Università di Catania

ASSOCIAZIONE per l'INFORMATICA UMANISTICA e la CULTURA DIGITALE

Consiglio Nazionale delle Ricerche

ME.TE. DIGITALI

MEDITERRANEO IN RETE TRA TESTI E CONTESTI

ATTI DEL XIII CONVEGNO ANNUALE
AIUCD 2024

28 - 30 MAGGIO

MONASTERO DEI BENEDETTINI

P.ZZA DANTE, 32 CATANIA

ISBN 978-88-942535-8-0

A cura di: Di Silvestro Antonio; Spampinato Daria (2024). Me.Te. Digitali. Mediterraneo in rete tra testi e contesti, Proceedings del XIII Convegno Annuale AIUCD, Catania 28-30 maggio 2024, Università di Catania.

Editing: Denise Bruno; Christian D'Agata; Laura Mazzagufo; Francesca Prado; Eliana Vitale; Alessandro Zammataro.

Ultimo accesso agli URL in data 15 maggio 2024.

Si prega di notificare all'editore ogni omissione o errore si riscontri: aiucd.segreteria [at] aiucd.org
Please notify the publisher of any omissions or errors found: aiucd.segreteria [at] aiucd.org

Il programma della conferenza AIUCD 2024 è disponibile online https://aiucd2024.unict.it/programma/
The AIUCD 2024 Conference Program is available online https://aiucd2024.unict.it/programma/

I contributi pubblicati nel presente volume hanno ottenuto il parere favorevole da parte di valutatori esperti della materia, attraverso un processo di revisione anonima mediante *double-blind peer review* sotto la responsabilità del Comitato di Programma di AIUCD 2024.
All the papers published in this volume have received favourable reviews by experts in the field of DH, through an anonymous double-blind peer review process under the responsibility of the AIUCD 2024 Program Committee.

Chair

Antonio Di Silvestro (Università di Catania)
Daria Spampinato (CNR Istituto di Scienze e Tecnologie della Cognizione)

Comitato di programma / *Program committee*

Emmanuela Carbé (Università di Siena)
Massimo Cultraro (CNR Istituto di Scienze del Patrimonio Culturale)
Christian D'Agata (Università di Catania)
Antonio Di Silvestro (Università di Catania)
Greta Franzini (Eurac Research)
Maurizio Lana (Università del Piemonte Orientale)
Cristina Marras (CNR Istituto del Lessico intellettuale europeo e Storia delle Idee)
Marco Mazzone (Università di Catania)
Ouafae Nahli (CNR Istituto di Linguistica Computazionale "Antonio Zampolli")
Marianna Nicolosi-Asmundo (Università di Catania)
Marina Paino (Università di Catania)
Giuseppe Palazzolo (Università di Catania)
Jonathan Prag (University of Oxford Merton College)
Daria Spampinato (CNR Istituto di Scienze e Tecnologie della Cognizione)
Rachele Sprugnoli (Università di Parma)
Francesco Stella (Università di Siena)

Segreteria scientifica / *Scientific Secretariat*

Liborio Barbarino (Università di Catania)
Denise Bruno (Università di Catania)
Giulia Cacciatore (Università di Catania)
Giuseppe Canzoneri (Università di Catania)
Elisa Conti (Università di Catania)
Milena Giuffrida (Università di Catania)
Miryam Grasso (Università di Catania)
Francesca Prado (Università di Catania)
Emilio M. Sanfilippo (CNR Istituto di Scienze e Tecnologie della Cognizione)
Eliana Vitale (Università di Catania)
Alessandro Zammataro (Università di Catania)

Comunicazione istituzionale: Claudia Cantale (Università di Catania) e Area Per la Comunicazione dell'Università di Catania (ACOM).
Institutional communication: Claudia Cantale (University of Catania) and the Area for Communication of the University of Catania (ACOM)

Supporto tecnico: Rosario Agrò, Area della Terza Missione dell'Università di Catania, per la consulenza e la progettazione grafica dei materiali informativi del convegno.
Technical support: Rosario Agrò, Third Mission Area of the University of Catania, for advice and graphic design of the conference information materials.

Enti organizzatori / *Organisers*
AIUCD; Università di Catania: Dipartimento di Scienze Umanistiche; CNR Istituto di Scienze e Tecnologie della Cognizione; CINUM: Centro di Informatica Umanistica dell'Università di Catania.

Supporter
CLARIN-IT; Neperia Group; Storage; programma Piaceri 2020-2022, Linea 1; Parmalat-Sole.

Chair di area/ Track chair

Le culture digitali nel Mediterraneo
Cristina Marras (CNR Istituto del Lessico intellettuale europeo e Storia delle Idee)
Paola Moscati (CNR Istituto di Scienze del Patrimonio Culturale)
Archivi ed edizioni digitali
Christian D'Agata (Università di Catania)
Greta Franzini (Eurac Research)
Analisi computazionale dei testi
Angelo Mario Del Grosso (CNR Istituto di Linguistica Computazionale "Antonio Zampolli")
Simone Rebora (Università di Verona)
Ontologie e Semantic Web
Marianna Nicolosi Asmundo (Università di Catania)
Francesca Tomasi (Università di Bologna)
Preservazione della memoria e del patrimonio digitale
Fabio Ciracì (Università del Salento)
Anna Maria Marras (Università di Torino)

Lista dei revisori /List of reviewers

Maristella Agosti (Università di Padova), **Stefano Allegrezza** (Università di Bologna), **Chiara Alzetta** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Liborio Barbarino** (Università di Catania), **Nicola Barbuti** (Università di Bari Aldo Moro), **Stefano Bazzaco** (Università di Verona), **Benedetta Bessi** (Università Ca' Foscari di Venezia), **Andrea Bolioli** (ricercatore indipendente), **Paolo Bonora** (Università di Bologna), **Federico Boschetti** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Dominique Brunato** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Flavia Bruni** (Università Gabriele D'Annunzio di Chieti-Pescara), **Marina Buzzoni** (Università Ca' Foscari di Venezia), **Alberto Campagnolo** (Universitè Catholique de Louvain/KULeuven), **Anna Cappellotto** (Università di Verona), **Emmanuela Carbé** (Università di Siena), **Vittore Casarosa** (CNR Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" – Università di Pisa), **Fabio Ciotti** (Università di Roma "Tor Vergata"), **Fabio Ciracì** (Università del Salento), **Elisa Conti** (Università di Catania), **Salvatore Cristofaro** (CNR Istituto per il Lessico Intellettuale Europeo e Storia delle Idee), **Christian D'Agata** (Università di Catania), **Elisa D'Argenio** (HUN-REN Hungarian Research Centre for Linguistics), **Mauro De Bari** (Università di Bari Aldo Moro), **Riccardo Del Gratta** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Angelo Mario Del Grosso** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Saulo Delle Donne** (Università del Salento), **Giorgio Maria Di Nunzio** (Università di Padova), **Antonio Di Silvestro** (Università di Catania), **Filippo Diara** (Università di Torino), **Giulia Fabbris** (Università Ca' Foscari di Venezia), **Riccardo Fedriga** (Università di Bologna), **Franz Fischer** (Università Ca' Foscari di Venezia), **Greta Franzini** (Eurac Research), **Francesca Frontini** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Daniele Fusi** (Stuttgart University & VeDPH – Università Ca' Foscari di Venezia), **Carola Gatto** (Università del Salento), **Lucia Giagnolini** (Università di Bologna), **Emiliano Giovannetti** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Milena Giuffrida** (Università di Catania), **Edmondo Grassi** (Università San Raffaele di Roma), **Miryam Grasso** (Università di Catania), **Alessandro Iannella** (Università di Cagliari - Università di Pisa – Università di Torino), **Paola Italia** (Università di Bologna), **Maurizio Lana** (Università del Piemonte Orientale), **Pietro Maria Liuzzo** (Bibliotheca Hertziana), **Dominique Longrée** (Université de Liège), **Francesco Mambrini** (Università Cattolica del Sacro Cuore di Milano), **Tiziana Mancinelli** (Istituto Italiano di Studi Germanici), **Anna Maria Marras** (Università di Torino), **Cristina Marras** (CNR Istituto del Lessico intellettuale europeo e Storia delle Idee), **Federico Meschini** (Università della Tuscia), **Alessio Miaschi** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Paolo Monella** (Università Sapienza di Roma), **Ouafae Nahli** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Marianna Nicolosi-Asmundo** (Università di Catania), **Giuseppe Palazzolo** (Università di Catania), **Valentina Pasqual** (Università di Bologna), **Gianluca Pavani** (Università di Roma "Tor Vergata"), **Giulia Pedonese** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Jonathan Prag** (University of Oxford Merton College), **Simone Rebora** (Università di Verona), **Giulia Renda** (Università di Bologna), **Roberto Rosselli Del Turco** (Università di Torino), **Enrica Salvatori** (Università di Pisa), **Emilio M. Sanfilippo** (CNR Istituto di Scienze e Tecnologie della Cognizione), **Eva Sassolini** (CNR Istituto di Linguistica Computazionale "Antonio Zampolli"), **Pietro Sichera** (CNR Istituto di Scienze e Tecnologie della Cognizione), **Daniele Silvi** (Università di Roma "Tor Vergata"), **Elena Spadini** (University of Basel), **Daria Spampinato** (CNR Istituto di Scienze e Tecnologie della Cognizione), **Linda Spinazzè** (Università Ca' Foscari di Venezia), **Rachele Sprugnoli** (Università di Parma), **Francesco Stella** (Università di Siena), **Cecilia Tamagnini** (Università di Bologna), **Timothy Tambassi** (Università Ca' Foscari di Venezia), **Francesca Tomasi** (Università di Bologna), **Marco Venuti** (Università di Catania), **Fabio Vitali** (Università di Bologna).

SOMMARIO

Lost in datification? The journey of data from the primary source to the final interpretation

Enrica Bruno[1], Sofia Baroncini[2], Francesca Tomasi[3]
[1] University of Bologna, Italy - enrica.bruno2@unibo.it
[2] Leibniz Institute for European History (IEG), DH Lab, Germany - baroncini@ieg-mainz.de
[3] University of Bologna, Italy - francesca.tomasi@unibo.it

ABSTRACT[1]

In the field of Digital Humanities, recent attention was given to the relationship between RDF triples and natural language in the context of natural language to RDF conversion of humanities texts. The rigid structure of ontologies obliges scholars to make critical choices during the formalization of data resulting from an interpretation of the cultural resource. This may result in crucial differences between the final RDF formalization and the natural language text in terms of how much of the final semantic content retains the original one and how much remains hidden due to the framework's formal structure. The verification of the adherence of structured data to the primary source is useful to test if the data model returns the semantic expressiveness of the primary source in order to pursue the specific goal driven by the computational process. In the context of humanities, in which the precision of information is fundamental for addressing sound analyses, this verification becomes crucial when derived data are treated as a substitute for the primary source in computational tasks. In this talk, we propose a three-step approach to verify the extent to which the RDF triples represent the respective content of the textual source from which they were generated within the limits of the modeling adopted. The approach is thus tested by proposing two case studies taken from two different cultural domains, namely literature and art history.

KEYWORDS

RDF data for humanities; RDF data quality; literature; art history.

1. INTRODUCTION

In the digital humanities field, growing attention is given to the issue of expressing the semantic content of texts (i.e., primary sources or their scholarly interpretation) in structured, computer-processable formats. Such formalization implies a selection of the information carried by the textual sources that will be expressed according to the specific perspective adopted during the modeling [19]. Nevertheless, even considering the adopted scope only, the process of transformation into structured data may lead to a manipulation of the primary information at various levels. Not only the actual content expressed may be misinterpreted, but also the ontological modeling adopted may affect its semantics. Furthermore, a user should contextualize data (i.e., how it was created and from which source) to avoid misleading interpretations during the data analysis. Assuming that structured data may be used as a faithful representation of the primary source in several computational analyses and information retrieval systems, the definition of the degree of difference between the resulting data and the source gains crucial importance. Despite several methodologies to verify ontology and triple readability being available [6, 18], to the extent of the author's knowledge, none of them addresses the topic of semantic alteration in the overall process of manually generating data from humanities texts. We state that this domain deserves particular attention to the accuracy of the created data for two reasons. First, due to the primary source complexity, the manual data creation does not guarantee its complete accuracy, as is generally accepted in the computer science field, since the transformation into data often requires a high degree of the annotator's interpretation, making the conversion a subjective, challenging task. Secondly, the consultation of the primary source in humanities practice has a crucial role in guaranteeing the trustworthiness of the analysis the humanist conducts. For this reason, in the context of the translation of the humanist's task to a computational one, the data should guarantee a level of accuracy equal to the primary sources from which it was extracted. Furthermore, this verification becomes crucial when derived data are treated as a substitute for the primary source in computational tasks.

In this talk, we propose a three-step approach to verify the extent to which the RDF triples represent the content expressed by the textual source from which they were generated within the limits of the modeling perspective adopted.

[1] Sofia Baroncini is responsible for Introduction and Conclusion. Enrica Bruno and Sofia Baroncini are responsible for the State of Arts, Methodology and Case Studies sections. Francesca Tomasi was the scientific supervisor providing critical revisions and feedback for the research and the entire writing process.

The main Research Question (RQ) can be expressed as follows: *to what extent do the RDF data and the resulting representations preserve the content of the textual source within the limits of the perspective adopted by the modeling?*

To answer this question, we identify three levels at which a semantic alteration may occur in the process of conversion of a humanist text into RDF data:

Meaning: is the semantic meaning expressed by the structured data adherent to the source (RQ1)?
Semantic modeling: are the modeling choices semantically expressive (RQ2)?
Contextualization: is the final RDF data contextualized, viz., is the reference to the provenance of data and statement in general indicated (RQ3)?

For each level, we provide an approach for evaluation adapted from metrics available in the state of the art, and we test it over two case studies from two diverse scenarios: the RDF conversion of 1) Italo Calvino's collection *Il castello dei destini incrociati* and 2) some of the art critique texts by the art historian Erwin Panofsky. The case studies were chosen as they belong to two different domains, namely literature and art history, and two respective types of sources (i.e. direct source and an interpretation of the primary source) to guarantee a wide range of applications in the broad humanities field.

2. STATE OF ART

Evaluation of quality is a crucial aspect when dealing with data. It aims at guaranteeing the correctness of data from multiple points of view, ranging from the logical aspects to the accuracy of the content expressed. The ISO standard related to data quality[2] defines it as the "degree to which data satisfy the requirements defined by the product-owner organization" in relation to 15 characteristics of the data. Furthermore, W3C provides a framework for data quality description[3], and multiple approaches for assessing semantic data quality data are present in the literature. Among them, worth mentioning [1], who extends data quality to every stage of the creation process, and [5], who provides a definition for 44 measures on the basis of the ISO definitions tested over large RDF datasets. Although many dimensions can be evaluated with an automatic or semi-automatic approach, some of them require human validation or the aid of domain experts, especially for provenance and contextual information criteria related to the specific use case [8, 4].

Several approaches providing ontology quality evaluation exist [6, 18] together with validator tools to ensure that ontologies are well-formed, consistent, and adhere to established methods (e.g. OWL API[4] Validator, Protégé Ontology Validator[5], and RDF Validator[6]). The validation of ontology semantics is usually done instead by involving domain experts as validators [17].

Various tools for converting text to RDF or the contrary are then currently available, showing an increasing accuracy in their results. Nevertheless, to the authors´ knowledge, they are not feasible for the task of validating the conversion of scholarly and literary text into RDF data. The quality of the conversion of text to RDF may be highly influenced by the writing style adopted and the absence of a specific structure, such as a database format [7, 16].

Furthermore, traditional RDF-to-text methods are based on hand-crafted and domain-specific rules of conversion. These methods provide a readability of the structured data [20, 11] where users are considered as testers and evaluators of the data model in order to verify an alignment between the user representation and the formal annotation [2], but it doesn't ensure the validity of the triple content. Despite algorithms based on showing promising results being available, their evaluation of correctness does not reach a sufficient score for the scope of this article [20]. For this reason, we adopt a human-based verification approach.

3. METHODOLOGY

As the verification occurs on the content level, the validators are experts in the domain of Digital Humanities, specifically semantic web technologies. The understandability of RDF data from generic users is not in the scope of this paper. The validation was performed by one validator for each case.

First step(RQ1): According to [1: 5] the semantic validity of triples is proved whether (i) it is available from a trusted source, (ii) it is common sense or (iii) the stated property can be directly measured. We adapt the metric by evaluating

[2] https://iso25000.com/index.php/en/iso-25000-standards/iso-25012
[3] https://www.w3.org/TR/vocab-dqv/
[4] https://owlapi.sourceforge.net/
[5] https://protegewiki.stanford.edu/wiki/Validation
[6] https://www.w3.org/RDF/Validator/

whether each triple from the case study expresses content that holds true when compared with the source text. The triple scores 1 if its content is fully represented, 0,5 if it is partially represented or inaccurate, and 0 if it is absent.

Second step (RQ2): Although several methodologies for ontological semantic validity based on experts' evaluation exist [6], our focus is to understand whether the adopted ontology is suitable to express the semantic content of the specific case study. To this end, we evaluate whether the type of entity assigned to each instance is suitable for describing the content of the phrase to which they refer. Therefore, all the triples having a relation for declaring the type of entity are extracted, and the accuracy of the type assigned is evaluated against the content provided by the text. The metric scores 1 if the class fully represents the content corresponding to the subject of the triple, 0,5 if it partially expresses it, and 0 if it does not express it.

Third step(RQ3): Contextual references concerning RDF triples are necessary to enable the verification of data [17], the assessment of reliability [9], and the analysis of the processes that generated the data [10]. For this respect, RDF data contextualization verifies whether the following aspects are made explicit through one or multiple statements:

- The entity responsible for the intellectual content of the resource (i.e., the creator)
- The indication of the source from which data were extracted
- Contextual information about the statement provenance

In this study, the evaluation is carried out on three possible levels (i.e., triple, entity, and graph), as our aim is to evaluate only whether the information is present. Similarly, the presence of contextual information expressed with different strategies (e.g. if the creator attribution is present both at the graph and statement level) does not affect the final score. Nevertheless, if the contextual information is provided at the assertion level, it should be verified whether every assertion examined has such information.
The first point scores 1 if the creator is stated and 0 if it is not. As regards the second point, the metric considers degrees of details, adding to the total score a) 0,5 if only the resource is cited, b) 0,25 if there is more specific information about the portion of text from which the triple is generated, and c) 0,25 if the text reference is present. Concerning the third point, the metric scores 0,75 if the responsibility of the statement is declared, as it is the fundamental aspect of provenance information. To this score, 0,25 is added if there is more information detail (e.g., the time when the statement was created).

4. CASE STUDIES

Here, we propose two case studies to test the three-step approach proposed to verify the extent to which the RDF triples properly represent the textual source from which they were generated. Both of the case studies are taken from two different cultural domains, namely literature and art history, and are resources manually created by domain experts. The main differences can be summarized in the table below (see Tab. 1).

BACODI	Iconology dataset
Literature	Art history
Direct source	Interpretation of the primary source
Description of the narrative-combinatorial relations between tarot cards in the collection's stories	Report of the work's interpretation (iconography and meanings)
Ontology *ad hoc*	Ontology based on theoretical approach, reused ontologies
Text with a defined structure	Discursive text

Table 1. Summary of the characteristics of the two case studies

BACODI (Base di Conoscenza dell'Ontologia dei *Destini incrociati* di Italo Calvino)

ODI (Ontologia dei Destini incrociati di Italo Calvino) and its corresponding Knowledge Base BACODI (Base di Conoscenza dell'Ontologia dei Destini incrociati di Italo Calvino) were created to represent and describe the first edition of Italo Calvino's *Il castello dei destini incrociati* [3] together with the description of the tarot cards used by the author to create twelve stories in the first homonymous collection of the work. In particular, BACODI stored the description of tarot cards considered both as cultural artifacts and as narrative instances in the text. The double descriptive dimension of each

tarot card combines the literary nature of the textual resource with its particular structure adopted by the author, who uses cultural objects from the artistic domain for storytelling[7].
The case study reported here is the fourth story of the collection, *Storia d'un ladro di sepolcri*, and in particular, the description of some tarot cards as narrative instances in the text, thus also considering their relationships within the story.

The Iconology dataset
The Iconology (ICON) dataset[8] represents the art interpretations by the art historian Erwin Panofsky expressed in four of his major contributions [12, 13, 14, 15]. For this purpose, an ontology based on his own theory was created (ICON ontology[9]). The interpretations concern ca. 400 artworks, mainly from the Western Early Modern period. Interpretations identify subjects and meanings depicted in the artworks according to the art historian and are distinguished on three levels of understanding, from a more superficial to a deeper one. In this way, objects, iconographies, and meanings are described, along with further details, if any (e.g., the textual or visual evidence supporting the recognition), and the provenance of the assertion.
The case study selected is the interpretation of a relief on the external wall of Modena Cathedral, which represents either Cupid or a personification of Death. The case is discussed in a chapter in which the tendency of the Middle Ages to read classical figures with a moral implication is treated.

5. RESULTS[10], CONCLUSION, AND FUTURE WORK

Table 2 shows the results of the analysis. Both datasets performed high scores in the first two steps of the evaluation, showing lower results only in the contextualization part (scores: 0,67 and 0,83). This is due to the fact that, on one hand, BACODI does not provide provenance information despite providing very precise indications of the creator of the content and of the text portion from which the information was extracted. On the other hand, the Iconology dataset provides provenance information, but it is not precise when indicating the text source, as only the reference to the overall book is provided.
Furthermore, despite showing similar results for the meaning aspect (0,95, and 0,96), details noted by the validators during the analysis need further consideration. In a few cases, the meanings noted in the triples were the result of a human interpretation of implicit knowledge. For instance, it is stated that the protagonist of Calvino´s story desires richness. Although this fact is never explicitly asserted by the text, it can be understood by reading the overall story. In other cases, the meanings are forged to facilitate modeling. For instance, in the Iconology dataset, it is stated that the artifact represents the action *grabbing with one hand,* whereas in the text, the phrasal verb *carrying* is used instead of *grabbing with*. This choice can be justified as a need to assimilate terms with more than one occurrence. As regards modeling choices, BACODI's results show an appropriate selection of the types for each instance, whereas the iconology dataset has little differentiation between entities represented in the artifact chosen (i.e. "torch", "putto", "ibis" and "wreath" are all in the class `icon:NaturalElement`). This is motivated by the different modeling scopes of the two cases. Whereas ODI was created to model Calvino´s specific book, the ICON ontology formalizes the domain of iconographical and iconological interpretations, aiming at being suitable for further domain descriptions.

	Results of step 1 (Meaning)	Results of step 2 (Semantic modeling)	Results of step 3 (Contextualization)
BACODI	0,95	1	0,67
Iconology dataset	0,96	0,85	0,83

Table 2. Results of the evaluation

[7] The complete documentation can be found at https://odi-documentation.github.io/materials/
[8] Available at https://iconology-dataset.streamlit.app/
[9] Available at https://w3id.org/icon/docs
[10] A complete overview of the results is available at https://doi.org/10.5281/zenodo.10973092

This talk focused on the semantic expressiveness of RDF triples, trying to question the extent to which data formalized represents the respective content of the native textual source. As the humanities texts are characterized by aspects of complexity (e.g., implicit, articulated, or undefined knowledge), their manual translation into data may be subject to variations at multiple levels. This aspect is crucial when considering the importance that the consultation of primary sources has for the humanities research advancement. The goal was to provide an approach to assess the extent to which semantic information manually extracted from humanities texts is valid, in relation to the layers of content, semantics, and contextualization. Being evaluated over case studies from two different domains and with different characteristics, we argue it can be applied to text belonging to different domains, either on a primary source or on scholarly literature. The fact that it is based on human evaluation allows us to verify also the modeling of implicit knowledge, which is likely to be embedded in complex texts such as literary and scholarly ones.

The current evaluation is limited to the extent to which the selected information provided by the text is correctly retained, expressed, and contextualized in data. Nevertheless, the semantic modeling of a resource aims not only at mirroring the plain text's content but also at enriching it with further knowledge that may be seen by a domain expert [19]. Moreover, a more extensive measurement involving multiple evaluators is needed to reduce the impact of evaluators' subjectivity on the results. Although the proposed approach does not aim to be an exhaustive solution to the challenging problem of derived data accuracy, it faces an initial critical reflection on the topic, proposing a practical method that can be applied in DH projects involving complex humanistic textual sources. Future work includes: 1) a more extensive evaluation including multiple evaluators and further domains, and 2) the extension of the approach to measure the added semantic enrichment that an expert of the considered domain may have included in the dataset during its creation process, as it constitutes a core aspect of the digital humanists' practice.

REFERENCES

[1] Assaf, Ahmad, and Aline Senart. 'Data Quality Principles in the Semantic Web'. In *2012 IEEE Sixth International Conference on Semantic Computing*, 226–229, 2012. https://doi.org/10.1109/ICSC.2012.39

[2] Bonora, Paolo, Martina Dello Buono, Francesca Giovannetti, and Francesca Tomasi. 'Tell Me the Truth. Validating the Semantic Alignment between the Annotation User Interface and the Knowledge Base'. In *Digital Humanities 2023: Book of Abstracts*, 202–4. Zentrum für Informationsmodellierung-Austrian Centre for Digital Humanities, University of Graz, 2023. https://doi.org/10.5281/zenodo.7961822.

[3] Calvino, Italo. *Il castello dei destini incrociati*. Torino: Einaudi, 1973.

[4] Daquino, Marilena, Valentina Pasqual, and Francesca Tomasi. 'Knowledge Representation of Digital Hermeneutics of Archival and Literary Sources'. *JLIS: Italian Journal of Library, Archives and Information Science = Rivista Italiana Di Biblioteconomia, Archivistica e Scienza Dell'informazione: 11, 3, 2020*, no. 3 (2020): 59–76. https://doi.org/10.4403/jlis.it-12642.

[5] Färber, Michael, Frederic Bartscherer, Carsten Menne, and Achim Rettinger. 'Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO'. Edited by Amrapali Zaveri, Dimitris Kontokostas, Sebastian Hellmann, and Jürgen Umbrich. *Semantic Web* 9, no. 1 (2017): 77–129. https://doi.org/10.3233/SW-170275

[6] Gangemi, Aldo, Carola Catenacci, Massimiliano Ciaramita, and Jos Lehmann. 'Modelling Ontology Evaluation and Validation'. In *European Semantic Web Conference*, 140–154. Berlin: Heidelberg: Springer Berlin Heidelberg, 2006. https://doi.org/10.1007/11762256_13

[7] Hassanzadeh, Kimia, Marek Reformat, Witold Pedrycz, Iqbal Jamal, and John Berezowski. 'T2R: System for Converting Textual Documents into RDF Triples'. In *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, 3:221–228, 2023. https://doi.org/10.1109/WI-IAT.2013.187

[8] Iyer, Vivek, Lalit Mohan Sanagavarapu, and Y. Raghu Reddy. 'A Framework for Syntactic and Semantic Quality Evaluation of Ontologies'. In *Secure Knowledge Management In The Artificial Intelligence Era*, edited by Krishnan Ram, H. Raghav Rao, Sanjay K. Sahay, Samtani Sagar, and Zhao Ziming, 73–93. Communications in Computer and Information Science. Cham: Springer International Publishing, 2022. https://doi.org/10.1007/978-3-030-97532-6_5

[9] McGlothlin, James P., Latifur Khan. 'Efficient RDF Data Management Including Provenance and Uncertainty'. In *Proceedings of the Fourteenth International Database Engineering and Applications Symposium*, 193–198. ACM, New York, 2010. https://doi.org/10.1145/1866480.1866508

[10] Moreau, Luc. 'The Foundations for Provenance on the Web'. *Foundations and Trends in Web Science* 2, no. 2–3 (2010): 99–241. https://doi.org/10.1561/1800000010

[11] Moussallem, Diego. 'Knowledge Graphs for Multilingual Language Translation and Generation'. *ArXiv Preprint ArXiv:2009.07715*, 2020.

[12] Panofsky, Erwin. *Meaning in the Visual Arts*. Garden City, NY: Doubleday, 1955.

[13] Panofsky, Erwin. *Renaissance and Renascences in Western Art*. New York: Harper & Row, 1972.

[14] Panofsky, Erwin. *Studies in Iconology: Humanistic Themes in the Art of the Renaissance.* Boulder, Colo: Westview Press, 1972.

[15] Panofsky, Erwin, and Fritz Saxl. 'Classical Mythology in Mediaeval Art'. *Metropolitan Museum Studies* 4, no. 2 (1933): 228–280. https://doi.org/10.2307/1522803

[16] Rincon-Yanez, Diego, and Sabrina Senatore. 'FAIR Knowledge Graph Construction from Text, an Approach Applied to Fictional Novels'. In *Proceedings of the 1st International Workshop on Knowledge Graph Generation from Text and the 1st International Workshop on Modular Knowledge Co-Located with 19th Extended Semantic Web Conference (ESWC 2022)*. Hersonissos, Greece, 2022.

[17] Sikos, L.F., and D. Philp. 'Knowledge Representation: A Survey of Data Models and Contextualized Knowledge Graphs'. *Data Sci. Eng.* 5 (2020): 293–316. https://doi.org/10.1007/s41019-020-00118-0

[18] Syed, Zafar Habeeb, Michael Röder, and Axel-Cyrille Ngonga Ngomo. 'Factcheck: Validating Rdf Triples Using Textual Evidence'. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018. https://doi.org/10.1145/3269206.3269308

[19] Tomasi, Francesca. *Organizzare La Conoscenza: Digital Humanities e Web Semantico*. Milano: Editrice Bibliografica, 2022. https://doi.org/10.53134/9788893573573

[20] Zhu, Yaoming, Juncheng Wan, Zhou Zhiming, Liheng Chen, Lin Qiu, Weinan Zhang, Xin Jiang, and Yong Yu. 'Triple-to-Text: Converting RDF Triples into High-Quality Natural Languages via Optimizing an Inverse KL Divergence'. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 455–464, 2019. https://doi.org/10.1145/3331184.3331232.