

The Linked Finding Aid as a Platform for Textual Research: The Case Study of the Giuseppe Raimondi Archive*

Francesca Giovannetti¹[0000-0001-6007-9118] and
Francesca Tomasi²[0000-0002-6631-8607]

¹ Department of Classical Philology and Italian Studies, University of Bologna, Italy
`francesc.giovannett6@unibo.it`

² Department of Classical Philology and Italian Studies, University of Bologna, Italy
`francesca.tomasi@unibo.it`

Abstract. This paper makes new suggestions for rethinking archival finding aids by means of linked open data. It does so by outlining some features of a conceptual model for writers' archives developed for the Digital Library of the Department of Classical Philology and Italian Studies of the University of Bologna.

The model, extending CIDOC-CRM/LRMoo, allows for the representation of complex collections of interrelationships between heterogeneous archival entities, especially texts. It also adopts named graphs as a way of enriching the finding aid with additional and possibly competing interpretations by researchers.

Through the case study of Giuseppe Raimondi's archive, the paper suggests how the adoption of linked open data can broaden the role of the digital finding aid to serve as a platform for archival and textual research.

Keywords: Digital Finding Aid · Linked Open Data · Writers' Archives.

1 Introduction

In his *Introduction to Archival Science*, Thomassen describes research on archives as “research on relations” [1]. The same point can be made about textual scholarship: creating a scholarly edition requires extensive research within archives to establish links between texts and documents. This paper considers how the role of the archival finding aid in the context of textual research can be rethought and expanded by means of linked open data and semantic web technologies.

Knowledge graphs have only recently been gathering attention as a way of publishing archives on the web. Compared to that of tree hierarchies, the data structure of knowledge graphs supports the creation of representations that can express higher orders of archival complexity. The descriptive potential of knowledge graphs is grounded on the semantic web architecture, which adopts the

* Section 1 is by F. Tomasi and Section 2 is by F. Giovannetti. Both authors contributed to Section 3.

Copyright 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Resource Description Framework (RDF) as its base, interoperable data model to convey information through semantic statements taking the form of subject-predicate-object expressions (see [2]).

The introduction of knowledge graphs into archival practice, however, does not merely provide archival practitioners with new technological tools but also challenges traditional approaches to archival representation both methodologically and conceptually. From a methodological point of view, the use of knowledge graphs as a method for archival representation prompts us to rethink the structure and organization of the archive towards a shift from neat hierarchies of records to fluid networks of logical interdependencies that can be arranged and rearranged into new representations. The semantic range of such interdependencies is virtually infinite, thanks to the possibility of combining predicates from different ontologies within a single graph and defining new predicates as needed.

From a conceptual point of view, reconfiguring the finding aid as a knowledge graph broadens its role in the context of textual research by allowing multiple interpretations by archivists and researchers to be incorporated into the archival representation as complex collections of interrelationships between heterogeneous entities (see [3]).

Our argument stems from a project undertaken by the Department of Classical Philology and Italian Studies (FICLIT)[4] and the Digital Humanities Advanced Research Centre (/DH.arc)[5] of the University of Bologna to expose as linked open data a selection of twentieth-century writers and intellectuals' archives.³ The Giuseppe Raimondi Archive was chosen as a pilot, representative case study for the project.

Giuseppe Raimondi (Bologna, 1898-1985) was an Italian writer. In the immediate aftermath of WWI, in 1918, he founded the literary journal *La Raccolta*, which published papers by European authors such as Vincenzo Cardarelli, Giuseppe Ungaretti, Guillaume Apollinaire and Blaise Cendrars, many of whom Raimondi met for the first time during the war. The archive, held at FICLIT, provides an example of what we might call 'multiple sedimentation': it contains heterogeneous material, both as to what regards carriers (notebooks, loose papers, albums, newspaper clippings, printed volumes) and document types (letters, drafts, notes, newspaper and journal articles, illustrated postcards, drawings, photographs).

Research on the archive have highlighted that the records it comprises are highly interconnected with one another as well as with material from other collections, and that the nature of such interrelations is heterogeneous. Consider, for example, a manuscript notebook, a newspaper article and a printed volume containing different versions of a text, possibly with internal corrections; these documents and their relationships span across various cultural heritage areas such as library, archival, museum and textual studies.⁴

³ See [6] for the list of personal archives held at FICLIT.

⁴ For an example of a subsequent scholarly reconstruction of interrelationships between heterogeneous archival documents in Giuseppe Raimondi's archive see [7].

Experiments with CIDOC CRM as a base ontology for representing archival information have already proved fruitful for demonstrating how classes and properties from CIDOC CRM could also be leveraged for the archival domain (see [8], [9], and [10]). However, none of these experiments have dealt specifically with the representation of the life cycle of writers' archives, and especially of the role of subsequent users-researchers as creators of additional interconnections between texts and documents. In addition, archival description practices to date have tended to focus on the representation of record sets rather than individual documents and texts.⁵ On the other hand, most existing digital scholarly editions prioritize a document-centred view of texts that uses TEI/XML markup [15] over LOD-based representations of textual phenomena and do not address the representation of the archival dimension in which texts participate.⁶

The primary goal of our project is to define a conceptual model that allows for the granular representation of complex interrelationships between heterogeneous documents and texts, such as those described above, within the finding aid. The model, extending CIDOC-CRM and LRMoo (formerly FRBRoo), leverages named graphs to enrich the finding aid with multiple and possibly competing readings by archivists and researchers.⁷ The section that follows presents selected features of the model through a practical example focusing on the representation of intertextual relationships as reconstructed by subsequent users-researchers.

2 Textual Research in the Finding Aid: An Example

Una forza per il poeta François Villon / [Giuseppe Raimondi]. - 1976. - 1 quaderno (7 p. ms., di cui alcune numerate irregolarmente, su 10 c.) ; 21 cm. ((In cop., di altra mano, anche: (Gelo invernale e nostalgia di legna accesa) Il giorno 7.6.76; I dattiloscritti sono inseriti ne "I tetti sulla città"; a c. [3v]: 24.5.1976; a c. [5v]: 3 luglio 1976. - Contiene anche: A proposito di tegole, di tetti e di fantasmi. - Ms.⁸

This archival note, taken from the existing finding aid of the Giuseppe Raimondi Archive, describes one of Raimondi's manuscript notebooks.⁹ The note reports four distinct titles: *Una forza per il poeta François Villon* (from now on T1), *Gelo invernale e nostalgia di legna accesa* (T2), *I tetti sulla città* (T3) and *A proposito di tegole, di tetti e di fantasmi* (T4). However, the note does not specify which texts are actually contained in the notebook nor the relationships between them, to the detriment of users and researchers. The description also includes three full dates, but they are not explicitly attributed to the corresponding texts.

Actually, the notebook only contains two of the mentioned texts, T1 (dated 24 May 1976) and T4 (3 July 1976). Both became part of T3, a collection of

⁵ Refer to [11] for an in-depth discussion of the topic.

⁶ One exception is the Paolo Bufalini's Notebook project, which describes intratextual relationships in LOD [12].

⁷ On the transition from FRBRoo to LRMoo see [13].

⁸ University of Bologna, BFICLIT, FR.A, QUADERNI.1 1976 03.

⁹ The existing finding aid of Giuseppe Raimondi's archive dates back to March 1993.

works by Giuseppe Raimondi, which was first published in September 1977. On the other hand, T2 is a variant version of T1 that was published as a newspaper article just a month after the creation of T1 (7 June 1976). The archive contains Giuseppe Raimondi’s copy of T3. Furthermore, FICLIT holds two additional copies of the same book in the personal archives of Clemente Mazzotta and Ezio Raimondi, one of which features noteworthy manuscript annotations. How can this scenario, which subsequent research on the archive reconstructed, be represented within the finding aid?

2.1 The Archivist’s Base Description of the Notebook

As anticipated above, our model, of which we only provide a brief account, adopts CIDOC-CRM as a basis for representing archival documents and order. Figure 1 and Figure 2 show a graphical representation of the archivist’s base description of the notebook (the graph is split into two parts for better reading). The following prefixes apply to all figures:

```
@base <https://w3id.org/ficlitdl/> .
@prefix fdlo: <https://w3id.org/ficlitdl/ontology/> .
@prefix crm: <https://w3id.org/ficlitdl/ontology/crm/> .
@prefix lrmoo: <https://w3id.org/ficlitdl/ontology/lrmoo/> .
@prefix pro: <http://purl.org/spar/pro/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix np: <http://www.nanopub.org/nschema#> .
```

CIDOC-CRM is the best candidate for integrating different conceptualizations into one model because it is event-based and already compatible with LRMoo. The notebook is modelled as an instance of E22 Human-Made Object that carries an F2 Expression incorporating two texts, T1 and T4 (Fig. 1, centre). Because it contains two distinct texts, the notebook is categorised both as a ‘notebook’ and as an archival ‘file’ through the *P2 has type* property (Fig. 1, top left). The notebook forms part of the archival unit, “QUADERNI.1 1976”, which is also an instance of E22.

One crucial feature of our model is that each physical document is tied to a IIIF manifest [14] (an instance of E73 Information Object) pointing to large-scale, zoomable facsimiles of the object. In a similar way, each expression is linked to a transcription, which can be encoded in TEI/XML (Fig. 1, bottom left). The direct linking of facsimiles and transcriptions to the archival knowledge base supports further research and can act as a basis for the production of new digital scholarly editions of our texts.¹⁰

The creation of the notebook is modelled as an instance of F28 Expression Creation, linked to a specific date and technique (handwriting). The role of author is assigned to Giuseppe Raimondi using the Publishing Role Ontology (PRO) model, which allows for the reification of roles in such a way that each role is always linked to a specific context [17]. In our case, Raimondi holds the

¹⁰ On the boundary between digital archive and edition see [16].

role of author in the context of the creation of the notebook. The role of author, the person holding such a role and the created document are linked together using the Role In Time class (Fig. 1, bottom right). Continuing Figure 1, Figure 2 shows how the Expression Creation activity is divided into two sub-activities, one for each text, each linked to a specific date, via the *P9 consists of* property (Fig. 2, top left).

The E13 Attribute Assignment class is used to model authorship attribution. The attribute assigned is the Role In Time, and the assignee is Giuseppe Raimondi. The agent responsible for the attribution (in this case an institution rather than an individual) and the time of attribution are linked to the instance of E13 (Fig. 1, bottom right).

2.2 Enhancing the Base Description with Subsequent Interpretations by Researchers

Among the objectives of our model is to support the incorporation of subsequent scholarly interpretations within the finding aid. Figure 3 shows a graph of statements, modelled according to LRMoo, that reconstructs the relationships between T1, T2 and T3 as established by a researcher analysing Giuseppe Raimondi's literary production. All three texts are modelled as instances of F2 Expression realizing the same work. The texts are linked to their carriers: T1 is carried by the notebook, while T3 is carried by multiple printed volumes available in the personal archives of Giuseppe Raimondi, Ezio Raimondi and Clemente Mazzotta. Because the volumes are from different archival collections, connecting them within the finding aid represents a fundamental step towards dismantling archival data siloes (see [18]).

Additional relationships inferred by the reasoner on the basis of our ontology are displayed in blue. These relationships, *FDLP2 has variant expression* and *FDLP3 is related by expression to*, automatically link together the alternate versions of a text and the physical documents containing such versions to facilitate search and retrieval.

The graph only shows a subset of possible text-to-text connections. Connections at a deeper level are also possible, such as fragment-to-fragment connections describing authorial changes from one version to another (fragments can be modelled as instances of E90 Symbolic Object belonging to an expression, while ontologies such as the Critical Apparatus Ontology (CAO) provide useful properties for the representation of corrections) [19]. For example one could represent authorial changes to the title, from the initial *Gelo invernale e nostalgia di legna accesa* to the final *Una forza per il poeta François Villon*, through the various intermediate stages. Using IIIF, it is also possible to establish links between circumscribed portions of the facsimiles and manuscript fragments.

2.3 Tying Each Set of Assertions to Their Provenance

In order to accommodate multiple perspectives in the finding aid, all collections of statements (in the case discussed above, there are two distinct graphs, the

archivist's and the researcher's) must be associated with provenance information. This allows the archival knowledge base to describe not only the archive but also the process of archival representation and to integrate more collections of statements over time.

Provenance information is modelled according to the Nanopublication framework [20]. The example below shows the basic structure of the nanopublication encapsulating the researcher's interpretation from Figure 3. It is composed of four graphs: 1. the graph of assertions being made by the researcher; 2. a graph describing the provenance of the assertions, 3. a graph describing the provenance of the publication itself; 4. the top graph combining the previous three graphs into a single nanopublication:

```
# Graph 1: the assertions being made.

:assertion-02 {
  # The researcher's reconstruction of the relationships between the
  texts (Fig. 2). }

# Graph 2: the provenance of the assertions.

:provenance-02 {
  :assertion-02 prov:generatedAtTime "2021-05-15T17:15:00Z"^^xsd:date ;
  prov:wasAttributedTo <person/francesca-giovannetti> . }

# Graph 3: the provenance of the nanopublication itself.

:pubinfo-02 {
  :nanopub-02 prov:generatedAtTime "2021-05-15T17:15:00Z"^^xsd:dateTime ;
  prov:wasAttributedTo <person/francesca-tomasi> . }

# Graph 4: the nanopublication and its components.

:head-02 {
  :nanopub-02 a np:Nanopublication ;
  np:hasAssertion :assertion-02 ;
  np:hasProvenance :provenance-02 ;
  np:hasPublicationInfo :pubinfo-02 . }
```

3 Concluding Remarks

Even in the digital environment, the finding aid remains a key tool for discovery and access. Its reconfiguration as an archival knowledge base has the potential to transform the finding aid into an expanding research platform where complex interrelationships between heterogeneous entities can be explicitly represented.

The primary goal of the project described in this paper is to define a conceptual model for representing writers' archives that allow for the expression of such interrelationships, with a special focus on connecting texts. Thanks to the use of named graphs, and nanopublications in particular, the model supports

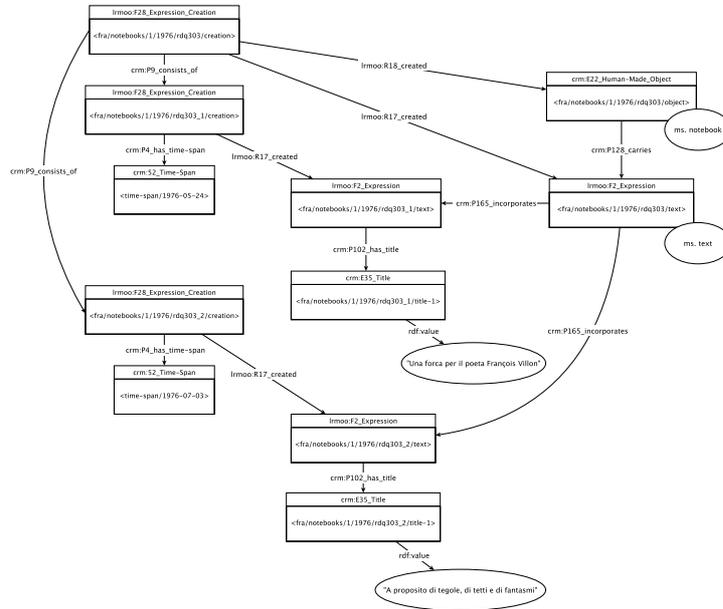


Fig. 2. The archivist's base description of the notebook (continued).

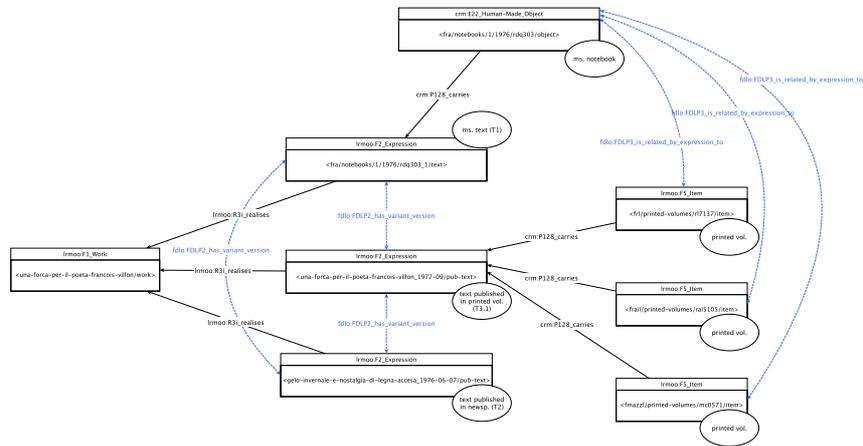


Fig. 3. The researcher's subsequent reconstruction of the interrelationships between the texts.

References

1. Thomassen, T.: A First Introduction to Archival Science. *Archival Science* **1**(4), 373–385 (2001)
2. Wood, D., Cyganiak, R., Lanthaler, M.: RDF 1.1 Concepts and Abstract Syntax. W3C Recommendation, <https://www.w3.org/TR/rdf11-concepts>. Last accessed 15 Jun 2021
3. Light, M., Hyry, T.: Colophons and Annotations: New Directions for the Finding Aid. *The American Archivist* **65**(2), 216–230 (2002)
4. FICLIT Homepage, <https://ficlit.unibo.it/it>. Last accessed 15 Jun 2021
5. DH.arc Homepage, <https://centri.unibo.it/dharc/en>. Last accessed 15 Jun 2021
6. I fondi archivistici e bibliografici del Dipartimento di Filologia Classica e Italianistica (FICLIT), <https://ficlit.unibo.it/it/biblioteca/collezioni/gli-archivi-culturali>. Last accessed 15 Jun 2021
7. Rossi, F., Wenzlawski, A.: Nello scrittoio di Giuseppe Raimondi: carte e libri di un letterato bolognese su Paul Valéry. In: Di Domenico, G., Sabba, F. (eds.) *Il privilegio della parola scritta: gestione, conservazione e valorizzazione di carte e libri di persona*, pp. 177–194. Associazione Italiana Biblioteche (AIB), Roma (2020). <https://doi.org/10.1400/276891>
8. Bountouri, L., Gergatsoulis, M.: Mapping Encoded Archival Description to CIDOC CRM. In: *Proceedings of the First Workshop on Digital Information Management*, pp. 8–25 (2011). <https://core.ac.uk/download/pdf/11888004.pdf>
9. Daquino, M., Tomasi F.: Linked Cultural Objects: dagli standard di catalogazione ai modelli per il web of data. *Spunti di riflessione dalla Fototeca Zeri. Umanistica Digitale* **1** (2017, October). <https://doi.org/10.6092/issn.2532-8816/7195>
10. Koch, I., Freitas, N., Ribeiro, C., Lopes, C. T., da Silva, J. R.: Knowledge graph implementation of archival descriptions through CIDOC-CRM. In: *International Conference on Theory and Practice of Digital Libraries*, pp. 9–16 (2019, September)
11. Yeo, G.: Contexts, Original Orders, and Item-Level Orientation: Responding Creatively to Users’ Needs and Technological Change. *Journal of Archival Organization*. **12**, 170–185 (2015). <https://doi.org/10.1080/15332748.2015.1048626>.
12. Daquino, M., Giovannetti, F., Tomasi, F.: Linked Data per le edizioni scientifiche digitali. Il workflow di pubblicazione dell’edizione semantica del quaderno di appunti di Paolo Bufalini. *Umanistica Digitale* **7**, 49–75 (2019). <https://doi.org/10.6092/issn.2532-8816/9091>.
13. Riva, P., Žumer, M.: FRBRoo, the IFLA Library Reference Model, and Now LR-Moo: A Circle of Development. IFLA WLIC – Kuala Lumpur, Malaysia – Transform Libraries, Transform Societies (2018). <http://library.ifla.org/id/eprint/2130>
14. International Image Interoperability Framework (IIIF) Homepage, <https://iiif.io>. Last accessed 16 Jun 2021
15. Text Encoding Initiative (TEI), <https://iiif.io>. Last accessed 16 Jun 2021
16. Eggert, P.: The Reader-Oriented Scholarly Edition. *Digital Scholarship in the Humanities* **31**(4), 797–810 (2016). <https://doi.org/10.1093/lc/fqw043>
17. Peroni, S., Shotton, D., Vitali, F.: Scholarly publishing and the Linked Data. Describing Roles, Statuses, Temporal and Contextual Extents. In: *Proceedings of the 8th International Conference on Semantic Systems (i-Semantics 2012)*, pp. 9–16. Association for Computing Machinery, New York (2016). <https://doi.org/10.1145/2362499.2362502>
18. Nichols, S.: Time to Change our Thinking: Dismantling the Silo Model of Digital Scholarship. *Ariadne* **58** (2009)

19. Giovannetti, F.: The Critical Apparatus Ontology (CAO). Modelling the TEI Critical Apparatus as a Knowledge Graph. In: Spadini, E., Tomasi, F., Vogeler, G. (eds.) *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, pp. 127–141. *Schriften des Instituts für Dokumentologie und Editorik*, 15 (2021) (in press)
20. Groth, P., Gibson, A., and Velterop, J.: The Anatomy of a Nanopublication. *Information Services and Use* **30**(1–2), 51–56 (2010)