EGITH - Engendering Data. Steps Towards an Ontology for the Representation of Gender in Cultural Heritage

Selenia Anastasi[1,*,†], Andrea De Domenico[2,†] and Marianna Nicolosi-Asmundo[3]

[1]*Università degli Studi di Genova, Scuola di dottorato in Digital Humanities, Dipartimento di Lingue e Culture Moderne*
[2]*Vrije Universiteit Amsterdam, School of Business and Economics, Ethics, Governance and Society*
[3]*Università degli Studi di Catania, Dipartimento di Matematica e Informatica*

Abstract

The neologism *en-gendering* indicates the process of generation, production, and constitution of gender [1]. From this point of view, gender is a semiotic device assumed by subjects in the formation of their own identity, adhering to the effects of meaning produced by the models and *habitus* of a given culture. Feminist scholars have long analysed the process in Art and Literature, but this knowledge is fragmented and not yet adequately supported by digital tools. In this paper we describe EGITH (En-gendering Identities Through History) an extension of GenderedCHContents, an ontology aligned with the Europeana Data Model (EDM), designed to capture concepts related to the construction of gender in the CH domain [2]. We further extend the ontology GenderedCHContents with fine-grained information, relative to the narrative and semiotic dimension of gender in CH. For the modelling of fictional characters we rely, on the one hand, on the semiotic of Charles Sanders Peirce and the Narrative Theory of Umberto Eco [3], while on the other, we rely on the tradition of feminist literary criticism. EGITH is populated with data concerning the main female characters in two well-known literary works of the Italian canon, Dante's *Divina Commedia* and Boccaccio's *Decameron*, and their literary critique.

Keywords
Ontologies and Vocabularies, Accessibility and inclusion in CH, Semiotics of Culture, Feminist Digital Humanities

1. Introduction and state of the art

Gender is (a) representation - which is not to say that it does not have concrete or real implications, both social and subjective, for the material life of individuals. On the contrary. The representation of gender is its construction-and in the simplest sense it can be said that all of Western Art and high culture is the engraving of the history of that construction [4].

In a narrow sense, Cultural Heritage (CH) is defined as everything that is preserved by memory institutions. They include, but are not limited to, History of Arts, manuscripts, historical relics, Archaeology, Design, Natural History collections, and Architecture. Most CH objects are rather mute evidence of past events that gain relevance from understanding the context of their origin and history [5], including the history of how gender has been conceived over time. Thus, CH is a system, in which the artefacts are captured in relationship with other objects and subjects, places, processes, fruition and preservation. In this perspective, issues relating to the representation of gender, (how gender is narrated in a museum, archive, or cultural site) have grown in importance during the past few years. While the earliest works focus on the development of special events and exhibitions featuring women, more recent works address the canonical construction of masculinity, femininity, and sexuality in history [6]. Most of this work has tended to be dedicated to developing guidelines for museology and curatorial activities, and less attention has been paid to the new social and epistemological practices of protecting and narrating memory, such as that offered by online digital repositories.

Archives and libraries are increasingly involved in the design of digital public spaces, which can be freely consulted. Several case studies are an example of this tendency. Since the 1980s, the Orlando Project founders bet on the potential of new electronic media as an antidote to the continuing marginalisation of female writing [7]. Similarly, the electronic archive of the Women Writers Online project [8] overcome the problems of inaccessibility and scarcity that had made women writers invisible. However, to the best of our knowledge, no archival project has thoroughly deepened the question of how to make accessible knowledge related to the formation of gender identities in history, including literary works, tangible and intangible cultural heritage.

In this contribution, we present **EGITH (En-Gendering Identities Through History)**[1] an ameliorated and extended version of the ontology GenderedCHContents, based on the Europeana Data Model (EDM). The latter expands *E89 propositional object*, which describes "immaterial items, including but not limited to stories, plots, procedural prescriptions, algorithms, laws of physics or images that are, or represent in some sense, sets of propositions about real or imaginary things and that are documented as single units or serve as topics of discourse"[2]. To draw attention to the narrative dimension, as a component of the CH domain, we chose to apply the model to describe Dante's and Boccaccio's female characters. The attempt can be regarded as a complex challenge on two main fronts: the first one has to do with the difficulty of adapting feminist epistemology and Gender Theory – mostly hostile to essentializing labels – to the descriptive aims of formal ontologies, the second one with the scant presence of Semantic Web resources for the description of literary characters. The latter issue led us to develop new entities, properly designed on Peirce's Theory of signs. Other semantic models for the representation of the literary and narrative domains are in [9], [10], [11], [12] and [13]. In Hastings and Schulz [9], attention is paid to the development of a model to represent the literary characters, in order to capture not only their ontology as fictional construction but also their relationship with History and writers. There the attributes associated with literary characters are expressed as property intersections thanks to the **as if** relationship. This ontology is based

---

[1]https://github.com/SeleniaAnastasi/EGITH_0.1.git
[2]see also: https://cidoc-crm.org/Entity/e89-propositional-object/version-6.2.1.

on BFO, but no semiotic theory is considered to deal with the relationship between the object (say the character) and its representation. [10] models the classic concepts of *fabula*, *syuzhet*, and narrations, following the Russian formalism, but leaving out fine-grained elements such as places, time, and characters. *et al.* [13] deepen beyond this work by conceptualising narrative structures in Digital Libraries. Damiano and Lieto [11] leverage on the importance of the narrative content description of artworks in Cultural Heritag, by implementing a formal ontology into the Labyrinth system. The ontology has been designed to support reasoning on the relations among characters, actions and stories. The ontology was the original starting point of [2], the ontology we propose to extend in this work, and it is also one of the few works addressing the problem of modeling narrative characters. Following the root traced by Gigliozzi in the 80's, another work stressing the topic of narrative characters in the Literary domain is [12]. Ciotti is also the first author highlighting the potentiality of the semiotic tradition in Applied Ontology, by developing the first draft of a formal ontology of characters. However, despite the fertile ground proposed by Ciotti, the draft ontology has not been further investigated so far. Finally, GenderedCHContents, the ontology that we extend here, is designed to describe figurative elements expressing gender-related attributes and properties in CH artefacts, but without any reference to its potential literary source.

To fill this gap, we rely on some well-known examples of feminist literary criticism and semiotics. Thus, we implemented new classes and properties to describe the typicality of the female repertoire. The conceptual model that we present here is designed to connect the literary characters not only to their narration, their author, and their possible iconographic representations but also to their historical existence. In the case of women, in some historical periods, this recognition is in fact made possible also thanks to literary testimonies, as in the cases proposed here to validate the ontology: Francesca da Rimini and Maria d'Aquino (Fiammetta). In addition, to emphasize the importance attributed to interpretation in the Eco's model, we enrich the ontology to express the relationship between a work and the scholar who provided the gender analysis.

In section 2 we discuss the issues concerning ontologies designed to describe the CH domain, focusing on the feminist perspective on data and metadata management [14].

In section 3 we explore the general theoretical model of literary characters and gender representation, as expressed respectively in Peirce's and de Lauretis's Semiotics. In the same section we discuss a detailed description of classes and properties adopted to model EGITH, related to literary characters, physical and narrative places, and gender-related attributes, updated with respect to [15] [2].

In section 4, finally, we propose an example of validation by populating the ontology with iconic characters from the Divina Commedia and the Decameron, as represented in their literary sources.

2. Ontologies, Gender and Cultural Heritage

New technologies have meant a renewal of the methodologies for the representation of documents and data, where data is framed as any object bearing meaning. Linked Open Data (LOD), ontologies and the Semantic Web move towards new digital systems such as the *knowledge site*,

defined as "a semantic environment that allows access to knowledge through the data contained in a given text, the strings interpreted within it and the relationship of these and other data on the Web" [16]. Many cultural institutions are now making cultural properties accessible online, by releasing their datasets as Linked Open Data and CH portals.

At various levels, several large-scale projects can be signalised in the direction of a semantic web for cultural heritage. As the most notable are the ArCo project [17] (the Italian Cultural Heritage knowledge graph), a network of ontologies that model the CH domain and a Linked Open Data (LOD) dataset based on the official General Catalogue of the Italian Ministry of Cultural Heritage and Activities, the ontologies Europeana Data Model (EDM) [18] and CIDOC - Conceptual Reference Model (CRM) [19]. However, the socio-technical changes lead by new technologies claim not only the development of new devices, but also new reflections on the relationship between data, technologies, and their epistemology.

According to Colella [20], in the field of CH some issues still need to be addressed:

- Gender-blind language in texts, documents and data management, that require an explicit representation of gender in the CH domain;
- Providing a model of gender analysis in data-management systems by developing guidelines and proper taxonomies;
- The involvement of gender epistemology and gender knowledge in digital and archival practices.

In this perspective, the centrality of data is not limited to their correct use and collection, but also to their organisation, accessibility and standardisation.

To meet this needs, for this project we rely on the ontology GenderedCHContents. The ontology we extend here is conceived in the OWL 2 language and was born as an extension of EDM, aligned with CIDOC-CRM, the international standard for the controlled exchange of information on cultural heritage focused on events. In addition to Europeana, GenderedCHContents uses the Open Archives Initiative Object Reuse and Exchange (OAI-ORE), the Simple Knowledge Organization System (SKOS) and Dublin Core vocabularies. GenderedCHContents contains twenty-two classes, defined as subclasses of the CIDOC CRM *E89* class, by modelling classes and properties to perform a gender analysis in iconographic sources. The theoretical background of the authors is Lorber's work on gender deconstruction [21] and O'Brien's Encyclopedia of Gender Terms [22].

3. Theoretical background of EGITH

3.1. Modeling literary characters

In *Lector in fabula* [3], Umberto Eco borrows Peirce's logical-semiotic model and the concept of possible worlds as a starting point for developing his Narratology. In Eco's model, in fact, semiosis is possible only in constant cooperation between an author who produces a text, its context (and encyclopedia), and the reader.

According to Peirce, "a sign, or *representamen*, is something which stands to somebody for something in some respect or capacity". Moreover, "[the sign] stands for that object, not in all

respects, but in reference to a sort of idea, which I have sometimes called the *ground* of the representamen" [23]. The object, to which the sign refers, is therefore not necessarily material, but is rather a correlative of the sign; it is the vehicle of a class of possible experiences.
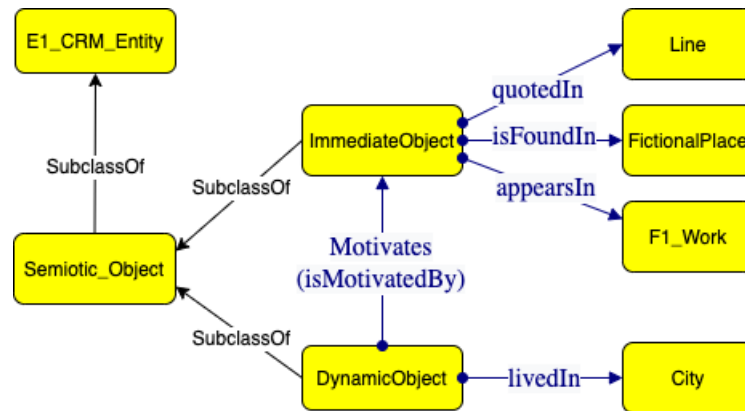

Figure 1: Classes and properties that describe the literary character, in accordance with Peirce's semiotic model and the spatial locations of the literary character: diegetic, extradiegetic and intradiegetic levels.

Thus, whenever we try to capture the process that goes from the sign to its possible interpretation, a triadic relationship is established between:

- A **Dynamic object**, a set of starting conditions, determined by the object itself.
- An **Immediate object**, the way in which the dynamic object is framed in its representation, in some aspect or capacity
- A **Ground**, the filter through which a subset of the qualities and properties are selected, among those possessed by the dynamic object. This category does not corresponds to a class in EGITH ontology, but rather emerges from the set of objects attributed to a subject, given a certain cultural interpretation and to the means of certain properties, in a triple. For example, for what concerns Francesca da Rimini in the Dante's *Commedia*, the ground can be described as follows: "The DynamicObject Francesca_da_Rimini motivates the ImmediateObject Francesca_da_Rimini_IO_1, represented in Dante by the set of objects X, Y, Z".

In this model, two are the main consequences: first, only partial knowledge about a subject is possible, given that it is always framed through some sort of filter (some author, some critic, some Work, etc.); second, although the semiosis is potentially unlimited (which means that we can potentially assign unlimited numbers of immediate objects to just one dynamic object), its limit lies precisely in the process of the new constitution of an interpretative *habitus* (namely, the set of ground established in a given work by means of a certain author in an historical period, and so forth).

3.2. Modeling representation of gender

Gender analysis is defined as "the systematic collection and examination of information on gender differences and social relationships to identify, understand, and correct inequalities based

on gender" [24]. Thus, gender is a descriptive tool applied to all aspects of cultural heritage to strengthen gender equality.

The semiotics of Teresa de Lauretis [4] also bases her hypothesis on the Peircian concept of *habitus*. According to her view, a semantic value can be attributed to gender. Such value is determined by the language and selects certain aspects and traits as relevant for the organisation of gender categories, thus helping also to determine our categorization of reality. This belief system, therefore, is historicized in the passage from generation to generation, building the more or less mobile boundaries of representations. In this model, a fundamental role is played by imagination, also called *scripts* [25]. The categories through which are created can be summarised in 4 conceptual macro-areas, corresponding to the classes we modelled in the EGITH ontology to define gender attributes (see Figure 2):

- **Affective interpretants**, the set of attributes and signs that determines the cultural division between feminine and masculine affectivity. For example, in Western culture, female emotional instability is classically opposed to male rationality.
- **Agency interpretants**, concerns the set of activities and social roles attributed to a certain gender, the space for action in the world, decision-making and self-determination processes. For example, the parental care role assigned to women.
- **Somatic interpretants**, the subject prefigured by contemporary feminist Semiotic is physically and bodily involved in the production of meaning, representation and self-representation [26]. That is, the culturally sexed set of bodily characteristics defining the dualism male-female.
- **Symbolic interpretants**, in Peirce's semiotics, a sign or interpretamen, can in turn refer to another sign. This is the case of characters who, as narrated, have themselves become metaphors, vehicles of meanings. A famous example in Dante is Beatrice, a symbol of the Platonism of courtly love.

3.3. Modeling locations

In terms of feminist epistemology, we can define our approach as "situated". Donna Haraway coined the term "situated knowledges" in a 1988 essay entitled *Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective* [27]. This notion has to do with points of view, the being in the world of the subject, the relationships with others, places, and times. Thus, situated knowledge representation prompts us to take particular account of the fictional places where the characters appear and where the interactions between characters take place. The basic assumption is that, in a storyline, the assignment of a specific place within the geography of the *fabula* is always justified, and often driven by gender prejudices, due to their historical time and their belonging to a certain poetic tradition. For instance, several authors highlight the strong metaphorical and role-dependent connection between female characters and their houses (i.e. Virginia Woolf's *A room of one's own*).

In Dante, by assigning so many powerful women to the circle of the lustful, the author reaffirms the idea that the domination of women over men, even in the sexual sphere, is perverse and wrong. The importance of places for the understanding of the characters narrated
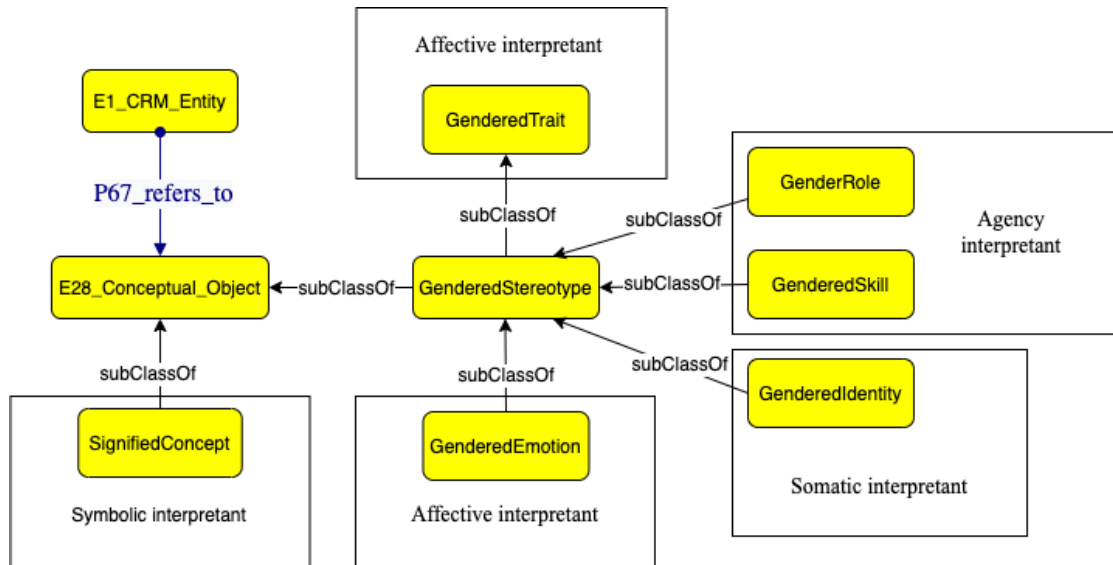
Figure 2: Classes that define gender attributes of the literary character, according to the category of affective, agency, symbolic and somatic interpretants.
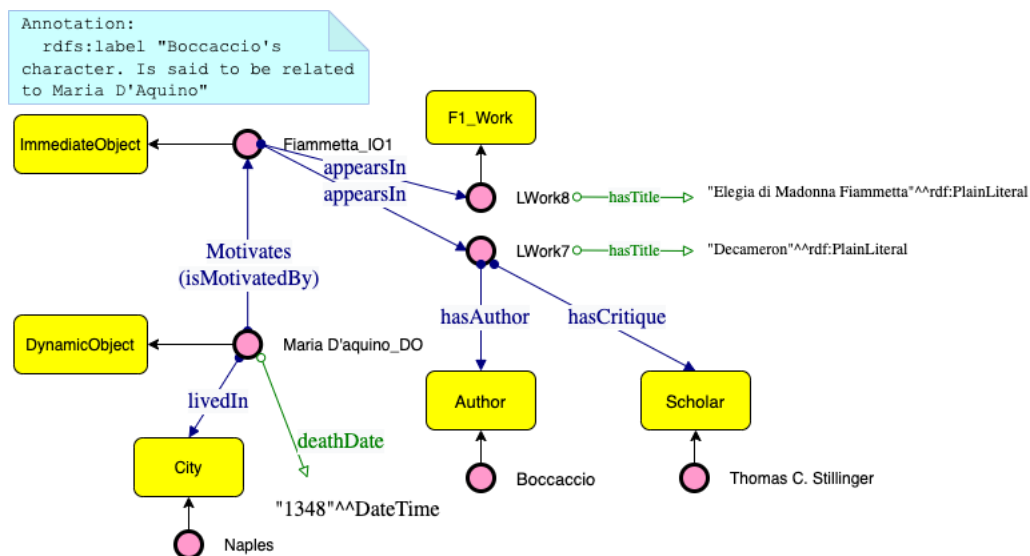

Figure 3: A model of Fiammetta in the *Decameron*, her correlation with the historical figure of Maria D'Aquino, the author and the scholar who made the feminist comment.

in the Commedia led us to consider the need to extend the ontology with the class *FictionalPlace*, dedicated to narrative places, which in turn contains instances such as Inferno, Paradiso and Purgatorio and their respective geographical subdivision into Cerchi, Balze and Cieli, or the Firenze's countryside for what concerns the *Decameron*. Thus, it will be possible to distinguish two cases: the one in which the characters are associated with the place where they physically appear within the narrative (i.e. Francesca da Rimini in the II Girone of the Inferno, see Fig.

4), but also on the level of the narration, that which in the Narrative Theory we can name as the intradiegetic level. To make this distinction clear, we used the *Line* class, to indicate the quotation of a character explicitly in the text.
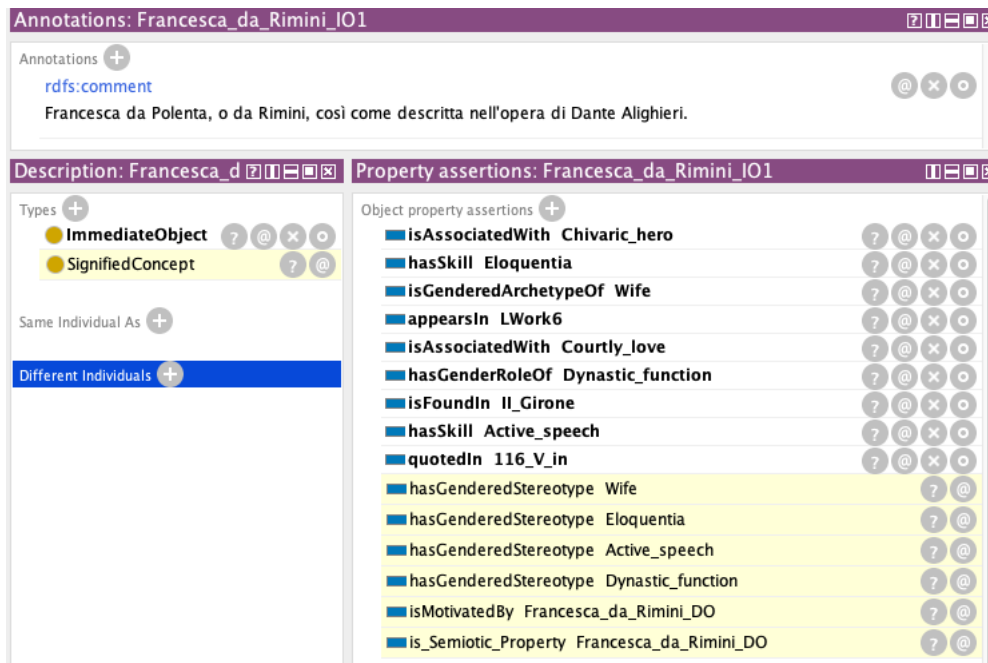

Figure 4: Francesca da Rimini's representation in EGITH ontology.

To model such a distinction, we further introduce two object properties: *quotedIn*, connecting the class *ImmediateObject* as domain and the class *Line* as its range, and the object property *isFoundIn*, used to express the level of the *fabula*, with the class *ImmediateObject* as its domain and the class *FictionalPlace* as its range (see Fig. 3). Moreover, as we can see in Fig. 4, thanks to the relation *isAssociatedWith*, which has the class *SignifiedConcept* as its domain, the reasoner was able to infer that Francesca_da_Rimini is a symbol for other concepts in herself, such as that classical *Dolce Stil Novo* (i.e. courtly love, chivalry hero). Along with standard inferences concerning characteristics of properties such as functionality and inverse properties, domains and ranges of properties, our ontology also includes SWRL rules in order to infer simple assertions concerning entities in relation to the verse, line, and fragments of the text where they are being quoted. The rules are listed in Fig. 5.

4. Case studies: *Francesca* and *Fiammetta*

In this paragraph, we briefly discuss the two case studies on which we have validated the ontology, the case of Francesca da Rimini as represented in Dante's work and of Fiammetta (based upon Maria d'Aquino), as represented in Boccaccio's work. The choice of the two characters is not accidental: in fact, we believe that the two figures possess common traits, which

Figure 5: Some SWRL rules concerning entities and places in the text where they are quoted.

are particularly exemplary for questioning the narrative on gender. In both cases, they are in fact two contemporaries of the authors, to whom we owe the historical testimony of the existence of the two women. The political centrality of the two female figures is mediated by the literary text, which is both narration and device of memory. Both Francesca and Fiammetta are described as two atypical women of their time: they are strong, assertive, charismatic and endowed with an explicit sensuality. Nevertheless, if a strong amorality is associated with Francesca da Rimini, Fiammetta's positive tone connotes the benevolent behavior of the character. For the modeling of gender traits and for further information about characters' actions and characterisation in the narratives, we drew on solid studies. On the one hand, we rely on the tradition of feminist literary criticism around Dante's and Boccaccio's works[28][29][30][31][32], while on the other, for the digital version of the two texts, on to web-based editions and critiques available in open access[3]. Starting with Maria d'Aquino, who lived in Naples and died in 1348[4], we modeled as DynamicObject the reference to her presumed historical existence. Her textual reference in Boccaccio's works is coded as the instance Fiammetta_IO1, inherent to the class *ImmediateObject*, to distinguish the character from a possible expression of Fiammetta in other works and from other authors. This choice was made because in each of Boccaccio's works Fiammetta is generally described in consistent terms, configuring her as a unitary character. In Fig. 3 we also show a few details about Fiammetta in relation to the literary scholars by whom the textual representation is discussed. Instances relating to the author of the original works that gave birth to the character, and the scholar who made the critique are connected to their textual instances, respectively, through the object properties hasAuthor and hasCritique. In the example of Francesca da Rimini, visible in Fig. 4, we describe in detail the modeling of the gender traits that emerge explicitly within Dante's text, and implicitly from the literary criticisms. The several inferences of the reasoner in yellow are due to the instantiation of the subproperties of hasGenderedStereotype, i.e. hasEmotions, hasGenderedIdentity, hasGenderRoleOf, hasSkill, isCharacterisedAs and isGenderedArchetypeOf, each of which expresses the relationship between the ImmediateObject (character) and the interpretants described in Fig. 2. Francesca is therefore characterized in gendered terms through her dynastic function, as she was given in marriage to Gianciotto dei Malatesta, to guarantee a long lasting peace between the two families. She is characterized also for her ability to express herself eloquently and being active in a speech in the V Canto of

---

the Inferno, as was not usual for women of that time. Finally, Francesca is in turn a vehicle of meaning as a SignifiedConcept, an inference which is guaranteed by the association between her and two concepts typical of the poetics of the *Dolce Stil Novo*, namely the *chivaric hero* and the *courtly love*.

5. Conclusions

In this work, we presented EGITH, a conceptualization and an OWL 2 ontology of characters based on fundamental notions in Semiotics and Narratology and their gendered representation according to *Feminist Literary* critique and the *Feminist Science Studies*. For the purpose of this paper, we focused exclusively on female characters, in order to emphasize gender stereotypes and several stylistic features in the Italian literary canon. Our ontology is modeled according to GenderedCHContents, an extension of the Europeana Data Model (EDM), to accurately represent gendered features in the Cultural Heritage domain. In particular, the ontology describes the connections between the character, the fabula, the character's historical correlates, the scholars who have analysed the work across gender categories, and the characteristics attributed to the characters through literary sources. New classes and properties are validated by representing two main female characters in Dante's Commedia and Boccaccio's Decameron. Finally, we plan to associate the ontology with tools for visualization and sharing of information about narratives and their feminist critique.

References

[1] T. de Lauretis, Soggetti eccentrici, Feltrinelli, 1999.
[2] I. Kyvernitou, A. Bikakis, An ontology for gendered content representation of cultural heritage artefacts, Digital Humanities Quarterly 11 (2017).
[3] U. Eco, Lector in fabula. La cooperazione interpretativa nei testi narrativi, Bompiani, Milano, 1979.
[4] T. de Lauretis, Technologies of Gender. Essays on Theory, Film, and Fiction, Indiana University Press Bloomington, 1987.
[5] M. Doerr, Ontologies for cultural heritage, in: Handbook on ontologies, Springer, 2009, pp. 463–486.
[6] A. Reading, Making feminist heritage work: Gender and heritage, in: The Palgrave handbook of contemporary heritage research, Springer, 2015, pp. 397–413.
[7] S. Brown, P. Clements, I. Grundy, S. Balazs, J. Antoniuk, An introduction to the orlando project, Tulsa Studies in Women's Literature 26 (2007) 127–134.
[8] J. Wernimont, J. Flanders, Feminism in the age of digital archives: the women writers project, Tulsa Studies in Women's Literature 29 (2010) 425–435.
[9] J. Hastings, S. Schulz, Representing literary characters and their attributes in an ontology., in: JOWO, 2019.
[10] V. Bartalesi, C. Meghini, D. Metilli, Steps towards a formal ontology of narratives based on narratology, in: 7th Workshop on Computational Models of Narrative (CMN 2016), Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

[11] R. Damiano, A. Lieto, Ontological representations of narratives: a case study on stories and actions, in: 2013 Workshop on Computational Models of Narrative, Schloss Dagstuhl Leibniz-Zentrum für Informatik, 2013, pp. 76–93.

[12] F. Ciotti, Toward a formal ontology for narrative, MATLIT: Materialidades da Literatura 4 (2016) 29–44.

[13] C. Meghini, V. Bartalesi, D. Metilli, Representing narratives in digital libraries: The narrative ontology, Semantic Web 12 (2021) 241–264.

[14] C. D'ignazio, L. F. Klein, Data feminism, MIT press, 2020.

[15] S. Anastasi, A. De Domenico, M. Nicolosi-Asmundo, Genderedontocomedy: Toward a gendered representation of literary characters in the dante's commedia, Copyright© 2022 AIUCD Associazione per l'Informatica Umanistica e la Cultura Digitale (2022) 76.

[16] F. Tomasi, Modelling in the digital humanities: conceptual data models and knowledge organization in the cultural heritage domain, Historical Social Research/Historische Sozialforschung. Supplement (2018) 170–179.

[17] V. A. Carriero, A. Gangemi, M. L. Mancinelli, L. Marinucci, A. G. Nuzzolese, V. Presutti, C. Veninata, Arco: The italian cultural heritage knowledge graph, in: International Semantic Web Conference, Springer, 2019, pp. 36–52.

[18] M. Doerr, S. Gradmann, S. Hennicke, A. Isaac, C. Meghini, H. Van de Sompel, The europeana data model (edm), in: World Library and Information Congress: 76th IFLA general conference and assembly, volume 10, IFLA, 2010, p. 15.

[19] G. Bruseker, N. Carboni, A. Guillem, Cultural heritage data management: the role of formal ontology and cidoc crm, Heritage and Archaeology in the Digital Age (2017) 93–131.

[20] S. Colella, «not a mere tangential outbreak»: gender, feminism and cultural heritage, IL CAPITALE CULTURALE. Studies on the Value of Cultural Heritage (2018) 251–275.

[21] J. E. Lorber, S. A. Farrell, The social construction of gender., Sage Publications, Inc, 1991.

[22] J. O'Brien, Encyclopedia of gender and society, volume 1, Sage, 2009.

[23] C. S. Peirce, Logic as semiotic: The theory of signs, Semiotics: An introductory anthology 344 (1985) 4.

[24] S. De Vido, Mainstreaming gender in the protection of cultural heritage, Cultural Heritage. Scenarios 2015-2017 (2017) 451–467.

[25] C. Demaria, A. Tiralongo, Teorie di genere: femminismi e semiotica, Bompiani, 2019.

[26] M. Leone, M. Pezzini, et al., Semiotica delle soggettività/semiotics of subjectivities, Aracne, 2013.

[27] D. Haraway, Situated knowledges: The science question in feminism and the privilege of partial perspective, in: Feminist theory reader, Routledge, 2020, pp. 303–310.

[28] T. Barolini, Dante and francesca da rimini: Realpolitik, romance, gender, Speculum 75 (2000) 1–28.

[29] B. Mellor, A. Patterson, Reading character: reading gender., English in Australia (1991) 4–23.

[30] T. C. Stillinger, The language of gardens: Boccaccio's 'valle delle donne', Traditio 39 (1983) 301–321. doi:10.1017/S0362152900009636.

[31] P. Levine, Why dante damned Francesca da Rimini, Philosophy and Literature 23 (1999) 334–350.

[32] B. L. Carey, Le donne di Dante: An historical study of female characters in the Divine

Comedy (2007).