Versione PROVVISORIA del contributo presentato al Convegno Annuale

MIMA: a data model to represent multi-disciplinary analysis on manuscripts. Use case on Pellegrino Prisciani's *Historiae Ferrariae*

Valentina Pasqual[1], Marilena Daquino[2], Francesca Tomasi[3]
[1] Università di Bologna, Digital Humanities Advanced Research Centre (/DH.arc), valentina.pasqual2(«»)unibo.it
[2] Università di Bologna, Digital Humanities Advanced Research Centre (/DH.arc), marilena.daquino2(«»)unibo.it
[3] Università di Bologna, Digital Humanities Advanced Research Centre (/DH.arc), francesca.tomasi(«»)unibo.it

ABSTRACT
This research aims to explore and advance scholars' understanding of the coexistence of multiple interpretations in a formal framework, by providing a practical solution to represent them. The state of the art, the adopted methodology and approach, and the results applied on a significant case study are shown. The research has been narrowed by working on a use case – the collection of lectures called *Scrivere, rappresentare, conoscere nel rinascimento. Pellegrino Prisciani, un intellettuale eclettico tra la corte e il mondo*. As a result, MIMA (Multi-disciplinary Interpretations model on Manuscript Apparatus) aims to formally represent these aspects by leveraging Semantic Web technologies and the systematic reuse of already existing ontologies.

PAROLE CHIAVE
Semantic web, cultural heritage, manuscript, data modelling, interpretation, contestability

1. INTRODUCTION
Despite Digital Humanities research widely addresses aspects related to knowledge representation of cultural heritage, there is a lack of formal representations addressing multi-disciplinary interpretations over artefacts along with their contestability. Considering this gap in literature, this research wants to investigate which are the underlying methods of experts' interpretation – extracted by the use case which presents philological, iconographical and palaeographic analyses over Pellegrino Prisciani's *Historiae Ferrariae* illuminated manuscript - and which is the best solution to formalise such information in an expressive and efficient way. It focuses on experts' comments peculiarities and shared patterns -, and the interpretative act underlying each of them, considering their coexistence into a single environment while preserving their contestability (Daquino, Pasqual and Tomasi 2020).

2. STATE OF THE ART
Dealing with cultural heritage and semantic technologies means that knowledge acquisition with domain experts is required (i.e. metadata definition, content interpretation). *What* an expert state about a cultural object has been broadly investigated by CH domain ontologies such as FRBR (IFLA Study Group 2008) – for bibliographic records, along with FRBR-aligned ontologies (i.e. FaBiO[1]); CIDOC CRM – designed as golden standard for representing cultural objects life cycle in museum domain (Doerr 2009) along with its extensions (e.g. CRMtex[2]); EDM – base ontology to represent descriptions provided by Europeana content providers (Charles and Isaac 2015); also handling with manuscript representation (e.g. STITCH[3], DM2E model[4], Sharing Ancient Wisdoms (SAWS[5]), Henry the III Fine Rolls (FRH3[6])).

For what concerns *how* experts state their interpretations, Van den Akker (2011) propose a model to express hermeneutic analysis on artworks about historical content. VIR[7] is a CIDOC extension model to represent visual recognition on iconographical objects. HiCo[8] expresses specific contextual information about an interpretative act on artefacts. *Bufalini*

[1] https://sparontologies.github.io/fabio/current/fabio.html
[2] http://www.cidoc-crm.org/crmtex/
[3] https://www.cs.vu.nl/STITCH/
[4] https://github.com/DM2E/dm2e-ontologies
[5] http://purl.org/saws/ontology
[6] https://data.kdl.kcl.ac.uk/dataset/frh3
[7] https://ncarboni.github.io/vir/
[8] http://hico.sourceforge.net/

Notebook[9] reuses HiCo and nanopublication[10] (Groth et al. 2010) to represent *Quaderno di appunti* dense annotation network together with its contextual information (Daquino, Giovannetti and Tomasi 2019).

3. METHOD AND APPROACH

This research is based on the design-science method proposed by (Hevner et al. 2004, 75–105), which seeks to extend the boundaries of human and organizational capabilities by creating new and innovative artefacts.

Considering the qualitative nature of this research, it has been approached firstly with document analysis (Bowen 2009) on source material, involving skimming (superficial examination to choose excerpts which clearly express disciplines points of view on the same manuscript element), reading (thematic analysis and pattern recognition within the data, seeking for categories for analysis), and interpretation (clarification of incomplete or implicit knowledge in the excerpts). Then, mind maps for each case study have been created to set all the pieces of content in a triple-fashion structure that will be refactored in ontological terms.

Data modelling activity consisted in translating mind maps in competency questions addressing aspects emerged from case studies (i.e. motivating scenarios). Then, a model that does not include any term belonging to any other existing ontology has been sketched, so as to represent pragmatically the scenario originally expressed in natural language. Finally existing ontologies literature has been reviewed and terms have been selected from existing ontologies so as to refactor the model.

MIMA[11] has been reapplied to case studies, first to test each discipline representativeness and then to test points of interaction between information. The model consistency has been tested through the use of a reasoner. Then a toy RDF dataset has been produced and queried (translating competency questions in SPARQL[12]) to test its expressivity.

4. RESULTS: MODEL AND CASE STUDY

The result is a four layered model organised in nanopublication structure, as shown in figure 1.

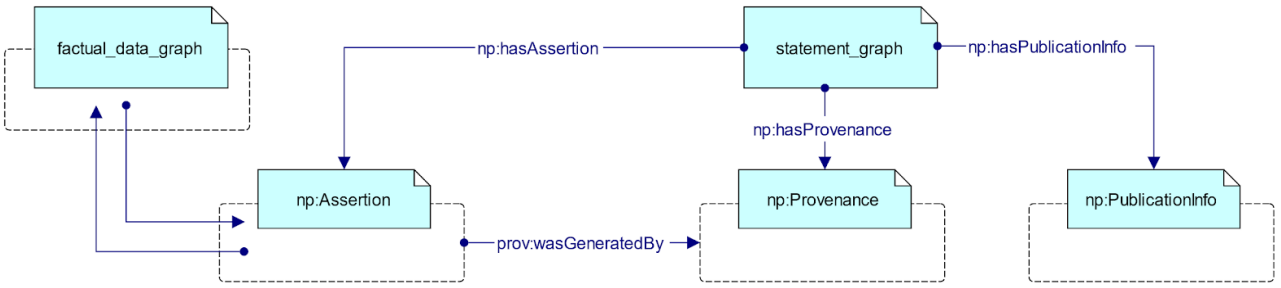

figure 1: MIMA macrostructure showing how RDF triples are stored and connected using named graphs and nanopublications

Figure 1 shows how the statement graph is connected with the four layers, expressed in MIMA as named graphs[13]. For each layer several ontologies have been reused, integrated, and extended if needed.

Layer 0: Factual data graph includes the information deemed not to be questionable. It includes bibliographic metadata (e.g. the edition of a work), the physical and logical description of artefacts (e.g. the folios). Artefacts are considered as an *unicum* (i.e. the manuscript physical structure, cited artworks) and have been modelled using CIDOC, VIR and CRMtex. Serial artefacts (i.e. books editions) have been modelled using FaBiO;

Layer 1: Assertion graph includes scholars' questionable and/or competing statements formulated over the artefact. Assertions on the iconographical apparatus have been modelled using VIR ontology; events, places, and periods involved in the manuscript history or described in the manuscript have been modelled with CIDOC; relations between artefacts

[9] http://projects.dharc.unibo.it/bufalini-notebook/introduction
[10] http://www.nanopub.org/nschema#
[11] https://mima-data-model.github.io/mima-documentation/
[12] https://www.w3.org/TR/rdf-sparql-query/
[13] https://www.w3.org/2009/07/NamedGraph.html

(e.g. manuscript sources and influences), have been described with CIDOC and OA Entry ontology[14]; handwriting analysis has been refactored with CRMtex and CIDOC.

Layer 2: Provenance graph includes contextual information on a statement (`np:Assertion prov:wasGeneratedBy np:Provenance`). Sources, motivations, types and performers of interpretative acts are represented by using terms of HiCo and PROV; the interpretation certainty degree is expressed reusing CWRC ontology[15] terms.

Layer 3: Publication information graph represents the meta-context of a statement that has been automatically or semi-automatically generated. Publication information has been addressed with terms from PROV ontology.

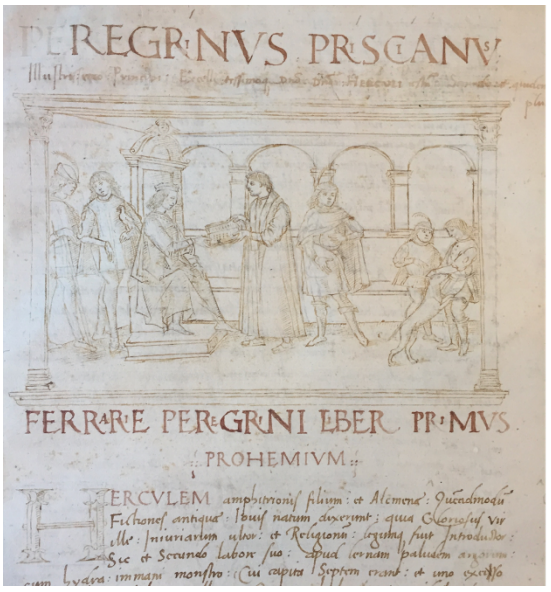

figure 2: Historiae Ferrariae, Vol. I, ASMo, ms. 129, proem

Consider the following scenarios in natural language (MS) referred to Historiae, Vol. I, proem - shown in figure 2.

- **Philologic statement** (MS1): The illustration represents Prisciani giving *Historiae* manuscript to Ercole I d'Este;
- **Palaeographic statement** (MS2): The monumentality of *Historiae* is conveyed by both 'capitale epigrafica' in the title and the illustration;
- **Iconographic statement** (MS3): The proem illustration depicts the consignment of *Historiae* to the commissioner by the author. The motif can be found also in *Orthopasca*.

Figure 3 summarises how the MS can be represented with MIMA and how MS1, MS2 and MS3 entities are shared among the graphs. Each MS (e.g. `philologic_statement` individual) is constituted by the respective assertion graph linked to its provenance information (respective provenance graph) and the manuscripts description (included in factual data graph). The pivotal element of the interaction between the assertions is the individual `hf_I_p1r_ill1` (instance of `vir:IC1_Iconographic_Atom`, representing the illustration in *Historiae* I, proem). Considering the disciplines different interests, each assertion is then specialised in its specific domain. For instance, MS3 states the illustration symbolic meaning (i.e. monumentality) to create a bridge between *Historiae* and *Ortophasca*; MS2 defines the illustration conceptual meaning to highlight the monumentality of a textual feature. The interpretation criterion (respectively `iconographicApproach` and `palaeographicApproach` individuals) and interpretation type (respectively `similatityDetectionBetweenArtworks` and `handwritingAnalysis` individuals) are stored in the respective provenance graphs to preserve this kind of peculiarities.

[14] http://oaentry-ontology.sourceforge.net/
[15] http://sparql.cwrc.ca/ontologies/cwrc#

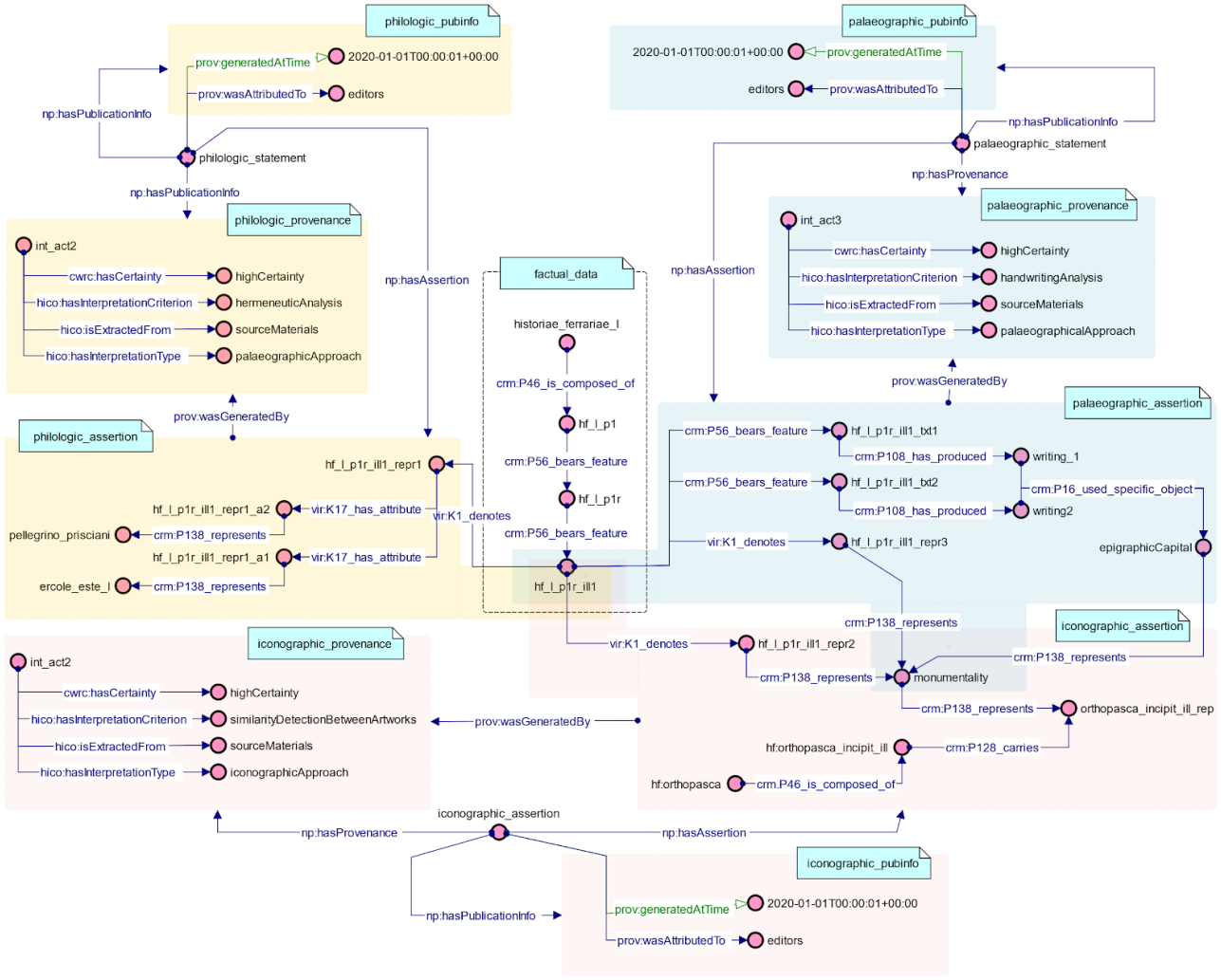

figure 3: illustrates MIMA applied on MS1, MS2 and MS3.

5. CONCLUSIONS

The case study wants to show how MIMA leverages potentialities of the Semantic Web technologies when re-organizing, disassembling and reassembling sources of information and connecting bits into a big network of entities with external and internal linking. MIMA allows different disciplines analysis to coexist in a single environment and to easily compare them over their four layers (e.g. *what* and *how* an expert states an interpretation). MIMA opens to layered data analysis (e.g. how disciplines interact between them? Are there shared patterns among them?) and visualisations (e.g. filtering cultural objects over layers). Even considering its verbosity (inherited both from CIDOC and nanopublications), MIMA offers an effective and reusable way to represent such a complex and multifaceted domain.

REFERENCES

[1] Bowen, Glenn A. 2009. 'Document Analysis as a Qualitative Research Method'. Qualitative Research Journal 9 (2): 27–40. https://doi.org/10.3316/QRJ0902027.

[2] Charles, Valentine, and Antoine Isaac. 2015. 'Enhancing the Europeana Data Model (EDM)'. White Paper. Europeana Foundation. https://pro.europeana.eu/post/enhancing-the-europeana-data-model-edm.

[3] Daquino, Marilena, Valentina Pasqual, e Francesca Tomasi. 2020. «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, n. 3. https://doi.org/10.4403/jlis.it-12642.

[4] Daquino, Marilena, Francesca Giovannetti, e Francesca Tomasi. 2019. «Linked Data per le edizioni scientifiche digitali. Il workflow di pubblicazione dell'edizione semantica del quaderno di appunti di Paolo Bufalini». Umanistica Digitale, n. 7 (dicembre). https://doi.org/10.6092/issn.2532-8816/9091.

[5] Doerr, Martin. 2009. «Ontologies for Cultural Heritage». In Handbook on Ontologies, a cura di Steffen Staab e Rudi Studer, 463–86. International Handbooks on Information Systems. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-92673-3_21.

[6] Groth, Paul, Andrew Gibson, and Jan Velterop. 2010. 'The Anatomy of a Nanopublication'. Information Services & Use 30 (1–2): 51–56. https://doi.org/10.3233/ISU-2010-0613.

[7] Hevner, Alan R., Salvatore T. March, Jinsoo Park, and Sudha Ram. 2004. 'Design Science in Information Systems Research'. MIS Quarterly 28 (1): 75–105. https://doi.org/10.2307/25148625.

[8] IFLA Study Group. 2008. 'Functional Requirements for Bibliographic Records'. Final Report 19. IFLA Series on Bibliographic Control. Munich: International Federation of Library Associations and Institutions. http://www.ifla.org/VII/s13/frbr.

[9] Van Der Akker, Chiel, Susan Legêne, Marieke van Erp, Lora Aroyo, Roxane Segers, Lourens van der Meij, Jacco van Ossenbruggen, et al. 2011. 'Digital Hermeneutics: Agora and the Online Understanding of Cultural Heritage'. In Proceedings of the 3rd International Web Science Conference, 1–7. WebSci '11. Koblenz, Germany: Association for Computing Machinery.