgii Reimeri.
- Lambin, D. (ed.) (1570). *T. Lucretii Cari De rerum natura libri IV*. 2a ed. Lutetiae: apud Ioannem Benenatum.
- Leo, F. (1909). «Doppelfassungen bei Iuvenal». *Hermes*, 44, 600-17.
- Lin, D. et al. (2020). «The TRUST Principles for Digital Repositories». *Scientific Data*, 7(1), 144. <https://doi.org/10.1038/s41597-020-0486-7>.
- Luck, G. (1972). «The Textual History of Juvenal and the Oxford Lines». *Harvard Studies in Classical Philology*, 76, 217-31.
- Mambrini, F. (ed.) (2010). *Lucretius. De rerum natura 4-5. Misisque Deoque. 4*. <https://www.mqdq.it/texts/LVCR|rena|004>; <https://www.mqdq.it/texts/LVCR|rena|005>.
- Manca, M. (2010). «Database and *Corpora* of Ancient Texts Towards the 'Second Dimension': Theory and Practice of MQDQ Project». Anreiter, P.; Kienpointner, M. (Hrsgg), *Latin Linguistics Today: Akten des 15. internationalen Kolloquiums zur Lateinischen Linguistik* (Innsbruck, 4-9 April 2009). Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck, 697-702.
- Martin, J. (ed.) [1934] (1969). *T. Lucreti Cari De rerum natura libri sex*. Lipsiae: in aedibus B.G. Teubneri.
- Martínez, C.F. (2016). «Claves para la interpretacion del carmen epigraphicum de Valeria, cui iuveni mores servarunt fata vetusta». *Zeitschrift für Papyrologie und Epigraphik*, 198, 53-8.
- Masaro, G. (2017). *Iscrizioni metriche e affettive della X regio augustea*. Roma: Aracne.
- Massaro, M. (2012). «Fra poesia e prosa affettiva in iscrizioni sepolcrali (a proposito di nuove raccolte territoriali iberiche in *CLE*)». *Epigraphica*, 74, 277-308.
- Mastandrea, P. (2004). «Poeti d'Italia in lingua latina. Un archivio elettronico da Dante al primo trentennio del XVI secolo». *Atti del Convegno "Il Latino*

- nell'età dell'Umanesimo*" (Mantova, 26-27 ottobre 2001). Firenze: Olschki, 35-50.
- Mastandrea, P. (2008). «Meccanismi della dizione poetica e critica del testo. Appunti per un seminario filologico». *Quaderni del Dipartimento di Filologia Linguistica e Tradizione classica dell'Università di Torino*, n.s. 7, 29-38.
- Mastandrea, P. (2009). «Gli archivi elettronici di *Musisque Deoque*. Ricerca intertestuale e cernita fra varianti antiche (con qualche ripensamento sulla tradizione indiretta dei poeti latini)». Zurli, Mastandrea 2009, 41-72.
- Mastandrea, P. (2012). «Filologia latina e testo elettronico. La ricerca dei prototipi letterari in poesia epigrafica». Palazzolo, N. (a cura di), *Diritto romano e scienze antichistiche nell'era digitale: atti del convegno di studio* (Firenze, 12-13 settembre 2011). Torino: Giappichelli, 231-53.
- Mastandrea, P. (2012). «1991-2011, un ventennio di evoluzione nella filologia 'pratica'». Ciotti, F.; Crupi, G. (a cura di), *Dall'Informatica umanistica alle culture digitali. Atti del convegno di studi in memoria di Giuseppe Gigliozzi* (Roma, 27-28 ottobre 2011). Roma: Sapienza Università Editrice, 151-66.
- Mastandrea, P. (2015). «Archivi elettronici di poesia latina e opzioni multiple di ricerca intertestuale», in «The Mechanic Reader: Digital Methods of Literary Criticism», num. monogr., *Semicerchio. Rivista di poesia comparata*, 53(2), 60-9.
- Mastandrea, P. (2017). «Sui principî della poesia, la ricerca intertestuale con strumenti elettronici», in «Tavola rotonda sul tema filologia digitale: problemi e prospettive (Roma 6 giugno, 2014)», num. monogr., *Contributi del Centro Linceo Interdisciplinare Beniamino Segre*, 135, 73-111.
- Mastandrea, P.; Spinazzè, L. (a cura di) (2011). *Nuovi archivi e mezzi d'analisi per i testi poetici latini*. Amsterdam: Hakkert.
- Mazzacane, R. (1984). «Varianti noniane di Lucrezio». Ferruccio, B.; Barabino, G. (a cura di), *Studi noniani*, vol. 9. Genova: Istituto di Filologia Classica e Medioevale, 149-77.
- Morelli, A.M. (2020). *Le iscrizioni metriche del Latium adiectum. Carmina Latina Epigraphica in Latio Adiecto Reperta (CLEiLAR)*. Tivoli: Tored.
- Munro, H.A.J. (ed.) [1886] (1978). *T. Lucreti Cari De rerum natura libri sex*, vol. 1. New York; London: Garland.
- Mynors, R.A.B. (ed.) (1969). *P. Vergili Maronis opera*. Oxonii: E typographeo Clarendoniano.
- Pagano, M. (2019). «Una iscrizione del V-VI secolo d.C.: dalla cattedrale di Vico Equense (NA)». *Epigraphica*, 81, 692-6.
- Pais, E. (ed.) (1884). *Corporis inscriptionum Latinarum supplementa Italica*. Romae: Ex typis Salviucci.
- Passarotti, M.; Sprugnoli, R. (2021). «The CIRCSE Collection of Linguistic Resources in CLARIN-IT». Monachini, M.; Eskevich, M. (eds), *CLARIN Annual Conference Proceedings* (Virtual Edition, 27-29 September 2021). Utrecht: CLARIN, 44-7.
- Pautasso, C.; Wilde, E.; Alarcon, R. (2013). *REST: Advanced Research Topics and Practical Applications*. New York; Heidelberg; Dordrecht; London: Springer.
- Pierazzo, E. (2015). *Digital Scholarly Editing. Theories, Models and Methods*. Farnham: Ashgate. Digital Research in the Arts and Humanities.
- Pikhaus, D. (1994). *Répertoire des inscriptions latines versifiées de l'Afrique romaine (Ier-VIe siècles)*. Vol. 1, *Tripolitaine, Byzacène, Afrique proconsulaire*. Bruxelles: Epigraphica Bruxellensia.
- Polaria, G. (ed.) (1973). *Publilii Optatiani Porfyrii Carmina*. 2 vols. Torino: Paravia.

- Puglia, E. (2021). «Sul testo della nuova iscrizione della cattedrale di Vico Equense». *Epigraphica*, 83, 618-20.
- Reeve, M.D. (1973). «Gladiators in Juvenal's Sixth Satire». *Classical Review*, 23, 124-5.
- Sabbadini, R. (ed.) (1930). *P. Vergili Maronis opera*. Romae: Typis Regiae officinae polygraphicae.
- Sabbadini, R.; Castiglioni, L. (eds) (1944). *P. Vergilii Maronis Aeneidos libri XII*. Augustae Taurinorum: In aedibus Paraviae.
- Sabbadini, R.; Castiglioni, L. (eds) (1945). *P. Vergilii Maronis Bucolica et Georgica*. Augustae Taurinorum: In aedibus Paraviae.
- Sanders, G. (1991). «Le dossier quantitatif de l'épigraphie latine versifiée». Sanders, G. (éd.), *Lapides memores. Païens et chrétiens face à la mort: le témoignage de l'épigraphie funéraire latine*. Faenza: Fratelli Lega, 207-19.
- Scaligero, G. (ed.) (1573). *Publii Virgili Maronis Appendix cum supplemento multorum antehac numquam excusorum poematum veterum poetarum [...]*. Lugduni: Apud Gulielmum Rovillium.
- Schmidt, M.G. (2014). «Carmina Epigraphica». Bruun, C.; Edmonson, J. (eds), *The Oxford Handbook of Roman Epigraphy*. Oxford: Oxford University Press, 764-82.
- Schneider, F; Kläy, A.; Zimmermann, A.B.; Buser, T.; Ingalls, M.; Peter Messerli, P. (2019). «How Can Science Support the 2030 Agenda for Sustainable Development? Four Tasks to Tackle the Normative Dimension of Sustainability». *Sustain Sci*, 14, 1593-604. <https://doi.org/10.1007/s11625-019-00675-y>.
- Spinazzé, L. (2011). «Risalire alle fonti: dall'edizione *Musisque Deoque* ai testimoni manoscritti». Mastandrea, Spinazzé 2011, 59-71.
- Stok, F. (2016). «Sabbadini editore di Virgilio». Stok, F.; Tomé, P. (a cura di), *La filologia classica e umanistica di Remigio Sabbadini*. Pisa: ETS, 241-58.
- Tucker, J. (2022). «Facing the Challenge of Digital Sustainability as Humanities Researchers.» *Journal of the British Academy*, 10, 93-120.
- Venuti, M. (2011). «L'apparato digitale di Virgilio». Mastandrea, Spinazzé 2011, 29-34.
- Wilkinson, M.D. et al. (2016). «The FAIR Guiding Principles for Scientific Data Management and Stewardship». *Scientific Data*, 3(1), nr. 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Willis, J.A. (1989). «Juvenalis Male Auctus». *Mnemosyne*, 42, 441-68.
- Winstedt, E.O. (1899). «A Bodelian Ms of Juvenal». *Classical Review*, 13, 201-5.
- Zenaro, S.; Del Grosso, A.M.; Boschetti, F.; Ranocchia G. (2022). «Verso la definizione di criteri per valutare soluzioni di *Scholarly Editing* Digitale: Il caso d'uso *GreekSchools*». Ciraci, F.; Miglietta, G.; Gatto, C. (a cura di), *AIUCD 2022 - Culture Digitali. Intersezioni, Filosofia, Arti e Media = Proceedings della 11ª conferenza nazionale* (Lecce, 1-3 giugno 2022). Lecce: Associazione per l'Informatica Umanistica e la Cultura Digitale, 20-5. <https://doi.org/10.6092/unibo/amsacta/6848>.
- Zurli, L.; Mastandrea, P. (a cura di) (2009). *Poesia latina, nuova e-filologia. Opportunità per l'editore e per l'interprete = Atti del Convegno Internazionale* (Perugia, 13-15 settembre 2007). Roma: Herder editrice e libreria.

La ricerca sulle ultime lettere della Resistenza dagli anni Cinquanta a oggi

Igor Pizzirusso

Istituto nazionale Ferruccio Parri, Milano, Italia

Abstract *Ultime lettere di condannati a morte e di deportati della Resistenza italiana database* is a no longer very recent tool, yet of great vitality and usefulness, as evidenced by the sporadic but constant additions of documents, access statistics and feedback from the public, often of non-specialists. The project stems from the book (and the related research) *Lettere di condannati a morte della Resistenza italiana*, published in the 1950s and edited by Piero Malvezzi and Giovanni Pirelli, but benefits from the new context of both technology and culture in the 2000s. This article, starting with a brief examination of the nature and contexts of production of such epistolary, aims to analyse these important differences and their consequences in terms of outcomes. Primarily, the database made it possible to trace more accurately the origin of the documents. But, above all, it has allowed to reconstruct more faithfully the texts (not infrequently subject to omissions or alteration at the transcription stage) and to add important 'context' information (such as writing supports or the way of writing and occupying the spaces on the sheet) in the frequent cases in which it has been possible to recover the autograph originals of the letters. This article will also highlight some of the evidence that the database has made possible through the inclusion of search filters and a form for searching for words and expressions in the text of the messages.

Keywords Resistance. World War II. Epistolary. Semantic web. Digital history.

Sommario 1 Gli epistolari della Seconda guerra mondiale. – 2 L'«Ultima lettera» di un condannato a morte. – 3 Le ultime lettere dei condannati a morte della Resistenza. – 4 Distribuzione geografica. – 5 Il database *Ultime lettere*. – 5.1 L'origine della banca dati. – 5.2 I *corpora* documentari. – 5.3 Le due ricerche. – 6 Il contenuto delle lettere. – 7 La materialità: l'importanza dei supporti. – 8 Conclusioni.



Peer review

Submitted 2023-02-23
Accepted 2023-06-12
Published 2023-08-02

Open access

© 2023 Pizzirusso |  4.0



Citation Pizzirusso, I. (2023). "La ricerca sulle ultime lettere della Resistenza dagli anni Cinquanta a oggi". *magazén*, 4(1), 121-144.

1 Gli epistolari della Seconda guerra mondiale

Nella produzione editoriale sul secondo conflitto mondiale, accanto a racconti e romanzi di autori quali Fenoglio, Calvino o Meneghello (per citarne solo alcuni), uno spazio importante, anche se non numericamente rilevante, è occupato dalle raccolte epistolari.

La più celebre è forse *L'ultimo fronte* (1971), in cui Nuto Revelli ha messo insieme le lettere scritte dagli alpini piemontesi sulle linee del Don, prima di cominciare la ritirata (e in molti casi perire) nel gelo russo. Le missive, chiuse in sedici sacchi, erano per altro state destinate al macero dalla burocrazia militare, con scarsa lungimiranza e poca consapevolezza dell'importanza di un corpus documentario, composto da

Diecimila lettere [scritte da], migliaia di uomini che parlano, che raccontano. Alcuni dicono quasi tutto, disegnano l'arco completo della loro vita militare; altri, con testimonianze frammentarie, ricostruiscono poche pagine della loro esperienza e non sempre le più importanti; altri ancora, e sono una folla, si inseriscono nel discorso per fermare soltanto un problema, uno stato d'animo, un momento. (Revelli 1971, XXXIX-XL)

Oltre alle lettere dal fronte, una produzione consistente riguarda le raccolte di documenti scritti dai militari durante la prigionia, che dunque non è una condizione sufficiente a interrompere il canale di comunicazione verso casa (Franzini 2004, 518) (o quanto meno a inibire il desiderio di farlo).¹ Si tratta tuttavia, nella grande maggioranza dei casi, di raccolte circoscritte all'ambito locale. Eppure nulla viene tolto al loro valore di testimonianza, di memoria, di documentazione, benché parziale.

Se dunque, come abbiamo visto, non mancano gli epistolari dal fronte o dall'internamento, altrettanta abbondanza non si riscontra per quanto riguarda la vita partigiana.² È d'altra parte perfettamente logico: far parte della Resistenza significa essere, di fatto, un

¹ Tra i più significativi epistolari di internati militari si possono annoverare: Serpelon, A. (a cura di) (1988). *Lettere di caduti e reduci del Cadore nella Seconda guerra mondiale*. Venezia: Istituto veneto per la storia della Resistenza; Zaghi, V. (1996). *Lettere dal Lager. Soldati e internati militari polesani nella Seconda guerra mondiale*. Rovigo: Minelliana; e (2004). *Lettere dai campi di battaglia e di prigionia*. Finale Ligure: Istituto Storico della Resistenza e dell'Età Contemporanea della Provincia di Savona. Relativamente agli epistolari dei deportati politici è giusto menzionare Borgato, D. (2002). *Non si poteva dire di no. Prigionia e Lager nei diari e nella corrispondenza di un'interdetta Venezia-Mauthausen-Linz 1944-1945*. A cura di M. Tommasi. Verona: Istituto veronese per la storia della Resistenza e dell'Età contemporanea.

² Per contro i militari della RSI scrivono più o meno regolarmente ai loro parenti, come scrive Franzini (2004, 518).

fuorilegge, un 'ribelle' destinato a una vita precaria e piena di incertezze, costantemente alla macchia e a caccia di nascondigli sicuri, con il rischio continuo di essere braccato e catturato dai nazisti o dai fascisti (Franzinelli 2004, 518). Le normali condizioni per comporre un epistolario, sia psicologiche che pratiche, vengono perciò a mancare. Oltre alle ovvie difficoltà logistiche nel far recapitare le lettere, bisogna anche considerare che qualsiasi messaggio destinato ai propri cari è un potenziale elemento di compromissione e di pericolo per il mittente e per i destinatari (518), questi ultimi oggetto di possibile rappresaglia nel caso in cui vengano trovati in possesso di documenti tali da far scaturire un'accusa di intelligenza con il nemico.

Il rischio è ben espresso da Nicolò Marino, in quello che è poi divenuto il suo testamento spirituale:³

Gianna carissima, finalmente dopo tanto tempo dalla mia partenza da Torino per Piani Auti, oggi ho la possibilità di farti avere mie nuove causa che scrivere è molto pericoloso per te e per me.⁴

Salvo rare eccezioni, dunque, l'esperienza partigiana sconsiglia il contatto epistolare con la famiglia, anche in caso di cattura e prigionia. Il controllo della corrispondenza da parte dei carcerieri è feroce, perché tra le righe di un banale (in apparenza) messaggio scritto a casa possono esserci indizi per trovare e catturare altri ribelli o anche solo strumenti di pressione sullo stesso prigioniero: più delle torture fisiche, può spesso spingere a tradire i compagni di lotta la minaccia che venga fatto del male ai propri cari.

Un caso del tutto peculiare sono infine i messaggi scritti nell'imminenza dell'esecuzione e di una morte che sia in qualche modo annunciata. Si tratta sempre di documenti infrequenti (e ne vedremo in seguito le ragioni), ma composti con uno stato d'animo e in una situazione che non di rado può far cadere le cautele e le remore che caratterizzano normalmente la condizione di prigionia.

2 L'ultima lettera' di un condannato a morte

Nella casistica contemplata negli epistolari, l'ultimo scritto di un condannato a morte è pertanto un messaggio del tutto peculiare. Come

³ Documenti scritti da persone che non si trovavano al momento della scrittura della lettera di fronte a una condanna a morte o a una consapevole certezza della morte imminente in considerazione del contesto e degli eventi. Per approfondire si veda «Testamenti spirituali». *Ultime lettere di condannati a morte e di deportati della Resistenza italiana*. https://www.ultimelettere.it/?page_id=105.

⁴ «Lettera di Nicolò Marino alla Moglie scritta in data 19-10-1943 da Forno Canavese». https://www.ultimelettere.it/?page_id=52&ricerca=494&doc=584.

scrive Gustavo Zagrebelsky nella sua prefazione al volume *Lettere di condannati a morte della Resistenza italiana 8 settembre 1943-25 aprile 1945* (Malvezzi, Pirelli 2002) si tratta di lettere

concepite nel momento più solenne della vita, un momento che non a tutti è dato di vivere, quando in piena lucidità e coscienza si è faccia a faccia con se stessi, in presenza della morte, erano indirizzate alla cerchia delle persone più vicine e care, in cui sono riposti gli affetti e da cui nasce l'impegno civile. Chiedono conforto, memoria e anche perdono per una scelta compiuta che è causa di dolore, spiegata e giustificata come decisione necessaria a un valore superiore. [...] questi testi sconvolgenti parlano di esseri umani negli ultimi istanti della loro vita, nell'attesa consapevole della fine per mano di altri esseri umani. Ogni facoltà spirituale deve essere stata provocata fino all'estremo. (Zagrebelsky 2002, V)

È dunque con estrema cautela, sostiene sempre Zagrebelsky, che bisogna approcciare volumi o fondi archivistici che raccolgono questo particolare tipo di documenti. Bisogna avere la consapevolezza che in questo, più che in qualsiasi altro caso, leggere e pubblicare questi scritti significa invadere la sfera privata di un individuo, accedendo (e diffondendo) i suoi ultimi e più intimi pensieri, destinati solo allo sguardo dei suoi cari o di una famiglia. È come essere autori di «qualcosa di simile a una profanazione» (VI).

D'altro canto, nella sua presentazione della banca dati *Ultime lettere*, Gianni Perona afferma che

l'ultima lettera di un condannato a morte, specialmente nei contesti di persecuzione a carattere politico o religioso, ha assunto ripetutamente nella storia europea un carattere di documento insieme privato e pubblico, di rivendicazione d'innocenza e di testimonianza, capace di trasmettere ai posteri un'immagine onorevole per la famiglia ed esemplare per i compagni di fede. Il raccogliere tali lettere approdò spesso - ad esempio presso i protestanti nelle guerre di religione - alla costituzione di un vero e proprio martirologio, e certamente a questo obiettivo mirarono anche quei militanti dei movimenti di resistenza che un po' dappertutto nell'Europa appena liberata, talora a guerra non ancora finita, ricercarono gli estremi messaggi dei loro compagni, e li riunirono e pubblicarono. (Perona 2006)

L'intento pubblico, ovvero costruire un pantheon di martiri civili per consolidare le fondamenta della memoria collettiva di una nazione, è quindi una motivazione sufficientemente forte per giustificare quella profanazione di cui parla Zagrebelsky.

3 Le ultime lettere dei condannati a morte della Resistenza

Come si è accennato nei paragrafi precedenti, nella Resistenza italiana le 'ultime lettere' sono relativamente poco consuete. S'è già descritta la pericolosità di scrivere messaggi dal carcere in un contesto persecutorio, ma si è anche sottolineato che in punto di morte questa inibizione va parzialmente a cadere, almeno in chi sta per terminare la sua vita e intende dare un ultimo saluto ai propri cari.

Quali sono dunque le ragioni della loro rarità? Ebbene, esse risiedono nelle stesse condizioni in cui le lettere sono state generalmente prodotte. La loro stesura presuppone innanzitutto delle condizioni almeno temporanee di legalità, in cui a un detenuto sia annunciata, con pur limitato preavviso, l'imminente fine (Perona 2006). Ciò, durante la guerra di liberazione, avviene in realtà molto di rado. La possibilità di scrivere un messaggio d'addio e di ricevere i conforti religiosi, ad esempio, è sistematicamente negata in occasione dei rastrellamenti (Franzini 2004, 536), dopo i quali i partigiani catturati vengono immediatamente passati per le armi. Solo in pochi casi (e decisamente non di prassi), chi viene catturato è sottoposto a un processo sommario, tenuto davanti a qualche improvvisato tribunale militare, al termine del quale viene emessa la sentenza di morte. E tuttavia, quand'anche nel condannato vi sia la consapevolezza del proprio destino, molto spesso costoro non sono in grado di comporre alcun messaggio, dopo giorni di torture e di prigionia in condizioni estreme (e ammesso che si possieda poi il livello di alfabetizzazione necessaria). Nei casi migliori e più fortunati, la trasmissione dell'ultimo pensiero è affidata all'opera generosa e volontaria di qualche trascrittore, come un parroco o un capellano militare.

Molto spesso, poi, ai morituri non sono concessi né gli strumenti (carta e penna o matita) né il tempo per la scrittura. Infine un certo peso, meno determinante ma comunque importante, è dato anche dalla mancata garanzia della trasmissione del messaggio ai destinatari: nel dubbio, alcuni scelgono di non scrivere; altri invece, desiderosi di comunicare con l'esterno informazioni importanti o di lasciare meramente una traccia di sé stessi, lo fanno ugualmente, magari cercando di far uscire gli scritti clandestinamente (con il rischio però di fidarsi delle persone sbagliate e di compromettere, come si è già accennato, sé stessi e i destinatari).

Emblematico è il caso di Andrea Mensa, che compone una serie di tre biglietti⁵ delle dimensioni di un odierno post-it, indirizzati ai compagni di lotta. Le sue parole servono alla causa della lotta partigiana

5 «Lettera di Andrea Mensa (Mirto) ai Compagni di lotta». https://www.ultimelettere.it/?page_id=52&ricerca=269&doc=730.

e hanno carattere eminentemente operativo, per questo l'autore sceglie un formato che possa essere nascosto facilmente e fatto quindi uscire dalla prigione con minor rischio per l'incaricato alla consegna.

Le lettere affidate ai canali 'ufficiali' sono invece sottoposte al vaglio della censura. Esiste quindi sempre il rischio che lo scritto non sia recapitato o che vengano operati decisi interventi sul contenuto. Avviene così, ad esempio, per la lettera che Ezio Cesarini (fucilato al poligono di tiro di Bologna il 27 gennaio 1944) scrive dal carcere al fratello Mario⁶ e che manca completamente di buona parte della seconda facciata; o per il messaggio che Pier Luigi Mazzoletti (ucciso nei pressi di Cormons il 2 ottobre 1944) compone per la moglie da Gradisca di Isonzo,⁷ nel quale sono presenti alcuni punti illeggibili perché cancellati dalla censura, come lo stesso autore precisa nel terz'ultimo foglio; o ancora lo scritto di Zosimo Marinelli (lui pure fucilato al poligono di tiro di Bologna il 27 gennaio 1944), anche in questo caso destinato alla moglie,⁸ dove sono omesse tre righe, alla stessa maniera di quanto avvenuto per Cesarini.

In altri casi i documenti potrebbero essere conservati dagli stessi aguzzini: l'originale della lettera di Giovanni Battista Vighenzi è stato infatti reperito tra le carte del Comandante delle SS Alois Thaler.⁹ Si tratta verosimilmente di episodi rari, vista la prevedibile sfiducia dei carcerati ad affidare queste lettere ai propri carcerieri o la scarsa propensione di questi ultimi a conservarle dopo averle trattenute alla censura o recuperate dal corpo dai cadaveri. La pietà e l'empatia mostrate da Thaler (ammesso che di questo si sia effettivamente trattato) non sono state storicamente frequenti, durante l'occupazione nazista in Italia.

Ci sono poi situazioni in cui si ha la certezza che lettere esistano e siano state scritte, ma a oggi risulta impossibile reperirle perché inserite nelle bare degli scriventi (è così per Giuseppe Bassani, Mario Molteni e Carlo Prina). Altre sono state poi certamente allegate alle pratiche per il riconoscimento partigiano: una ricerca nei fascicoli del fondo Ricompart¹⁰ sarebbe in questo senso opportuna.

6 «Lettera di Ezio Cesarini al Fratello Mario scritta in data 27-01-1944 da Carceri di Bologna». http://www.ultimelettere.it/?page_id=52&ricerca=107&doc=142.

7 «Lettera di Pier Luigi Mazzoletti alla Moglie scritta in data 2-10-1944 da Gradisca d'Isonzo». http://www.ultimelettere.it/?page_id=52&ricerca=204&doc=114.

8 «Lettera di Zosimo Marinelli alla Moglie scritta in data 27-01-1944 da Carceri di Bologna». http://www.ultimelettere.it/?page_id=52&ricerca=115&doc=146.

9 «Lettera di Giovanni Battista Vighenzi (Sandro Biloni) alla Moglie scritta in data 27-04-1945». http://www.ultimelettere.it/?page_id=52&ricerca=619&doc=787.

10 Con la sigla Ricompart si indica il fondo Archivio per il servizio riconoscimento qualifiche e per le ricompense ai partigiani, che conserva la documentazione prodotta dalle Commissioni regionali istituite nell'immediato dopoguerra (1945 e 1948) e dalla Commissione unica nazionale istituita nel 1968. Il fondo archivistico era conservato presso il Ministero della Difesa, che lo ha versato all'Archivio Centrale dello Stato

E tuttavia, come sostiene Gianni Perona (2006), se ne ricaverebbe un incremento numerico poco significativo in rapporto al numero complessivo di vittime: probabilmente di qualche decina fra i circa 25.000 civili e partigiani disarmati uccisi dai nazifascisti (escludendo quindi i partigiani morti in battaglia, con i quali il conto salirebbe a circa ottantamila).¹¹

Un simile scenario aiuta a comprendere molto bene perché questi esemplari di documento siano così rari, perché la loro ricerca sia stata (e sia tuttora) così faticosa e perché, spesso, di eventi molto importanti e tragici, quali ad esempio la deportazione politica e razziale, ci è stata trasmesso solo qualche frammento: un biglietto avvolto intorno a un sasso e gettato dal vagone di un trasporto tedesco diretto a Mauthausen, come fatto da Tullio Degasperi;¹² o un messaggio inciso con una punta di spillo tra le prime e le ultime pagine di una bibbia reperita sul luogo della sua esecuzione, come nel caso di Willy Jervis;¹³ a volte solo un nome e un cognome che, segnando le spoglie, permettessero l'identificazione futura. E spesso nemmeno quella, come dimostrano i molti casi di 'anonimi' delle Fosse Ardeatine,¹⁴ che, nonostante gli sforzi per decifrare i documenti ormai usurati e quasi illeggibili, rimangono purtroppo ignoti.

4 Distribuzione geografica

A conferma di quanto esposto finora si può prendere in considerazione la distribuzione geografica degli scritti, così ineguale e disequilibrata. Se la differenza tra nord e sud è tutto sommato comprensibile, analizzando l'evoluzione della Seconda guerra mondiale, stupisce di più notare come l'area nordoccidentale sia estremamente preponderante rispetto a quella nordorientale.

negli anni 2009-12. I documenti sono oggi digitalizzati nella banca dati *Partigiani d'Italia* (<https://partigianiditalia.cultura.gov.it/archivio/>).

11 Le risultanze del lavoro condotto nel dopoguerra dalle Commissioni riconoscimento qualifiche partigiani e patrioti indicano in 62.070 il numero dei partigiani caduti e in 14.350 i civili uccisi dai nazifascisti. Dai dati dell'*Atlante delle stragi naziste e fasciste in Italia* (www.straginazifasciste.it) risultano 15.500 civili e oltre 7.000 partigiani uccisi dopo il disarmo e la cattura.

12 Ad esempio «Lettera di Tullio Degasperi (Ivan) a moglie e familiari scritta in data 1-02-1944 dalla tradotta ferroviaria per Mauthausen». http://www.ultimelettere.it/?page_id=52&ricerca=181&doc=483.

13 «Lettera di Guglielmo Jervis (Willy) alla Moglie e ai figli scritta in data 5-08-1944». http://www.ultimelettere.it/?page_id=52&ricerca=252&doc=858.

14 Benché ormai vi sia assoluta certezza sull'identità di tutti i 335 fucilati alle Fosse Ardeatine, alcuni dei messaggi reperiti sui cadaveri rimangono di autore indefinito, sia perché le salme non sono riconoscibili, sia perché gli stessi messaggi non sono decifrabili, anche a causa dell'usura.

Sia che consideriamo il luogo di nascita degli scriventi che il luogo di morte, il Piemonte è la regione più rappresentata: ben 130 lettere fra quelle finora reperite sono state scritte da questa regione. Molta incidenza è data dall'area di Torino e dal ruolo dei cappellani militari alle Carceri nuove, che offrivano loro stessi carta e penna ai condannati al momento di ricevere la loro confessione, persuadendoli a scrivere un ultimo messaggio ai propri cari. Non è infatti un caso che per molti dei morituri incarcerati a Torino si conoscono e conservano anche più lettere (fino a sei, come nel caso di Ferreira)¹⁵ composte tutte poche ore prima dell'esecuzione. La loro trasmissione fino ai giorni nostri deve poi molto all'opera di Padre Ruggero Cipolla, cappellano militare delle Nuove dal novembre 1944, e al volume da lui pubblicato a guerra appena finita (Cipolla 1945).

Le lettere scritte in Veneto sono invece solo 24 (ma sono 40 i veneti che riescono a comporle). Qui la presenza nazista è più numerosa, funzionale a presidiare le vie di comunicazione con (e di fuga verso) la madrepatria e a monitorare il fronte jugoslavo, che a differenza di quello francese è decisamente più 'caldo', almeno fino allo sbarco alleato in Costa Azzurra dell'estate 1944. Questo si estrinseca in parte con una maggiore ferocia repressiva, tale da negare anche concessioni minime ai morituri. È però vero che, analizzando l'andamento degli episodi di violenza letale nella zona (Dogliotti 2016, 113-14), emerge che la metà delle stragi e delle uccisioni avviene da gennaio a maggio 1945, quindi nell'ultimo periodo di guerra e a ridosso o in concomitanza con la ritirata. È, questo, un frangente spesso confuso, nel quale l'azione dell'occupante si fa più erratica e disorganica, in cui il controllo centralizzato del territorio lascia spazio all'azione dei singoli reparti o dei singoli soldati. Sotto una lente simile, si può dunque effettivamente affermare che la ferocia repressiva è maggiore e le concessioni ai prigionieri (condannati a morte o meno che siano) sono tra le ultime priorità dell'occupante.

Diverso problema sembra essere invece quello dell'area meridionale. Qui è lecito ipotizzare che la difficoltà delle comunicazioni abbia pesato durante la guerra sul destino dei messaggi che possono essere stati inviati da alcuni fra le migliaia di caduti originari del sud e delle isole, e che una rinnovata campagna di ricerche riporti alla luce una documentazione troppo dispersa per poter essere reperita in tempi brevi. E tuttavia si tratterebbe potenzialmente, anche in questo caso, di pochissimi ritrovamenti (Perona 2006).

15 «Pietro Ferreira». https://www.ultimelettere.it/?page_id=35&ricerca=215.

5 Il database *Ultime lettere*

5.1 L'origine della banca dati

La banca dati *Ultime lettere*¹⁶ è stata pubblicata dall'Istituto nazionale Ferruccio Parri nell'ormai lontano 2006. Si è trattato, all'epoca, di un avvenimento piuttosto peculiare e innovativo, perché per la prima volta l'Istituto metteva online uno strumento digitale atto a restituire non un catalogo bibliografico o archivistico o una schedatura di dati semplici, bensì di un prodotto che avesse anche (se non soprattutto) un forte carattere divulgativo. Partendo dal presupposto della pubblicazione cartacea curata da Pieri Malvezzi e Giovanni Pirelli nel 1952 (Malvezzi, Pirelli 1952), che ha avuto tanto successo da essere poi ristampata nel corso degli anni un'altra ventina di volte,¹⁷ la banca dati *Ultime lettere* ne riproduce in linea di massima la struttura, con un breve profilo biografico del morituro e la trascrizione integrale del testo della lettera. A esse però, il database aggiunge alcuni elementi che sono tipici di un prodotto digitale. Innanzitutto la scansione dei documenti originali, laddove è stato possibile recuperarli, o delle loro trascrizioni. In secondo luogo la mappatura di alcuni dati significativi degli autori delle lettere, che è possibile suddividere per luogo di nascita o morte, studi e professioni, ma anche esperienze militari e appartenenza alle formazioni partigiane. Si può così scoprire, ad esempio, che circa il 25% dei condannati a morte o deportati lavorava come operaio, ma altrettanti appartenevano alle cosiddette 'professioni liberali' (architetti, avvocati, ingegneri, insegnanti, medici, ecc.) o al novero degli ufficiali o degli impiegati pubblici e privati.¹⁸ Emerge poi che grossomodo uno su quattro era (o era stato) arruolato nelle forze armate (con una grande preponderanza dell'esercito).¹⁹ O che il livello di istruzione fosse generalmente sopra alla media, quando noto, con una grande prevalenza di diplomati e laureati.²⁰ A ben riflettere è un fatto logico: maggiore è l'alfabetizzazione, maggiore è la capacità (e la volontà) di scrivere; e tuttavia i numeri delle vittime di cui non si hanno notizie sul percorso

16 I dati esposti nel paragrafo precedente, così come i documenti menzionati in quello prima, sono contenuti nella banca dati *Ultime lettere di condannati a morte e di deportati della Resistenza italiana* (www.ultimelettere.it). Per comodità, d'ora in avanti la chiameremo semplicemente *Ultime lettere*.

17 Quattordici edizioni nei *Saggi* Einaudi, quattro negli *Struzzi*, almeno una nei *Tabacchi*, nei *Millenni* e negli *ET Saggi*.

18 «Professione». http://www.ultimelettere.it/?page_id=140.

19 «Appartenenza alle Forze Armate». http://www.ultimelettere.it/?page_id=127.

20 «Studi». https://www.ultimelettere.it/?page_id=141.

di studi (il 60%) rende difficile tracciare una statistica significativa. Un dato interessante è infine l'età al momento della morte: per circa la metà degli autori dei messaggi era compresa tra i 16 e i 25 anni.²¹

5.2 I corpora documentari

Si è citato poco sopra il volume di Malvezzi e Pirelli e non certo per caso. L'idea iniziale, quella da cui trae origine la realizzazione della banca dati *Ultime lettere*, è proprio rendere disponibili online gli ultimi scritti di partigiani, antifascisti e deportati politici e razziali conservati nel fondo donato all'Istituto nazionale da Piero Malvezzi negli anni Ottanta,²² in cui è conservato tutto il materiale della ricerca che sta alla base della pubblicazione cartacea. A questo materiale, nel database si aggiunge la raccolta effettuata da Mimmo Franzinelli nel 2003-04,²³ anch'esso propedeutico a un libro, ovvero *Ultime lettere di condannati a morte e di deportati della Resistenza* (Franzinelli 2005).

A queste due donazioni archivistiche si è poi aggiunto un numero non trascurabile di ulteriori acquisizioni, avvenute sia contestualmente alle operazioni di *data entry* all'interno di *Ultime lettere* (esaminando altri fondi archivistici depositati in Istituto o negli istituti a esso associati),²⁴ sia dopo la pubblicazione della banca dati stessa, grazie alla solerzia di alcuni famigliari o archivi privati che hanno inviato copia digitale del loro posseduto.

In altri casi, l'ulteriore supplemento di ricerca - che perdura tutt'oggi, malgrado le scarse disponibilità di tempo e di risorse - ha consentito di reperire documenti originali e autografi laddove, in precedenza, erano note solo trascrizioni a macchina o, nel migliore dei casi, copie fotografiche in bianco e nero. Questo aspetto sarà analizzato più nello specifico in seguito.

5.3 Le due ricerche

La ricerca pionieristica di Giovanni Pirelli e di Piero Malvezzi e la seconda indagine condotta da Mimmo Franzinelli hanno comunque

²¹ «Età». https://www.ultimelettere.it/?page_id=114.

²² Archivio dell'Istituto nazionale Ferruccio Parri, Fondo Malvezzi Piero Lettere di condannati a morte della Resistenza italiana e europea.

²³ Archivio dell'Istituto nazionale Ferruccio Parri, Fondo Raccolta Franzinelli/Ultime lettere di condannati a morte e di deportati della Resistenza 1943-45.

²⁴ Per maggiori dettagli sulla rete Parri consultare «Chi siamo». *Portale Rete Parri*. <https://www.reteparri.it/chi-siamo/>.

prodotto la parte più consistente delle lettere pubblicate nel database, anche se in modi che non potrebbero essere più diversi tra loro.

Pur se donato all'Istituto negli anni Ottanta, il Fondo Malvezzi è composto da documenti raccolti per la maggior parte negli anni Cinquanta e primi anni Sessanta. Tra le 'ultime lettere' in esso conservate abbondano le trascrizioni a macchina; solo nei fascicoli relativi agli anni Settanta è possibile trovare qualche rara fotografia in bianco e nero. Il fondo donato da Mimmo Franzinelli è invece composto per lo più da supporti digitali. Ci sono moltissime scansioni di originali e, nei casi in cui questo non si verifica, esiste chiara indicazione di dove reperirli.

5.3.1 Differenze tecnologiche

Questa dissonanza tra i due corpora ha ragioni ovvie e meno ovvie, connesse al diverso momento in cui essi sono stati raccolti. Innanzitutto, vi sono questioni meramente tecniche. Negli anni Cinquanta le trascrizioni erano l'unica possibilità per riuscire ad avere copia dei documenti. Le macchine fotografiche erano appannaggio ancora di pochi e i vari ricercatori che collaboravano con Malvezzi e Pirelli non ne possedevano. Negli anni Duemila invece siamo già in piena epoca digitale; anche la fotografia ormai si sta svincolando dalla pellicola. Nella peggiore delle ipotesi, ci sono comunque le copie fotostatiche, che permettono - con tutte le cautele del caso per non rovinare i documenti - di ottenere una copia fedele dell'originale.

5.3.2 Contesto socioculturale della ricerca di Malvezzi e Pirelli

Non è però solo la tecnica a marcare la differenza. Un ruolo fondamentale lo ha anche il diverso contesto socioculturale. Negli anni Cinquanta, Malvezzi e Pirelli non chiedevano neppure ai propri ricercatori di recuperare gli originali,²⁵ ben consci del valore che questi documenti avevano per chi li conservava e del fatto che, di fronte a una simile istanza, molti di loro si sarebbero inibiti e avrebbero finito per non rendere disponibili le lettere nemmeno per una trascrizione. Esisteva infatti allora un legame molto stretto tra il possessore del documento e il documento stesso, perché molto stretto e diretto era il legame con chi lo aveva scritto: figli, genitori, mogli o mariti, fratelli o sorelle. La lettera era una reliquia, l'ultimo lascito tangibile del congiunto defunto.

25 Emerge chiaramente dal carteggio tra loro e con i loro collaboratori, conservato proprio nel Fondo Malvezzi Piero Lettere di condannati a morte della Resistenza italiana e europea.

Il concetto è espresso in modo perfetto proprio in uno dei documenti, scritto dal diciottenne Mario Marveggio ai genitori: «questo mio scritto sarà il ricordo che voi gelosamente custodirete, in cielo ci rivedremo». ²⁶ Spesso, per il suo contenuto, si tratta anche di un conforto per i destinatari, laddove il condannato sia riuscito a inserire espressioni consolatorie capaci, almeno in parte, di lenire il dolore della perdita e lasciare di sé un ricordo positivo, se non addirittura eroico, nei congiunti (Franzinelli 2004, 537). Il ventenne Walter Del Carpio, ad esempio, scrive: «E tu babbo, sappiti rassegnare come già lo sono io, vedi che la mia mano non trema». ²⁷ Anche Ivan Silvestrini, coetaneo dei Del Carpio, non trema, anche se in lui non vi sono parole di rassegnazione: «Vengo fucilato ma non tremo e come non tremo io non lo dovete nemmeno voi». ²⁸ La rassegnazione c'è invece in Rino Gualandi, ma legata alla fede, che in molte lettere (e lo vedremo meglio in seguito) rappresenta un architrave del discorso: «Siate sereni. Abbiate fede, pregate per me. Ci rivedremo in cielo». ²⁹

Più semplicemente si tratta, in tutti i casi, dell'ultimo lascito dei propri cari, dell'ultima traccia lasciata nella vita terrena. È ancora il caso di Gualandi a renderlo esplicito, nel racconto della sorella Natalina:

Passo tre giorni convinta che mio fratello sia deceduto nel mitragliamento, ma trovo strano il comportamento di mio padre; non una parola riguardo la disgrazia, non tocca cibo e sta giorno e notte rincantucciato nella stalla completamente distrutto; cerco di avvicinarlo e lo trovo che bacia e bagna di lacrime un piccolo foglio che appena mi vede tenta di nascondere; gli chiedo il perché di questo ed è così che papà grida come impazzito che Rino è stato fucilato, mi dà il foglio e mi dice che quello è l'addio di Rino e mentre leggo queste righe comprendo benissimo il comportamento di papà. (Bonicalzi, Scardovi 1997, 115)

A ciò si aggiunge un secondo elemento, che deriva non dall'atteggiamento dei possessori del documento, bensì da quello dei richiedenti.

²⁶ Nell'originale «chuesto mio scritto sarà il richordo che voi gelosamente chustodete, in celo cirivedremo». «Lettera di Mario Marveggio ai genitori scritta in data 24-02-1944 da Lucca». http://www.ultimelettere.it/?page_id=52&ricerca=117&doc=853. Marveggio è di natali umili e probabilmente analfabeta. Questa sua ultima lettera è verosimilmente anche la prima mai scritta.

²⁷ «Lettera di Walter Del Carpio (Scott) ai suoi cari scritta in data 25-01-1945 da Pieve di Tecco». http://www.ultimelettere.it/?page_id=52&ricerca=596&doc=696.

²⁸ «Lettera di Ivan Silvestrini ai Genitori scritta in data 2-05-1944 da Fabriano». http://www.ultimelettere.it/?page_id=52&ricerca=156&doc=345.

²⁹ «Lettera di Rino Gualandi ai familiari da Rio delle Vallette». https://www.ultimelettere.it/?page_id=52&ricerca=198&doc=110.

Non era infatti tra gli scopi della pubblicazione cartacea quello di diffondere gli originali; e, per altro, sarebbe stato molto complicato riuscire a farlo, per le ragioni tecniche e tecnologiche descritte in precedenza. L'obiettivo, in primis proprio di Malvezzi e Pirelli, era soprattutto memoriale, idealizzante; l'intento era costruire un martirologio del partigianato che fosse edificante, che tramandasse esempi di nobiltà e purezza d'animo contro l'oppressore nazifascista. Il fulcro era il testo (o parti di esso), non il contesto. Tale atteggiamento si inserisce perfettamente nel clima che permea la stagione immediatamente successiva alla fine della guerra, in cui si è cercato (con successo) di imporre la memoria e il racconto della Resistenza come fondativi dello Stato post-fascista e repubblicano attraverso una narrazione priva di aspetti critici, enfatica ed eroizzante, non di rado oltre i limiti della ricostruzione fantasiosa.³⁰

Questa è anche la ragione per cui a volte i trascrittori hanno espunto alcune parti dello scritto o ne hanno modificato determinate parole durante la loro opera. È un fatto di cui ci si è potuti accorgere solo dopo aver recuperato gli originali dei documenti presenti nel fondo Malvezzi. Si è notato così che in qualche occasione è stata omessa una frase contenente espressioni d'ira e di rancore o pretese di vendetta. In altre non figura un momento di debolezza che induce il condannato a morte a chiedere ai propri famigliari di intercedere presso il comando nemico per ottenere clemenza. In altre ancora, infine, scompaiono brani legati alla quotidianità, come la supplica di avere un cambio di biancheria dopo una settimana trascorsa in carcere o note quasi testamentarie come la preghiera di lasciare dei soldi o degli oggetti in eredità a qualcuno. Forse anche queste semplici note si temeva che potessero guastare in qualche modo l'immagine esemplare che si voleva trasmettere.

5.3.3 La ricerca di Franzinelli

Il contesto della ricerca coordinata inizialmente da Mimmo Franzinelli (e finanziata dall'Istituto nazionale Ferruccio Parri, che poi l'ha proseguita autonomamente) è invece sostanzialmente diverso. I documenti si trovano spesso in archivi pubblici o comunque accessibili, come quelli degli istituti associati all'Istituto nazionale Ferruccio Parri. Se invece permangono presso i famigliari, si tratta di parenti di seconda o terza generazione, con un legame meno viscerale con la lettera originale e anche con una differente consapevolezza del valore storico e documentale della stessa. Un principio che è per altro

³⁰ Si vedano a tal proposito i recenti studi di Luigi Borgomaneri e Santo Peli, in particolare Peli 2022.

alla base della stessa raccolta operata da Franzinelli e dai ricercatori che hanno collaborato con lui.

L'ultima lettera diventa dunque fonte storica, da recuperare nella sua integrità innanzitutto testuale, senza censure o omissioni di sorta. Ogni passaggio del testo, ogni parola o virgola, restituisce il quadro complesso della condizione di morituro, che sia per condanna o per deportazione, e del contesto storico, sociale e culturale in cui la sua vicenda personale si inserisce. Diventa così emblematico, ad esempio, che il comunista Giordano Cavestro usi la parola «camera-ta» per parlare dei suoi commilitoni partigiani³¹ (parola che diviene «compagni» nella pubblicazione di Malvezzi e Pirelli): forse l'intento di Cavestro è ironico, forse è un tentativo di sfuggire alla censura (ma poco sotto la parola «compagni» è invece usata senza timore) o forse - più probabilmente - permane l'imprinting culturale fascista, difficile da cancellare soprattutto nei momenti di grande tensione quale può essere quello che precede l'esecuzione della pena capitale.

Se, com'è evidente, il digitale offre diversi vantaggi in termini di facilità, velocità e precisione nella riproduzione delle lettere autografe, occorre giustamente rilevare come lo scenario degli anni Duemila non sia tutto benefico ed efficiente. La distanza temporale ha infatti fatto spesso perdere le tracce dei documenti, smarriti in chissà quali cassette famigliari. Il reperimento di originali, così come di nuovi scritti, è quindi assai più complicato oggi, malgrado tutte le innovazioni possibili e la diversa sensibilità sulla natura di fonte storica che si respira oggi.

In soccorso giungono fortunatamente le ricerche locali, che negli anni hanno aiutato a diffondere e preservare almeno la natura testuale dei documenti, anche se spesso senza alcuna citazione archivistica. Risalire il percorso di conservazione si è rilevato spesso arduo, ma comunque possibile, come dimostra il caso di alcuni anonimi delle Fosse Ardeatine. I corpi delle vittime della strage del 24 marzo 1944 sono infatti portati all'Università di Macerata affinché a eseguire l'autopsia sia il prof. Attilio Ascarelli. Indosso ai cadaveri ci sono ovviamente tutti i loro effetti personali e quelli di coloro per i quali non è possibile risalire a un'identità certa confluiscono in quello che a tutt'oggi è il Fondo Ascarelli, conservato per l'appunto nell'archivio dell'Ateneo maceratese. Nel 1965 Ascarelli pubblica un volume (Ascarelli 1965) in cui, tra le altre cose, riporta alcuni stralci di questi documenti scritti da mano ignota. Non essendo uno storico, il medico giustamente non indica con precisione quali estratti provengono da quali lettere. Nel loro volume del 1999, Mario Avagliano e Gabriele Le Moli (Avagliano, Le Moli 1999, 19-20) pubblicano dunque quegli

31 «Lettera di Giordano Cavestro (Mirko) ai Compagni di lotta scritta in data 4-05-1944 da Parma». http://www.ultimelettere.it/?page_id=52&ricerca=130&doc=349.

estratti senza consultare i documenti originali ma indicando comunque la fonte letteraria che, per quanto poco conosciuta, ha comunque permesso ai ricercatori dell'Istituto nazionale Parri di risalire all'archivio dell'Università di Macerata e agli originali in esso conservati.

Malgrado il buon esito della ricerca, la vicenda dimostra una evidente vulnerabilità derivante dal fare troppo affidamento sulle ricerche e sulle pubblicazioni di carattere più locale: errori e imprecisioni possono perpetuarsi nel tempo, fino a solidificarsi e a diffondersi. Nel caso degli 'anonimi' delle Fosse Ardeatine, l'errore di Ascarelli (se così possiamo definirlo, visto che non si tratta di uno storico né la sua opera ha intenti storiografici) viene ripreso da Avagliano e Le Moli e - in un primo momento - anche dall'Istituto nazionale Parri, che solo in una seconda fase ha recuperato gli autografi e li ha inseriti nel database *Ultime lettere*.

Il tema del reperimento dei documenti autentici è dunque fondamentale e dirimente. Capita - benché raramente - che anche i messaggi originali non nascano dalla penna dei morituri, ma siano riportati da trascrittori, visto che - come si è descritto in precedenza - non sempre le vittime hanno la capacità di scrivere o sono nelle condizioni per farlo, dopo giorni di continue sevizie. Se in questi ultimi esempi non è possibile risalire all'intenzione primigenia dell'autore, perché la mediazione è già nell'originale arrivato fino a noi, nei casi in cui (e sono la stragrande maggioranza) sia stato lo stesso condannato a stilare il suo ultimo scritto, l'autografo diventa essenziale per comprendere e interpretare il più fedelmente possibile il senso e il significato delle sue parole e del suo ultimo discorso in relazione con il contesto storico, come si farebbe con qualsiasi altra fonte documentaria.

6 Il contenuto delle lettere

Uno degli aspetti più interessanti della banca dati *Ultime lettere* riguarda proprio questo punto.

Come si è spiegato in precedenza, trattandosi di un prodotto di schedatura digitale, il database effettua un'ovvia serie di categorizzazioni in merito sia alle biografie dei morituri che ai documenti. È dunque possibile analizzare - tra le altre cose - l'estrazione sociale degli scriventi, l'età, la professione, l'appartenenza politica o il grado di istruzione; ma anche la tipologia e il grado di trasmissione del testo delle lettere, se (e quanti) autografi originali, copie fotostatiche, trascrizioni a mano o a macchina sono giunte fino a noi, depositate in archivi pubblici o privati.³² Questo secondo aspetto può essere di

³² «Filtri di trasmissione e tradizione del testo». https://www.ultimelettere.it/?page_id=111.

grande interesse per i filologi, mentre il primo risulta statisticamente poco significativo per trarre conclusioni in riferimento all'intero movimento partigiano, giacché - come si è descritto nella prima parte del contributo - il numero di lettere è davvero troppo esiguo in rapporto alle migliaia di aderenti alla Resistenza. E tuttavia, i risultati possono essere utili a capire qualcosa di più sul corpus documentario nel suo complesso.

Lo strumento che però risulta maggiormente significativo in rapporto all'epistolario e al contenuto delle lettere è senza dubbio il filtro di ricerca semantica, che permette cioè di scandagliare l'intera raccolta digitalizzata cercando determinate parole o frasi nel testo.³³ I dati ricavati sono estremamente utili e intriganti. Essi descrivono innanzitutto uno scarso rilievo dell'ideologia. Termini come 'comunismo' o 'comunista' risultano in soli 4 messaggi, benché quasi la metà degli scriventi appartenesse a quell'area dottrinale. Anche i concetti di 'libertà', 'onore' o 'dovere' sono tendenzialmente poco usati (rispettivamente 45, 32 e 51 volte). Riflessione a parte merita il concetto di 'patria', che ricorre in 67 messaggi su 753 (8%), scritti da 61 condannati. Fra questi, il 44% appartiene alle Forze Armate, che è una percentuale molto maggiore del totale dei militari censiti nel database (27%). È inevitabile considerare come quel concetto fosse sentito molto più vicino a chi avesse una formazione militare, anche forse in virtù dell'uso fattone dal fascismo. Non è un caso che tra le lettere dei condannati della RSI il suo utilizzo sia molto più frequente; e non è nemmeno un caso che, nei messaggi dei partigiani, ricorra molto più spesso il termine 'Italia' (117 volte, cioè quasi il doppio).

Il motivo per cui la sfera ideologica è dunque tanto ridimensionata risiede evidentemente nella terribile contingenza della morte annunciata e imminente, da cui scaturisce il valore prettamente affettivo e privato del contenuto dei messaggi. Scrive Franzinelli:

Sulla soglia della morte scatta il bisogno di spiegare *perché* la propria vita viene stroncata, o anche solo di respingere il marchio di bandito sanzionato da chi ha decretato la condanna. (Franzinelli 2004, 518)³⁴

Ricorrono quindi spesso le parole di saluto ('addio' è in 154 lettere), la consapevolezza della fine ('morire' è scritto 77 volte, 'muoio' 164, 'condannare' / 'condanna' / 'condannato' 99) e, più raramente, una rivendicazione di innocenza (63 volte). Grande rilievo ha poi la sfera

³³ «Ricerca avanzata». https://www.ultimelettere.it/?page_id=134.

³⁴ La tendenza è confermata anche negli scritti giunti clandestinamente e non transitati dalla censura, che - quantunque importante - non risulta in questo caso un fattore determinante.

religiosa: il 'destino' è invocato 115 volte (in varie declinazioni), 'Dio' in 207, 'Prego' in 98. Una menzione a parte meritano le parole legate al concetto di 'perdono', che ricorrono in 216 messaggi: si tratta in molti casi di una richiesta spirituale, ma non di rado essa si indirizza anche (o soltanto) ai famigliari, ai quali il condannato domanda indulgenza per la scelta fatta e le sofferenze causate, pur non rinnegando (quasi mai) l'adesione al movimento partigiano, nonostante questo lo abbia portato all'estremo sacrificio.

Un'ultima considerazione va spesa per vendette e torture. Malgrado il momento e i prevedibili rancori verso eventuali delatori o traditori, raramente (9 volte) nelle lettere viene chiesta una rivalsa o una punizione nei confronti del nemico o, più nello specifico, di chi ha determinato la propria cattura o la morte. A dispetto dei sentimenti più laici, permane forse nei morituri il desiderio di abbandonare la vita terrena senza macchie o rancori.³⁵ Anche delle torture non si parla spesso, ma quando lo si fa le parole risultano particolarmente incisive, anche quando non si usa espressamente il termine. Così Mario Pasi fa avere clandestinamente ai compagni di lotta un bigliettino in cui scrive «mandatemi del veleno non resisto più»;³⁶ mentre Giovanni Venturini racconta alla madre che «Ormai sono ridotto a misera cosa, non sono più un uomo³⁷ e qualche volta piango dal dolore dei miei piedi che non mi serviranno più».³⁸

35 Anche in questo caso c'è equilibrio tra lettere uscite clandestinamente e passate attraverso la censura, a riprova che, ancora, questa non era un fattore determinante al punto da incidere così tanto su alcuni aspetti del contenuto.

36 «Lettera di Mario Pasi (Montagna) ai Compagni di lotta da Carceri delle SS di Belluno». https://www.ultimelettere.it/?page_id=52&ricerca=120&doc=148

37 Durante le sevizie Venturini viene sottoposto a evirazione.

38 «Lettera di Giovanni Venturini (Tambia) alla Madre da Edolo». https://www.ultimelettere.it/?page_id=52&ricerca=676&doc=850

7 La materialità: l'importanza dei supporti

Non è però solo in ragione dell'integrità testuale che gli originali hanno grande valore. Spesso sono altrettanto importanti le informazioni di contesto che essi restituiscono. Un caso emblematico è costituito dalla lettera di Alfredo Formenti,³⁹ uno dei rari casi di copie fotografiche presenti nel Fondo Malvezzi. Nella copia in bianco e nero le macchie a metà del foglio possono apparire d'inchiostro; dopo aver recuperato l'originale risulta invece evidente come si tratti di sangue, verosimilmente di Formenti stesso [fig. 1].

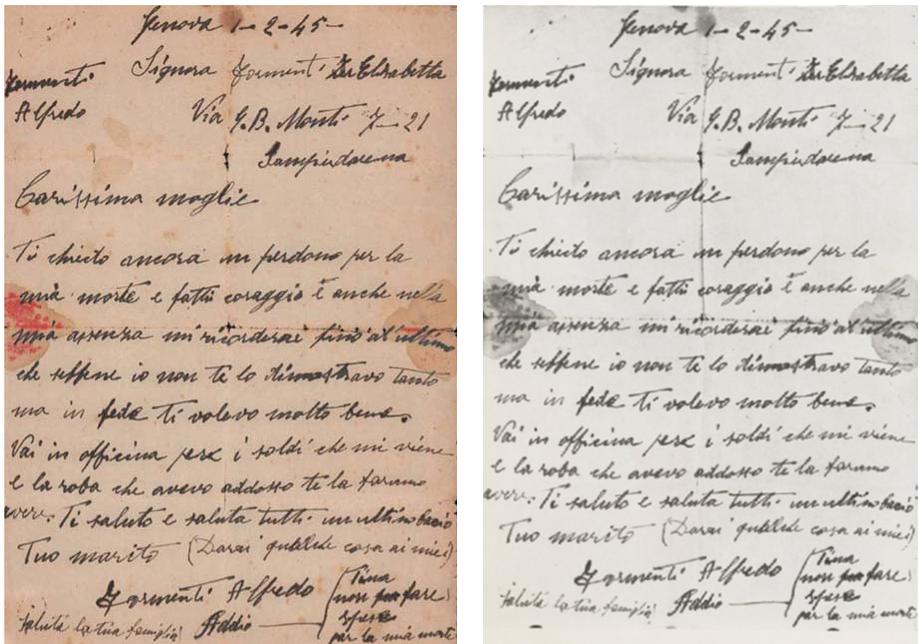


Figura 1 A sinistra la lettera originale di Alfredo Formenti; a destra la copia fotografica in bianco e nero

Un altro dato importante che il documento originale fornisce riguarda il modo in cui la scrittura riempie lo spazio, spesso più ordinata e ariosa all'inizio e poi via via sempre più frettolosa, sincopata e schiacciata in concomitanza con la fine del foglio. Occorre infatti sempre considerare la scarsa disponibilità dei supporti su cui scrivere e che spesso il massimo di cui si disponeva erano due facciate di un

³⁹ «Lettera di Alfredo Formenti (Brodo) alla Moglie scritta in data 1-02-1945 da Genova». http://www.ultimelettere.it/?page_id=52&ricerca=424&doc=360.

biglietto. A volte però nemmeno quello. In diversi casi, infatti, abbiamo messaggi scritti su ciò che si aveva in tasca al momento dell'arresto. Esterino Rech e Secondo Brignolo [fig. 2] vergano quindi le loro ultime parole su dei libri di preghiere; Willy Jervis fa lo stesso, ma non avendo a disposizione una penna incide l'interno della copertina di una bibbia con una punta di spillo. Strumento simile è usato probabilmente da Ignazio Vian su una pagnotta [fig. 3]. Epimenio Liberi, ucciso alle Fosse Ardeatine, ha con sé un foglio pentagrammato in quanto maestro di musica di professione [fig. 4]. Gianfranco Mattei scrive la sua ultima lettera sul retro di un assegno [fig. 5], mentre Umberto Fogagnolo (ucciso a piazzale Loreto, a Milano) lo fa sul lato posteriore di una fotografia. A Ivo Pasquinelli, in modo quasi ironico, è concessa una cartolina con lo slogan fascista «Vinceremo». Guido Valota infine compone un lungo messaggio su un manifesto di carico, concessogli probabilmente da qualche anima pia nell'ultimo giorno di prigionia alla Caserma Umberto I di Bergamo, prima di essere tradotto a Mauthausen.



Figura 2 L'ultimo messaggio di Secondo Brignolo

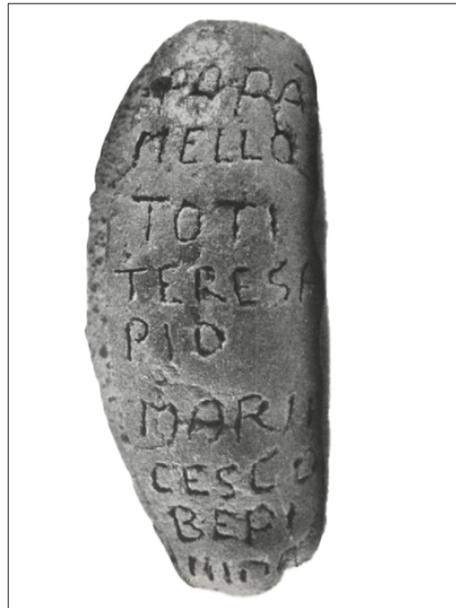


Figura 3 La pagnotta incisa da Ignazio Vian

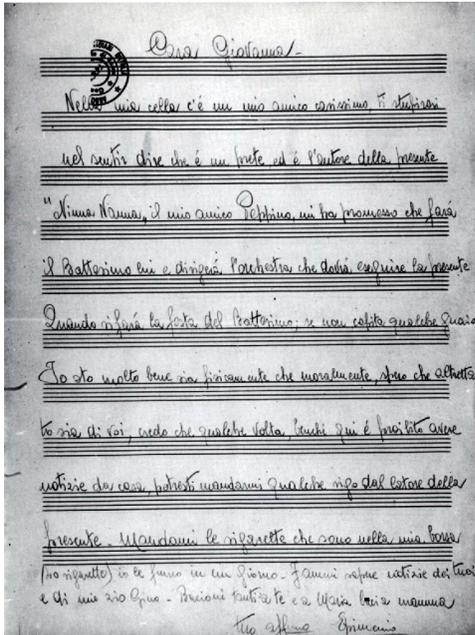


Figura 4 Il pentagramma su cui Epimenio Liberi scrive le sue ultime parole

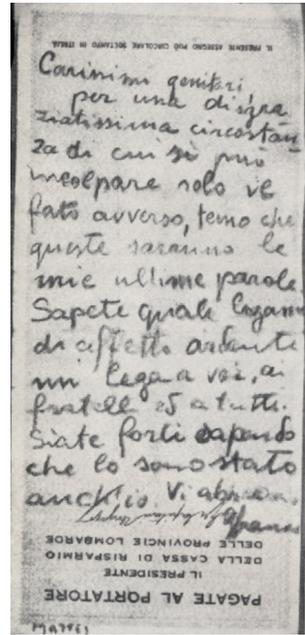


Figura 5 Gianfranco Mattei scrive la sua ultima lettera sul retro di un assegno

In altri casi invece le tasche sono vuote o si ha il desiderio di lasciare una traccia più tangibile e pubblica del proprio passaggio o del proprio lascito. Giovanna Bordignon Sereni, ad esempio, incide una sorta di graffito sul muro della cella nella Risiera di San Sabba, dove tiene un resoconto dettagliato delle date di deportazione dei propri famigliari [fig. 6]. L'ultima riga del testo, che recita «Non potiamo kapire ke fine a fata kuesta signora»,⁴⁰ è stata postillata da un altro prigioniero, probabilmente sloveno o croato. Di carattere meno pratico e più ideologico ciò che invece il generale Sabato Martelli Castaldi lascia nelle carceri di Via Tasso, a Roma: «Quando il tuo corpo non sarà più, il tuo spirito sarà ancora più vivo nel ricordo di chi resta. Fa che possa essere sempre di esempio».⁴¹

Insomma, tutto concorre, al pari se non più delle parole stesse, a trasmettere l'estrema drammaticità e precarietà del momento, in cui v'è la sospensione di ogni diritto più elementare, come la possibilità di comunicare con i propri cari e informarli del proprio triste destino.

⁴⁰ «Lettera di Giovanna Bordignon Sereni a... da Risiera di San Sabba». https://www.ultimelettere.it/?page_id=52&ricerca=202&doc=112.

⁴¹ «Lettera di Sabato Martelli Castaldi (Tevere) a... da Carceri di Via Tasso, Roma». https://www.ultimelettere.it/?page_id=52&ricerca=262&doc=209.

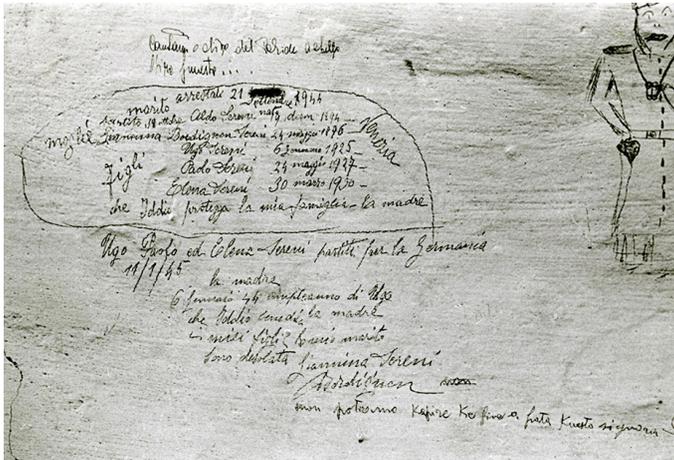


Figura 6 L'incisione di Giovanna Bordignon Sereni

8 Conclusioni

Come si evince, un primo e fondamentale elemento di interesse della banca dati *Ultime lettere* è rappresentato proprio dall'aver messo a disposizione, gratuitamente e sul web, un archivio inventato⁴² capace di radunare in un unico corpus documentario virtuale un patrimonio che possiede un estremo valore civile e morale, fondamentale tassello nella costruzione di una religione civile⁴³ italiana purtroppo spesso (e sempre più) messo in discussione. Naturalmente la pubblicazione di un prodotto digitale di per sé non è sufficiente a garantirne l'ampia fruizione e il coinvolgimento del pubblico (Cauvin 2016, 177-81); e tuttavia, i percorsi biografici segnati dalle difficoltà e dal martirio, uniti alle parole e alle testimonianze di chi tale martirio lo ha subito, esercitano ancora oggi – pur a distanza di anni – una presa emotiva molto forte, che rende *Ultime lettere* uno strumento utilissimo per essere utilizzato con gli studenti delle scuole secondarie o in iniziative destinate alla cittadinanza, come ad esempio *reading* o spettacoli teatrali.⁴⁴

In secondo luogo, per quanto le categorie e i filtri di ricerca non siano in grado di dare riscontri significativi sul movimento partigiano

⁴² Una definizione approfondita di «archivi inventati» si trova in Rosenzweig 2011.

⁴³ Sulla definizione e sul tema si vedano i lavori di Maurizio Ridolfi, in particolare Ridolfi 2021.

⁴⁴ Esempio recente è l'iniziativa realizzata il 14 aprile 2023 presso la Casa della memoria di Milano (<https://www.reteparri.it/eventi/canti-lettere-della-resistenza-reading-partecipato-9463/>).

nel suo complesso (visti i numeri troppo esigui del campione rispetto alla totalità degli aderenti alla Resistenza), essi possono ugualmente delineare un quadro di tendenze e relazioni tra alcuni fattori sociali e demografici, quali età, genere, istruzione, alfabetizzazione, età e professione. Uno spunto, da collegare poi con altre schedature più complessive come quella di «Partigiani d'Italia».

Come enunciato nel paragrafo precedente, non è poi da sottovalutare l'importanza che il digitale può avere nel restituire la materialità di oggetti difficili da raggiungere o in qualche caso perduti. Può sembrare un controsenso, e ovviamente stiamo parlando di una materialità che è esclusivamente visiva, ma grazie all'opera di reperimento e digitalizzazione dei documenti originali è possibile restituire la piena integrità della fonte, sia testuale che contestuale, allo scopo di inquadrarla al meglio sia nel particolare momento vissuto dallo scrivente (la prigionia e l'imminenza della morte) che nel periodo di riferimento.

Infine, l'aspetto forse più rilevante (e ancora oggi non comune in strumenti analoghi) riguarda l'analisi semantica del corpus documentario, che rappresenta la frontiera tecnica e tecnologica verso la quale ogni prodotto di Digital History dovrebbe tendere oggi. Avere la possibilità di cercare parole o frasi all'interno di raccolte testuali coerenti per tema, epoca o esperienze è un sicuro valore aggiunto per la ricerca, sia in campo storico che sociologico. Potrebbero così emergere prevalenze linguistiche o ideologiche tali da far comprendere meglio mentalità e situazioni sulle quali, spesso, si ha una percezione troppo intermediata e poco diretta. Da questo punto di vista, la banca dati *Ultime lettere* rappresenta una selezione forse troppo specifica e limitata nel numero, per quanto assolutamente esaustiva della particolare forma epistolare nel periodo preso in esame. Inoltre, il suo corpus documentario potrebbe fornire spunti interessanti nella relazione con altri corpora analoghi, ma di diversa epoca storica (come ad esempio il Risorgimento), dando così conto di permanenze, distanze o rielaborazioni potenzialmente dal grande valore storico e filologico.

Bibliografia

- Ascarelli A. (1965). *Le Fosse Ardeatine*. Roma: Canesi.
- Avagliano, M.; Le Moli, G. (1999). *Muoi innocente. Lettere di caduti della Resistenza a Roma*. Milano: Mursia.
- Bonicalzi, M.; Scardovi, E. (1997). *Storie, per non dimenticare*. A cura di A. Leoni. Longastrino: Ed. Spazio Cultura.
- Borgomaneri, L. (2015). *Li chiamavano terroristi. Storia dei Gap milanesi (1943-1945)*. Milano: Unicopli.
- Cauvin T. (2016). *Public History. A Textbook of Practices*. New York: Routledge.
- Cipolla, R. (1945). *I miei condannati a morte. Lettere e testimonianze*. Torino: SATET.
- Flores, M.; Franzinelli, M. (2021). *Storia della Resistenza italiana*. Bari-Roma: Laterza.
- Franzinelli, M. (2004). «Ultime lettere. Scritti di fucilati e deportati della Resistenza». *Italia Contemporanea*, 237.
- Dogliotti C. (2016). «Cronografia. Territori e fasi della politica del massacro». Fulvetti, G.; Pezzino, P. (a cura di), *Zone di guerra, geografie di sangue. L'Atlante delle stragi naziste e fasciste in Italia (1943-1945)*. Bologna: Il Mulino, 95-126.
- Franzinelli, M. (2005). *Ultime lettere di condannati a morte e di deportati della Resistenza*. Milano: Mondadori.
- Malvezzi, P.; Pirelli G. (1952). *Lettere di condannati a morte della Resistenza italiana*. Milano: Einaudi.
- Peli, S. (2015). *Storia della Resistenza italiana*. Torino: Einaudi.
- Peli, S. (2022). *La necessità, il caso, l'utopia. Saggi sulla guerra partigiana e dintorni*. Pisa: BFS edizioni.
- Perona, G. (2006). «Presentazione». *Ultime lettere di condannati a morte e di deportati della Resistenza italiana*. http://www.ultimelettere.it/?page_id=53.
- Revelli, N. (1971). *L'ultimo fronte. Lettere di soldati caduti o dispersi nella Seconda guerra mondiale*. Torino: Einaudi.
- Ridolfi, M. (2021). *Le feste nazionali*. Bologna: Il Mulino.
- Rosenzweig, R. (2011). «The Road to Xanadu. Public and Private Pathways on the History Web». Rosenzweig, R., *Clio Wired. The Future of the Past in the Digital Age*. New York: Columbia University Press, 203-36.
- Zagrebel'sky G. (2002). «Prefazione». *Lettere di condannati a morte della Resistenza italiana*. Milano: Einaudi, V-IX.

Designing for Audience Engagement

Exploring the Use of Online Metrics in the GLAM Sector

Ellen Charlesworth
Durham University, UK

Claire Warwick
Durham University, UK

Leonardo Impett
University of Cambridge, UK

Andrew M. Beresford
Durham University, UK

Abstract Many heritage and cultural professionals have emphasised the ability of online content to reach beyond museums' traditional audiences, yet one of the largest surveys to date shows no significant change in the demographic breakdown of online and on-site visitors. This paper aims to investigate the discrepancy between the data and the narrative, using two case studies to illustrate the pitfalls of using common social media metrics such as 'views', 'likes', and 'shares'. Drawing from the authors' previous sector-wide study, the paper highlights what shapes these numbers, and questions how useful they are as a measure of 'audience engagement'. Finally, it looks to computational methods to explore nuanced alternatives that could supersede these metrics in the coming decade.

Keywords Social media. Audience engagement metrics. Digital GLAM. British Museum. Sentiment analysis. Topic modelling. Network analysis.

Summary 1 Introduction. – 2 Funding Pressures. – 3 The Limitations of Engagement Metrics. – 4 Case Studies. – 4.1 The British Museum. – 4.2 The Linen Hall Library. – 5 Alternatives. – 5.1 Sentiment Analysis and Topic Modelling. – 5.2 Network Analysis. – 6 Conclusions.



Peer review

Submitted 2023-04-28
Accepted 2023-07-04
Published 2023-08-02

Open access

© 2023 Charlesworth, Warwick, Impett, Beresford | © 4.0



Citation Charlesworth, E.; Warwick, C.; Impett, L.; Beresford, A.M. (2023). "Designing for Audience Engagement. Exploring the Use of Online Metrics in the GLAM Sector". *magazén*, 4(1), 145-174.

DOI 10.30687/mag/2724-3923/2023/01/005

1 Introduction

The past three years have seen a dramatic rise in the number of gallery, library, archive, and museum (GLAM) organisations uploading resources online. Initially seen as a way to mitigate the impact of the COVID-19 pandemic and meet a perceived demand for cultural content, many organisations were able to pivot to digital by uploading previously digitised resources to social media (NEMO 2020). This increased online output has led to an enormous range of free-to-access content online. Yet, over the same period, museums also faced increased financial pressures resulting from the pandemic closures of 2020 and 2021 (Gnezdova, Osipov, Hriptulov 2022). This paper makes the argument that this financial crisis, which has exacerbated existing scrutiny of spending on culture and heritage, has shaped the way that we discuss digital adoption and online audiences.

During the COVID-19 pandemic many GLAM professionals argued that online content could transform audience engagement and reach beyond traditional audiences (Noehrer et al. 2021; Samaroudi, Echavarría, Perry 2020). This idea has been widespread since museums first started launching initiatives on the web. As Avenier summarised,

It is now generally acknowledged that museums have gained significantly from the giant technological strides associated with the growth of the Internet. Nobody indeed disputes that the widespread dissemination of information and knowledge about museum collections is an added step towards the democratization of culture. (Avenier 1999, 31)

Yet, this narrative has been heavily criticised. In 2011, the demographic analysis performed by Culture24

led the group to question whether the cultural sector is in fact attracting new audiences online, as has traditionally been assumed, or is simply engaging with the same audiences that they interact with offline. (Finnis, Chan, Clements 2011, 22)

A decade later, the Cultural Participation Monitor – a project which included a series of surveys with over 6,000 correspondents – corroborated this finding, and found that those who engaged online were “more likely to have been regular in-person arts attenders”, with “more than half of audiences engaging with Museums and

This work was supported by Arts and Humanities Research Council, grant no. AH/R012415/1. The Authors would also like to thank the editors and reviewers of *magazén*, including Dr Stefania De Vincentis for her invaluable guidance and support.

Heritage online are aged 55+” (Walmsley et al. 2022). Conducted by the Audience Agency and the Centre for Cultural Value, this work has highlighted the continuing discrepancy between the rhetoric of broadening access and organisations’ actual ability to reach new audiences (Audience Agency 2021).

Gathering data to contextualise why this narrative is still so prevalent and identifying the biases in our discussions of online content is an integral part of establishing a sustainable design process and fostering eventual improvement. However, the act of measuring engagement is fraught with difficulty. Methods such as surveys and interviews are shaped by self-selection bias in respondents and do not reflect wider – often less engaged – audiences (Gran et al. 2019; Bethlehem 2010). There is a desperate need in the sector for large-scale quantitative metrics to create a benchmark against which GLAM professionals can evaluate survey results and their own anecdotal experience (NEMO 2020). This need has been repeatedly highlighted over the past decade which has seen government investment in the Culture Metrics and Culture Counts initiatives in the UK (Culture Metrics 2022; Arvanitis et al. 2016). The pandemic further highlighted this need with the Network of European Museum Organisations stating that a “sound metric to benchmark online visits is missing”; this resulted in new investment in inDICES’ Europeana Impact Framework by the EU’s Horizon programme (inDICES 2023; NEMO 2020, 3).

Yet taken in isolation, quantitative metrics have been found to alter how we conceive of value and are a simplification of a vast array of outcomes and motivations which are unique to each organisation (Espeland, Sauder 2016). This paper will therefore explore how quantitative metrics from social media platforms may be misleading and explore what kind of metrics may be more useful to GLAM organisations. Through an investigation of two online responses to the UK lockdowns in 2020, the paper will examine how institutional aims inform the way audiences are conceptualised and incorporated in online metrics commonly used by GLAM organisations.

The UK case studies – a national London-based museum and an independent library in Northern Ireland – have been chosen to illustrate the breadth of ways that organisations can engage with online audiences. These two examples were both identified from the authors’ larger study of 315 UK GLAM organisations, and have been chosen to highlight how the sector’s diverse online landscape is poorly captured by simple quantitative measures and social media metrics (Charlesworth et al. 2023).¹ Using museums websites’ and five

1 The study took a sample from the “Arts Council Accredited List of Museums” which, despite the title, includes a number of Libraries, Galleries, and Archives (Arts Council England 2021).

social media platforms, our analysis of social media metrics during the study highlighted how misleading the measures can be. By exploring how content is disseminated on platforms such as YouTube and Facebook, we will explore whether these numbers reflect audience engagement, and make the argument that alternative methods such as network analysis, sentiment analysis, and topic modelling will provide more nuanced measures of engagement.

Beginning with a summary of the funding pressures that incentivise audience engagement, the paper explores why quantitative metrics, including those from social media, are increasingly important to the sector. It then goes on to explore whether commonly used online metrics, designed by and for social media platforms, are accurate measures of audience engagement and serve the needs of GLAM institutions. It details the limitations of commonly used measures of engagement – views, likes, shares, and comments – and debunks common misinterpretations of the figures. Finally, we ask whether these metrics are of use to GLAM institutions at all and explore potential alternatives that, with further research and development, could supersede the metrics available on social media platforms, to create a more nuanced method to evaluate audience engagement.

2 Funding Pressures

There is a broad range of literature in museum studies and cultural policy that explores the many factors that contributed to the increasing importance of engagement. Postmodernist thinking redefined museums' role in society and their relationships with visitors (Speight 2010; Hooper-Greenhill 1992); this saw the adoption of the idea that museum experience is shaped by an individual's prior knowledge and broader social context (Falk, Dierking 1995; 2000). Postcolonial critics, most prominently Edward Said, problematised the role of culture as an agent of empire, prompting a questioning of the role of colonial power relations had on the interpretation of objects (Barringer, Flynn 1997; Said 1993). The impact that these discussions had on audience engagement is too large to cover in a single paper. Instead, this paper will focus on how engagement has been conceptualised by the UK government and funding bodies. This approach aims to contextualise why there is a prevailing emphasis on online content's ability to broaden GLAM audiences, exploring how shifting funding priorities have shaped GLAM organisations' goals and the ways in which they measure impact and engagement.

The majority of GLAM institutions in the UK are either directly supported by government institutions or are non-profit organisations reliant on substantial grants to stay open (Ballatore, Candlin 2022, 224). This financial arrangement has undoubtedly had some influence on

the organisational aims of these institutions, and the focus on broadening their appeal is partially motivated by an increasing pressure to justify public expenditure in the sector through tangible economic and social outcomes (Bailey-Ross 2014; Anderson 2013). This can be observed in the way museums' organisational practices have been shaped by funder's emphasis on evaluating 'impact' and 'engagement' (Aroles, Hassard, Hyde 2022); as Belfiore argues, 'impact' has become a proxy for the term 'value' and is indicative the way in which funding is increasingly framed as an 'investment', which expects to see 'returns' and 'healthy dividends' (2015, 101, 106). These returns may not be strictly financial and GLAM organisations are regularly conceptualised as vehicles for government policy. Black notes that where organisations have traditionally received public funds

there has been an expectation at both the national and local levels that all such bodies, including museums, will actively support relevant political initiatives. (Black 2010, 130; Newman, McLean, Urquhart 2005)

A political focus on 'engagement' can be traced back to 1990s, when there was growing concern at a perceived decline in public dialogue (Ashley 2014). This was considered symptomatic of declining levels of civic engagement there was a fear that citizens were not actively participating in their communities or in the democratic process (Adler, Goggin 2005, 236). Partially attributed to an increase in multiculturalism in the UK, the discussion saw growing demands for recognition of minority groups, and resultingly, a renewed focus within government on improving social integration and inclusion (Newman, McLean 2004). In this context, cultural participation was seen as a means to increase civic engagement: in part, by opening democratic public discussions to previously excluded voices; but also by fostering social inclusion and encouraging dialogue between communities (Long 2013; Putnam 1995). Engagement in this context is not an end, but a mechanism to enact political and societal change.

The rhetoric around engagement today highlights that cultural institutions continue to be seen as a means to encourage social inclusion, even if it is as agents of positive change and advocates of "social justice" (Ünsal 2019). Arts Council England, a major funder of libraries, museums, and galleries places an emphasis on "diversity [...] and creating more pathways for a wider range of people to become part of the arts and culture sector" (2023a). Yet, the efficacy of using cultural organisations in this way is still contested. Culture alone cannot be expected to repair social exclusion an issue that has its roots in issues with economic integration, housing, and the welfare system (Newman, McLean, Urquhart 2005). Newman and McLean have problematised public investment in museums on the basis that they

improve social inclusion, noting that participation in cultural institutions is not the same as participation in culture (2004). In addition, despite a turn towards participatory practices (Simon 2010), the power structures and hierarchies of long-standing western institutions exacerbate existing asymmetrical power relations between communities, complicating GLAM organisations' role as facilitators of open dialogue (Ashley 2014, 263; Schuch, Harden, Smith 2023, 313).

It is unsurprising that within this funding context, made more acute by the pandemic, so many GLAM professionals framed the digital shift as an opportunity to diversify audiences and reach beyond museums' traditional visitors (Samaroudi, Echavarria, Perry 2020; Mantell, Turpin 2020). This is important in understanding the use of online platforms as the ability to diversify and grow audiences is a significant appeal of online content for most GLAM institutions (Noehrer et al. 2021, 2).

3 The Limitations of Engagement Metrics

Given the funding context discussed, it is clear that 'audience engagement' in the GLAM sector is one of the means through which institutions try to achieve social inclusion and civic engagement. It is these, rather than audience engagement itself, that we would ideally like to measure. Yet the 'social inclusion' generated through a GLAM organisation is difficult to quantify and almost impossible to capture at scale - an issue that similarly plagues measures of audience engagement. The ubiquitous audience survey has major limitations, especially online. To get a large sample is difficult, is often a labour-intensive process, and surveys are often designed for a specific project preventing comparison between institutions. In addition, as surveys are reliant on people voluntarily filling them out, there is a self-selection bias amongst respondents that distorts our understanding of general visitors' views and demographic groups (Nuccio, Bertacchini 2021; Gran et al. 2019; Bethlehem 2010).

The Culture Metrics project launched in 2014 aimed to address these issues, involving cultural organisations in the process of establishing standardised metrics for the GLAM sector. A versatile survey was designed to work across multiple contexts; participants scored different aspects of an event or visit on a sliding scale. In describing the project, Arvanitis et al. (2016) noted the benefit of this approach:

The use of standardised metrics across organisations potentially allows for comparisons among them and among different kinds of events and over time.

The Culture Metrics team developed a platform in partnership with Culture Counts which automated the analysis of the results, and in

2016, Arts Council England announced that using the Culture Counts framework would become “a mandatory requirement” for organisations that received over £250,000 a year (Gilmore, Glow, Johanson 2017; Arts Council England 2016, 2). This decision, which was later revoked, was met with vocal opposition. Phiddian et al. noted that the issue was not with the metrics themselves but in the way that they were being used as to assess artistic quality:

The categories are nuanced enough to provide usable feedback for practitioners and bureaucrats with the time and desire to think hard about what the numbers mean. But, they remain essentially marketing analytics rather than a window on artistic value. (Phiddian et al. 2017, 178)

What is a helpful tool for well-meaning practitioners can, in other hands, be used to ineffectually rank relative performance.

The Culture Counts example highlights the importance of context for any metric, and the goals of a project should inform the methods we use to evaluate success. As Phiddian et al. argue, the adoption of framework across all grant recipients would “imply a spurious homogeneity of purpose in the arts” (Phiddian et al. 2017, 178). As Glass notes, “the importance of data isn’t the data by itself. It’s the possibility of self-awareness and self-reflection that it brings to bear to make us better” (2015, 5). This partially explains the GLAM sector’s frequent use of surveys, which can give additional information about audiences’ motivations for visiting and answer why an experience was positive or negative (Villaespesa 2013).

Yet, the most accessible and widespread metrics available to GLAM organisations are simple quantitative measure provided by social media platforms. In the UK, it is more common for a likely for GLAM organisations to have a Facebook account than their own website (Charlesworth et al. 2023, 8). Larkin, Ballatore, and Mityurova found that 77% of museums had a Facebook account, and 67% had Twitter (2023, 6). This high uptake was true across small and medium sized museums, making measures of social media engagement – such as ‘likes’, ‘views’, and ‘shares’ – an already widespread form of standardised metric.

As we conducted follow-up interviews to the study of 315 UK GLAM organisations, social media metrics were regularly raised when describing success, and digital teams were proud of their posts with high levels of social media engagement. This was tempered by an understanding that some posts – such as site closures and event advertisements – would not perform well, despite their utility to visitors. However, this nuance is rarely conveyed in the reporting of social media numbers. Annual reports across the GLAM sector regularly highlight social media successes, including both organisations used here

as case studies. The British Museum highlights that “Social media continues to enable us to reach audiences directly” and celebrates its “nearly six million followers across all platforms” with little elaboration (British Museums 2022, 6). The Linen Hall Library was similarly “pleased to report that it [their Instagram] reached 1,000 subscribers by year end” (Linen Hall Library 2020, 9). This kind of reporting is understandable as Arts Council England actively encourage organisations to use social media channels “to show how public funding enables your work and positively impacts the lives of your audience and your local community” (Arts Council England 2023b). Considering this tendency to equate high social media metrics with impact, it is worth examining exactly what these numbers represent and the context in which they were developed.

Social media metrics were designed with commercial interests in mind. The ways in which we interact with the internet are both facilitated and mediated by companies such as Meta and Google. It is for them and their paying users – that is advertisers – that the ranking of content based on ‘click throughs’, ‘likes’, and ‘shares’ were developed (Gillespie 2010). In a study of social media advertising, Voorveld et al. define social media engagement as “the emotional, intuitive experiences or perceptions” that occur on social media platforms, and cover a broad range of motivations including “satisfying the need to find useful information, fill empty moments, and do or share something with others” (2018, 40). Pertinently for this discussion, Voorveld et al. conclude that “information might increasingly be a ‘by-product’ of social media use rather than a central feature” where exchanging information “is a means to an end rather than a goal in itself” (2018, 51).

This paper cannot comprehensively cover the vast body literature dedicated to identifying the most effective advertising, persuasion, and e-commerce strategies based on these metrics (Rautela, Sharma, Virani 2021). However, there is a consensus that the most widely shared content provokes emotions; alarmingly, Ji et al. found that Facebook posts that roused negative emotions were more likely to be shared than positive posts (2019). It is therefore questionable whether these same metrics accurately reflect the meaningful inter-community connections that GLAM institutions hope to facilitate. By borrowing quantitative metrics from social media platforms, GLAM institutions are adapting tools designed to commodify attention. Therefore, it is worth pointing out the obvious: social media engagement does not necessarily correlate with engagement in the sense it is used in the GLAM sector.

There are remaining issues even if a metric was found to accurately reflect audience engagement. If a metric was used during the development process to test projects and exhibitions, GLAM institutions may inadvertently optimise content to maximise the metric,

as opposed to maximising engagement. This issue is summarised in Goodhart's law, which was succinctly described by Strathern, "When a measure becomes a target, it ceases to be a good measure" (1997, 308). The principle is most commonly observed in education, where students learn to pass the exam, not understand the material (Fire, Guestrin 2019). This could result in GLAM online content that has high levels of engagement on paper, but individuals may not have the intended experience. This issue can only be combatted with extensive user testing and verbal feedback, a process that should not be side-lined in favour of cheaper and more readily available quantitative metrics.

Despite the need for a sound benchmark to evaluate online audiences, this section has highlighted some of the problems that sector faces to reach this goal (NEMO 2020, 3). In the next section, we will explore how these problems emerge across two case studies. By investigating how different types of institutions are served by social media metrics, we highlight the types of engagement they capture and question how useful this is to GLAM organisations.

4 Case Studies

The utility of a metric is reliant on what an organisation is trying to learn by gathering that data. Therefore, in order to understand the breadth of ways an audience engagement metrics can be used, the case studies draw from the two dramatically different institutions. Informed by the authors' previous large-scale analysis of GLAM organisations in the UK, the two examples - a huge museum, and a small library - have entirely different institutional aims (Charlesworth et al. 2023).² Using these case studies, this section explores how institutional aims inform the way audiences are conceptualised and incorporated in the online design choices and digital strategies of GLAM organisations. Crucially, an in-depth analysis of the reception of two online exhibitions enables us to ask whether 'click-throughs', 'likes', 'shares', and 'comments' are useful metrics for evaluation.

² Theses sizes are based on the guidelines produced by the Mapping Museums Project, which categorised organisations by the number of yearly visitors; "small (0-10,000 visits); medium (10,001-50,000 visits); large (50,001 to one million visits); and huge (over one million visits)" (Candlin et al. 2019, 57).

4.1 The British Museum

The first case study will explore how a single video uploaded by a huge national institution was received by audiences during the pandemic. In the study of 315 UK museums, the British Museum was the largest organisation - based on the annual visitor numbers - in the sample. It had by far the largest social media following and at the time of writing the British Museum has 1.76 million followers on Facebook, 2 million on Twitter, and 2.1 million followers on Instagram (Charlesworth et al. 2023, 10).

With an established online presence before the outbreak of COVID-19, the museum was able to adapt particularly quickly to the closure of its physical site on March 18, 2020 (British Museum 2020a). The first national lockdown in the UK - beginning on the 23rd of March with most restriction ending on the 4th of July - prolonged what was originally envisaged as a short temporary closure, and the museum would not reopen until August 27 (British Museum 2020c). On announcing its closure, the museum had already outlined its digital strategy going forward. The Director of the British Museum, Hartwig Fischer, stated,

We'll be updating and adding to digital content during the period we're closed to allow visitors to stay in touch with the Museum. We'll share our collections, research and programmes in new ways that will not require a trip to the Museum. (Brown et al. 2021; British Museum 2020a)

Alongside his statement, the website was updated so that the homepage displayed a large banner detailing a list of digital activities. The page prompted visitors to 'stay connected' in multiple ways: it invited them to take a virtual tour on the website; look at the collections online; use the resources for schools; listen to the British Museum's podcast; get in touch through social media; or explore the museums' content on Google Street View and Google Arts and Culture.

Over the course of the first lockdown period, the website was regularly updated with new content - predominantly published within the framework of existing digital initiatives at the museum. Unfortunately, metrics such as the number of views, dwell time, and audience analytics of the website and Google Arts and Culture content are not publicly available. However, the reception of online content over time can be monitored in part on social media platforms, for which data is available through an application programming interface (API). In addition, basic metrics and the relative success of a piece of content is easily viewable on the platforms themselves. One the most successful pieces of content the British Museum produced, was an old video tour of an exhibition uploaded to YouTube in May 2020. This video

has been chosen as the focus of this case study, as it received large amounts of engagement on social media and remains the single most watched piece of content on the museum's YouTube channel in 2023.

"Vikings Live: a tour from the British Museum" is an hour and a half long video that was originally created in 2014 and re-uploaded on May 27, 2020 (British Museum 2020b). Advances in technology since the video was filmed, mean that the production values - particularly the resolution and sound quality - are notably lower than more recent uploads on the museum's YouTube channel. The presenters, Bet-tany Hughes and Michael Wood, take turns in interviewing key staff members and visiting experts who explain the process of setting up the exhibition as well as the history of the physical objects. These segments are interspersed with footage of historical re-enactments, questions from the public, and graphical overlays that show migration patterns and the location of archaeological sites.

Originally created for broadcast in cinemas, the video itself was not created for YouTube and therefore is atypical of the platform, which is dominated by individual content creators, not institutions (Xiang 2022). With a run time of 89 minutes, the duration of "Vikings Live" is much longer than the average YouTube video which is less than 15 minutes long (Rieder, Coromina, Matamoros-Fernández 2020; Che, Ip, Lin 2015). Yet despite content designed for other platforms often underperforming compared to "native" YouTube content, by March 2023 "Vikings Live" had been viewed over 8 million times (Rieder, Matamoros-Fernández, Coromina 2018; Arthurs, Drakopoulou, Gandini 2018, 6). Does that make the "Vikings Live" video a success? By contextualising this figure, this section aims to highlight what influences the number of views a video might receive and investigate whether the metric has any use as an evaluative tool for researchers and GLAM professionals.

The number of views on YouTube can be accessed by researchers in two ways: either through YouTube's application programming interface; or for small scale research, it is possible to manually check the view count displayed below the video. The first challenge this poses is that unless you are an administrator of the channel that uploaded the video, the number only reflects a snapshot of the number of views at the time it is checked. This means that researchers are unable to chart a video's popularity in retrospect. Fortunately, over the course of 2020, a significant volume of GLAM digital content was preserved using the Internet Archive. As a non-profit organisation that aims to preserve websites and digital artifacts, their tool, the Wayback Machine, enables researchers to see snapshots of museum websites and social media taken throughout the pandemic period. Anyone is able to add a webpage to the archive, and between May 30, 2020 and March 14, 2023 "Vikings Live" was archived using the Wayback Machine over 70 times (British Museum 2020b). Although

not comprehensive, these archived versions of the YouTube page enable us to roughly reconstruct the views the video received in the three years since it was first uploaded [fig. 1a-b].

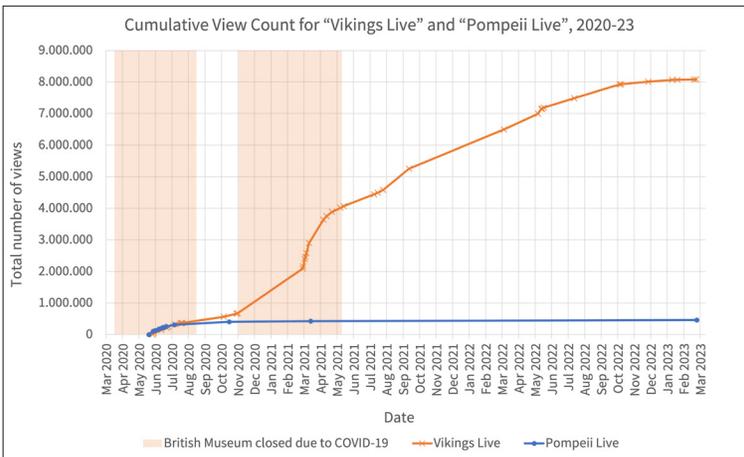
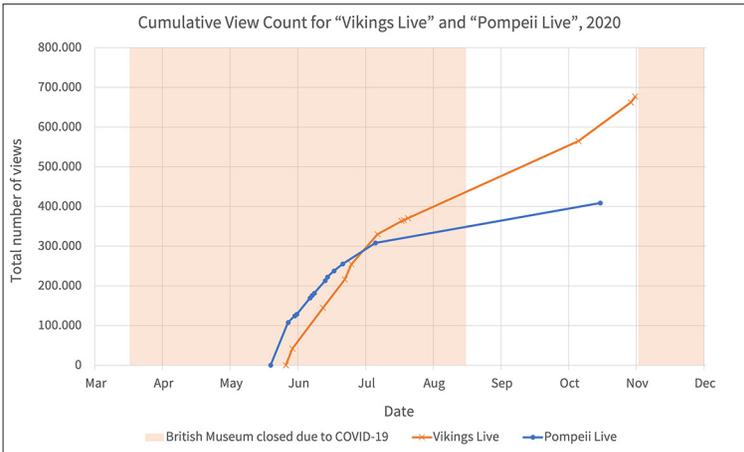


Figure 1a-b Graphs showing the cumulative number of views of two videos, "Vikings Live" and "Pompeii Live", a) over the course of 2020 and b) over the course of three years, from 2020 to 2023

As figure 1a shows, "Vikings Live" was uploaded in the middle of the UK's first national lockdown. In uploading a virtual tour near the beginning of the pandemic, the British Museum likely benefited from the influx of online visitors looking for cultural content. Larger organisations were better able to quickly adapt to lockdown conditions than their smaller counterparts and increase the number of videos

they were uploading to YouTube faster (Charlesworth et al. 2023, 13). ICOM found that large museums were more likely to provide digital services, while a UNSECO survey of over 50,000 museums, found that large museums – many of whom already had a bank of digital resources – saw the largest rise in visitor numbers to their websites over the first lockdown period (ICOM 2021, 18; UNESCO 2020). The British Museum was similarly able to quickly respond to demand for at-home content, utilising their pre-existing YouTube channel which already had an established follower base. Comparatively, organisations without existing audiences or staff expertise struggled to establish themselves in highly competitive online market (NEMO 2020; Arthurs, Drakopoulou, Gandini 2018). This provides the first indications that these metrics are not reflective of the content itself, but are heavily influenced by how, when, and by whom the content is uploaded.

Notably, despite the exceptional circumstances in which the video was uploaded to YouTube, its initial reception was in line with other videos on the British Museum’s channel. In the first three months after it was uploaded, the number of views remained below 500,000. This is not unusual, which is best illustrated when compared to the view count of “Pompeii Live from the British Museum”. This video, uploaded a week before “Vikings Live” on May 20, 2020, provides a useful benchmark for comparison. Filmed in 2013, it is remarkably similar to “Vikings Live”, sharing both a format and presenters. Yet in 2023, it still has less than 500,000 views and is only the 24th most popular video on the museum’s channel. Does this mean that “Vikings Live” is more engaging or a better video? By breaking down the number of views changed over time, it is possible to see why this kind of evaluation can be misleading.

When initially uploaded to YouTube, the two videos both performed similarly. On first being released, the number of views for both videos increased rapidly before slowing down over the summer of 2020. Taken in isolation this early spike in audience views could be assumed to be shaped by the larger pandemic audiences, however, Kamiyama and Muratahas identified this lognormal distribution of views is typical of YouTube videos (2019, 1103). This should not be falsely attributed to audience viewing habits or the quality of engagement the video inspires, instead, it is the result of the way YouTube’s recommendation algorithm prioritises recently uploaded videos (Gregersen, Ørmen 2021; van Es 2020). The recommendation algorithm suggests videos for users to watch next, and combined with their ranking of search results, heavily influences how many people see a video (Arthurs, Drakopoulou, Gandini 2018). The effect of these systems can be seen placing the British Museum’s videos metrics in a wider context [fig. 1b]. During the second lockdown period, “Pompeii Live” did not see any increase in the number of views, meaning that the mere presence of house-bound audiences during the lockdowns did not automatically result in larger audiences for GLAM institutions.

In contrast, the second lockdowns saw an enormous increase in views for “Vikings Live”. The incomplete nature of the Wayback Machine’s archive means that there is little data from this changed over November and December of 2020. However, the stark contrast between the Pompeii and Vikings exhibition is indicative of the way videos are popularised on YouTube. While the specifics of the recommendation and ranking algorithms of YouTube have not been made publicly available, there are two factors that most likely contribute to this effect. Firstly, at the start of the first pandemic, “Vikings Live” proved the slightly more popular video, making it more likely to be recommended to YouTube users – an effect that is compounded over time. The more views a video has, the more likely it is to be recommended, the more it is recommended the more likely it is to be viewed. If the video is regularly recommended but users do not click on it, over time it is recommended less often; this is partially why outrageous titles and staged picture thumbnails are so prevalent on the platform (van Es 2020; Shimono, Kakui, Yamasaki 2020). The ranking of videos, both through the search and recommendation systems, massively affect the click-through rate on a new video; the first video in a list is almost six-times more likely to be clicked on than the last (Zhou, Khemmarat, Gao 2010). It is probable that these differences over time had contributed to the dramatically different viewing figures for “Pompeii Live” and “Vikings Live”.

Although there is a wealth of literature on these effects regarding viral videos, beauty content, and vlogs, it is relatively rare for the videos of GLAM institutions’ to be affected in such a way. The “Vikings Live” video proved such an extreme outlier that it was had to be removed from calculated averages over the authors’ survey of 315 UK museums (Charlesworth et al. 2023, 10). This highlights how limited ‘views’ are as a tool for evaluation on social media platforms: not only is the figure not a reflection of the content’s quality, it also does not reliably reflect the type of content that will do well in the future.

It may be possible to attribute some of the popularity of “Vikings Live” to the subject matter. During the pandemic, Vikings were on the UK syllabus and a topic that was taught in primary schools (Department of Education 2013). Yet in a large study of what drives the number of views on videos, Zhou et al. found that the recommendation system had far more long-term impact than promoting a video or embedding it on other social media sites (2016, 6052). Therefore, it is unlikely that people looking for Viking education content or following links from the British Museum website directly caused the high number of views on “Vikings Live”. It is however possible that this was a contributing factor in the recommendation algorithm pushing “Vikings Live” and not “Pompeii Live”.

As ongoing work on the YouTube algorithms is not made public, we can only speculate as to why certain topics and videos are chosen,

and the details of how it functions are not understood by users, or by the uploaders of videos. There is a wealth of speculation about the algorithm published on the platform itself, and network of self-styled algorithmic experts who sell advice on how to game the system have emerged on the platform (Bishop 2020; 2018). However, even amongst experts and specialists, the workings of the algorithm are a mystery (Cunningham, Craig, Silver 2016). Even the British Museum, a huge national institution with a dedicated digital team, has been unable to optimise their videos for YouTube’s algorithm. In their video posted on March 9, 2023, Nick Harris, a senior producer commented,

And if you could give the new videos a like when they come out, we would really appreciate it as it massively helps with the algorithm, which - to be honest - hasn’t been that kind to us of late. (British Museum 2023)

Ultimately, the unpredictability of the algorithm, and its lack of correlation to the quality of the content or social engagement makes it useless for evaluation. The type of content museums are more likely to make - irregularly uploaded, long-form content - is not optimised for sharing on YouTube (van Es 2020). By breaking down the view count of the “Vikings Live” video and comparing it to the similar video “Pompeii Live”, we have shown how unpredictable metrics on social media platforms can be. Whether an individual video is popular or not, is heavily influenced by factors beyond an organisation’s control, and as such, they cannot be used to compare between organisations or exhibitions over time by themselves. This severely limits their usefulness in evaluating how the design and curatorial decisions have impacted the reception of the video.

4.2 The Linen Hall Library

In order to understand what constitutes a useful metric, we will use another case study of an entirely different type of GLAM institution to highlight how diverse the use cases for an audience engagement metric would be. To this end, a smaller organisation was chosen from the 315 museums in the authors’ large-scale study. The Linen Hall Library in Belfast has been chosen as it is both small and independent, the most common type of organisation in the UK GLAM sector, making it more representative of the sector as a whole. The library was chosen ahead of the other organisations for two reasons. Firstly, the library’s social media following is above average for a small GLAM organisation, but it is not so popular as to be an outlier whose methods would be unapplicable to other organisations. Secondly, the library’s website and social media presence were regularly archived on the

Wayback Machine over the course of the pandemic. Combined with their easily accessible annual reports, this makes tracking changes to their website and digital strategy easier than for other organisations of a similar size.

The Linen Hall Library has diversified its income streams and relies on memberships, shop sales, and grants (Linen Hall Library 2020, 13). The library's funders include the Belfast City Council, and the Department for Communities in Northern Ireland, and as such, its commitment to the Northern Irish public is a central aspect of the organisation's aims and strategy. Outreach and engagement feature heavily in the way projects are framed, and there is a particular focus on the local. Visitors to the gift shop are charmingly described as "lovers of all things literary and local", but this seems equally apt for their wider audiences (Linen Hall Library 2019, 8). The people and histories in the site's immediate vicinity are placed at the heart of events, workshops, outreach programme, corporate partnerships, media relations, and the collection itself, with "Irish & Local studies" books making up the majority of item requests. The library also frequently collaborated with other GLAM and educational institutions based in Northern Ireland: National Museums Northern Ireland, the Nerve Centre, The Public Record Office of Northern Ireland, Conflict Textiles, Belfast's Grand Opera House, and the Conflict Archive on the Internet at Ulster University (Linen Hall Library 2019). The library is only one actor in a vibrant cultural ecosystem that aims to serve the diverse communities of Belfast.

When the library was forced to close on March 17 2020, their immediate response was two-fold: firstly, by the end of March, groups that facilitated social interaction were moved online, including the English and Irish book club; and secondly, the library highlighted local artists through a series of online exhibitions on the library's social media channels (Linen Hall Library 2019, 2). In contrast to the British Museum, this flurry of activity did not utilise old digital assets but sought to provide a space to meet for the library's existing community. This is typical of smaller organisations' responses to the pandemic, especially those with limited digital expertise (NEMO 2020; ICOM 2020). Live events were prioritised and by the end of the first lockdown a LGBT+ history group was established, alongside a new series of lectures on the library's political collection.

However, there were unanticipated but welcomed consequences to the rapid adoption of the digital. The library's 2020 report states, "Audiences have been developing rapidly from a national international perspective and numbers have averaged around 80 attendees per session" (Linen Hall Library 2020, 2). In the annual report for 2020, the library noted that by October it had hosted "20 online events with an (international) audience engagement of c.12,000 people" (2020, 7). This figure is larger than the library's annual visitor numbers, and

like many GLAM institutions in 2020, the digital content was “reaching new audiences far beyond our traditional catchment area” (Linen Hall Library 2020, 7). The Centre for Cultural Value and Audience Agency found that this apparent broadening of audiences during lockdown was the result of individuals who were already heavily engaged with their local GLAM organisations digitally visiting institutions geographically distant from their own location. There was no significant change in the demographic groups accessing cultural content; visitors were “often older, professional, had children and were based in larger cities (especially London), in line with what we would expect from existing studies” (Walmsley et al. 2022, 10).

The Linen Hall Library’s online audience highlights an interesting conundrum for GLAM professionals; to what extent is it worth pursuing higher audience numbers if the majority of online visitors are unable to travel to the physical site? Whereas tracking international audience engagement may benefit the British Museum - in 2019 64% of their visitors were international tourists - this is less true of the Linen Hall Library (Bailey 2023). GLAM organisations have consistently struggled to monetise their online audiences (Anderson 2018, 91), with the profitability of image licensing being repeatedly questioned (Grosvenor 2018; Tanner 2004), and print on demand services remaining the reserve of large institutions (Valeonti et al. 2019; Valeonti et al. 2018). To start monetising online audiences in this way requires a sustained investment in digital infrastructure, with an online shop also necessitating large upfront costs. It is therefore questionable whether it is worth the required resources to monetise online visitors if this does not directly serve the organisations’ target audience.

The Linen Hall Library’s online events have since been uploaded to YouTube, and it is clear they facilitated rich discussions about Northern Irish history and engaged local audiences as well. However, the depth of engagement is not captured in online metrics, with the majority of content not receiving any comments on social media or being viewed widely. This case study highlights that rich enrichment for a small audience is not currently measured in quantitative metrics. Considering the role many GLAM institutions play in their local communities, this makes it difficult for staff who did not attend every event to pinpoint which were the most successful, and to compellingly convey that information to funders. If organisations have the resources, interviews and surveys can help at this small scale, but is there any way the quantitative methods can be improved? This final section of the paper asks how we might begin to measure engagement in a way that is more meaningful for GLAM organisations.

5 Alternatives

By exploring social media metrics, our paper has built on the discussion around the use of quantitative measures in the cultural sector and highlighted the specific limitations presented by platforms metrics such as ‘views’, ‘likes’, and ‘shares’. While of limited use themselves, these numbers can contribute to far more sophisticated measures of engagement. This section will explore methods that could create more nuanced evaluations of online content and have been trialled at a small scale in the GLAM sector. It will briefly outline the benefits and drawbacks of three approaches: sentiment analysis, topic modelling, and network analysis. This section will not explore these computational methods in depth, but instead aims to highlight potential avenues for further research and experimentation in the GLAM sector that address some of the concerns raised regarding quantitative metrics.

5.1 Sentiment Analysis and Topic Modelling

Sentiment analysis is a machine learning tool that aims to quantify the emotional tone or opinion of a given text – usually tweets, comments, or reviews (Valdivia, Luzon, Herrera 2017). A text is then often identified as having a positive or negative sentiment, and more sophisticated approaches have been developed that label texts with greater granularity, for example ‘bored’, ‘confused’, or ‘inspired’ (Zhou, Ye 2020, 5; Gerrard, Sykora, Jackson 2017). Performing this kind of analysis on the comments of GLAM online resources can summarise large amounts of audience feedback and captures more nuanced information about the reception of online content.

In 2013, Villaespesa analysed tweets from the “Art in Action” festival at the Tate Modern. A quantitative analysis of the entire data set was combined with qualitative analysis, including sentiment analysis, on a subsection of the tweets. Each tweet in this subset was coded with a predetermined category using a qualitative data analysis software. Villaespesa then investigated visitors’ motivations and themes; although she noted that this was a time-consuming process, “automatic coding does not disclose the reason why the experience was positive or negative” (Villaespesa 2013). More recently, Arias preformed a qualitative analysis of a controversial curatorial initiative at the Manchester Art Gallery (2020). Initially utilising the metrics available through Twitter and the “predetermined metadata categories”, she supplemented this data with thematic analysis conducted using the software NVivo (Arias 2020, 135). This process of coding the topics present in the text requires close reading and human interpretation.

By using topic modelling – an unsupervised machine learning method that identifies clusters of similar words – researchers can automate identifying themes within the data set and the subject of a tweet, reducing the amount of human time needed to perform an analysis. By adopting a fully automated process, it will be possible to massively increase the size of data sets. Both of the aforementioned studies focused on a discreet time period and a single Twitter hashtag. Yet as Arvanitis, Gilmore, Florack, Zuanni noted when discussing the potential and pitfalls of machine learning, “longitudinal data tracking and the use of more data points (e.g., both in terms of the range of events and audiences evaluated) could ensure a safer and more insightful analysis and comparison of data” (Arvanitis et al. 2016). Automation makes this long-term monitoring possible and would enable the identification of sector-wide trends in online audiences. This could help inform GLAM organisations’ digital strategies by creating a benchmark and enable them to address the discrepancy between the rhetoric of broadening access and organisations’ ability to reach new audiences online (Walmsley et al. 2022).

A significant barrier to creating this benchmark is the difficulty in running an analysis on the content of smaller organisations who do not have large online audiences. Chatzopoulou, Sheng, and Faloutsos found that an average YouTube video receives one interaction – that is a comment or a like – for every 400 views (2010, 2). With half of museum YouTube channels in the UK having less than 46 subscribers, the majority of videos receive no comments at all (Charlesworth 2023, 11). The low number of comments on the Linen Hall Library Facebook posts is also indicative of this issue; it is clear that for most GLAM institutions there is simply not enough interaction on posts to evaluate audience engagement using sentiment analysis or topic modelling. This issue is compounded by a lack of expertise and resources in the sector; Nesta has found that “the majority of arts and cultural organisations still do not use data for important purposes such as understanding their audiences better through data analysis and profiling” (2017, 5).

By performing topic modelling and sentiment analysis on a data set that spans multiple organisations, larger trends can be identified to help provide insight into the audiences of small and medium sized institutions that do not have the resources to analyse their own data. Using this kind of data set could identify success digital strategies and help improve the content of GLAM organisations which do not have large online audiences. Improvements to both methods have made conducting such work significantly easier in recent years, making a large-scale endeavour feasible.

5.2 Network Analysis

Network analysis could prove similarly beneficial and would nuance our understanding of content's online reception by capturing the interconnected nature of organisations and their audiences. Many small organisations are highly networked, and this type of collaboration increased over the pandemic period (Crooke 2020). Linen Hall Library forms part of a rich tapestry of organisations, events, and projects, that share funding, volunteer bases, and audiences. Looking to these larger cultural ecosystems, and charting the way in which individual exhibitions, events, and even people are linked, it is possible to identify forms of value that are not captured by most online metrics. Network analysis excels at identifying actors who facilitate collaboration or disseminate information and has been used to identify key social media influencers at international and local levels (Ingenhoff, Calamai, Sevin 2021). It is able to chart the way information is spread and identify the most important actors in a network and is used both within individual organisations and across entire sectors.

In the museum sector, the same methods have been applied at a smaller scale. La Magnética conducted an analysis of the Twitter response to "Ask a Curator Day", which saw them analyse 47,546 tweets posted over three days (2014). The team used different measures of centrality to calculate Twitter users' relevance and influence in the communities participating in #AskACurator, and where possible, sections of the data illustrating key accounts in different countries were illustrated in network graphs. They note that this kind of analysis,

over a longer period of time allow[s] us to have a deeper understanding of our museum communities. How our followers relate to us, how they relate to each other, who are the most influential users in a group, etc. And this understanding leads to a better definition of our strategy and a better way to assess our Social Media efforts' results. (La Magnética 2014)

Kydros and Vrana similarly performed network analysis on museums using both Twitter and Instagram (Kydros, Vrana 2021; Vrana et al. 2019). In their study of Twitter, Vrana, Kydros, Kerhis, Theocharidis, Karavasilis looked at the tweets of the top 25 museums in Europe for a two-month period in 2019. They were able to identify users that facilitated exchange between international communities; unsurprisingly, 34% of the 30 most important users were official museum accounts, yet six nodes (20%) were "plain persons" (Vrana et al. 2019, 581). This approach provides an understanding of who information is reaching and via which avenues, enabling organisations to better adapt their digital strategies and identify potential collaborators.

Network analysis also facilitates a broader overview of interactions. The recent work of Chang et al. explores the Instagram followers of five South Korean museums (2022). They focused on categorizing types of visitors and identified six groups using community extraction, each with its own characteristics. Some groups were heavily orientated around a single central user - often a gallery - while others were far less centralised, with connections between users more evenly distributed amongst the group (Chang et al. 2022, 56:10). The different hashtags, subjects, and ways of disseminating information were popular in each group, enabling organisations to create content that better targets different types of users.

The potential use of network analysis is three-fold: as discussed, it can be applied to social media networks to identify key actors and better understand audiences; but it can also be used more broadly to map formal structures (Cheong, Cheong 2011). Marshall and Staeheli note that

network analysis can illustrate the ways in which some local organizations are more successful at attracting donors, and positioning themselves as key intermediaries between international funding and local partners. (2015, 60)

This could be a prove a helpful tool for funders by identifying organisations who are repeatedly overlooked and underfunded.

Network analysis nuances the social media metrics currently available to GLAM organisations, and computational advances - combined with increasing awareness of the potential of digital technologies in the sector - mean that these methods are likely to become more accessible in the coming years. Chang et al. highlight the benefits these computational methods have over traditional approaches;

visitor analysis on SNS [Social Networking Services] can be continuously updated after it has been designed, unlike an interview or observational survey that must be conducted individually and is time-consuming. (2022, 56:16)

Conducted on a large scale may require a collaborative effort and the backing of larger GLAM organisations and research institutions, but it is possible that the sector may be able to develop sustainable evaluative tools that better capture online audience engagement.

6 Conclusions

Recent research has highlighted a discrepancy between the rhetoric of broadening access and organisations' ability to broaden the demographic make-up of their audiences (Walmsley et al. 2022). Yet, researchers and GLAM professionals are unable to investigate why this is the case or improve their digital strategies without sufficient data and evaluative tools (UNESCO 2020).

Currently, the metrics widely available on social media platforms and most Web 2.0 applications are ineffective tools for measuring audience engagement. By contextualising why GLAM organisations are monitoring their audiences and placing their institutional aims in their wider funding context, the stark difference between what metrics are implied to indicate and what they actually reflect has become clear. 'Views', 'likes', 'click-throughs', 'shares', and the number of comments, are designed to facilitate a certain type of engagement - motivated by commercial interests - that bears little similarity to the type of open dialogue and meaningful participation GLAM organisations are trying to facilitate.

Through the use of two case studies, this paper has explored how we can contextualise these metrics and nuance what is inevitably a simplification of enormously complex outcomes (Espeland, Sauder 2016). It has highlighted that they are ineffective as evaluative tools. The numbers are shaped by factors that have nothing to do with the quality of the content itself and success on a platform cannot be reliably replicated, and if these metrics are used during the design process, GLAM organisations may optimise their content for the platform's algorithms rather than audience engagement.

It is therefore important to contextualise findings with audience surveys and interviews, but there is also hope that more sophisticated methods for measuring large-scale engagement online will become widely accessible of the next few years. Sentiment and network analysis alongside topic modelling can provide more nuanced understandings of the complex interactions and connections GLAM institutions aim to foster. With the uptake of digital technologies amongst GLAM organisations set to continue, this paper's exploration of how common metrics - and potential alternatives - can be adapted to meet different institutional needs, will become of increasing importance to researchers, GLAM practitioners, and designers alike.

Bibliography

- Adler, R.P.; Goggin, J. (2005). "What Do We Mean By 'Civic Engagement'?" *Journal of Transformative Education*, 3(3), 236-53. <https://doi.org/10.1177/1541344605276792>.
- Anderson, G. (2013). "Reflections on Organizational Transformation in the Twenty-First Century". Janes, R.R. (ed.), *Museums and the Paradox of Change*. London: Routledge, 192-204. <https://doi.org/10.4324/9780203383964>.
- Anderson, S. (2018). "Visitor and Audience Research in Museums". Drotner, K. et al. (eds), *The Routledge Handbook of Museums, Media and Communication*, New York: Routledge, 80-95.
- Arias, M.P. (2020). "From Takeover to Debacle: An Analysis of the Nymphgate Network Using Twitter Data". *Museum and Society*, 18(2), 132-50. <https://doi.org/10.29311/mas.v18i2.3275>.
- Aroles, J.; Hassard, J.; Hyde, P. (2022). "'Culture for Sale'. The Effects of Corporate Colonization on the UK Museum Sector". *Organization Studies*, 43(3), 347-68. <https://doi.org/10.1177/0170840621989001>.
- Arthurs, J.; Drakopoulou, S.; Gandini, A. (2018). "Researching YouTube". *Convergence: The International Journal of Research into New Media Technologies*, 24(1), 3-15. <https://doi.org/10.1177/1354856517737222>.
- Arts Council England (2016). *Arts Council Response to Quality Metrics Trial*. <https://web.archive.org/web/20230621212612/https://www.artscouncil.org.uk/sites/default/files/download-file/Arts%20Council%20Response%20to%20Quality%20Metrics%20trial%20FINAL.pdf>.
- Arts Council England (2021). *List of Accredited Museums in the UK, Channel Islands, and Isle of Man*. <https://web.archive.org/web/20210725083629/https://www.artscouncil.org.uk/supporting-arts-museums-and-libraries/uk-museum-accreditation-scheme>.
- Arts Council England (2023a). *Strategy 2020-2030: Outcomes*. <https://web.archive.org/web/20230627085343/https://www.artscouncil.org.uk/lets-create/strategy-2020-2030/outcomes>.
- Arts Council England (2023b). *Making the Case: Say It Loud on Social Media!* <https://web.archive.org/web/20230622072508/https://www.artscouncil.org.uk/make-case-creativity-and-culture/making-case-say-it-loud-social-media>.
- Arvanitis, K. et al. (2016). "Data Culture and Organisational Practice". *MW2016: Museums and the Web* (Los Angeles, 6-9 April 2016). <https://web.archive.org/web/20230626201522/https://mw2016.museumsandtheweb.com/paper/data-culture-and-organisational-practice/>.
- Ashley, S.L.T. (2014). "'Engage the World': Examining Conflicts of Engagement in Public Museums". *International Journal of Cultural Policy*, 20(3), 261-80. <https://doi.org/10.1080/10286632.2013.808630>.
- Audience Agency (2021). *Digital Audience Survey. Findings*. <https://web.archive.org/web/20230626201618/https://www.theaudienceagency.org/evidence/digital-audience-survey-findings>.
- Avenier, P. (1999). "Putting the Public First: The French Experience". *Museum International*, 51(4), 31-4. <https://doi.org/10.1111/1468-0033.00227>.

- Bailey, M. (2023). "Two Years on, London Museums Are Still Missing Visitors. What's Happened?". *The Art Newspaper*, 27 March. <https://web.archive.org/web/20230626201709/https://www.theartnewspaper.com/2023/03/27/two-years-on-london-museums-are-still-missing-visitorswhats-happened>.
- Bailey-Ross, C. (2014). *Radical Trust Works. An Investigation of Digital Visitor Generated Content and Visitor Engagement in Museum Spaces* [PhD dissertation]. London: University College London.
- Ballatore, A.; Candlin, F. (2022). "A Geography of UK Museums". *Transactions of the Institute of British Geographers*, November, tran.12578. <https://doi.org/10.1111/tran.12578>.
- Barringer, T.; Flynn, T. (1997). "Introduction". Barringer T.; Flynn, T. (eds), *Colonialism and the Object: Empire Material Culture and the Museum*. London: Routledge, 1-8.
- Belfiore, E. (2015). "'Impact', 'Value' and 'Bad Economics'. Making Sense of the Problem of Value in the Arts and Humanities". *Arts and Humanities in Higher Education*, 14(1), 95-110. <https://doi.org/10.1177/1474022214531503>.
- Bethlehem, J. (2010). "Selection Bias in Web Surveys". *International Statistical Review*, 78(2), 161-88. <https://doi.org/10.1111/j.1751-5823.2010.00112.x>.
- Bishop, S. (2018). "Anxiety, Panic and Self-Optimization. Inequalities and the YouTube Algorithm". *Convergence: The International Journal of Research into New Media Technologies*, 24(1), 69-84. <https://doi.org/10.1177/1354856517736978>.
- Bishop, S. (2020). "Algorithmic Experts. Selling Algorithmic Lore on YouTube". *Social Media + Society*, 6(1), 205630511989732. <https://doi.org/10.1177/2056305119897323>.
- Black, G. (2010). "Embedding Civil Engagement in Museums". *Museum Management and Curatorship*, 25(2), 129-46. <https://doi.org/10.1080/09647771003737257>.
- British Museum (2020a). *Home Page of the British Museum Website*. <https://web.archive.org/web/20200318142350/https://www.britishmuseum.org/>.
- British Museum (2020b). *Vikings Live. A Tour from the British Museum*. https://web.archive.org/web/20230501000000*/https://www.youtube.com/watch?v=NuL0Q0tsid4.
- British Museum (2020c). *One Day to Go Until Reopening*. <https://twitter.com/britishmuseum/status/1298515601892298759?s=20>.
- British Museum (2022). "The British Museum Annual Report and Accounts 2021 to 2022". Corporate report HC 438. London: Department for Digital, Culture, Media & Sport. <https://www.gov.uk/government/publications/the-british-museum-annual-report-and-accounts-2021-to-2022>.
- British Museum (2023). *New Season of Curator's Corner and Other Channel Updates*. <https://web.archive.org/web/20230326120638/https://www.youtube.com/watch?v=vrtDCfpXHsQ>.
- Brown, J. et al. (2021). *Coronavirus. A History of English Lockdown Laws*. <https://commonslibrary.parliament.uk/research-briefings/cbp-9068/>.

- Candlin, F. et al. (2019). "The Missing Museums. Accreditation, Surveys, and an Alternative Account of the UK Museum Sector". *Cultural Trends*, 29(1), 50-67. <https://doi.org/10.1080/09548963.2019.1690392>.
- Chang, M. et al. (2022). "Identifying Museum Visitors via Social Network Analysis of Instagram". *Journal on Computing and Cultural Heritage*, 15(3), 1-19. <https://doi.org/10.1145/3505635>.
- Charlesworth, E. et al. (2023). "Understanding Levels of Online Participation in the U.K. Museum Sector". *Museum Management and Curatorship*: 1-24. <https://doi.org/10.1080/09647775.2023.2188478>.
- Chatzopoulou, G.; Sheng, C.; Faloutsos, M. (2010). "A First Step towards Understanding Popularity in YouTube". *2010 INFOCOM IEEE* (San Diego, 15-19 March, 2010). New York: IEEE. <https://doi.org/10.1109/INFCOMW.2010.5466701>.
- Che, X.; Ip, B.; Lin, L. (2015). "A Survey of Current YouTube Video Characteristics". *IEEE MultiMedia*, 22(2), 56-63. <https://doi.org/10.1109/MMUL.2015.34>.
- Cheong, F.; Cheong, C. (2011). "Social Media Data Mining. A Social Network Analysis of Tweets During the 2010-2011 Australian Floods". *PACIS 2011 Proceedings* (Brisbane, 7-11 July 2011). Brisbane: Queensland University of Technology. <https://aisel.aisnet.org/pacis2011/46>.
- Crooke, E. (2020). "Communities, Change and the COVID-19 Crisis". *Museum and Society*, 18(3), 305-10. <https://doi.org/10.29311/mas.v18i3.3533>.
- Culture Metrics (2016). *Culture Metrics: A Shared Approach to Measuring Quality*. <https://web.archive.org/web/20220814211345/https://www.culturemetricsresearch.com/>.
- Cunningham, S; Craig, D.; Silver, J. (2016). "YouTube, Multichannel Networks and the Accelerated Evolution of the New Screen Ecology". *Convergence: The International Journal of Research into New Media Technologies*, 22(4), 376-91. <https://doi.org/10.1177/1354856516641620>.
- Department for Education (2013). *National Curriculum in England: History Programmes of Study*. <https://web.archive.org/web/20230605074601/https://www.gov.uk/government/publications/national-curriculum-in-england-history-programmes-of-study>.
- Es, K. van (2020). "YouTube's Operational Logic. 'The View' as Pervasive Category". *Television & New Media*, 21(3), 223-39. <https://doi.org/10.1177/1527476418818986>.
- Espeland, W.; Sauder, M. (2016). *Engines of Anxiety. Academic Rankings, Reputation, and Accountability*. New York: Russell Sage Foundation.
- Falk, J.; Dierking, L. (2000). *Learning from Museums: Visitor Experiences and the Making of Meaning*. Plymouth: AltaMira Press.
- Falk, J.; Dierking, L. (1995). "Recalling the Museum Experience". *Journal of Museum Education*, 20(2), 10-13. <https://doi.org/10.1080/10598650.1995.11510292>.
- Finnis, J.; Chan, S.; Clements, R. (2011). "Let's Get Real: How to Evaluate Online Success?". *Action Research Project*. London: Culture24. <https://web.archive.org/web/20230626204427/https://www.keepandshare.com/doc/3148918/culture24-howtoevaluateonlinesuccess-2-pdf-september-19-2011-11-15-am-2-5-meg?da=y&dnad=y>.
- Fire, M.; Guestrin, C.; (2019). "Over-Optimization of Academic Publishing Metrics. Observing Goodhart's Law in Action". *GigaScience*, 8(6), giz053. <https://doi.org/10.1093/gigascience/giz053>.

- Gerrard, D.; Sykora, M.; Jackson, T. (2017). "Social Media Analytics in Museums. Extracting Expressions of Inspiration". *Museum Management and Curatorship*, 32(3), 232-50. <https://doi.org/10.1080/09647775.2017.1302815>.
- Gillespie, T. (2010). "The Politics of 'Platforms'". *New Media & Society*, 12(3), 347-64. <https://doi.org/10.1177/1461444809342738>.
- Gilmore, A.; Glow, H.; Johanson, K. (2017). "Accounting for Quality: Arts Evaluation, Public Value and the Case of 'Culture Counts'". *Cultural Trends*, 26(4), 282-94. <https://doi.org/10.1080/09548963.2017.1382761>.
- Glass, S. (2015). "Insights First, Data Second". Audas, J. (ed.), *Making Digital Work: Data*. London: Digital R&D Fund for the Arts, 4-5. <https://web.archive.org/web/20230621202054/https://www.artscouncil.org.uk/media/6814/download?attachment>.
- Gnezdova, J.V.; Osipov, V.S.; Hriptulov, I.V. (2022). "Creative Industries. A Review of the Effects of the COVID-19 Pandemic". Osipov, V.S. (ed.), *Post-COVID Economic Revival*, vol. 2. Cham: Springer International Publishing, 159-71. https://doi.org/10.1007/978-3-030-83566-8_10.
- Gran, A. et al. (2019). "A Digital Museum's Contribution to Diversity. A User Study". *Museum Management and Curatorship*, 34(1), 58-78. <https://doi.org/10.1080/09647775.2018.1497528>.
- Gregersen, A.; Ørmen, J. (2021). "The Output Imperative. Productivity and Precarity on YouTube". *Information, Communication & Society*, 1-18. <https://doi.org/10.1080/1369118X.2021.2006745>.
- Grosvenor, B. (2018). "How Abolishing Museum Image Fees Could Boost Audiences". *The Art Newspaper*, 14 February 2018. <https://web.archive.org/web/20230621165326/https://www.theartnewspaper.com/2018/02/14/how-abolishing-museum-image-fees-could-boost-audiences>.
- Hooper-Greenhill, E. (1992). *Museums and the Shaping of Knowledge*. London: Routledge.
- ICOM, International Council of Museums (2021). "Museums, Museum Professionals and Covid-19: Third Survey". Museums, Museum Professionals and Covid-19. Paris: ICOM. https://web.archive.org/web/20230626201847/https://icom.museum/wp-content/uploads/2021/07/Museums-and-Covid-19_third-ICOM-report.pdf.
- inDICES. (2023). *inDICES: Measuring the Impact of Digital Culture*. <https://web.archive.org/web/20230626195913/https://indices-culture.eu/>.
- Ingenhoff, D.; Calamai, G.; Sevin, E. (2021). "Key Influencers in Public Diplomacy 2.0. A Country-Based Social Network Analysis". *Social Media + Society*, 7(1), 205630512098105. <https://doi.org/10.1177/2056305120981053>.
- Ji, Y.G. et al. (2019). "Functional and Emotional Traits of Corporate Social Media Message Strategies. Behavioral Insights from S&P 500 Facebook Data". *Public Relations Review*, 45(1), 88-103. <https://doi.org/10.1016/j.pubrev.2018.12.001>.
- Kamiyama, N.; Murata, M. (2019). "Reproducing Popularity Distribution of YouTube Videos". *IEEE Transactions on Network and Service Management*, 16(3), 1100-12. <https://doi.org/10.1109/TNSM.2019.2914222>.
- Kydros, D.; Vrana, V. (2021). "A Twitter Network Analysis of European Museums". *Museum Management and Curatorship*, 36(6), 569-89. <https://doi.org/10.1080/09647775.2021.1894475>.

- La Magnética (2014). #AskACurator through Social Network Analysis. <https://web.archive.org/web/20230626202922/https://www.lamagnetica.com/es/blog/askacurator-through-social-network-analysis/>.
- Larkin, J.; Ballatore, A.; Mityurova, E. (2023). "Museums, COVID-19 and the pivot to social media". *Curator: The Museum Journal*: 1-18. <https://doi.org/10.1111/cura.12558>.
- Linen Hall Library (2019). *Annual Report 2019*. Belfast, NI: Linen Hall Library. <https://web.archive.org/web/20230402185353/https://www.linenhall.com/wp-content/uploads/2023/01/annual-report-2019.pdf>.
- Linen Hall Library (2020). *Annual Report 2020*. Belfast, NI: Linen Hall Library. <https://web.archive.org/web/20230402103654/https://www.linenhall.com/wp-content/uploads/2023/01/annual-report-2020.pdf>.
- Long, S. (2013). "Practicing Civic Engagement. Making Your Museum into a Community Living Room". *Journal of Museum Education*, 38(2), 141-53. <https://doi.org/10.1080/10598650.2013.11510765>.
- Mantell, O.; Turpin, R. (2020). *Summary Report, COVID-19 Cultural Participation Monitor*. London, UK: The Audience Agency. <https://www.theaudienceagency.org/asset/2434>.
- Marshall, D.J.; Staeheli, L. (2015). "Mapping Civil Society with Social Network Analysis. Methodological Possibilities and Limitations". *Geoforum* 61: 56-66. <https://doi.org/10.1016/j.geoforum.2015.02.015>.
- NEMO, Network of European Museum Organisations (2020). *Survey on the Impact of the COVID-19 Situation on Museums in Europe. Final Report*. https://www.ne-mo.org/fileadmin/Dateien/public/NEMO_documents/NEMO_COVID19_Report_12.05.2020.pdf.
- Newman, A.; McLean, F. (2004). "Presumption, Policy and Practice. The Use of Museums and Galleries as Agents of Social Inclusion in Great Britain". *International Journal of Cultural Policy*, 10(2), 167-81. <https://doi.org/10.1080/1028663042000255790>.
- Newman, A.; McLean, F.; Urquhart, G. (2005). "Museums and the Active Citizen. Tackling the Problems of Social Exclusion". *Citizenship Studies*, 9(1), 41-57. <https://doi.org/10.1080/1362102042000325351>.
- Nesta (2017). "Digital Culture 2017". London: Nesta. https://web.archive.org/web/20230626181420/https://media.nesta.org.uk/documents/digital_culture_2017.pdf.
- Noehrer, L. et al. (2021). "The Impact of COVID-19 on Digital Data Practices in Museums and Art Galleries in the UK and the US". *Humanities and Social Sciences Communications*, 8(1), 236. <https://doi.org/10.1057/s41599-021-00921-8>.
- Nuccio, M.; Bertacchini, E. (2021). "Data-Driven Arts and Cultural Organizations. Opportunity or Chimera?" *European Planning Studies*, 30(9), 1-18. <https://doi.org/10.1080/09654313.2021.1916443>.
- Putnam, R.D. (1995). "Bowling Alone. America's Declining Social Capital". *Journal of Democracy*, 6(1), 65-78. <https://doi.org/10.1353/jod.1995.0002>.
- Rautela, S.; Sharma, S.; Virani, S. (2021). "Influence of Customer Participation in New Product Development. The Moderating Role of Social Media".

- International Journal of Productivity and Performance Management*, 70(8), 2092-2112. <https://doi.org/10.1108/IJPPM-05-2020-0241>.
- Rieder, B.; Coromina, O.; Matamoros-Fernández, A. (2020). "Mapping YouTube". *First Monday*. <https://doi.org/10.5210/fm.v25i8.10667>.
- Rieder, B.; Matamoros-Fernández, A.; Coromina, O. (2018). "From Ranking Algorithms to 'Ranking Cultures'. Investigating the Modulation of Visibility in YouTube Search Results". *Convergence: The International Journal of Research into New Media Technologies*, 24(1), 50-68. <https://doi.org/10.1177/1354856517736982>.
- Said, E.W. (1993). *Culture and Imperialism*. London: Random House.
- Samaroudi, M.; Echavarria, K.R.; Perry, L. (2020). "Heritage in Lockdown. Digital Provision of Memory Institutions in the UK and US of America during the COVID-19 Pandemic". *Museum Management and Curatorship*, 35(4), 337-61. <https://doi.org/10.1080/09647775.2020.1810483>.
- Schuch, J.C.; Harden, S.B.; Smith, H.A. (2023). "Museums Using Community Dialogue as a Building Block toward Cultural Competence". *Journal of Latinos and Education*, 22(1), 302-16. <https://doi.org/10.1080/15348431.2020.1749632>.
- Shimono, A.; Kakui, Y.; Yamasaki, T. (2020). "Automatic YouTube-Thumbnail Generation and Its Evaluation". *Proceedings of the 2020 Joint Workshop on Multimedia Artworks Analysis and Attractiveness Computing in Multimedia* (Dublin, 8 June 2020). Dublin: ACM, 25-30. <https://doi.org/10.1145/3379173.3393711>.
- Simon, N. (2010). *The Participatory Museum*. Santa Cruz: Museum 2.0.
- Speight, C. [2010] (2016). "Museums and Higher Education: A New Specialist Service?". Cook, B.; Reynolds, R.; Speight, C. (eds), *Museums and Design Education: Looking to Learn, Learning to See*. London: Routledge, 11-28.
- Strathern, M. (1997). "'Improving Ratings'. Audit in the British University System". *European Review*, 5(3), 305-21. [https://doi.org/10.1002/\(SICI\)1234-981X\(199707\)5:3<305::AID-EURO184>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2-4).
- Tanner, S.G. (2004). "Reproduction Charging Models & Rights Policy for Digital Images in American Art Museums: A Mellon Foundation Funded Study". London: King's Digital Consultancy Services. <https://web.archive.org/web/20230621165626/https://kclpure.kcl.ac.uk/portal/en/publications/reproduction-charging-models-am-p-rights-policy-for-digital-images>.
- UNESCO (United Nations Educational, Scientific and Cultural Organization) (2020). "Museums around the World in the Face of COVID-19". 0000373530. Paris: UNESCO Digital Library. https://web.archive.org/web/20230626202631/https://unesdoc.unesco.org/ark:/48223/pf0000376729_eng.
- Ünsal, D. (2019). "Positioning Museums Politically for Social Justice". *Museum Management and Curatorship*, 34(6), 595-607. <https://doi.org/10.1080/09647775.2019.1675983>.
- Valdivia, A.M.; Luzon, V.; Herrera, F. (2017). "Sentiment Analysis in TripAdvisor". *IEEE Intelligent Systems*, 32(4), 72-7. <https://doi.org/10.1109/MIS.2017.3121555>.
- Valeonti, F. et al. (2018). "Reaping the Benefits of Digitisation: Pilot Study Exploring Revenue Generation from Digitised Collections through Technological Innovation". *Proceedings of the Conference on Electronic Visualisation*
-

- and the Arts* (9-13 July 2018) Swindon: BCS Learning & Development Ltd, 56-63. <https://doi.org/10.14236/ewic/EVA2018.11>.
- Valeonti, F. et al. (2019). "Examining Mobile Print-on-Demand as an Alternative to Image Licensing for Monetising Digitisation to Promote OpenGLAM". *Proceedings of MW19, the 23rd Annual MuseWeb Conference* (Boston, 2-6 April 2019). Boston: MuseWeb. <https://web.archive.org/web/20230626202143/https://discovery.ucl.ac.uk/id/eprint/10075084/>.
- Villaespesa, E. (2013). "Diving into the Museum's Social Media Stream: Analysis of the Visitor Experience in 140 Characters". *MW2013: Museums and the Web* (Portland, 17-20 April 2013). <https://web.archive.org/web/20230626201331/https://mw2013.museumsandtheweb.com/paper/diving-into-the-museums-social-media-stream/>.
- Voorveld, H.A.M. et al. (2018). "Engagement with Social Media and Social Media Advertising. The Differentiating Role of Platform Type". *Journal of Advertising*, 47(1), 38-54. <https://doi.org/10.1080/00913367.2017.1405754>.
- Vrana, V. et al. (2019). "A Network Analysis of Museums on Instagram". Kavoura, A; Kefallonitis, E.; Giovanis, A. (eds), *Strategic Innovative Marketing and Tourism*. Cham: Springer International Publishing, 1-10. https://doi.org/10.1007/978-3-030-12453-3_1.
- Walmsley, B. et al. (2022). *Impacts of Covid-19 on the UK Cultural Sector and Where We Go from Here*. https://web.archive.org/web/20230626200558/https://www.culturehive.co.uk/wp-content/uploads/2022/01/Culture_in_Crisis.pdf.
- Xiang, Y. (2022). "YouTube and the Protocological Control of Platform Organisations". *Qualitative Research in Accounting & Management*, 19(3), 348-72. <https://doi.org/10.1108/QRAM-04-2021-0060>.
- Zhou, J.; Ye, J. (2020). "Sentiment Analysis in Education Research. A Review of Journal Publications". *Interactive Learning Environments*: 1-13. <https://doi.org/10.1080/10494820.2020.1826985>.
- Zhou, R.; Khemmarat, S.; Gao, L. (2010). "The Impact of YouTube Recommendation System on Video Views". *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement* (Melbourne, 1-3 November 2010). Melbourne: ACM, 404-10. <https://doi.org/10.1145/1879141.1879193>.
- Zhou, R. et al. (2016). "How YouTube videos are discovered and its impact on video views". *Multimedia Tools and Applications*, 75, 6035-58. <https://doi.org/10.1007/s11042-015-3206-0>.

Semestral journal

Venice Centre

for Digital and Public Humanities

Università Ca' Foscari Venezia



Università
Ca' Foscari
Venezia