

# The BIFLOW-*Toscana* Bilingue Catalogue

A digital representation of the socio-cultural  
history of translation in the Tuscan  
Middle Ages (1260–1430)

*Tiziana Mancinelli and Antonio Montefusco*<sup>1</sup>

## ABSTRACT

This article aims at presenting the BIFLOW-*Toscana* Bilingue catalogue, the final outcome of the BIFLOW (Bilingualism in Florentine and Tuscan Works) ERC project, which investigates the forms and modes of the transmission of texts circulating in several languages in medieval Tuscany between the end of the thirteenth and the middle of the fifteenth century (translations of non-classical texts produced or realized in Tuscany between 1260 and 1430). The construction of a digital catalogue comprehends the theoretical framework and methodological backgrounds of the many fields involved in the project; more crucially, this research focused on the tradition and translations of texts transmitted in manuscripts seeks to reshape humanistic methods and methodologies in order to find the best solutions for representing it with computational models.

---

THE BIFLOW-*TOSCANA* BILINGUE CATALOGUE<sup>2</sup> WAS CONCEIVED with the intention OF digitally representing the social and cultural history of translation in the Tuscan Middle Ages. It is one result of an extensive collective research project focused on this theme.<sup>3</sup> The aim of the BIFLOW-

1. While this article was conceived and discussed by the co-authors, sections 1–3 were written by Antonio Montefusco, and sections 4–7 were written by Tiziana Mancinelli.
2. The Website is temporarily hosted on Tiziana Mancinelli's GitHub account @ [https://tmancinelli.github.io/biflow\\_website/](https://tmancinelli.github.io/biflow_website/). It will soon be transferred to [catalogobiflow.vedph.it](https://catalogobiflow.vedph.it).
3. Biflow-Bilingualism in Florentine and Tuscan Works (ca. 1260–ca. 1416) is a research project funded by ERC (StG 2014 n637533: <https://cordis.europa.eu/project/id/637533>) and hosted by University Ca' Foscari of Venice and EHESS-

*Toscana Bilingue* catalogue is to provide scholars with an advanced research tool, both in the fields of social and cultural history and Digital Humanities, and to offer a model applicable to other contexts and issues (especially in Translation Studies). Before presenting our digital model, we set out the scholarly framework used for our traditional study and describe how the processes of reshaping humanistic methods and methodologies were conducted. First, we will briefly present the theoretical-critical assumptions of our approach to medieval translation; and second, we will identify the *corpus* in its historical-geographical context. On this basis we will then define the catalogue project, which finds digital realization through data modeling based first on CIDOC CRM and eFRBRoo and further inspired by the *Biblissima* ontology (4–7). While the work on the data model of the catalogue unfolded parallel to the hermeneutical model and served to clarify it, here we will explore these models separately so as to give a clearer overview of each.

## 1. Medieval translation as an object of socio-cultural history

“The idea that translation has an history is an old one, but until quite recently this history was an academically marginal activity, pursued on the fringes of literary and religious history” (BURKE and PO-CHIA HSIA 2007, 1). With this provocative statement, Peter Burke finally invited literary and other historians to reflect systematically on translation as an object of study in cultural history. This invitation can be considered a specific application of a program of “social history of language, a social history of speech, [and] a social history of communication” (BURKE and PORTER 1995, 1). A *historical turn* has taken place in Translation Studies at least since the 1970s, with a new focus on regimes of translation (what the translators actually do) and audience. There is still a big gap in historical studies, however, for there is a lack of interest in issues of translation even among cultural historians.

The history of translation in medieval Italy was aptly read in the context of Translation Studies by Gianfranco Folena in his groundbreaking 1973 paper, published in book form two decades later. Folena’s study, however, remained strongly linked to an historical-linguistic approach. His analy-

---

Ecole d’Hautes Etudes en Sciences Sociales of Paris. An update about initiatives and publications may be found here: <https://biflow.hypotheses.org>.

sis was centered around the lexicon of translation (*traducere* and *tradurre* appeared at the beginning of the fifteenth century, at the same time as a series of equivalents in the main European languages). The widening of perspective that came from *Translation Studies* — especially those linked to this theoretical moment, as per Holmes' 1988 reconstruction — allowed Folena to define a precise nomenclature for the various types of translation, classified according to the languages of arrival and of departure. He distinguished between 'vertical' and 'horizontal' translations, the former being from Latin to vernacular languages, and the latter being between vernacular languages. In this innovation, there was an important recovery of translations of medieval texts (non-classical), until then largely neglected by scholars (FOLENA 1993).

Earlier scholarly efforts had tackled the vernacularization movement in Italy mainly for linguistic and stylistic reasons, in association with important philological achievements. The most important scholars in this field were undoubtedly Francesco Maggini (1952) and Cesare Segre (1963). According to both, the study of translation was aimed at measuring the contribution the relationship of Medieval Culture with Classical Latin literature had made to the establishment of Italian prose. The idea is that Italian literary prose had been formed from a transfer of the structures of Latin literature derived from a specific line of scholarly literature, going back to Benvenuto Terracini and to earlier scholarly work from the Accademia della Crusca.

A particularly important consequence, and one of hermeneutical importance, was the absolute primacy given to the *volgarizzamenti* from classical texts. The study of translation therefore also suffered from the attraction, fatal for various reasons, to the orbit of reflection on the origins of Humanism. According to Segre, "la varia fortuna e il vario atteggiarsi dei volgarizzamenti ci portano piano, se non al centro, nella prossima periferia di quello che sarà l'Umanesimo" (1963, 49). Some alternative ideas, however, were also voiced. Aldo Schiaffini proposed searching for the infrastructure of literary prose in the medieval version of classical rhetoric, the *ars dictaminis* (SCHIAFFINI 1969). Don Giuseppe De Luca highlighted the enormous mass (he referred to it as "oceanica") of religious vulgarizations, particularly important in the fourteenth century (1977). A shift of focus from syntax to audience would have perhaps enhanced our understanding of a vast number of translated texts, which makes us immediately understand how the act of translation was much more actively exercised on medieval texts than on classical ones (CERULLO and LEONARDI 2017, xi). These philological-linguistic approaches have remained dominant over the last forty years,

producing a significant array of new editions of texts within the framework of systematic projects of study of the phenomenon, to which we shall return briefly (ARTALE 2003; GUADAGNINI and VACCARO 2004).

It seems evident that a *stricto sensu* historical-cultural treatment of translation activity in the Italian Middle Ages is necessary. The general picture is well known: Jeanette Beer is right to say that “at no time in the history of the West has translation played a more vital role than in the Middle Ages” (1978, 1; see also BEER 2019). The renewal of ideas on translation, which takes place in Humanism at the same time as the affirmation of the new lexicon of the translator (Leonardo Bruni’s *De recta interpretatione* is emblematic: see BRUNI 2004), has often been the foundation for hasty and inaccurate notions of a Middle Ages devoid of a precise vision of the translating activity. Not all approaches to translation in the Middle Ages (at least, in the Italian Middle Ages) can be reduced to Girolamo’s theoretical elaboration, which, among other things, presents a plural and incongruous praxis with this elaboration (CHIESA 1987).

Once we move from a linguistic-philological standpoint to a cultural-historical one, the center of reflection must accordingly shift from translation to the activity of translating. Such an activity contributes to building a complex and multi-layered culture from the point of view of identity (translation involves, according to Steiner, the activity of self-understanding). At the same time, translation must be studied in terms of the relationship with other cultures, implying what Phym defined as ‘negotiation’, that is, the exchange of ideas and the consequent change in their meaning within a different cultural framework (STEINER 1975; PHYM 1993). To identify more precisely what we mean by ‘translation activities’, a review of Peter Burke’s concept of ‘cultures of translation’ may prove useful. Derived from anthropology, it refers to the set of processes that accompany cultural exchanges, including developing shared conventions that govern translation practices. In this sense, developing the earlier suggestion of Cesare Segre, at an early stage translation is both a ‘situazione mentale’ (mental situation) and an ‘attività specifica’ (specific activity) (SEGRE 1963, 49). In other words, within these exchanges (*mental situation*), translation between languages (*specific activity*) is of paramount importance, because it is the prerequisite for widening the readership. This happens both in vertical translations (from a scholarly language to a popular language), which allow the illiterate to access textual content, and in horizontal translations (between two languages of the same level), which facilitate cultural exchange. A similar process occurs in reverse (from a popular language to an international language, for example, from the vernacular to Latin), which enables the

largest dissemination of texts. Further, an activity — here the activity of translation — needs shared *rules*, tools (*means and poetics*), purposes (*ends and strategies*), as well as specifically-dedicated cultural operators (*translators and scribes*). On this level, the Middle Ages and early modern culture share a certain *mouvance* of the relationship between source-text and translated text: despite the refinement of translation theories between Humanism and the Renaissance, in order to attain an ideal, univocal relationship between the texts included in the translation dossier and the object of this activity, an automated, large-scale book market would be needed (BENJAMIN 2000). In the Middle Ages, as in the Early Modern period, the negotiation between the two texts involved in the process is quite unstable. Philological studies on medieval *volgarizzamenti* have revealed the great diffusion of anonymity in manuscript transmission. More generally, when taking place between texts very close in time (i.e., from a medieval text to a medieval translation), translation activity remarkably tends to weaken our notions of author, translator, and copyist. For this reason, it becomes very important to identify all the people involved in the activity (the *cultural operators*) and to understand their contribution (i.e., a copyist who intervenes by changing the text of a vulgarization through a new check of the source-text; see CORNISH 2011).

## 2. Which *corpus* for a social history of medieval translation?

Identification of the regional context was natural: Tuscany is the place where vernacular writing spreads and opens up great poetic and philosophical debates, albeit in a context — or perhaps precisely because of it — in which there is no university (COCCIA and PIRON 2008). The most urgent need, however, was to carve out a representative *corpus* according to our primary historical-social interest, with the intention of escaping the hegemony of studies on translations of the classics and the birth of Humanism. Reference has already been made to the voluminous nature of the translations from medieval texts; yet these translations also have their own individuality and specificity. An illustrious critical tradition has suggested this for a long time: suffice it to mention Folena, for whom “*si possono dunque considerare i volgarizzamenti dai classici come legati fra loro da un rapporto significativo e non puramente estrinseco, e come un filone di particolare importanza, che si distingue dalla massa dei volgarizzamenti, talora*

pregevolissimi e superiori a questi, dal latino della tradizione cristiana e medievale e da quello dei moderni” (1993, 42; see also DIONISOTTI 1967).

In Tuscan cultural history, two dates are particularly significant in the history of the translation of medieval texts. In the 1260s, several translations of texts written by leading Latin author Albertano da Brescia and focused on the use of words in public contexts and addressed to a lay and municipal audience, were produced within a very short timeframe. The middle of the thirteenth century coincides with a series of significant changes in the institutions of Tuscan cities (and Italian cities in general), which concern the enlargement of public life and the consequent need to consolidate the culture of a wider social group. This also led to an intensification of access to education and writing, with regional peculiarities. The interest in Albertano’s works is particularly significant, because his translated works continued to be read and copied in widely circulated miscellanies during the fourteenth century (TANZINI 2012). The success of these works and of their vernacular versions stems from the mix between the rhetorical education and moral principles they entail. This approach, according to Paul Gehl, has been characteristic of the teaching of grammar in Florence since the 1240s. Florentine teachers, in fact, refined the teaching of Latin by recovering the monastic approach and trying to combine grammar and morals. In this practice they remained far from humanistic innovations, and, at the same time, made the Florentine cultural context unique (GEHL 1993). The case of Albertano and his success is particularly due to this peculiar scholastic framework; the translation *vagues*, in a certain way, followed this line.

Although Florentine and Tuscan Humanism continued to have their own characteristics within the larger framework of European Humanism (for the continuity of Latin / vernacular bilingualism and the importance of political issues, see TANTURLI 1998), a strong socio-cultural fracture in the history of translation was created between the years of the Council of Constance (when Dante’s *Commedia*, a pivotal work of vernacular literature, was reworked in Latin) and the 1430s, when, as mentioned before, profound innovations intervened in translation theory, especially in relation to classical languages (FERRANTE 2010).

Not only did we need to respond to a *desideratum* of criticism but also to identify a *corpus* that allows us to more directly reflect on translation as an object of history *iuxta propria principia*. We therefore began by considering the concept of ‘translation’ in a broad sense, thus we include both *volgarizzamenti* and translations that involve passages between different languages

(Latin and the vernacular), as well as *compendia* or re-elaborations that involve inter-linguistic passages and allow us to clearly identify a basic text. Secondly, the chronological cutout also enabled us to define the type of source text, so we have privileged texts written and translated within a short time span. The source texts are medieval texts produced since 1200, and to be included in the *corpus*, the dossier must also include a text (source or translation) produced in Tuscany. By imposing this limitation, we have had the opportunity to more accurately investigate the manuscript traditions of the various texts in the different languages and, where possible, assess their interactions. In this way, we were able to deepen our understanding of the phenomenon of translation in a multilingual context. Accordingly, the different disciplinary stances of our work could find a unified lens of analysis on the phenomenon of translation through the catalogue approach, conducted in parallel with team-work, which identified a series of case studies that we considered particularly significant.

### 3. From existing catalogues to the *BIFLOW-Toscana Bilingue* Catalogue

Digital tools for the investigation and deepening of our understanding of translation in the Middle Ages are few and consist essentially of informative yet static databases, since they only transfer existing repertoires into a digital environment. Particularly representative of this phenomenon is the digital catalogue *TradLat: traductions latines d'œuvres vernaculaires*.<sup>4</sup> The repertory contains translations of vernacular Latin texts from the Middle Ages. The information is organized in a double list, i.e., by authors and, where the author is anonymous, by texts; the list gives access to information sheets with essential bibliographical information. Let us take the example of the entry dedicated to Boethius:

Boèce (ca. 480–524), *Consolatio Philosophiae* : \*Traduite en catalan, puis retraduite en latin. Voir Francesca Ziino, *Una traduzione latina del "Boezio" catalano*, dans *Romania*, t. 119, 2001, p. 465–482 [signalé par Fabio Zinelli].\* Pierre de Paris, *commentaire français à la Consolatio Philosophiae de Boèce*, composé vers 1309 en complément de sa traduction

4. See <http://tradlat.irht.cnrs.fr//Repertoire,13>; see also BRUN, DUVAL, FERY-HUE, and GADRAT 2005.

française de Boèce (ms. : Vatican, Vat. lat. 4788 qui conserve trad. fr. et commentaire fr.), traduit en latin au 15<sup>e</sup> siècle (ms. : Nice, Bibl. mun. 42).

(A. VERNET, in <http://tradlat.irht.cnrs.fr/Corpus-des-auteurs-connus-A-et-B>).

No search tools have yet been developed.

A very large *corpus* is covered by *Transmédie*, directed by Claudio Galderisi (2011). This repertoire gathers all of the translations made in French and in Occitan from the eleventh to the fifteenth century; so far, the repertoire is included in a voluminous book and in three volumes (GALDERISI 2011a). However, the *Transmédie* project is due to be put online on the CESCO website of the University of Poitiers. The entries included in the repertory deal with the texts involved in the process of translation; the information is discursive and concerns the history of a text and its translation, along with details of the manuscript's circulation and the bibliography (see GALDERISI 2011). Each entry is organized starting from a summary of the content of the source-text. From this point of departure, in cascade, the philological and critical information on translations into Gallo-Roman languages follow. Here is an example:

RICHARD DE SAINT-VICTOR, *Beniamin minor ou De duodecim patriarchis ou De preparatione animi ad contemplationem*, XII<sup>e</sup> s.

Richard de Saint-Victor est un moine d'origine anglaise : il entre vraisemblablement à l'école de Saint-Victor vers l'année 1155. Sous-prieur en 1159 au plus tard, prieur en 1162, il y a sans doute enseigné. Cette œuvre propose une lecture allégorique de l'histoire de Jacob, de ses deux épouses (Léa et Rachel) et du plus jeune de ses douze fils, Benjamin : il s'agit de décrire l'élévation de l'âme vers Dieu.

ÉDITION: J. Châtillon, M. Duchet-Suchaux et J. Longère, *Richard de Saint-Victor, les douze patriarches ou Benjamin minor*, Paris, Editions du Cerf, 1997, 374.

TRAVAUX: J. Châtillon, 'Richard de Saint-Victor, mort en 1173', in *Dictionnaire de spiritualité*, XIII, 1988, 594–654 ; Dominique Poirel, 'L'école de Saint-Victor au Moyen Âge : bilan d'un demi-siècle historiographique', *Bibliothèque de l'école des Chartes*, 156, 1998, 187–207.

Traduction lorraine, ANONYME (MOINE D'UNE ABBAYE DU NORD-EST?)



Le manuscrit fr. 24864, copié à la fin du XIV<sup>e</sup> siècle, reproduirait, selon G. Hasenohr, le texte d'une traduction lorraine datant de la fin du XIII<sup>e</sup> siècle (manuscrit non retrouvé).

MS.: Paris, BnF, fr. 24864, XIV<sup>e</sup> s.

TRAVAUX: G. Hasenohr, "Sur une ancienne traduction lorraine (XIII<sup>e</sup> s. ?) du Beniamin minor", *Revue d'histoire des textes*, 21, 1991, 237–242".  
(GAUTHIER 2011)

The impressive mass of indexes prepared is very promising with regard to search tools, but it is very difficult at present to predict the digital tools that will be perfected for this very rich repertoire. All in all, we can say that these two catalogues are characterized by completeness and an organization of the *corpus* by the hermeneutical category of texts. The interpretative element is difficult to deduce from the single entry. The database, moreover, is not able to explain *per se* the process of the translation.

The database that is closest to ours is *DiVo-Dizionario dei Volgarizzamenti*, a project that largely inspired our catalogue. Indeed, *BIFLOW-Toscana Bilingue* is to some extent a continuation and development of a part of this complex Dictionary-Catalogue. The *DiVo* combines two datasets: an exclusively lexical one, and a philological-textual one. The *corpus* considered includes all medieval Italian translations of Latin texts from Antiquity to the Early Middle Ages (and specifically up to Gregorio Magno) that are published in a critical or acceptable modern edition. The project was born in the context of editing the lexical entries of the *Tesoro della Lingua Italiana delle Origini*, the historical dictionary of the Italian language before the fourteenth century, realized at the *Accademia della Crusca* (<http://tlio.ovi.cnr.it/TLIO/>). The first aim is therefore the collection and lemmatization of the *volgarizzamenti*; the vernacular text, marked *ad hoc*, is also associated with the Latin original (in the context of the *corpus Clavo-Classici Latini Volgarizzati*).<sup>5</sup> The texts inserted and lemmatized in the *corpus* are associated with bibliographic and philological files and organized in a repertory with the name *DiVo DB*. Its entries give information about the author of the translation, the dating, the language of the text, the identification of the literary genre, and finally the history of the transmission and the bibliography. Another important qualifying point concerns the cataloguing of

5. See the website of *Corpus DiVo*: [http://divoweb.ovi.cnr.it/\(S\(gz3ivtjurhbb2hqiyyvrgk45\)\)/CatForm01.aspx](http://divoweb.ovi.cnr.it/(S(gz3ivtjurhbb2hqiyyvrgk45))/CatForm01.aspx); see also DOTTO 2012.

the manuscript witnesses. The codices are listed and briefly described: they are searchable both through a free field filter and through an advanced search using several filters simultaneously. In this solid digital representation of the historical-linguistic approach to translation, *DiVo* also represents a significant *overture* to the history of the manuscript tradition of the medieval translations.

As we have seen, the available tools either did not meet our requirements from a hermeneutical and digital point of view or are not focused on the social history of translation. For this reason, while developing and integrating some elements of *DiVo* we drew on a new digital resource according to current technologies. Our aim was to study the phenomenon of translation in a given context, i.e. Medieval Tuscany, as well as within a defined historical-cultural framework, that is, the affirmation of translation practices before Humanism, and to represent this in a digital environment. Digital representation must be innovative (to become a model for other research), comprehensive (covering the whole *corpus*), and exhaustive (providing all the data and analysis tools). Therefore, it must provide all the elements of the phenomenon in its complexity, allowing the user to have all paramount aspects easily available, starting from the composition of the texts, the crafting of the translation(s), and finally their parallel circulation in the various language versions (original, translation, rewriting). In this way, the phenomenon is no longer merely a linguistic history study, but an object of social and cultural history in a given context. In the construction of a hermeneutical scheme, we therefore made use of the questions elaborated by Burke to investigate the translation systems prevailing in a given period: Who translates? With what intentions? What? For Whom? In what manner? With what consequences?

From these questions, some classes of information can be extrapolated and organized according to a philological-historical hierarchy that we considered appropriate to our *corpus*. The heart of the information that is systematically and completely recorded is made up of the following three classes:

1. What? The texts (the source and the translated versions);
2. Who? The authors; the translators; the copyists;
3. In what manner? The manuscripts; the relationships between the texts (the graph of versions).

The decision to study translation as a form of dissemination of works in a bilingual context has led us to focus on the relationship between the

source-text and the various derived versions. The fact that the latter are often subject to reworking, sometimes based on a return to the source text, has also led to careful study of the circulation of these texts (through manuscripts). For this reason, the basic element of the catalogue entry is what we call the ‘textual dossier’, which brings all these elements together. They must be represented in two parallel ways, analytically and synthetically; this synthesis can be visually captured with a graph that reorganizes the relationship between the texts through memorable formulas (acronyms).<sup>6</sup>

Burke’s other questions are important, but at times they are ambiguous:

4. For Whom (in the double sense of the recipients and patrons)?
5. With what intentions and with what consequences?

Nevertheless, these questions constitute a guide for the analysis of the material reorganized in the entry, which will then be recorded in special sub-entries, of a discursive type (focused on content and textual history). The user we have in mind is the student/scholar interested in the cultural history of pre-humanist Tuscany. The user will wish to find a series of precise and controlled details, mostly research-based and not derived only from secondary literature, for points 1–3, and possible insights for points 4–5. The catalogue must consequently be constructed so as to represent the variety of aspects of this realm because every time the user finds a single piece of information, this information must be inserted into a complex and stratified network of elements linked together in the textual dossier. We are convinced that such a model can serve to retrieve and contextualize every piece of information inserted — which in itself can be of the inert type: philological, paleographic, codicological, prosopographic — within the complex phenomenon that is the object of our study, namely translation.

6. The Acronym is composed by six letters identifying the author and the text; e.g., Aldobrandino da Siena’s *Régime du corps* is indicated with AldSieRC; the different versions of the source-text, French in this case, are AldSieRC-A, AldSieRC-B, etc.; the various versions of the different acts of translations are set out with minuscules: AldSieRC-a, AldSieRC-b, etc. The codices, which are not yet linked to the different versions, are placed under the categories of TradLat and TradVolg.

## 4. Modeling Medieval translation

Modeling has been considered one of the main activities of Digital Humanities.<sup>7</sup> It has also been a standard method within the fields of both Humanities and Humanities Computing. Nevertheless, modeling is not an easy task. The digital shift for scholarly research in the humanities requires a change in both its methods and methodological approaches and, consequently, in our understanding of our objects of analysis. The activity of modeling can be explained by the heuristic process of transforming concepts of a special type of digital data. Manipulating and processing humanities data can be undertaken not only with the purpose of reusing and curating digital research in the long term, but also as a way of extracting new knowledge through the possibility of modeling and describing the data. The centrality of data modeling to all forms of digital scholarship is at the core of the approaches used to construct the *BIFLOW-Toscana Bilingue* catalogue. The process of constructing this model and shaping our theoretical assumptions comes from a deep examination of existing studies and the research questions largely explained above. This digital project is also designed to create new opportunities for interpretation and analytical perspectives. A digital project needs to be based on a clear model, which means a detailed interpretation of the object of study. According to Eide and Ore, “To better understand an object of study, it can be useful to construct a model of it. The model can, for instance, be on paper, in a computer, or realized in some physical material” (2019, 178). Hence it is not the cutting-edge technology that makes the project innovative, but, first and foremost, how the project has been modeled. Still, technologies are crucial. How we say something and how the medium embeds the message is fundamental to a scholarly project; indeed, as McLuhan famously stated as long ago as 1964, “the medium is the message” (7). Through a transdisciplinary collaboration, scholars can build high quality digital resources and tackle essential scholarly problems and/or issues about their objects of analysis.

7. Since the publication of *Humanities Computing* (McCARTY 2005), modelling has become a field of study in Digital Humanities. This topic was tackled in a recent volume (FLANDERS and JANNIS 2019), which conveys different disciplinary perspectives as well as a discussion on several methods adopted within the Digital Humanities. For more recent work on this subject, see CIULA, EIDE, MARRAS, and SAHLE 2018.

In this particular case, the work with the *BIFLOW-Toscana Bilingue* catalogue adopted *Semantic Web*<sup>8</sup> technologies and *Linked Open Data*, which has allowed us to fully map out entities and their relationships using principles of abstraction in order to represent multilingualism and translation in the Middle Ages. The description of relationships gives consistency to the meaning of the data with respect to the context. The network of relations can be formalized through statements in the Resource Description Framework (RDF)<sup>9</sup> that allow the data to be transformed into computable content. This raises an important issue for humanists, namely that they must take great care when modeling data and delivering it on the Web.

In order to achieve high quality digital resources, modeling must always be undertaken with publishing in mind, the latter being our final outcome and of utmost importance. Digital projects are generally delivered on the Web and for a user (our ideal reader) with a web interface that allows users to access content along with collected and modeled data. The problem of the current Web is precisely that it is a network of documents written for human users, and as far as we can refine linguistic analysis techniques, we are still far from the possibility that a machine can understand the meaning of such a text. The purpose of the Semantic Web is to expose our contents directly in structured data, readable by a machine. Hence, the Web is no longer a document-centered database, rather it is becoming a data-centered architecture of knowledge with highly interactive applications that connect the user to one or more data sources. This is the environment in which humanists publish their data and share knowledge. It is therefore imperative that we, as humanists, care about the Web and its standards, for our cultural heritage will be shaped by it. ‘Semantic Web’ is a formal description of knowledge as a set of concepts within a domain in which their relationships are explicitly expressed and structured. XML metadata could be insufficient for representation of any kind of metadata which focuses on the description of documents. Additionally, Linked Open Data “is about using the Web to connect related data that wasn’t previously

8. “Semantic Web” is connected to Web 3.0. As the consortium W3C notes, “the Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries” (November 7, 2011). The term ‘Semantic Web’ was originally coined by Tim Berners-Lee in 2001 to describe “a web of data that can be processed by machines”; see also BERNERS-LEE, HANDLER, and LASSILA 2001.

9. See <https://www.w3.org/RDF/>.

linked, or using the Web to lower the barriers to linking data currently linked using other methods”.<sup>10</sup>

## 5. Methodology

Our aims are twofold: firstly, to create models with explicit descriptions and definitions to be shared with other scholars; secondly, to formalize structured data in ways that enable its exchange and connection to other projects. Therefore, the methodologies adopted follow both the need to model the object of study and to apply Semantic Web and Linked data technologies that are currently at the center of the challenges of scholarly research in Digital Humanities in the field of metadata. Ontologies are the means by which the Semantic Web shares models. In computer science, ontologies are conceptualization models formally structured depending on “the relevant entities and relations that emerge from its observation, and which are useful to our purposes” (GUARINO ET AL. 2009, 2–3). An ontology is therefore a computational artefact realized by a process of abstraction of a certain area of interest and can become quite large and complex. Eide and Ore, quoting Gruber, the first to introduce this concept to computer science, point out: “The term is borrowed from philosophy, where an ontology is a systematic account of Existence. For AI systems, what ‘exists’ is that which can be represented. [ . . . ] In computer science and also in digital humanities, the term is often used as a near synonym to ‘data model’, ‘thesaurus’ and even to ‘closed vocabulary’” (FLANDERS and JANNIS 2011).

An ontology is composed by a set of classes and properties that represents concepts and their relationships. Those classes and properties are built through the basic framework of the Semantic Web, RDF, and OWL (*Web Ontology Language*),<sup>11</sup> an additional semantic layer that allows us to set relationships among classes. OWL also allows a reasoner or any semantic application to browse the graph of resources. Our model was based on pre-existing models: reusing the variety of ontologies already in existence can be a useful way to build others and to describe a specific domain, hence our decision to draw on models published and shared for a given domain by similar projects in order to construct our own.

This rather useful and flexible method of building ontologies can be seen as an explicit formalization of a specific domain. We would suggest that development through modularity (CUENCA GRAUA, HORROCKS,

10. See <http://linkeddata.org>; see also NEEDLEMAN 2011.

11. See <https://www.w3.org/OWL/>.

KAZAKOV, and SATTLER 2008) as we have pursued it with the *BIFLOW-Toscana Bilingue*, could be beneficial to many projects that use ontologies and Linked Data. We have explored the methodology of modularity to describe data modeling through Semantic Web. In computer science, modularity is defined as “the degree to which a system’s components may be separated and recombined, often with the benefit of flexibility and variety in use”.<sup>12</sup> This analytic approach was adopted to represent several concepts involved in the *BIFLOW-Toscana Bilingue* catalogue. Whilst an ontology usually contains a large number of classes for a specific domain, this methodology has been used because translation in the Middle Ages involved specific areas within distinct domains of knowledge that required a range of radically different skill sets.<sup>13</sup> Therefore, while each domain is bounded by a specific disciplinary field (e.g., paleography, philology, codicology, etc.), the domains can also be linked together, thus enabling a wider understanding of the topicality, connectedness, and structure of the whole ontology. While metadata is an important issue for the catalogue, this methodology allows us to manage the complexity by breaking it down into key categories for analysis. The choice of an ontology is not taken for granted. As the next section shows, state-of-the-art catalogues based on data models related to the *BIFLOW-Toscana Bilingue* project are also using other standards and technologies, while for certain aspects a complete absence of standards remains the norm.

## 6. Metadata through catalogue projects

Metadata represents the inherent characteristics of a catalogue and is considered key to accessing resources, despite the absence of shared standards in cultural institutions’ metadata policies. Interoperability and integration are the main issues to consider for accessibility, particularly as regards cultural heritage information, which comes in various forms. If a good practice for interoperability and interchange of data entails reusing models and metadata from other projects, existing catalogue projects can give an overview of the current state of the art. Firstly, as regards the *BIFLOW-Toscana Bilingue* project (as explained above), there are no scholarly projects that address their data model in terms of translation. Attempts have been made to theorize ontologies in articles published some time ago (e.g., MARAIS 2013); however, as yet, we are unaware of any actual projects using such standards. As far as the description of the materiality and physicality of

12. This definition is from Wikipedia: <https://en.wikipedia.org/wiki/Modularity>.

13. See also PIERAZZO and STOKES 2010.

objects, many projects adopt TEI (Text Encoding Initiative)<sup>14</sup> standards in both bibliographical and manuscript description projects. TEI is still a standard that uses XML, a hierarchical metadata model that can be a fitting data structure for many levels of annotation. Many institutions have adopted the TEI Guidelines as one of its modules is dedicated to shaping *Manuscript description*. An example from the *Italian Central Institute of Cataloguing* (ICCU) is *Manus OnLine* (MOL)<sup>15</sup>, a national catalogue of manuscripts that describes and exports manuscript descriptions into TEI XML documents. Some limits regarding the TEI representation for codicological aspects were already pointed out by Barbero and Trasselli (2015). Nonetheless, institutions using TEI descriptions are spread across the map: many Oxford libraries have chosen TEI for their online catalogues including the Bodleian Library as well as the Cambridge University Library and the British Library.<sup>16</sup> Likewise TEI has served as the foundation for other interesting projects such as *Bibliophilly*<sup>17</sup> (University of Pennsylvania) and *Manuscriptorium*,<sup>18</sup> the *European* sub-aggregator for the sphere of historical resources. Additionally, work on catalogue projects such as *Manuscripta: A Digital Catalogue of Manuscripts in Sweden*<sup>19</sup>, has spurred attempts to use new methodologies and create guidelines for medieval manuscripts.

Nevertheless, there are an increasing number of projects currently working on a model to describe a specific domain for manuscripts using a Semantic Web. The ontology from the *Biblistima* project brings together descriptions from about forty, primarily French, manuscript catalogues. This prototype combined data from the Mandragore and Initiale databases into an RDF-based framework, using an ontology modeled on CIDOC CRM and eFRBRoo (Erlangen Functional Requirements for Bibliographic Records object-oriented).<sup>20</sup> While eFRBRoo is largely used in information

14. As noted on the TEI website, “The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form” ( <https://tei-c.org/>).

15. See <https://manus.iccu.sbn.it/>; see also BARBERO and TRASSELLI 2015.

16. See [https://wiki.tei-c.org/index.php/TEI\\_manuscript\\_catalogues](https://wiki.tei-c.org/index.php/TEI_manuscript_catalogues).

17. See *BiblioPhilly: An Interface For The Bibliotheca Philadelphiensis Project* @ <http://bibliophilly.library.upenn.edu/>.

18. See <http://www.manuscriptorium.com/en/tei-p5-enrich-schema-en>.

19. See <https://www.manuscripta.se/>; for reference on the guidelines, see <https://www.rj.se/en/anslag/2011/greek-manuscripts-in-sweden--a-digitization-and-cataloguing-project/>.

20. For information on the Functional Requirements for Bibliographic Records see <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>; for developments of this project see <http://erlangen-crm.org/efrbr/>.



and bibliographic studies applied to library cataloguing, CIDOC CRM<sup>21</sup> is a standard used for information related to cultural heritage but also in the world of libraries, archives and research institutions. However, the lack of existing projects, or a specific ontology for the description of manuscripts, is still a gap that the Digital Humanities community needs to fill. Another project from an Oxford institution that is shifting from TEI metadata to Semantic Web standards is *Mapping Manuscript Migrations*, an international project designed to track the ownership and provenance of medieval manuscripts using data from several databases related to manuscript history.<sup>22</sup> The variety of projects attempting to define a model can lead to a lack of common standards on the issue, as well as to a proliferation of projects using standards designed for a range of conceptual domains with disparate purposes (e.g., FRBRoo or CIDOC CRM) and approaches that differ radically in style. As far as the *Manus* catalogue metadata are concerned, it is also worth mentioning the proposal by Anna Bellotto to migrate the TEI manuscript description using a subset of CIDOC CRM (BELLOTTO 2020).<sup>23</sup>

## 7. The BIFLOW Ontology

“The main purpose of developing an ontology is to get a better understanding of the domain in question and to create a tool for analyzing data concerning the domain”.<sup>24</sup> The BIFLOW-*Toscana Bilingue* project encompasses a variety of aspects within its specific domain. Here we describe the structure of the RDF/OWL ontology and how the model facilitates the expression and analysis of implicit and hidden information/associations existing in the sources. In order to handle the variety of the features involved, we define classes and properties that import and extend other formal ontologies already in existence, such as eFRBRoo,<sup>25</sup> CIDOC CRM (Comité International pour la Documentation-Conceptual Reference Model), and *Bibliissima* ontology.<sup>26</sup> We shall discuss each stage in turn, focusing exclu-

21. For CIDOC CRM documentation see [http://cidoc-crm.org/docs/cidoc\\_crm\\_version\\_6.0.pdf](http://cidoc-crm.org/docs/cidoc_crm_version_6.0.pdf).

22. See <https://mappingmanuscriptmigrations.org/>.

23. For several recent proposals on migrating TEI/XML into ontologies see CIOTTI and TOMASI 2016–2017 and also CIOTTI 2018.

24. See EIDE and ORE 2019, 184.

25. See FRBRoo 2.2 @ [http://www.ifla.org/files/assets/cataloguing/frbr/frbroo\\_v2.2.pdf](http://www.ifla.org/files/assets/cataloguing/frbr/frbroo_v2.2.pdf).

26. For documentation see @ <https://doc.bibliissima.fr/ontologie/bibma/>.

sively on the main classes and properties in order to elucidate the logic behind our ontology. Before explaining the peculiarities and complexity of this ontology, it is worth elaborating on the concept of the ‘dossier’. Each element of the dossier has been atomized to enable its retrieval and to understand how it interacts with the others. We have already set out the dossier and what a scholar should analyze in order to examine the ways in which texts and manuscripts circulating in multiple languages in medieval Tuscany were transmitted. A dossier is an act of interpretation operated by an editor that collects information and versions of texts regarding a specific work. The dossier forms the basis of our computational model, even if it is not expressed as RDF/OWL class directly.

The BIFLOW ontology maps out two main units of analysis that are clearly distinct from one another: firstly, the text with its linguistic and literary aspects (including language, genre, authorship, etc.); secondly, the material and historical characteristics of the manuscripts in which the texts are presented. Consequently, the core of the ontology is based on four main classes: *Work*, *Expression*, *Manuscript*, and *Person*. The theoretical assumptions and the following classes were also inspired by Peter Burke’s questions raised in “Cultures of Translations” and noted here in section three. The questions What?, Who?, and In what manner? were very helpful starting points for building a theoretical framework. When Antonio Montefusco finalized the dossier, the very first computational structure envisioned took into consideration the established models developed within the fields of Digital Humanities, with Burke’s questions in mind. In particular, the eFRBRoo model helped us to understand the different levels of abstraction and to adequately represent such complexity. The structure of Functional Requirements for Bibliographic Records standard (FRBR), and in particular its version eFRBRoo<sup>27</sup> (FRBR has since been changed and is now more connected to the CIDOC CRM model), was conceived for information and bibliographic studies applied to library cataloguing. This is a model that embraces four levels: *Work*, *Expression*, *Manifestation* and *Item*.<sup>28</sup> The two first classes allowed us to clarify the semantic and

27. See <http://www.cidoc-crm.org/frbroo/ModelVersion/frbroo-v.3.0>.

28. Since 1990, IFLA (*International Federation of Library Associations*) has aimed to provide a model to describe bibliographic records: “the purpose of the study is to delineate in clearly defined terms the functions performed by the bibliographic record with respect to various media, various applications, and various user needs. The study is to cover the full range of functions for the bibliographic record in its widest sense — i. e., a record that encompasses not only descriptive elements, but access points (name, title, subject, etc.), other “organizing” elements (classification, etc.), and annotations” (<https://www.ifla.org>).

conceptual ambiguities and to identify and divide the different versions of the texts into hierarchies: *Biflow Work* is a class that is equivalent to *F1 Work*, defined in the eFRBRoo guidelines as a class that “comprises distinct concepts or combinations of concepts identified in artistic and intellectual expressions [. . .] [that] may appear in the course of the coherent evolution of an original idea into one or more expressions that are dominated by the original idea”.<sup>29</sup> Whilst some properties of *Work* class are equivalent to eFRBRoo, some additional proprieties are also created *ad hoc* for the BIFLOW ontology. For instance, the BIFLOW property is represented by the property *has genre*, which expresses the relationship with a literary genre. Each *Work* then can be categorized by its genres. A *Genre* is also a class of the BIFLOW ontology and can identify a particular type or style of literature that a scholar can recognize. Some equivalent properties with the *F1 Work* class include *was created by*, which expresses the relation to the author who created the work. The property is equivalent to *P94i\_was\_created\_by* from the CIDOC CRM class *E28 Conceptual Object*. This class has a sub-class called *E89 Propositional Object*, which is a super class of *F1 Work*. The range of that property is *E65\_Creation*, which has a propriety *P140\_assigned\_attribute\_to* that expresses a statement of responsibility of the creation of a knowledge, in the case of CIDOC CRM, and precisely of a *Work* in the BIFLOW ontology. The property that links *Work* with *Expression* is *has representative expression*. This describes “the most characteristic expression of the instance of *F1 Work*”, meaning that the editor identifies a specific version of a text to be connected to the class *Work* (see Fig. 1).

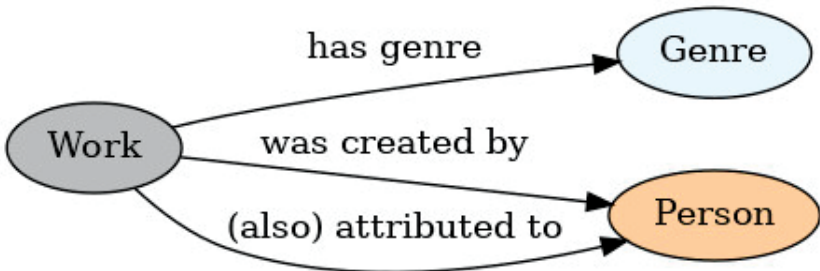


Figure 1. A graphical representation of the model *Work*.

29. For the current version, see FRBRoo v. 3.0 and [http://www.cidoc-crm.org/frbroo/fm\\_releases](http://www.cidoc-crm.org/frbroo/fm_releases).

*Expression* is essentially the main class of the BIFLOW ontology. *Expression* is a class that covers the concept of text described as linguistic code, rhetorical structure, and a series of words separated but also linked to the description of the materiality of a document that contains the text. This class is used to identify the version of a work that could be written by a copyist (which is linked to *Manuscript* class) or written in another language by a translator. *Expression* is also an act of interpretation defined as the class *Interpretation Act* of HiCo (Historical Context Ontology).<sup>30</sup> This ontology explicitly describes how ‘tackling an object of interest in a particular historical’ context could be an interpretation by ‘a person’: i.e., in this particular case of the BIFLOW ontology, the editor of the dossier who critically selects the information regarding a particular *Work*. Therefore, “an interpretation act is a situation in which an agent defines some useful information — about the context of an object — as RDF triples starting from the ‘content’ of an object”.<sup>31</sup> *Expression* is also equivalent but not identical to class in eFRBRoo. The peculiarity of this project obtains in the overlap in meaning of those two classes: *Interpretation Act* and *Expression* (see Fig. 2).

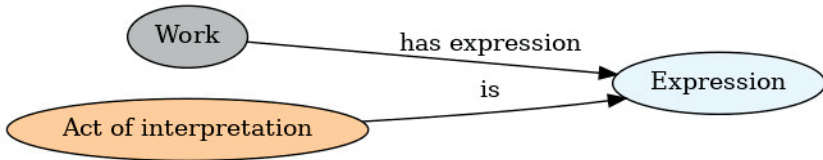


Figure 2. A graphical representation of the model *Act of interpretation*.

*Interpretation Act* is the high-level description of the eFRBRoo model and also of the BIFLOW ontology. An *Expression* can be connected with another *Expression* if it derives from it. ‘Derived from’ is a very fundamental property of an *Expression* that connects another *Expression*. Hence, it is a hierarchical structure as an *Expression-1* (a version of a text) that can be connected to another *Expression-2* (another text) that means that *Expression-2* can be a translation derived from one or more previous *Expressions*. If these do not exist, it means that the expression is first in the ‘chain of Expressions’ (see Fig. 3).

30. This ontology was developed by Marilena Daquino, Silvio Peroni, and Francesca Tomasi: see <http://hico.sourceforge.net/>.

31. For a definition, see <http://hico.sourceforge.net/#d4e3404>.

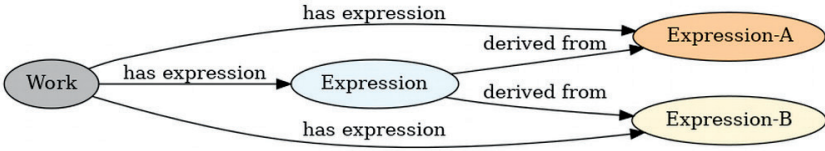


Figure 3. A graphical representation of the model *Expressions*.

Here one can see a very simple representation of the work *Liber* by Angela da Foligno (see Fig. 4):

The screenshot shows a website interface with the following elements:

- Navigation bar: Home, About, Filtra categorie, Ontologia Biflow, Team, Pubblicazioni, Eventi-Blog, and a search icon.
- Metadata: Autore: Angela da Foligno; Genere: Letteratura religiosa, Autobiografia spirituale, Mistica.
- Buttons: Anteprima di stampa, Risorse RDF.
- Section: Opera e relative versioni, containing a graph with nodes AngFolLIB-A, AngFolLIBa, and AngFolLIBb.
- Section: Versioni, containing a list of three entries:
  - Codice: AngFolLIBa; Titolo: Liber [Memoriale + Instructiones o Exhortationes]
  - Codice: AngFolLIBa; Titolo: [Qui comincia] la vita di sancta Agnola da Fuligno sechondo la experientia che ebbe delle chose di dDio
  - Codice: AngFolLIBb; Titolo: [Questa si e] una bella ed utile dottrina a dare conoscimento all'anima della via che de' tenere a volere venire a stato di salute

Figure 4. An example of the website interface.

The opportunity to explicitly describe one or more properties and types of relationship provides an adequate semantic data representation of the relationships among the texts and the representation of those versions of texts in a graph. We can see the first version of the text as the first *Expression* of a *Work*, even if there may be many ‘first’ such *Expressions*.

Using OWL, the BIFLOW ontology sets a few mandatory properties for *Expression* resources, which are needed to define and structure the basic features of the text’s version. This class comprises many aspects describing the linguistic and philological aspects of a text. The classification of the

language used for each *Expression* is a fundamental requirement for mapping out the act of translation. It defines the languages used in a version of a text and comprises the natural languages in the sense of concepts. We use a class called *Language* to describe the definition of the languages introduced by the editor as well as more general codes, such as those defined in ISO 639:3.<sup>32</sup> This international standard provides an enumeration of “languages as complete as possible, including living and extinct, ancient and constructed”.<sup>33</sup> Despite the latest ISO evolution, one can still find problems connected with identifying specific old idioms. This class *Language* is equivalent to the CIDOC CRM *E56\_Language*. Even though a text can include different types of languages, one can predominate. *Expression* may also have defined ‘other languages’ that can be used in a version of a text. Since the specifications tend to be less rigorously defined, one of the primary classes, *Language*, has crucial importance in the text in terms of the quantity of text written in that particular language. The resources described by the catalogue of the *BIFLOW-Toscana Bilingue* project are objects characterized by multiple levels of complexity: non-immediate compositionality, references to both internal and external resources, variability of authorship identity, manuscript descriptions, and so on. Every class of a particular *Language* is connected to an *Expression* and can be defined as a translation by a different language and by the connection to a translator.

The BIFLOW ontology allows us to model important aspects of the textual tradition, manuscript production, and their mutual relationships, as well as the prosopography of the people involved as authors, translators, or copyists. The variety of aspects within the BIFLOW project come together to integrate different models as a framework for a project-specific extension. The very first actor in this ontology is the creator of the *Work*. The class *Person* identifies key narrators of each artistic process such as authors, copyists, and translators. Each person described is an agent and defines the act of creating, translating, and copying a text. There are several standards available that can be used for this particular class. *Person* is declared an equivalent class to *F10\_Person*, that is, ‘class comprises real persons who live or are assumed to have lived’. It is also an equivalent class to the CIDOC CRM current *E21\_Person* and the FOAF (Friend Of A Friend) *Person* class.<sup>34</sup> Those models are chosen for describing people and their relations with the three classes *Work*, *Expression*, and *Manuscript*. A *Person*

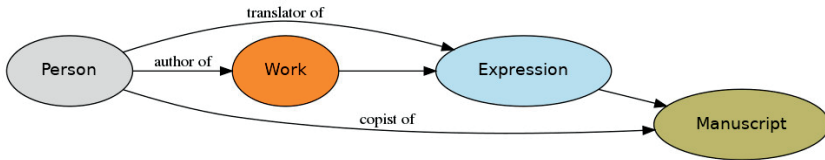
32. See [https://en.wikipedia.org/wiki/ISO\\_639-3](https://en.wikipedia.org/wiki/ISO_639-3).

33. See [https://en.wikipedia.org/wiki/ISO\\_639-3](https://en.wikipedia.org/wiki/ISO_639-3).

34. FOAF: Friend of a friend: see <http://xmlns.com/foaf/0.1>.

is the creator, the author, and is connected to the class *Work* through the property *has\_author*. A *Person* linked to an *Expression* is defined by the property *translator* and identifies the act of translation. A copyist is connected to the class *Manuscript* and this relation is described in the act of physically writing a text through the property *copyist*. Additionally, people and places have been aligned with existing authority lists, such as the Virtual International Authority File (VIAF, <https://viaf.org/>) and GeoNames (GN, <http://www.geonames.org/ontology>).

A large number of medieval texts are anonymous. Some texts circulated without a named author or translator or details of anyone who might have copied it. An anonymous author is still a person but is an unknown person. In the BIFLOW ontology, anonymous authors are handled as blank nodes, and those elements assume a meaning in relation to *Work*, *Expression*, and *Manuscript* and their properties. In this way, different anonymous authors are identified as unique nodes in the RDF graphs. Blank nodes in RDF “are graph nodes that represent a subject (or object) for which we would like to make assertions but have no way to address with a proper URI (a resource)” (SEGARAN ET AL. 2009, 67) (see Fig. 5).<sup>35</sup>



**Figure 5.** A graphical representation of the model *Manuscript*.

The model also defines further specific classes and properties related to the conceptual domain of *manuscripts* as inspired and imported by the *Biblistima* ontology project. Hence, in BIFLOW every *Expression* is not connected with a class called *Manifestation* as in other bibliographical entries described by eFRBRoo, but rather is connected to a *Manuscript* class.

An *F4 Manifestation singleton* class is an equivalent class<sup>36</sup> as well as the CIDOC CRM *E22 Man-Made Object*. The former is defined as a class

35. See also SEGARAN, EVANS, and TAYLOR 2009, 67.

36. In OWL, `owl:equivalentClass` “is a [ . . . ] property that links a class description to another class description. See <https://www.w3.org/TR/owl-ref/#equivalentClass-def>.”

that “comprises physical objects [. . .] that were produced as unique objects, with no siblings intended in the course of their production” and the latter as a class that “comprises physical objects purposely created by human activity”.<sup>37</sup> Both definitions can be applied to the BIFLOW class *Manuscript*. However, by naming a class *Manuscript*, the intention is to emphasize the importance of conceptualizing a domain so often represented in a wholly varied and fragmented way. In order to examine the various manuscript forms circulating during the Middle Ages, the BIFLOW ontology dedicates a specific class to describing documentary sources: *Manuscript*. Following the *Biblissima* project<sup>38</sup> and inspired by the TEI manuscript description module, while assessing various scholarly contributions scattered in various publications (PAGE ET AL. 2019), this class provides a broader description of several features belonging to a medieval manuscript and linked to many other different classes which, for example, can be related to information regarding the library where the document and manuscript identifier are preserved. In order to fully understand the phenomenon of medieval translation, its witnesses need to be examined in their materiality. This class includes any type of writing that is not printed (autographs, documents, deeds, inventories, etc.). In doing so, it addresses many elements previously discussed concerning a manuscript description. For instance, how to identify a composite and homogeneous manuscript, the dating, the history of the manuscript as well as its binding, its *mise en page* and *mise en text* as well as its graphical system. Thus, many are classes connected to *Manuscript*, starting from the class that describes a single *folio* or group of *folios*. The class *folio* distinguishes between manuscripts and loose, unbound *folios*. Those *folios* are in the class called *Localisation*, which is the equivalent of the *Biblissima* class *folio*, a subclass of the CIDOC CRM class *E84\_Information\_Carrier* and the *Biblissima* class *Component*. This class is also linked to *Expression*. Hence for some concepts a subset of the *Biblissima* and CIDOC CRM ontologies are taken into account, such as dimension, material, and place. However, some other classes inspired by the TEI module including *History*, *Binding*, and *Collation* are currently defined as BIFLOW classes although we are considering using the VisColl (Modeling and Visualizing the Physical Construction of Codex Manuscripts)<sup>39</sup> model for codicological aspects. The categories and the structure of this project will allow users

37. For the definitions, see <http://www.cidoc-crm.org/frbroo/ModelVersion/frbroo-v-3.0>.

38. See ERUNZEANU, RÉGIS, and MACDONALD 2016.

39. See <https://viscoll.org/>.



to retrieve data from a single category (each expression, or people involved in the process) or cross-reference data for more in-depth results. The project is provided with a SPARQL Endpoint that will allow users to query catalogue data through the ontology.

## 8. Conclusion

Translation, and particularly medieval translation, is a complex phenomenon. To define it with the precision and formalization necessary for its digital representation, it was crucial to adopt a hermeneutical model that allows us to describe translation in Middle Ages in its aspect of ‘specific activity’. The focus of the survey, compared to the previous secondary literature as well as to the already existing digital tools, has been shifted accordingly, from the lexicon of the texts involved to the activity of translation. The aim of the work of the BIFLOW team and of the catalogue is, in fact, to describe translation as a social-historical phenomenon. We found a useful interpretative tool for this description in the innovative approach that the historian Peter Burke used for the study of language and translation in modern times.

The BIFLOW-*Toscana Bilingue* project opens up new opportunities for interpretation and analysis based on an innovative historical-social approach. In this case, the adoption of Semantic Web technologies and Linked Open Data, which allowed us to fully map out entities and their relationships in order to represent multilingualism and translation in the Middle Ages, was fundamental. This raises an important issue for humanists who need to care about data modeling and delivering it on the Web. Abandoning the centrality of lexical investigation, texts and their history of production and transmission play a crucial role in the BIFLOW data model. Manuscripts are essential documents for deepening the investigation into the transmission of text, especially in relation to its production and circulation in space and time. Whereas translation in the Middle Ages is defined by manuscript transmission and the network of factors detected by the BIFLOW data model, translation studies are considerably more extensive and involve a wide variety of components. However, each item of information formally structured can deepen the examination previously undertaken by the researcher-editors of the *Biflow-dossiers*.

The data modeling for the BIFLOW project has highlighted the importance of sharing models for the sustainability and interchange of data, particularly in the realm of medieval studies. The range of models applied in

our project has also shown how our digital scholarship research is based on ‘Communities of Practice’ and on shared methods and models. In the *Biblissima* ontology, the model of manuscripts was a crucial data model for understanding the complexity of the class *Manuscript* as a specific conceptual domain. Likewise, eFRBRoo and CIDOC CRM ontologies were used consistently in our model. Frameworks such as RDF and OWL, in which all those models are created, have enabled us to expand and describe the peculiarities of our catalogue in a flexible way. The model of the BIFLOW ontology has been designed but is still being finalized because many of the aspects contained in it do not yet have a standard. This underlines the importance of implementing, for instance, an ontology that can represent a precious valorization for the description of manuscripts and serve as a powerful research tool for a range of different projects within the field of medieval manuscripts.

Additionally, what we aim to achieve in this first phase may act as the beginning of a new one. This project can be expanded and refined in both its model and in the way that we visualize the data. Built in RDF and OWL, the BIFLOW ontology could further be expanded to all standards of Semantic Web (for instance TEI), and it could be used together with other formal ontologies. Lastly, a long-term aim involves connecting the documentary and critical editions of the texts in the *BIFLOW-Toscana Bilingual* project. One final observation about the challenge of digitally representing the phenomenon of medieval translation as a ‘specific activity’ is necessary. The user of the *BIFLOW-Toscana Bilingual* catalogue can find specific information on each act of translation in a specific context; at the same time, thanks to data models expressed in the ontology, they will be able, starting from a single item of data (within the identified ontological classes: work, expression, people) to identify stratified research paths and to reconstruct relevant historical-social phenomenon. Thus, it will also be important to implement interoperability projects with the main databases concerning the disciplinary fields covered by the *BIFLOW-Toscana Bilingual* catalogue (e.g., *Mirabileweb*,<sup>40</sup> *Corpus Corporum*,<sup>41</sup> etc.). In this way, the phenomenon of translation will be fundamental to the enrichment of scholarly analysis and will facilitate access to scholarly research of the cultural history of the Italian Middle Ages.

*University of Venice*

40. See <http://www.mirabileweb.it/>.

41. See <http://www.mlat.uzh.ch/MLS/>.

## Works Cited

- ARTALE, Elena. 2003. "I volgarizzamenti del corpus TLIO". *Bollettino dell'Opera del Vocabolario italiano* 8: 299–377.
- BARBERO, Gigliola and Francesca TRASSELLI. 2014–2015. "Manus OnLine and the Text Encoding Initiative Schema". *TEI: Journal of the Text Encoding Initiative* 8 @ <https://journals.openedition.org/jtei/1054>.
- BEER, Jeanette. 1989. *Medieval Translators and their Craft*. Kalamazoo: Medieval Institute Publications.
- BEER, Jeanette, ed. 2019. *A Companion to Medieval Translation*. Amsterdam: Arc Humanities Press.
- BELLOTTO, Anna. 2020. "Medieval manuscript descriptions and the Semantic Web: Analysing the impact of CIDOC CRM on Italian codicological-paleographical data". *Digital Humanities Quarterly* 14.1 @ <http://www.digitalhumanities.org/dhq/vol/14/1/000449/000449.html>.
- BENJAMIN, Walter. 2000. *L'opera d'arte nell'epoca della sua riproducibilità tecnica*. Torino: Einaudi.
- BERNERS-LEE, Tim, James HANDLER, and Ora LASSILA. 2001. "The Semantic Web". *Scientific American* 284–5: 34–43.
- BIZER, Christian, Tom HEATH, and Tim BERNERS-LEE. 2011. "Linked Data: The story so far". *International Journal on Semantic Web and Information Systems* 205–27 @ <http://linkeddata.org/docs/ijswis-special-issue>.
- BRUN, Laurent, Frédéric DUVAL, Françoise FERY-HUE, and Christine GADRAT. 2005. "Vers un inventaire informatisé des traductions latines d'œuvres vernaculaires". *Scriptorium* 59.1: 90–108.
- BURKE, Peter and Ronnie PO-CHIA HSIA, eds. 2007. *Cultural Translation in Early Modern Europe*. Cambridge: Cambridge University Press.
- BURKE, Peter and Roy PORTER, eds. 1995. *Languages and Jargons: Contributions to a Social History of Language*. Cambridge: Polity Press.
- CERULLO, Speranza and Lino LEONARDI. 2017. *Tradurre dal latino nel Medioevo italiano*. Firenze: Edizioni del Galluzzo per la Fondazione Ezio Franceschini, 2017.
- CHIESA, Paolo. 1987. "Ad verbum o ad sensum? Modelli e coscienza metodologica della traduzione tra tarda antichità e alto medioevo". *Medioevo e Rinascimento* 1: 1–51.
- CIOTTI, Fabio and Francesca TOMASI. 2016–2017. "Formal Ontologies, Linked Data, and TEI Semantics". *Journal of the Text Encoding Initiative* 9 @ <https://journals.openedition.org/jtei/1480>.
- CIOTTI, Fabio. 2018. "A formal ontology for the Text Encoding Initiative". *Umanistica Digitale* 2.3 @ <https://umanisticadigitale.unibo.it/article/view/8174>.
- CIULA, Arianna, Øyvind EIDE, Cristina MARRAS, and Patrick SAHLE, eds. 2018. "Models and Modelling between Digital and Humanities—A Multidisciplinary Perspective". *Historical Social Research* 43.4: 343–61.
- COCCIA, Emanuele and Sylvain PIRON. 2008. "Poésie, sciences et politique. Une génération d'intellectuels italiens (1290–1330)". *Revue de Synthèse* 129.4: 551–86.

- COPELAND, Rita. 1995. *Rhetoric, Hermeneutics, and Translation in the Middle Ages: Academic Traditions and Vernacular Texts*. Cambridge: Cambridge University Press.
- CORNISH, Alison. 2011. *Vernacular Translation in Dante's Italy: Illiterate Literature*. Cambridge: Cambridge University Press.
- CUENCA GRAU, Bernardo, Ian HORROCKS, Yevgeny KAZAKOV, and Ulrike SATTLER. 2008. "Modular Reuse of Ontologies: Theory and Practice". *Journal of Artificial Intelligence Research* 31: 273–318.
- DE LUCA, Giuseppe. 1977. *Scrittori di religione del Trecento. Volgarizzamenti*, I–IV. Torino, Einaudi.
- DIONISOTTI, Carlo. 1967. *Geografia e storia della letteratura italiana*. Torino, Einaudi.
- DOTTO, Diego. 2012. "Note per la lemmatizzazione del corpus DiVo". *Bollettino dell'Opera del Vocabolario Italiano* 17: 33–64.
- EIDE, Øvind and Christian-Emil ORE. 2019. "Ontologies and data modeling". *The Shape of Data in the Digital Humanities: Modeling Texts and Text-based Resources*, edited by Julia FLANDERS and Fotis JANNIDIS, 178–203. London: Routledge.
- FERRANTE, Gennaro. 2010. "Forme, funzioni e scopi del tradurre Dante da Coluccio Salutati a Giovanni da Serravalle (con edizione delle dediche della 'Translatio Dantis')". *Annali dell'Istituto Italiano per gli studi storici* 25: 147–82.
- FLANDERS, Julia, and Fotis JANNIDIS, eds. 2019. *The Shape of Data in the Digital Humanities. Modeling Texts and Text-based Resources*. London: Routledge.
- FOLENA, Gianfranco. 1993. *Volgarizzare e tradurre*. Torino, Einaudi.
- FRUNZEANU, Eduard, Régis ROBINEAU, and Elisabeth MACDONALD. 2016. "Bibli-sima's Choices of Tools and Methodology for Interoperability Purposes". *CIAN-Revista de Historia de las Universidades* 19.1 @ <https://e-revistas.uc3m.es/index.php/CIAN/article/viewFile/3146/1783>.
- GALDERISI, Claudio. 2011. "Accessus au corpus Transmédié. Exhaustivité, subjectivité, interactivité . . . et catastrophes". *Translations médiévales: cinq siècles de traductions en français au Moyen Âge (XIe–XVe siècles): Étude et répertoire*. Vol. 2, *Le Corpus Transmédié: Répertoire*, 5–52. Turnhout: Brepols.
- GAUTHIER, Patrick. 2011. "Richard de Saint-Victor. Benjamin minor". *Translations médiévales cinq siècles de traductions en français au Moyen Âge (XIe–XVe siècles)*, edited by Claudio GALDERISI, Vol. 2, no. 487. Turnhout: Brepols.
- GEHL, Paul. 1993. *A Moral Art: Grammar, Society, and Culture in Trecento Florence*. Ithaca: Cornell University Press.
- GUADAGNINI, Elisa, and Giulio VACCARO. 2014. "Un contributo allo studio del 'volgarizzare e tradurre': il progetto DiVo". *Lingua testi culture. L'eredità di Folena vent'anni dopo*, edited by Sandra COVINO, Pier Vincenzo MENGALDO, and Lorenzo RENZI, 91–105. Padova: Esedra.
- GUARINO, Nicola, Daniel OBERLE, and Steffen STAAB. 2009. "What Is an Ontology?" *Handbook on Ontologies*, edited by Steffen STAAB and Rudi STUDER, 1–17. Berlin, Heidelberg: Springer.
- HOLMES, James S. 1988. "The Name and Nature of Translation Studies". *Translated! Papers on Translation and Translation Studies*, 66–80. Amsterdam: Rodopi.

- MAGGINI, Francesco. 1952. *I primi volgarizzamenti dei classici*. Florence: Le Monnier.
- MARAIS, Kobus. 2013. "Exploring a Conceptual Space for Studying Translation and Development". *Southern African Linguistics and Applied Language Studies* 31: 403–14.
- MCCARTY, Willard. 2005. *Humanities Computing*. London: Palgrave.
- MCLUHAN, Marshall. 1962. *The Gutenberg Galaxy: The Making of Typographic Man*. Toronto: University of Toronto Press.
- NEEDLEMAN, Mark. 2011. "Linked Data: What Is It and What Can It Do?". *Serials Review* 37.3: 234.
- PIERAZZO, Elena and Peter STOKES. 2010. "Putting the Text back into Context: A Codicological Approach to Manuscript Transcription". *Codicology and Palaeography in the Digital Age 2*, edited by Franz FISCHER, Christiane FRITZE, and Georg VOGELER, 397–429. Norderstedt: Books on Demand.
- PHYM, Anthony. 1993. "Negotiation Theory as an Approach to Translation History: An inductive lesson from the 15th Century Castille". *Knowledge and Translation*, edited by Yves GAMBIER and Jorma TOMMOLA, 27–39. Turku: University of Turku Center for Translation and Interpreting.
- SCHIAFFINI, Aldo. 1969. *Tradizione e poesia nella prosa d'arte italiana dalla latinità medievale al Boccaccio*. Roma: Edizioni di Storia e Letteratura.
- SEGRE, Cesare. 1963. *Lingua, stile e società. Studi sulla storia della prosa italiana*. Milano: Feltrinelli.
- SEGARAN, Toby, Colin EVANS, and Jamie TAYLOR. 2009. *Programming the Semantic Web*. Boston: O'Reilly Media, Inc.
- STEINER, George. 1975. *After Babel*. Oxford: Oxford University Press.
- TANTURLI, Giuliano. 1998. "Continuità dell'Umanesimo civile da Brunetto Latini a Leonardo Bruni". *Gli umanesimi medievali*, a cura di C. LEONARDI, 735–78. Firenze: SISMEL.
- TANZINI, Lorenzo. 2012. "Albertano e dintorni. Note su volgarizzamenti e cultura politica nella Toscana tardo-medievale". *La parola utile, Saggi sul discorso morale nel Medioevo*, edited by Duilio CAOCCI, Rita FRESU, Patrizia SERRA, and Lorenzo TANZINI, 161–217. Roma: Carocci.
- VITI, Paolo, ed. 2004. *Leonardo Bruni, La perfetta traduzione*. Napoli: Liguori.