

# Modeling Medieval Romances

Kelly Christensen

École nationale des chartes, 65 Rue de Richelieu, 75002 Paris, [kelly.christensen@chartes.psl.eu](mailto:kelly.christensen@chartes.psl.eu)

This document explains and justifies the architecture of a data model whose objective is to help build knowledge about traditions of chivalric literature in the Middle Ages. The building blocks of this knowledge are Medieval works and their expressions in various text versions, as well as both extant and missing inscriptions of those versions on physical objects, including manuscripts and fragments. By structuring this information in a relational framework, linking abstract textual traditions and concrete archival evidence, the data model aims to let users curate large-scale collections of information from a diverse corpus, including Medieval French, Middle Dutch, and Old Norse.

The model needs to be able to execute the following user tasks:

- Return a collection of resources (manuscripts, works, etc.) according to search criteria.
  1. Ex. All the manuscripts that contain at least one witness of David Aubert’s 1462 prosified version of *Renaut de Montauban*.  $\rightarrow [\mathbf{Doc}_1, \mathbf{Doc}_2, \dots \mathbf{Doc}_i]$
  2. Ex. All works that are grouped within the cycle about the knight Renaut de Montauban.  $\rightarrow [\mathbf{Work}_1, \mathbf{Work}_2, \dots \mathbf{Work}_i]$
- Return a single resource according to search criteria.
  1. Ex. The five-volume witness of David Aubert’s version of *Renaut de Montauban* that is partly conserved in Paris and partly in Munich.  $\rightarrow [\mathbf{Wit}_x]$
  2. Ex. The earliest known version of the work *Renaut de Montauban*.  $\rightarrow [\mathbf{Text}_y]$
- Return a tree of all the ancestors, descendents, and/or neighbors linked to a single resource or a collection of resources based on search criteria.
  1. Ex. All the witnesses of David Aubert’s version of *Renaut de Montauban*, along with all the witnesses’ page ranges as well as the bibliographic references to the physical documents that contain those pages.  $\rightarrow [(\mathbf{Wit}_1, [\mathbf{Pag}_a, \mathbf{Doc}_a]), (\mathbf{Wit}_2, [\mathbf{Pag}_b, \mathbf{Doc}_b]), [\mathbf{Pag}_c, \mathbf{Doc}_c]), \dots]$

## 1 Precedent Models

The data model must be adaptable yet consistent. Users should be able to request collections of linked resources regardless the genre or language of the works involved. Furthermore, they should be able to combine repertories through shared vocabularies and metadata fields. As such, we face an age-old dilemma. We are pulled between being specific and idiosyncratic, on the one hand, and being simple and generalizable on the other. To confront this issue, we first turn to other successful data models. From those precedents, we adopt a hierarchical approach and certain ontological distinctions, notably distinctions between a *Work* (the story), a *Text* (a telling of the story), and a *Witness* (a written version of that telling).

### 1.1 Functional Requirements for Bibliographic Records

One of the most robust bibliographic data models is the Functional Requirements for Bibliographic Records (FRBR), which slowly solidified during the last quarter of the twentieth century and is still today used to structure bibliographic databases.<sup>1</sup> In 1961, information scientists from around the world gathered in Paris to discuss best practices for cataloguing. Following their Paris meeting, the International Federation of Library Associations (IFLA) supported further discussions and members developed theoretical frameworks for organizing bibliographic data. Finally, in 1997, the IFLA approved the FRBR.<sup>2</sup> The FRBR articulates a generalizable hierarchy of four related entities: Works (**W**), Expressions (**E**), Manifestations (**M**), and Items (**I**). One of the examples in the 2009 corrected edition demonstrates how these entities relate to one another through the case of Johann Sebastian Bach’s Goldberg variations.<sup>3</sup>

<sup>1</sup>R. Bennet, B. F. Lavoie, and E. T. O’Neill. “The Concept of a Work in WorldCat: An Application of FRBR”. in: *Library Collections, Acquisitions, and Technical Services* 27.1 (2003).

<sup>2</sup>I. S. G. on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records*. Tech. rep. International Federation of Library Associations and Institutions, 1998. URL: <https://repository.ifla.org/handle/123456789/830>.

<sup>3</sup>We have augmented the example by adding an Item. I. S. G. on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records*. Tech. rep. International Federation of Library Associations and Institutions, 2009. URL: <https://repository.ifla.org/handle/123456789/811>, p. 58.

- **W<sub>1</sub>** J. S. Bach’s Goldberg variations
  - **E<sub>1</sub>** performances by Glenn Gould recorded in 1981
    - \* **M<sub>1</sub>** recording released on 33 1/3 rpm sound disc in 1982 by CBS Records
    - \* **M<sub>2</sub>** recording re-released on compact disc in 1993 by Sony
      - **I<sub>1</sub>** copy held at Cook Music Library, Bloomington, Indiana, USA (WOODWARD CD .B118 K1.988-35)

In the FRBR model, a *Work* is “a distinct intellectual artistic creation” and can have a title, a date, and a form or genre.<sup>4</sup> An *Expression* is an “intellectual or artistic realization” of the *Work* in some specific form, such as alphanumeric or musical notation, sound, image, or movement.<sup>5</sup> Manifestations are the concrete, “physical embodiment” of the *Expression* of a *Work*.<sup>6</sup> For instance, if an *Expression* of a *Work* is notated in Vedic Sanskrit around 900 BCE, a *Manifestation* of that literary *Expression* would be a physical manuscript. Finally, an Item is a “single exemplar of a manifestation,” which is a concept best suited for mass-produced resources, such as copies of an edited book or, as in the example above, copies of a CD-ROM.<sup>7</sup>

## 1.2 *Textrad*

More recently this century, Patrick Sahle has developed a set of entity relationships that have significantly impacted the fields of philology and textual criticism.<sup>8</sup> At the core of his wheel model (*Textrad*), which we illustrate as a triangle in Figure 1.1, Sahle deconstructs the idea of a text into three primary dimensions: text as a story’s ideal content [*Inhalt*], not mediated by anything (**I**); text as a story’s linguistic content [*sprachlichem Gehalt*], mediated through human language (**S**); and text as a physical document [*Dokument*] (**D**) on which a story’s content is materially represented.

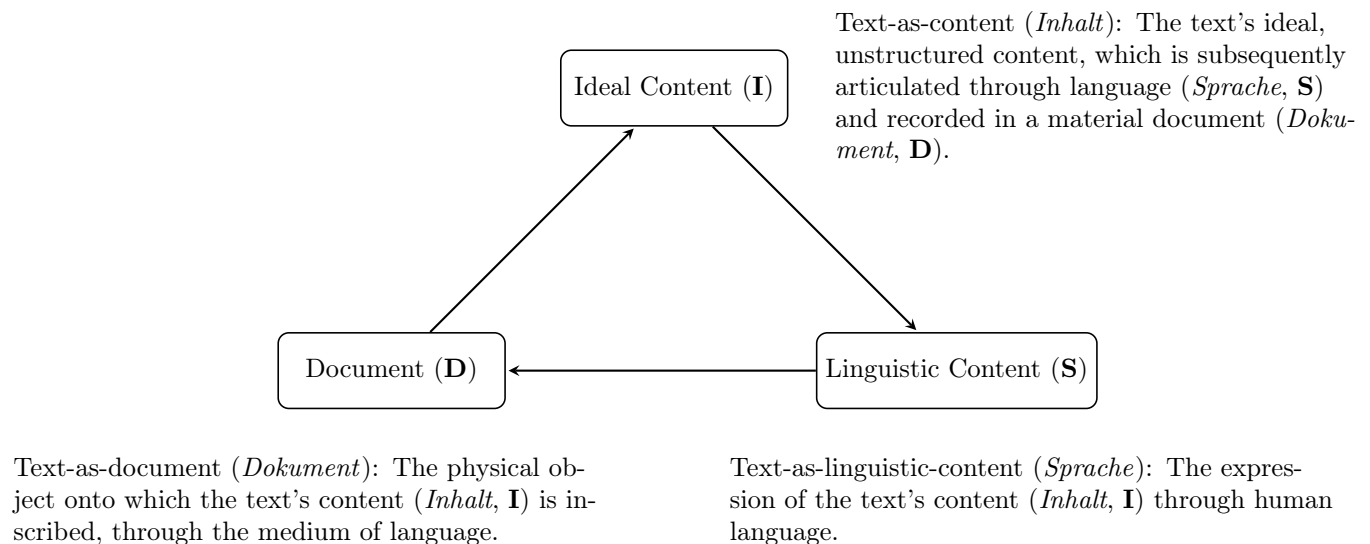


Figure 1.1: Main spokes of Sahle’s *Textrad*

Between the main spokes of the text-wheel (*Textrad*), as seen in Figure 1.2, Sahle includes three other dimensions: text as a set of signs (*Zeichen*), text as a structured sequence of content (*Werk*), and text as a version (*Fassung*). Sahle’s embedded dimensions make certain assertions about texts. For example, changes in a text’s linguistic content (**S**), such as translating *Beowulf* from Old English to contemporary Italian, will produce a new version (**F**) of the text, which will in turn produce a different physical document (**D**). However, changing *Beowulf* from Old English to contemporary Italian does not necessarily change the work (**W**) *Beowulf* itself, which still features the same organizing structure of the content of the *Beowulf* story (**I**).

Sahle avoids ascribing the term “text” to any one entity within the text-wheel. However, as Frédéric Duval notes, many scholars in the fields of philology, textual criticism, and scholarly editions habitually rely on the term “text” as well as “work” and “document.” Attempting to make explicit many scholars’ and editors’ implied typologies, Duval summarizes the state of the field as such:

“[*W*]ork designates the author’s text, eventually the text corresponding to the author’s intention, and implies authenticity; *text* denotes the linguistic sequence, which is attested in the document that is transmitting the

<sup>4</sup>Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records*, p. 17.

<sup>5</sup>Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records*, p. 19.

<sup>6</sup>Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records*, p. 21.

<sup>7</sup>Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records*, p. 24.

<sup>8</sup>P. Sahle. *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 3: Textbegriffe und Recodierung. [Finale Print-Fassung]*. Vol. 9. Norderstedt: BoD, 2013. URL: <https://kups.ub.uni-koeln.de/5353/>.

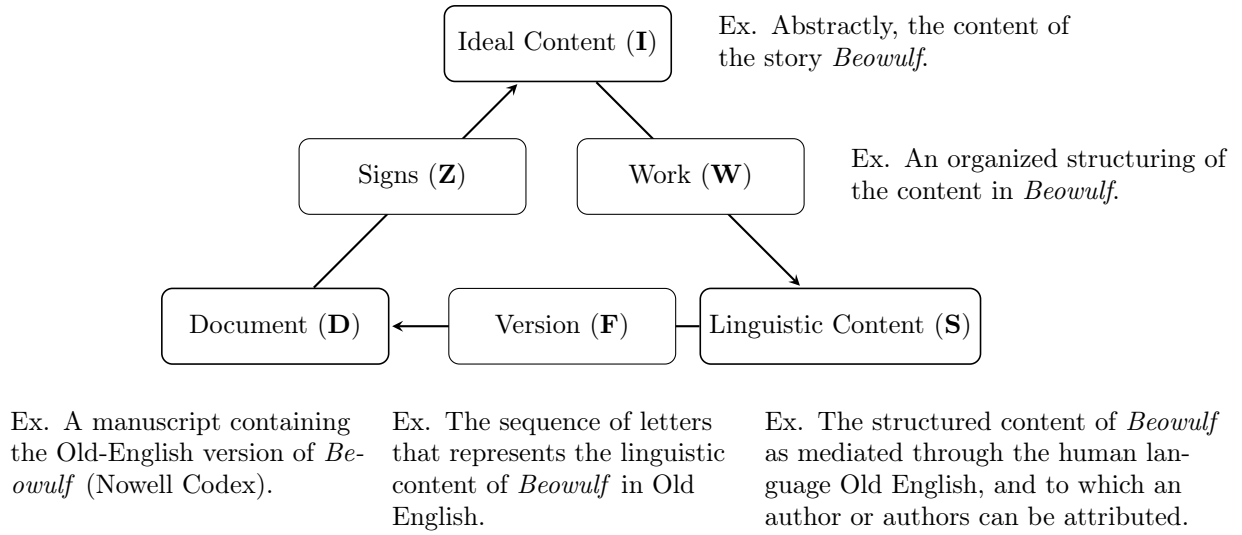


Figure 1.2: All of Sahle's *Textrad*

work; finally *document* is a physical manifestation of a text.”<sup>9</sup>

In Figure 1.3, we overlay the typologies of Sahle's text-wheel with Duval's summary of the mainstream typology used in textual criticism and scholarly editing.

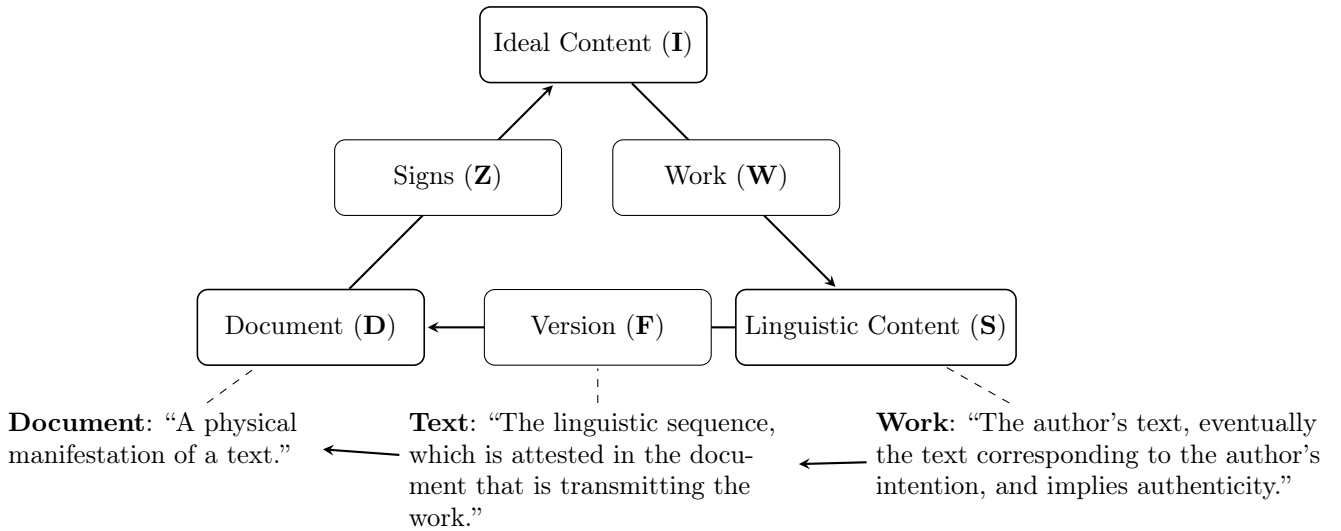


Figure 1.3: Overlap between Duval's summary and Sahle's *Textrad*.

In seeking to resolve terminological discrepancies between the fields of textual criticism and scholarly editing, Duval introduces a new discrepancy between his concept of *work* and Sahle's *Werke* concept. Given his focus on articulating the tripartite *work-text-document* concerns of textual criticism and scholarly editions, this is not a problem for Duval. Sahle's concept of a work (*Werke*), which is not yet mediated through any human language or literary style, is not typically the subject of scholarly editions or close textual readings. Such endeavors focus on what Sahle would call linguistic content (S).

## 2 Precedent Models' Limitations

To stress test the FRBR and *Textrad*, neither of which were specially designed for Medieval literature, we explore two cases. The first is a work about the knight Renaut de Montauban. In the 1460s, the work was recomposed in a multi-volume manuscript, which introduces some complexity in relating a text version (*Fassung*) to multiple documents. Second, we underscore the challenge of relating text versions to physical documents through the case of a lost manuscript, which, before being dismembered, allegedly transmitted a version of the *Chanson d'Otinel* and a version of the *Chanson*

<sup>9</sup>“*work* désigne le texte de l’auteur, éventuellement le texte correspondant à la volonté de l’auteur, et implique la notion d’authenticité ; *text* dénomme la séquence linguistique attestée dans un document transmettant l’œuvre ; enfin *document* est une manifestation physique d’un text” F. Duval. “Pour des éditions numériques critiques. L’exemple des textes français”. In: *Médiévales* 73 (2017), pp. 13–29. ISSN: 07512708, 17775892. URL: <http://www.jstor.org/stable/44988396> (visited on 06/04/2024), p. 16.

*d'Aspremont*. Parts of both are conserved today in two different nineteenth-century collections of Medieval fragments. Through the data model, we want to be able to recognize what scholars argue, which is that the fragments were once transmitted together in a now lost manuscript.

## 2.1 *Renaut de Montauban* and multi-volume witnesses

Let us start with the easier case, concerning a work about the legendary knight Renaut de Montauban. Sahle's *Werke* concept is helpful here in that it defines a work as the ordering of a story's abstract content (*Inhalt*) into a narrative structure. The sequence of events whose order defines the *Work Renaut de Montauban* begins with a backstory that contextualizes the main conflict.<sup>10</sup> The narrator explains that four brothers, Aymon, Beuves, Girart, and Doon, once fought together against the emperor Charlemagne. Beuves, who is duke of Aigremont and vassal of Charlemagne, refused some of the duties the emperor had imposed, which provokes the latter's fury. The brothers came to Beuves's aid and joined his conflict with Charlemagne. *Renaut de Montauban* begins with this history because the work's main storyline focuses on how a new generation of brothers, Aymon's sons Renaut, Alard, Guichard, and Richard, again push back against Charlemagne and how the powerful emperor pursues revenge.<sup>11</sup>

In various languages and forms, many people have composed and recomposed the *Work* known by its French title as *Renaut de Montauban*. The earliest instance of such a composition, what Sahle calls the text-as-linguistic-content [*sprachlichen Gehalt*] and the FRBR call an *Expression* of the *Work*, dates from the end of the twelfth century and was expressed in French alexandrine verse. Following the argument Duval makes to rely on familiar terminology, but with greater attention to precise definitions, we use the term *Text* to refer to the idea of the articulation of a *Work* in some language and form, Sahle's text-as-linguistic-content.

Our test case concerns a *Text* of *Renaut de Montauban* that was written down about three centuries after the first known *Text* was inscribed. While working for the Burgundian duke Philippe le Bon between 1459 and 1465, David Aubert adapted *Renaut de Montauban* into contemporary prose. He structured his text's linguistic content inside evenly distributed chapters, each about 8 to 12 leaves long, and in five manuscript volumes. Each volume was about the same length, between 350 to 399 leaves, and featured nearly the same number of illuminations, between 47 and 53.<sup>12</sup> Copies of those volumes exist in Paris and Munich; the first four are in the Bibliothèque nationale de France and the fifth volume is in the Bayerische Staatsbibliothek. Crucially, in the terminology of the *Textrad*, this particular *Version* ([*Fassung*]) of Aubert's *Text* was intentionally produced in five physical documents; it does not exist today in five manuscripts because of some process of deconstruction after its production.

Aspect	FRBR	<i>Textrad</i>	LostMa
The subject matter of <i>Renaut de Montauban</i> .	NA	text-as-content ( <i>Inhalt</i> , <b>I</b> )	NA
The ordered series of episodes about Renaut's revolt, starting with the backstory of the older generation.	<i>Work</i>	text-as-work ( <i>Werke</i> , <b>W</b> )	<i>Work</i>
David Aubert's prosified French-language version.	<i>Expression</i>	text-as-linguistic-content ( <i>Sprache</i> , <b>S</b> )	<i>Text</i>
A five-volume copy of the prose version.	<i>Manifestation</i>	text-as-version ( <i>Fassung</i> , <b>F</b> )	<i>Witness</i>

Table 2.1: FRBR and *Textrad* modeling *Renaut de Montauban*.

Both the FRBR and Sahle's *Textrad* are capable of modeling some of what we have described thus far. As Table 2.1 demonstrates, both the FRBR and Sahle find that the term *Work* appropriately describes the ordered series of episodes that define Renaut's story of revolt. To describe instances of the *Work*, the FRBR and Sahle each prefer abstract concepts defined by their mediation through some form of human expression, *Expression* in the FRBR and *Sprache* in the *Textrad*. In our model, and in line with Duval, we hold that the term *Text* is best suited to indicate Aubert's specific telling of the *Work*, as articulated in a specific human language and literary style. Finally, what Sahle calls the *Fassung* or text version we call the *Witness*. It is standard amongst scholars of Medieval literature to call the version of the *Text* that was manifestly inscribed onto something, and which features its own text variants and formatting idiosyncracies, a *Witness*.

### 2.1.1 *Text* v. *Witness*

What is the difference between the *Text* and the *Witness*, or the text-as-linguistic-content and the text-as-version, as Sahle puts it? The former is indifferent to versions of spelling and formatting, much in the way the *Work* is indifferent

<sup>10</sup>Whether this opening section constitutes a prologue, in alignment with the generic expectations of a prologue for *chansons de geste*, is the subject of scholarly debate. P. Leverage. "Unrolling Memories : Audience Inscription in the Renaut de Montauban Prologue". In: *Romania* 118.469 (2000), pp. 206–218. DOI: 10.3406/roma.2000.1526. URL: [https://www.persee.fr/doc/roma\\_0035-8029\\_2000\\_num\\_118\\_469\\_1526](https://www.persee.fr/doc/roma_0035-8029_2000_num_118_469_1526).

<sup>11</sup>Having made peace, duke Aymon brings his four sons, Renaut, Alard, Guichard, et Richard, with him to meet emperor Charlemagne in Paris. After the meeting, Renaut plays chess with Charlemagne's nephew, but the game devolves into a dispute. The knight ultimately slays the nephew. Fearing the emperor's vengeful wrath, Renaut and his brothers flee Paris on the back of a magical horse, Bayard. Adventures ensue.

<sup>12</sup>D. Quérue. "Formules narratives rubriques et enluminures: l'exemple de l'histoire de Renaut de Montauban de David Aubert (XVe siècle)". In: ed. by C. Heck. 2007, pp. 131–145.

to language and form. By transcribing here the first line of Aubert’s *Text*, we would either be choosing which *Witness* to copy, as is the case with critical editions, or creating our own text version. Variations at the minute level of linguistic expression and dialect distinguish one text-as-version (*Witness*) from another. For example, the second volume of two different multi-volume *Witnesses* of Aubert’s *Text* each present the same linguistic content, as seen in Tabel 2.2, but each one has its own way of transforming linguistic content into written language.

Manuscript shelfmark	First line
BNF Arsenal 5073	Qui a veu l’histoire de Maugis d’Aigremont poeut avoir leu comment Vivien.
BNF français 19174	Qui a veue l’histoire de Maugis d’Aigremont bien au long, peult avoir veu comment Vivien.

Table 2.2: First lines of the second volume of two *Witnesses* to Aubert’s *Text*, *Renaut de Montauban*.

Both transcribed sequences of characters in Table 2.2 are each part of a different text-as-version or *Witness*, both of which are part of the same text-as-linguistic-content or *Text*, in this case, Aubert’s *Renaut de Montauban*.

### 2.1.2 Cycle

Neither the FRBR nor the *Textrad* have a concept well suited for describing a group of related *Works*. Both models’ ontological categories reach only to the extent of a single *Work*. Yet when treating similar content, such as the adventures of Renaut de Montauban and his family, sometimes *Works* of chivalric literature cohere into a collection known as a *Cycle*. While the earliest instance of *Renaut de Montauban* appeared in the twelfth century, other *Works* were created in the centuries that followed, which revisited characters central in *Renaut de Montauban*. As Gaëtan Augustine notes in his dissertation, the later *Works* do not all return to the core *Work*’s central themes, namely revolt and imperial tyranny, but they nevertheless form a *Cycle* by building a world around figures central to the *Work Renaut de Montauban*.<sup>13</sup> The precedents, FRBR and *Textrad*, are not equipped to model such metadata.

### 2.1.3 The Archival Item and its Pages

A more significant incompatability between the *Renaut de Montauban* test case and the existing data models arises when we introduce archival evidence. The five-volume *Witness* to Aubert’s *Text* that is partly preserved in Paris and partly in Munich does not directly relate to what Sahle calls the text-as-document (*Dokument*) and what the FRBR call the *Item*. Both concepts, especially in the FRBR, are meant to fully describe a uniquely produced object. However, the written *Witness* to an author’s *Text* does not always share the same boundaries as an *Item* in the archive.

We need an intervening entity to connect the multi-volume *Witness* to the five physical documents that substantiate it. Moreover, that intervening entity needs to have a discrete beginning and an end, representing one continuous set of leafs or pages in the *Archival Item*. Let us call the intervening entity *Pages*, by which we mean one continuous set of pages or leafs in a physical document. Figure 2.1 illustrates how we would model the multi-volume *Witness* of Aubert’s *Text*, taking into account the fact that the beginning and end of an *Item* in the archive is not always synchronous with the beginning and end of a *Witness*.

It is necessary to introduce the intervening *Pages* entity between the *Witness* and the *Item*. The reason is that an archival *Item* may contain more than one *Witness*. This happens to not be the case with *Renaut de Montauban*. Each of the *Items* contains nothing but its part of the *Witness*. However, our second test case does involve manuscripts that transmit *Witnesses* of more than one *Work*, which will demonstrate the need for further development of the preexisting models.

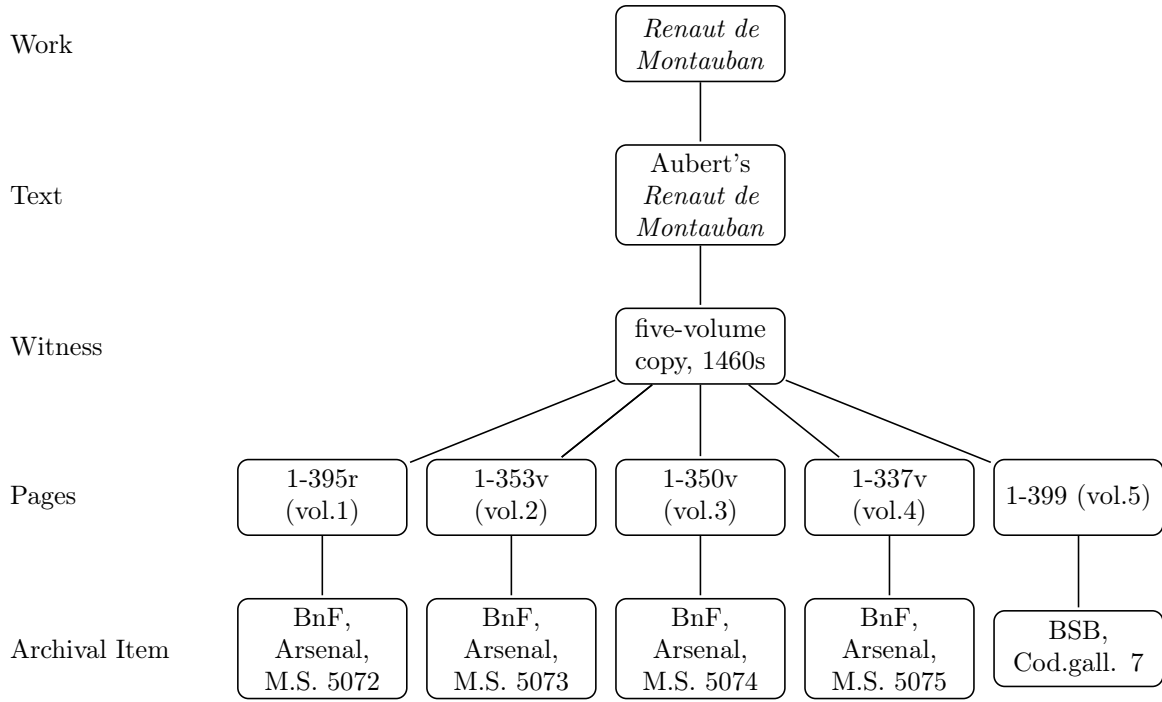
## 2.2 Chanson d’Aspremont and the lost manuscript

In April 1586, a notary named François Daunys certified who owned and leased what properties around the neighboring villages of Le Mazet, Fournels, and La Vachellerie, in what is today the Lozère *département* of France.<sup>14</sup> In the years that followed, notaries and others in the region kept track of the legal agreement, which eventually made its way into the Archives départementales de la Lozère. Around 1883, archivist Ferdinand André noticed the cover protecting the 1586 contract was an old bifolio, a piece of parchment folded in half, that had been cut from the quire of a manuscript. André then observed that each half of the bifolio contained part of a *chanson de geste*, which he presumed dated from the thirteenth-century. André communicated his discovery to his colleagues and the fragments, being separated from the sixteenth-century legal document, were sent to the Bibliothèque nationale de France (BNF), where they are conserved today in a nineteenth-century collection of Medieval fragments.

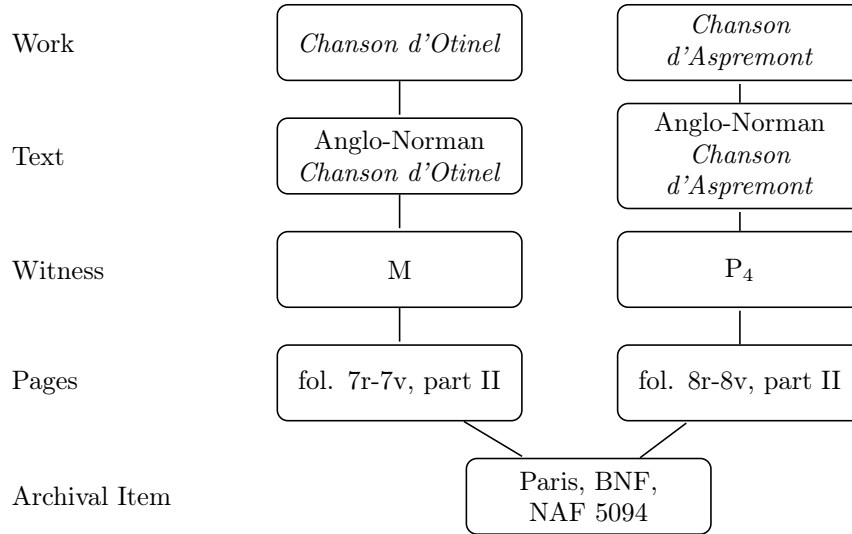
The folded parchment that André discovered makes up the seventh and eighth folios of BNF, nouvelles acquisitions français (NAF) 5094. Each side’s contents are not of the same *chanson*, though they do come from the same original manuscript. Thus, in our adapted version of the FRBR data model, informed by Sahle’s *Textrad*, the composite BNF

<sup>13</sup>G. Augustine. “Le cycle de Renaut de Montauban, un exemple de l’écriture cyclique au Moyen Âge”. 2020ORLE3174. PhD thesis. 2020. URL: <http://www.theses.fr/2020ORLE3174/document>.

<sup>14</sup>J.-B. Camps. “La Chanson d’Otinél. Édition complète du corpus manuscrit et prolegomènes à l’édition critique”. 2016PA040173. PhD thesis. 2016. URL: <http://www.theses.fr/2016PA040173>, p. cxiii.



NAF 5094 manuscript would upwardly relate to two *Pages* entities, one which begins on 7r and ends on 7v, and the other which begins on 8r and ends on 8v. Each *Pages* entity would then relate to its own *Witness*, because each folio is the fragment of a different *Text* of a different *Work*. Figure 2.2 illustrates these relationships.



### 2.2.1 The Attested Document

There is a glaring problem with the modeling in Figure 2.2. As it stands, The provisional model suggests that everything in BNF NAF 5094 was transmitted together, the *Chanson d'Otinél*, the *Chanson d'Aspremont*, and a fragment of the *Roman de la Rose* which follows on folios 9 and 10. We can prevent such an incorrect inference by noting the *Archival Item* is a “false” manuscript [*factice*], meaning it consists of writings that have been belatedly compiled together. However, preventing a false positive still does not solve the problem because we want to also know that the biofolio now on the seventh and eighth pages of BNF NAF 5094 was part of one manuscript. We need another entity, the *Attested Document*.

While the *Archival Item* links *Witnesses* M and P<sub>4</sub> by virtue of their current conservation status, the *Attested Document* unites them by virtue of scholarly argumentation that claims they were initially transmitted together. More detail about our test case of the *Chanson d'Aspremont* reveals why an *Attested Document* entity is so crucial. In the 1880s, while François André was sharing his discovery about the *Chanson d'Otinél* and *Chanson d'Aspremont* fragments in the Archives départementales de la Lozère, archivist Paul Meyer noted that the archives of a nearby French *département*,

the Puy-de-Dôme, also had a fragment of the *Chanson d'Aspremont* that was being used to cover some records in their collection. Unlike the Lozère archives, the Puy-de-Dôme archives did not give the Bibliothèque nationale de France their fragment of the *Chanson d'Aspremont*. Today, one can consult it at the Archives départementales de Clermont-Ferrand under the shelfmark I F2.

The updated provisional data model illustrated in Figure 2.3 allows us to recognize the attested coexistence of the three fragments in a now-lost manuscript. Because the missing document is not conserved as such anywhere, it lacks a clear identifier like the *Archival Items* that have shelfmarks. To compensate, we generate a name that concatenates the *Attested Document's* alleged parts. The known copies of the two *chansons de geste*, *Witnesses* M and P<sub>4</sub>, bypass their *Pages* entities and meet one another directly in the *Attested Document*.

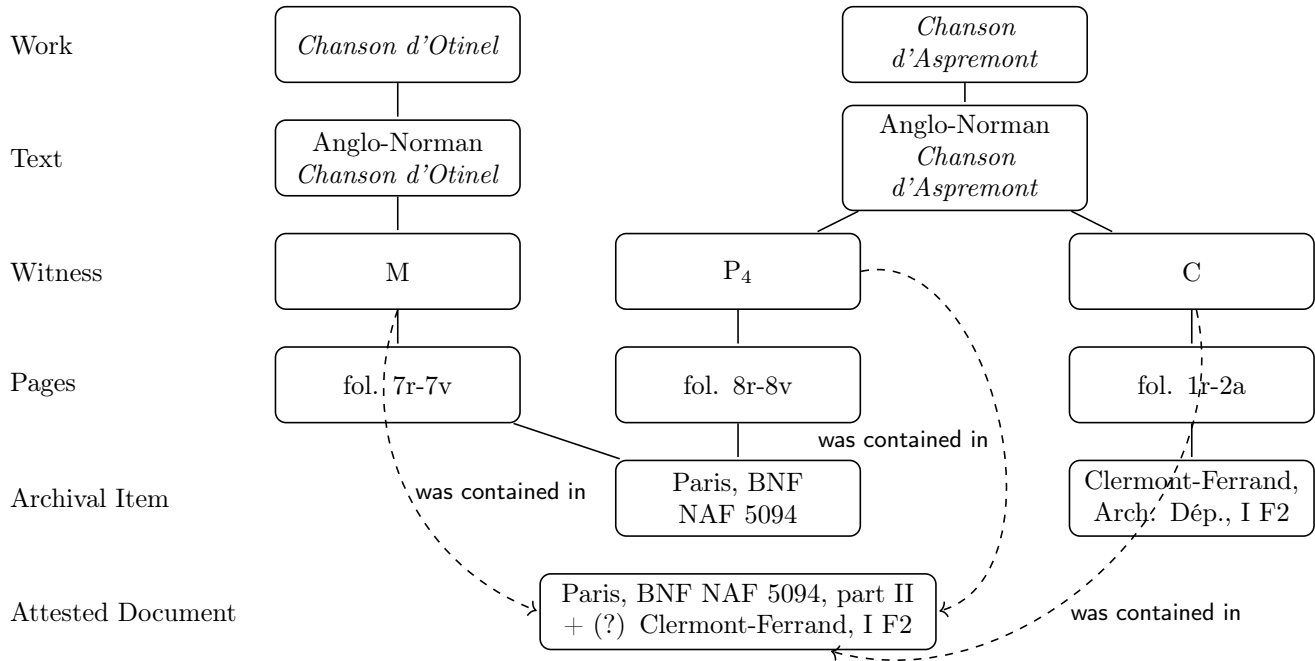


Figure 2.3: Provisionary model of *chansons de geste* in BNF, NAF 5094 and Clermont-Ferrand I F2.

The model shown in Figure 2.3 perpetuates sigla (P<sub>4</sub>, C) that philologists have historically given to each fragment and communicates that both *Witnesses* were likely transmitted together in the same manuscript (the *Attested Document*). It does not, however, document the hypothesis that P<sub>4</sub> and C are in fact two parts of the same *Witness*. While it is good to demonstrate that a manuscript contained copies of both the *Chanson d'Otinel* and the *Chanson d'Aspremont*, this is not sufficient. We do not want to lose the related theory that the verses of the Anglo-Norman *Chanson d'Aspremont*, which are preserved today in two different fragments, are in fact different parts of one copy. Put another way, the model as is risks suggesting that the *Attested Document* transmitted one copy of the Anglo-Norman *Chanson d'Otinel* and two copies of the Anglo-Norman *Chanson d'Aspremont*. The evidence does not suggest this latter claim.<sup>15</sup>

## 2.2.2 Inter-*Witness* relationships

The P<sub>4</sub> *Witness* in BNF NAF 5094 does not start at the beginning of the *Text*, meaning the Anglo-Norman version of the *Chanson d'Aspremont*. As Jean-Baptiste Camps writes in his dissertation, the Paris fragment is likely missing about 85 verses of the beginning, after which it presents 395 verses of the *Text*.<sup>16</sup> The Clermont-Ferrand (C) fragment presents the end of the *Text*; it would total 384 verses if not for edges of its pages being cut. In putting the two fragments together, they miss about 85 verses at the beginning and 9606 in the middle. Their content does not overlap and, based on handwriting and language, there is reason to believe they were produced as one copy (*Witness*) of the Anglo-Norman *Chanson d'Aspremont*.

Why not create another entity, the *Attested Witness*, like the *Attested Document*?<sup>18</sup> The latter is justified because the *Attested Document* is ontologically different than the *Archival Item*. The *Archival Item* is a physical object in the world, which can be photographed and damaged. The *Attested Document* is an historical claim asserting that a text object was once produced and existed in the world, but has not been conserved as such. An *Attested Witness* and a regular *Witness* are too ontologically similar; they are both philological claims that the sequence of characters inscribed on the pages of a

<sup>15</sup>Some scholars do argue that, originally, the manuscript for which *Witness* P<sub>4</sub> was written was at some point damaged and it lost some of the leaves that presented the end of the *Chanson d'Aspremont*, making the original *Witness* into a fragment. Thus, the *Witnesses* P<sub>4</sub> and C are not of the same original document because *Witness* C was written after P<sub>4</sub> to replace some of the lost leaves. However, this theory, while interesting for the study of the *Chanson d'Aspremont*, does not undermine the argument for why we need an *Attested Document* entity because it still posits the existence of a lost manuscript that compiled the *Witnesses* together.

<sup>16</sup>Camps, “La Chanson d'Otinel. Édition complète du corpus manuscrit et prolégomènes à l'édition critique”, p. xcvi.

<sup>18</sup>An example of such an entity is suggested in the first row of Table 2.3, bearing the name “P<sub>4</sub> + C.”

<i>Witness</i>	initial lacuna	$P_4$ text	middle lacuna	C text
$P_4 + C$		395		377 [384]
Ch	84	415	9606	388

Table 2.3: Reproduction of the comparison between the number of verses in the complete *Witness* Ch and the fragments  $P_4$  and C, originally published in the dissertation of Jean-Baptiste Camps<sup>17</sup>

document intentionally represents the linguistic content of a *Text*, or of a text-as-linguistic-content as Sahle puts it. The difference between the attested *Witness* “ $P_4 + C$ ” and the *Witnesses*  $P_4$  and C is not categorical. Both are essentially historical claims. It does not justify the creation of a new entity.

A more efficient way to make legible in the data model the theoretical *Witness* “ $P_4 + C$ ” is to build that information into attributes of the *Witnesses* that constitute it. The provisional model in Figure 2.4 shows how a *Witness* that is a fragment can relate to another *Witness* fragment via the attribute “is preceded by fragment.” In the *Chanson d’Aspremont* case, the *Witness* C would refer to  $P_4$  in its field “is preceded by fragment.” *Witness*  $P_4$  would not have any value assigned to the attribute “is preceded by fragment” because it is the root of a sequence of *Witness* fragments.<sup>19</sup>

However, we also know that *Witness* M of the *Chanson d’Otinél* was positioned before the *Chanson d’Aspremont* in the *Attested Document*. To show this, we link the root fragment *Witness*  $P_4$  to the *Witness* that precedes it, M, with the attribute “is preceded by witness.” The relationships between *Witnesses* as established by scribes consulting different manuscripts as models when composing a new version of a *Text*, the subject of stemmatology, are not the focus of these inter-*Witness* relationships. As discussed in Section 3.3, the proposed relational framework is focused on curating and enriching data about the literary content and archival evidence of the corpus. How that evidence, the *Witnesses*, were composed is best modeled in a graph framework.

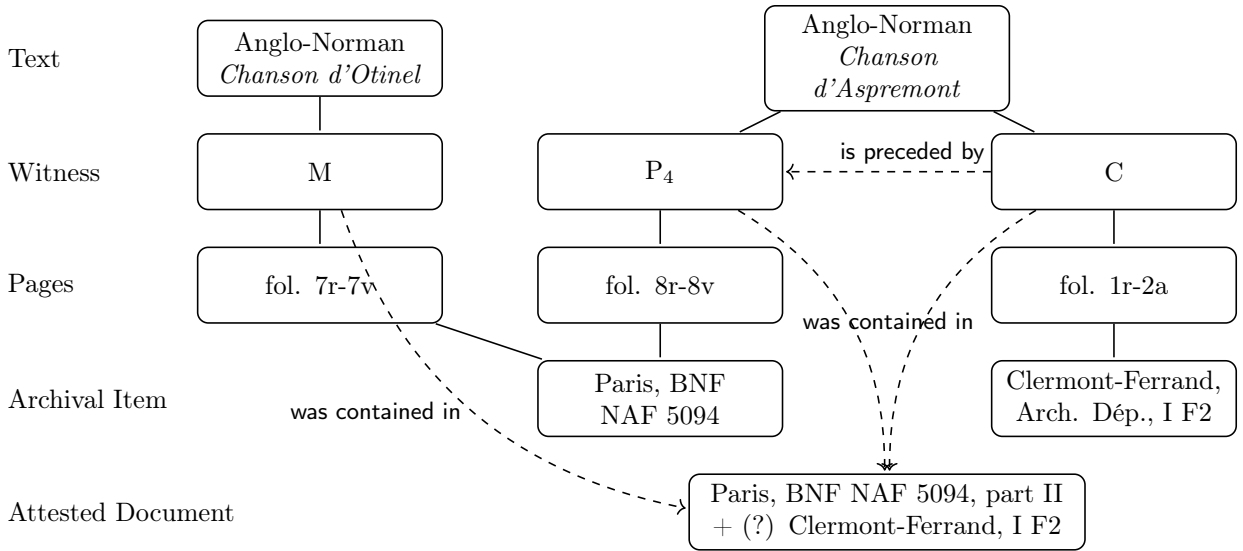


Figure 2.4: Relationships between *Witnesses* and neighboring entities in provisional model.

### 3 Our Model

Having explored existing data models and tested their limitations against several case studies from the Middle Ages, we propose a model with the entities and relationships represented in Figure 3.1. To begin, we present all the entities we believe necessary for a data model tasked with organizing and delivering information about texts and traditions in Medieval literature, particularly texts about chivalric tales and legends. Then, we list all the attributes each entity features. Compared to the former, this latter section is more likely to need further modifications based on the idiosyncratic qualities of different literary corpora.

#### 3.1 Intertextuality

On the top level of Figure 3.1, running from left to right, are the three abstract entities, *Cycle*, *Work*, and *Text*, that manage information about intertextuality in our corpus. *Works* and *Cycles* can be nested together within the scope of

<sup>19</sup>The fact that the  $P_4$  *Witness* would not have any value in its “is preceded by fragment” attribute, yet it has the status of “fragment,” could mean one of two things: (a) it is the only known fragment of that *Witness*, or (b) that it is the root in a sequence of fragments that constitute one attested *Witness*. To determine which is the case, one would need to identify roots of sequences by grouping all the *Witnesses* based on other fragment *Witnesses* referenced in the attribute “is preceded by.”



a *Cycle*, based on the *Works*’ narrative content; both entities have the attribute “is part of,” which points to a *Cycle* entity. *Works* can also be modeled on other works, as in the case of a new *Work* compiling and reworking the episodes and characters from two or more pre-existing *Works*. The data model also registers intertextuality amongst *Texts*, which can be modeled on one another as translations, prosifications, abbreviations, elaborations, versifications, or other forms of adapting the expression (*Text*) of a common *Work*. The nature of a *Text*’s relationship to a model *Text* can be inferred through differences and similarities in the two entities’ attributes, such as their language and literary form.

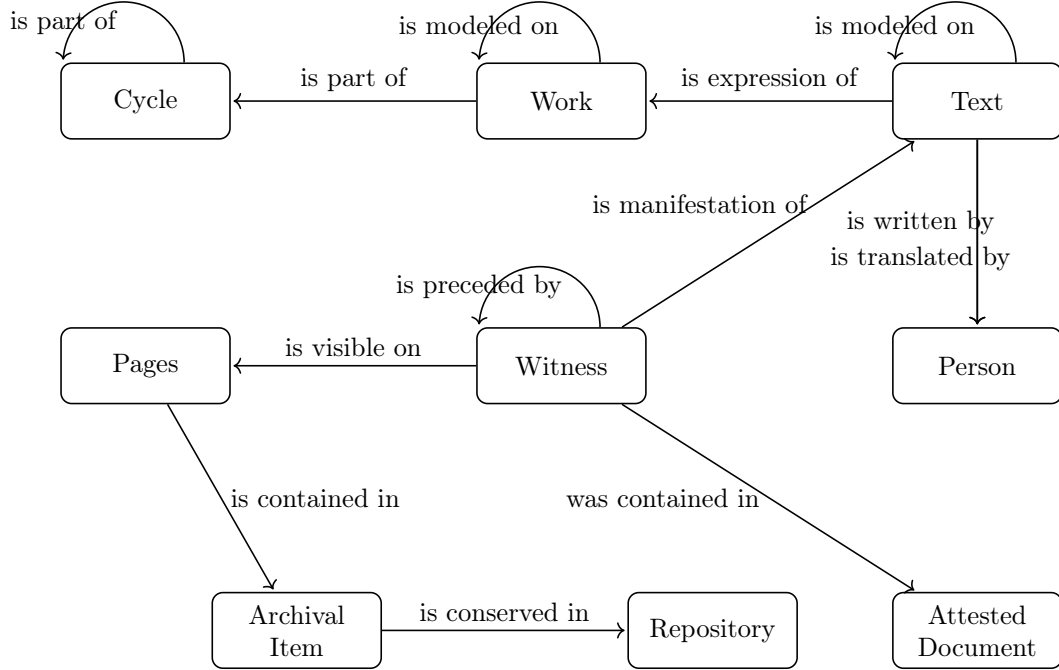


Figure 3.1: Proposed Entities.

### 3.2 Archival Evidence

Stemming from an extant *Witness* are potentially four types of relationships to other entities. First and foremost, a *Witness* must relate to one and only one *Text*. By definition, a *Witness* is an extant version of a *Text*’s linguistic content, spelled out in a sequence of characters, as demonstrated in Table 2.2. Therefore, the *Witness* must be the manifestation of a *Text*. Second, the *Witness*, by virtue of being extant, must be visible on the *Pages* of an *Archival Item*. When the version of a *Text* has survived through fragments, one of the fragmented version’s *Witnesses* can relate to another fragment through the attribute “is preceded by fragment.” An example of this inter-*Witness* relationship is demonstrated in the case of the *Chanson d’Aspremont* and the Table 2.3. Finally, an extant *Witness* can have formerly been contained in a document that has not survived but to whose historical existence scholars attest based on philological and codicological evidence.

*Pages* represents an uninterrupted set of folios and, when digitized, images of an *Archival Item* on which the text content of a *Witness* can be read. In our data model, a *Pages* entity, which depends on an extant *Witness*, must be contained in an *Archival Item*. The *Archival Item* can relate to multiple *Pages* entities. They represent a unique set of leafs or pages in an *Archival Item*, and they present the content of only one *Witness*. Lastly, the *Archival Item*, being an object one can consult and which continues to persist, contrary to the *Attested Document*, must be conserved in a *Repository*.

### 3.3 Genealogy of a *Text*

While the proposed data model manages metadata about intertextuality within the corpus, its relational framework does not register how *Witnesses* derive from one another. This choice reflects an assumption about intertextuality, meaning *Works*’ and *Texts*’ models, and stemma, meaning *Witness*’ models. On the one hand, we assume a relatively consistent consensus has been reached on the models of *Works* and *Texts*. We presume scholars and archivists have largely accepted as historical fact the assertion that one *Text* is a translation of another *Text*, that a *Work* is a compilation of other *Works*, that one *Text* is the prose version of another *Text*, and so on. Through analyses of the content and literary form of extant *Witnesses* that are examples of *Texts* and *Works*, such claims of genealogy and intertextuality tend not to attract as much skepticism as claims scholars posit about which scribe derived their *Witness* (Sahle’s text-as-version) from the example of an older *Witness*.

Such stemmatological claims are crucial, and we want to model them for the corpus. However, in addition to establishing a relationship between two *Witness* entities, it is also critical to cite the source of that attested dependence. Furthermore, we want to permit conflicting stemma without over complicating the model’s relationships. While it is

possible to model such connections in a relational framework, exploiting those connections to respond to users' queries risks pushing the model to become over complex. The choice becomes one between exploiting one over-complicated model or harmonizing and jointly exploiting two simpler models.

Rather than reinvent the wheel and model stemma in a complex relational framework, we propose pairing the former with a graph framework, which philologists have used for decades.<sup>20</sup> The relational framework's main tasks are, therefore, to generate entities relevant to our corpus, store metadata about them, and establish intertextual relations at the level of the abstract, narrative content, meaning the *Work* and *Text*. In conjunction, the graph database's *Witness* nodes will bear the same identifier as their counterpart in the relational database and thus can be enriched with metadata.

The project *OpenStemmata* has already designed a workflow through which contributors can encode stemma that have been published in a scholarly context or otherwise submitted by scholars. For example, Giovanni Palumbo and Paolo Rinoldi, who, while developing a critical edition of the French *Chanson d'Aspremont*, produced a stemma of the first part of the *Text*. Their stemma was encoded and uploaded to the *OpenStemmata* database, and we have represented it here in Figure 3.2. The graph's gray node  $P_4$  is equivalent to the *Witness*  $P_4$  in our earlier *Chanson d'Aspremont* test case, as seen in Figures 2.2, 2.3, and 2.4.

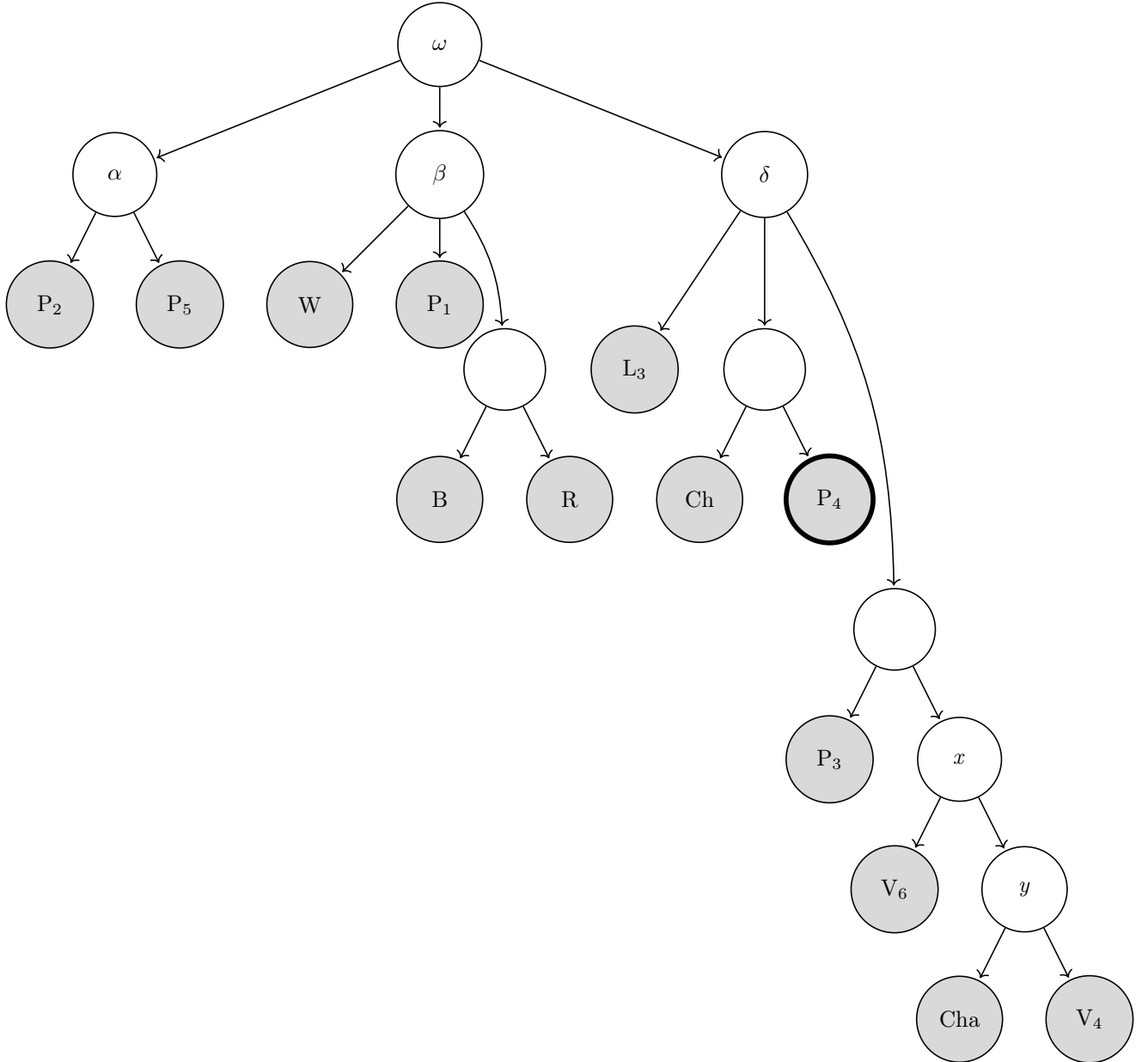


Figure 3.2: Stemma of extant *Witnesses* (gray) and hypothetical intermediaries (white) of the first part of the French *Chanson d'Aspremont* by G. Palumbo and P. Rinoldi.

We propose outsourcing the data model's *Text* genealogies to the *OpenStemmata* project's established workflow, encoding standards, and open-source repository. Subsequently, we reconcile the *Witness* nodes in the *OpenStemmata* database, further valorizing and enriching those publications, with records registered in our proposed relational data

<sup>20</sup>J. van Zundert. "5 Computational methods and tools". In: *History, Methodology, Digital Approaches*. Ed. by P. Roelli. Berlin, Boston: De Gruyter, 2020, pp. 292–356. ISBN: 9783110684384. DOI: doi:10.1515/9783110684384-006. URL: <https://doi.org/10.1515/9783110684384-006>.

model. Finally, having linked the two record types, such as the *Witness* P<sub>4</sub> of the Anglo-Norman *Chanson d’Aspremont*, the data model can deliver thoroughly enriched and linked records based on users’ requests.

### 3.4 Reconciling and citing records

Finally, we propose a *Reference* table (see Figure 3.3), which associates certain entities with bibliographic resources. This last addition serves two purposes. First, it associates an entity with a unique identifier in an authoritative aggregator, such as WikiData, Biblissima, and VIAF (Virtual International Authority File). This association helps reconcile potentially duplicate records in the data model and helps reconcile records between the relational data model and an *OpenStemmata* graph. Second, *Reference* associates entities with scholarly citations. In addition to making the data model interoperable with other databases, including VIAF, the *Reference* entity also allows us to enrich records with a scholarly bibliography.

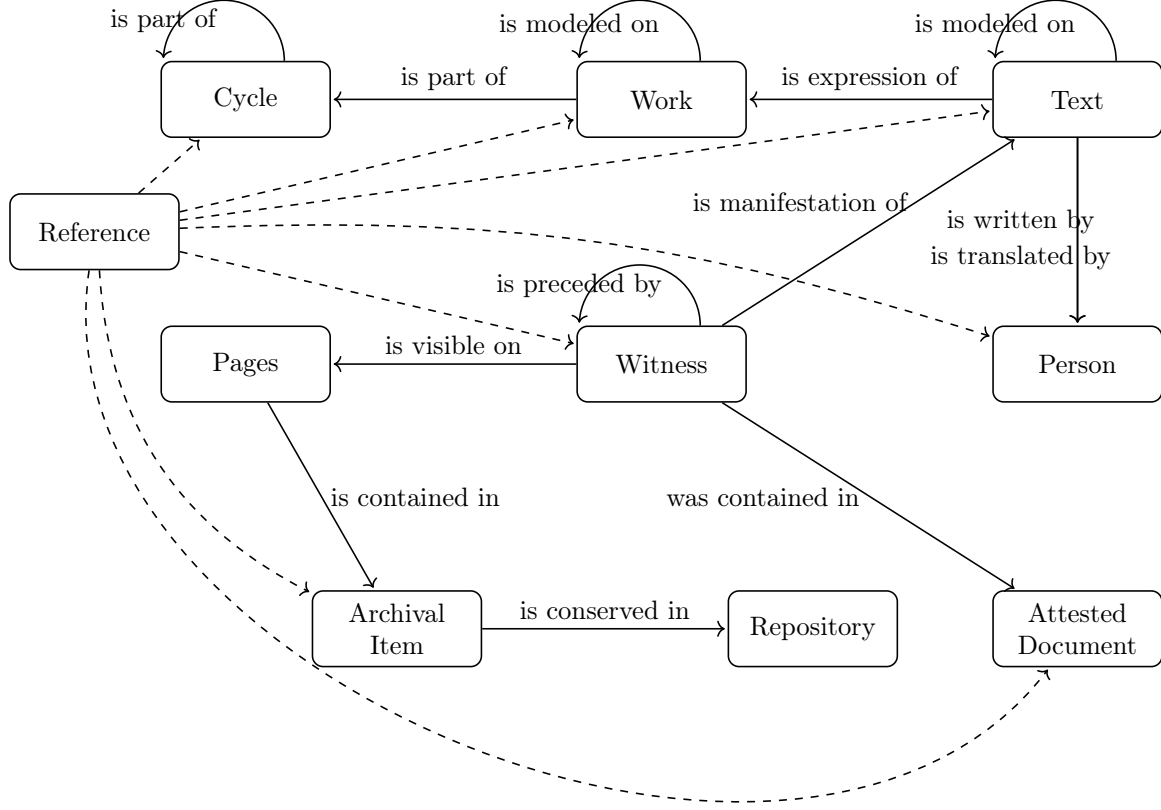


Figure 3.3: Proposed Entities with *Reference* table.

*Reference* can reconcile conflicting *Cycle*, *Work*, *Text*, *Person*, and *Archival Item* entities and link them to universal unique identifiers. The *Archival Item* entity should ideally have an Archival Resource Key provided by its *Repository*, which should help avoid duplicate records of the same manuscript. The entities *Cycle*, *Work*, *Person*, and *Text* overlap with entities in the VIAF database; our model’s *Person* is equivalent to VIAF’s *Personal Names* and our model’s *Cycle*, *Work*, *Text* records can find an equivalent under the VIAF’s *Work* and *Expression* records. The latter two entities in the VIAF model are borrowed from the FRBR.

For example, the *Work* *Renaud de Montauban*, seen in Table 3.4b, has the VIAF identifier 174185484, as seen in the last row of Table 3.4c. By linking a *Work* record to a *Reference* record, which in this case associates *Renaud de Montauban* with an identifier in the VIAF database, we can enrich the *Work* with all the linked data available in VIAF. Furthermore, while our record for *Renaud de Montauban* has one title, its link to the VIAF database via the *Reference* entity associates the *Work* with other names by which people might identify it, including *Renaud de Montauban*, *Reinolt von Montelban*, and *Quatre fils Aymon*, further improving the data’s interoperability and resilience against duplication.

ID	title	is part of
1	Renaut de Montauban	

(a) *Cycle* record for *Renaut de Montauban*.

ID	title	is part of	is modeled on
2	Renaut de Montauban	1	

(b) *Work* record for *Renaut de Montauban*.

entity type	entity ID	unique identifier	identifier source	permalink	citation
Cycle	1				G. Augustine. “Le cycle de Renaut de Montauban, un exemple de l’écriture cyclique au Moyen Âge”. 2020ORLE3174. PhD thesis. 2020. URL: <a href="http://www.theses.fr/2020ORLE3174/document">http://www.theses.fr/2020ORLE3174/document</a>
Cycle	1	5045	Arlima	<a href="https://arlima.net/no/5045">https://arlima.net/no/5045</a>	
Cycle	1	Q59212800	WikiData	<a href="https://www.wikidata.org/wiki/Q59212800">https://www.wikidata.org/wiki/Q59212800</a>	
Work	2	318	Arlima	<a href="https://arlima.net/no/318">https://arlima.net/no/318</a>	
Work	2	Q115962675	WikiData	<a href="https://www.wikidata.org/wiki/Q115962675">https://www.wikidata.org/wiki/Q115962675</a>	
Work	2	174185484	VIAF	<a href="http://viaf.org/viaf/174185484">http://viaf.org/viaf/174185484</a>	

(c) *Reference* records for *Renaut de Montauban*.

## 4 Entities and Attributes

### 4.1 Cycle

Definition: General theme that a group of *Works* can share.

Attributes:

- **title** (text, req., uniq.): Received name of the *Cycle*, either in the language of the first known *Text* to treat the matter or in the language most used in scholarship.
- **is part of** (foreign key, **Cycle**, opt., uniq.): The meta-*Cycle*, of which the *Cycle* is a part.

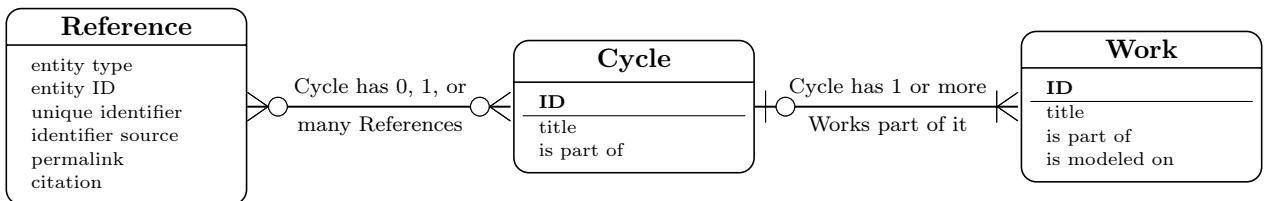


Figure 4.1: *Cycle* entity relationships.

### 4.2 Work

Definition: Content of a story, which has a recognizable structure and can be recounted in different ways while remaining the same story.

Attributes:

- **title** (text, req., uniq.): Received name of the *Work*, either in the language of the first known *Text* to treat the matter or in the language most used in scholarship.
- **is part of** (foreign key, **Cycle**, opt., uniq.): The *Cycle*, of which the *Work* is a part.
- **is modeled on** (foreign key, **Work**, opt., repeat.): If a reworking of an anterior *Work*, the *Work* on which it is modeled.

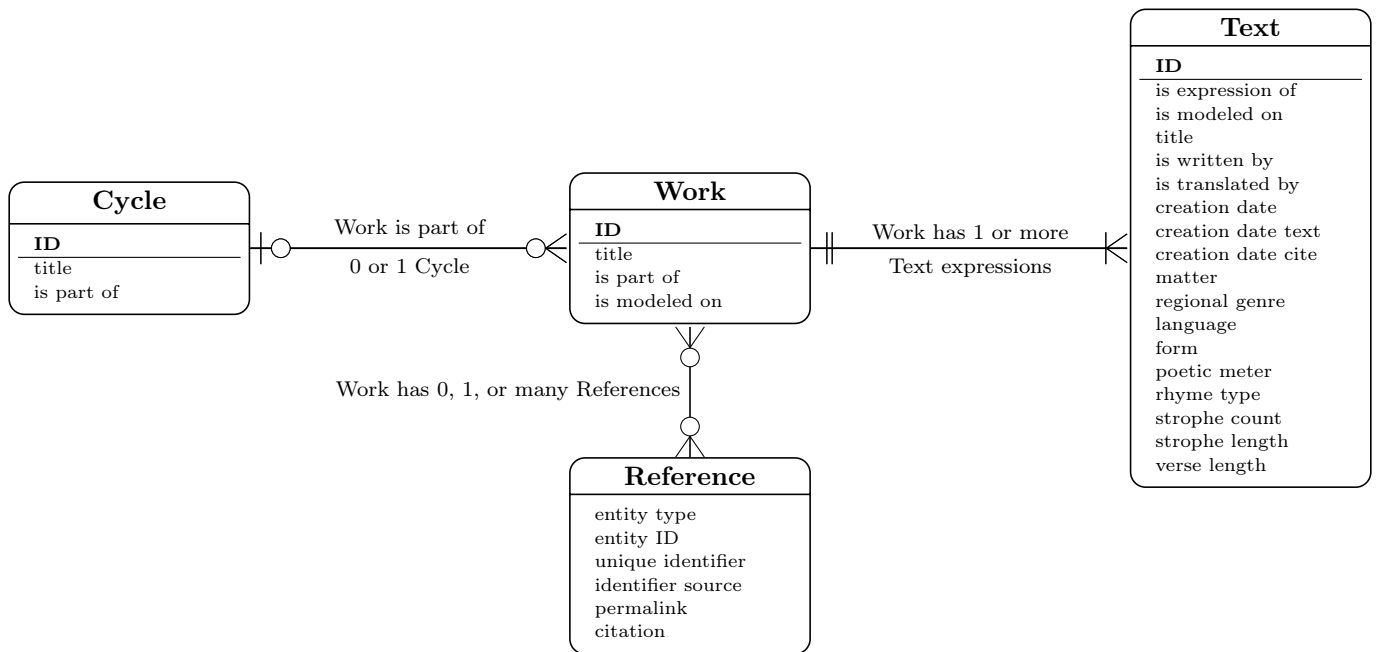


Figure 4.2: *Work* entity relationships.

### 4.3 Text

Definition: Formulation of a *Work* in human language, whose literary form and style can be detected and whose creation can be attributed to one or more individuals.

Attributes:

- **is expression of** (foreign key [**Work**], req., uniq.): The *Work* that the *Text* articulates.
- **is modeled on** (foreign key, [**Text**], opt., uniq.): If the *Text* is derived from another *Text*, a reference to the model *Text*.
- **title** (text, req., uniq.): Either the given title of the *Text*, as provided by the creator, or the standardized title most used in scholarship to refer to the *Text*.
- **is written by** (foreign key [**Person**], opt., repeat.): The individual accredited with composing the *Text*.
- **is translated by** (foreign key [**Person**], opt., repeat.): When the *Text* is a translation of another *Text*, the individual accredited with creating the translation.
- **creation date** (list[date], opt., uniq.): A list of two or one dates; the first date is either the earliest or the only date associated with the *Text*'s creation, and, in the case of a range, the second date is the latest date associated with the creation.
- **creation date text** (text, opt., uniq.): The date associated with the *Text*'s creation as it is written in a scholarly source.
- **creation date cite** (text, opt., uniq.): A citation of the source that provided the date of creation.
- **matter** (terms, opt., repeat.): The matter treated in the *Text*, as defined by Jean Bodel.
  - **Britain**: Matter of Britain, which includes stories about Tristan and King Arthur.
  - **France**: Matter of France, which includes stories about Charlemagne.
  - **Rome**: Matter of Rome, which includes stories about Troy, Rome, Alexander, and antiquity.
- **regional genre** (terms, req., uniq.): Literary genre attributed to the *Text*.

- Relevant to French tradition
  - \* `chanson de geste`
  - \* `roman`
- Relevant to Iberian tradition
  - \* `romancero`
  - \* `novela`
  - \* `cantar de gesta`
  - \* `mester de juglaria`
  - \* `mester de clerecia`
  - \* [needs development]
- Relevant to Italian tradition
  - \* `cantare`
  - \* `poema cavalleresche`
  - \* [needs development]
- Relevant to Islandic tradition
  - \* `riddarasögur`
  - \* `riddararímur`
  - \* [needs development]
- Relevant to Middle Dutch tradition
  - \* `ridderepiek`
  - \* `ridderroman`
  - \* `rijmkronieken`
  - \* [needs development]
- Relevant to Middle English, Middle Irish, Middle Welsh traditions
  - \* `romance`
  - \* [needs development]
- Relevant to Middle High German tradition
  - \* `Roman`
  - \* [needs development]

• **language** (terms, req., uniq.): ISO code of the primary language through which the *Text* expresses the *Work*.

- `cat`: Catalan
- `dum`: Middle Dutch
- `enm`: Middle English (1100-1500)
- `frm`: Middle French (ca. 1400-1600)
- `fro`: Old French (842-ca. 1400)
- `fro_ENG`: Anglo-French
- `fro_ITA`: Franco-Italian
- `fro_PRO`: Franco-Occitan
- `ghg`: Hiberno-Scottish Gaelic, Early Modern Irish
- `glg`: Galician
- `glg_POR`: Galician-Portugese
- `gmh`: Middle High German (ca. 1050-1500)
- `gml`: Middle Low German
- `isl`: Islandic
- `ita`: Italian
- `mga`: Middle Irish (900-1200)
- `non`: Old Norse
- `non_DAN`: Old East Norse, Old Danish (800-1100)
- `non_SWE`: Old Swedish (800-1500)
- `oco`: Old Cornish
- `por`: Portugese
- `pro`: Old Occitan, Old Provençal (to 1500)
- `spa`: Spanish or Castilian
- `wlm`: Middle Welsh

• **form** (terms, req., uniq.): Whether the *Text* is formed in prose, verse, or a mix of both.

- `prose`
- `verse`
- `mixed`

• **poetic meter** (terms, opt., uniq.): If the *Text* is in verse, the type of poetic meter.

- **French alexandrine**: line consisting of 2 half-lines each of 6 syllables, total of 12 syllables.
- **dodecasyllabe**: line consisting of 12 syllables.
- **decasyllabe**: line consisting of 10 syllables.
- **octosyllabe**: line consisting of 8 syllables.
- **hexasyllabe**: line consisting of 6 syllables.
- **pentasyllabe**: line consisting of 5 syllables.

• **rhyme type** (terms, opt., repeat.): If the *Text* is in verse, the types of rhyme used.

- **alliteration**: Rhyme is allowed between words that either start with the same consonant sound or with the same vowel, such as “*With floures fele, fair under fete*” in Middle English.<sup>21</sup>
- **assonance**: Rhyme is allowed between words that have a repeated vowel sound, such as “a” in “*Vio puertas abiertas e uços sin cañados / alcandaras vazias sin pielles e sin mantos*” in Castilian.<sup>22</sup>

<sup>21</sup>S. Davis. “Guillaume de Machaut, Chaucer’s “Book of the Duchess”, and the Chaucer Tradition”. In: *The Chaucer Review* 36.4 (2002), pp. 391–405. ISSN: 00092002, 15284204. URL: <http://www.jstor.org/stable/25096182> (visited on 06/20/2024), p. 396.

<sup>22</sup>J. Gornall. “Assonance in the Hispanic “Romance”: Precept and Practice”. In: *The Modern Language Review* 90.2 (1995), pp. 363–369. ISSN: 00267937, 22224319. URL: <http://www.jstor.org/stable/3734546> (visited on 06/20/2024), p. 364.

- **end-rhyme**: Rhyme is allowed between words that have an ending that sounds the same, such as “*ihesu guz son ihesu goþe / bløt mit hiærta mæþ þino bloþe*” in Old Norse.<sup>23</sup>
- **generic**: “[R]hyme is allowed between any one member of a phonetic group and is itself or any other member of the same group,” such as ‘b,’ ‘g,’ ‘d’ in Old Irish.<sup>24</sup>
- **strophe count** (integer, opt., uniq.): If the *Text* is in verse and strophic, the number of strophes.
- **strophe length** (integer, opt., uniq.): If the *Text* is in verse and strophic, the number of lines in each strophe.
- **verse length** (integer, opt., uniq.): If the *Text* is in verse, the number of verses a complete version (*Witness*) of it should have.

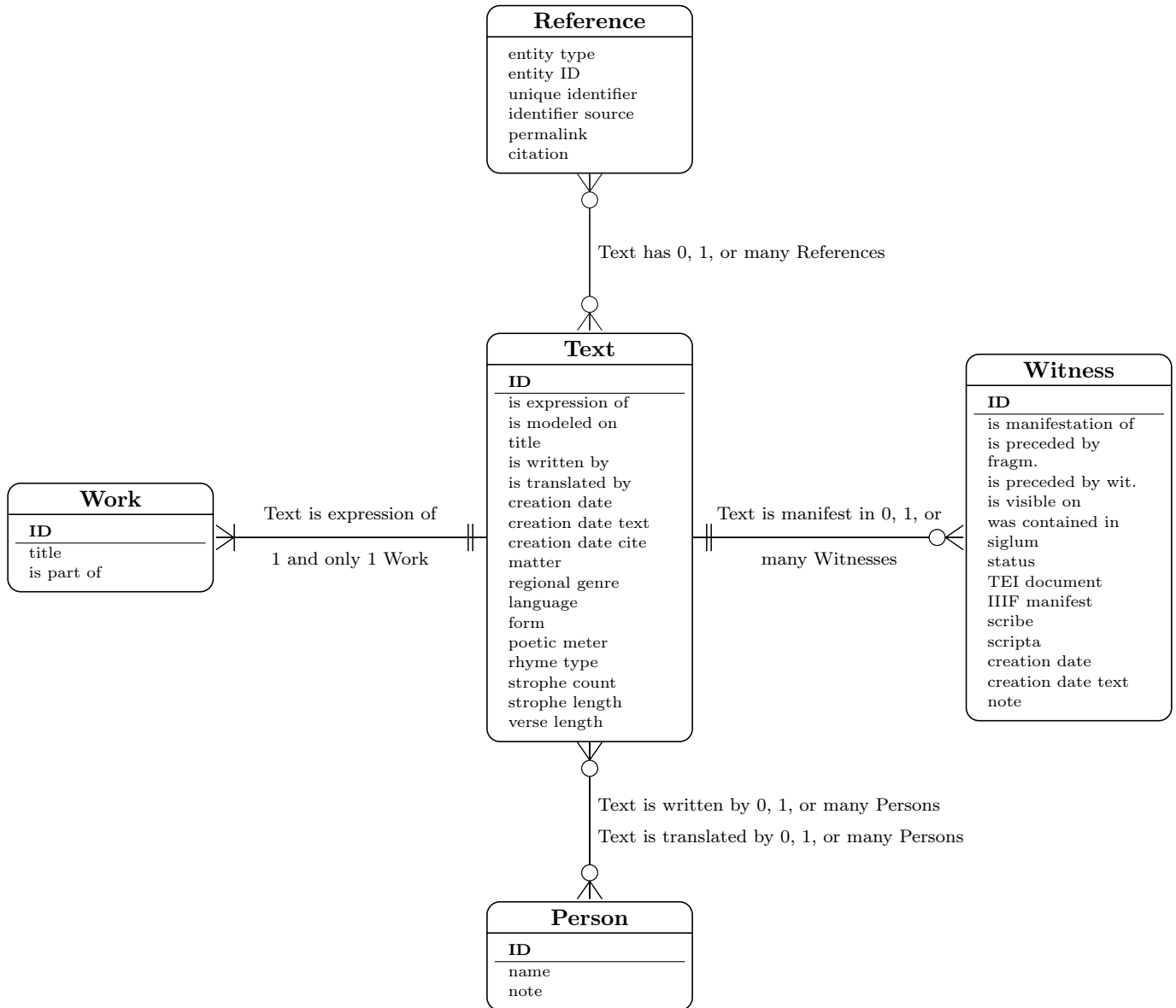


Figure 4.3: *Text* entity relationships.

## 4.4 Person

**Definition:** An individual bearing some responsibility for a *Text* in the data set, either as a creator/author or as a translator, in the case of a *Text* that is the translation of a model *Text* in another language.

**Attributes:**

- **name** (text, req., uniq.): The recieved name of the individual.

<sup>23</sup>W. Layher. “The Big Splash: End-Rhyme and Innovation in Medieval Scandinavian Poetics”. In: *Scandinavian Studies* 80.4 (2008), pp. 407–436. ISSN: 00365637. URL: <http://www.jstor.org/stable/40920821> (visited on 06/20/2024), p. 423.

<sup>24</sup>M. McKie. “The Origins and Early Development of Rhyme in English Verse”. In: *The Modern Language Review* 92.4 (1997), pp. 817–831. ISSN: 00267937, 22224319. URL: <http://www.jstor.org/stable/3734202> (visited on 06/20/2024), p. 822.

- **note** (text, opt., uniq.): Optional notes to help identify the individual.

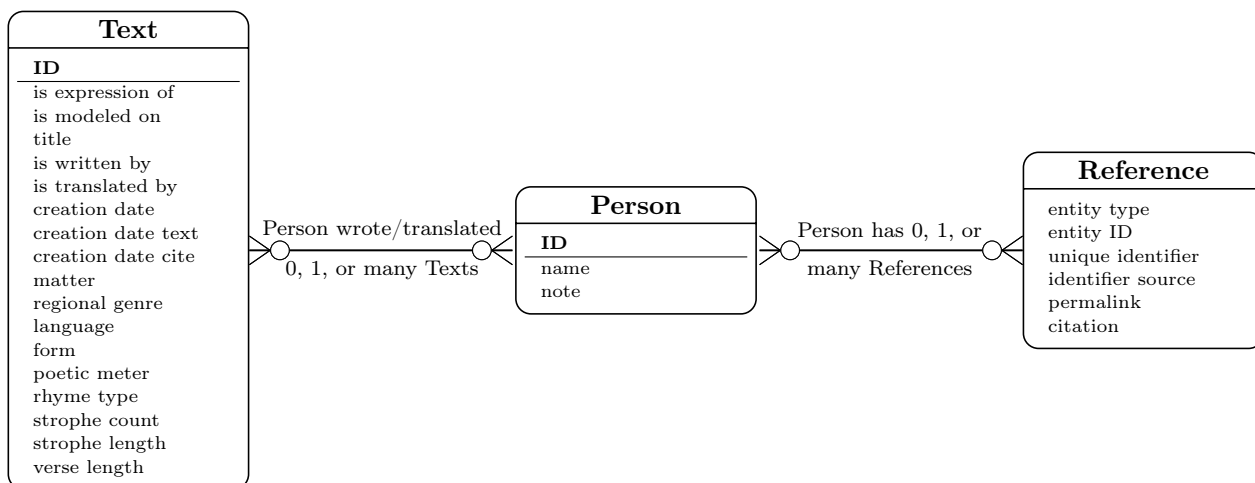


Figure 4.4: *Person* entity relationships.

## 4.5 Witness

Definition: An extant manifestation of a *Text* in a defined sequence of characters, which have been inscribed on a physical document.

Attributes:

- **is manifestation of** (foreign key [**Text**], req., uniq.): The *Text* whose linguistic content the *Witness* manifests in writing.
- **is preceded by fragm.** (foreign key [**Witness**], opt., uniq.): If the *Witness* is a fragment, a reference to another *Witness* fragment, with which the *Witness* is thought to have been produced in an original document (*Attested Document*), and which presents an earlier part of the *Text* than its own part.
- **is preceded by wit.** (foreign key [**Witness**], opt., uniq.): If the *Witness* was contained in an *Attested Document* and is not preceded by another fragment of its same *Text*, a reference to the *Witness* or to the root of a sequence of fragmented *Witnesses* that used to precede it in the *Attested Document*.
- **is visible on** (foreign key [**Pages**], opt., repeat.):
- **was contained in** (foreign key [**Attested Document**], opt., uniq.):
- **siglum** (text, opt., repeat.): Identifier used by scholars to indicate the extant *Witness* within the *Text*'s tradition.
- **status** (terms, req., uniq.): Status of the extant *Witness*.
  - **complete**: All pages containing the main textual content of the *Witness*, excluding dedications and decorations, have survived.
  - **mutilated**: The *Witness* presents parts of its *Text* on a set of surviving pages ( $\geq 3$  pages).
  - **fragment**: Only a few pages of the *Witness* survive ( $\leq 2$  pages).
  - **citation**: The *Witness* testifies to the existence of a *Text*, through citation, but does not present all of the latter's linguistic content.
- **TEI document** (text, opt., uniq.): Reference to a TEI-XML file representing the text content of the *Witness*.
- **IIIF manifest** (text, opt., uniq.): Reference to a IIIF manifest JSON file representing the digitized images of the *Witness*'s *Pages*.
- **scribe** (text, opt., repeat.): When known, identifying information about the scribe alleged to have written the text version.
- **scripta** (terms, opt., uniq.): When known, the name of a regional writing style, similar to the written version of a spoken dialect, as defined by Louis Remacle.<sup>25</sup>

<sup>25</sup>“Je désigne la langue vulgaire écrite au moyen âge par le néologisme *scripta*. L'expression ‘la scripta’ est synonyme de l'allemand ‘die Schriftsprache.’” L. Remacle. *Le Problème de l'ancien wallon*. Presses universitaires de Liège, 1948. URL: <https://doi.org/10.4000/books.pulg.347>, p. 24.



- **bourg.**: Bourguignon
- **champ. mérid.**: Champenois méridional
- **champ. setp.**: Champenois septentrional
- **frc.**: francien (Paris)
- **lorr.**: Lorrain
- **lorr. mérid.**: Lorrain méridional
- **lorr. sept.**: Lorrain septentrional (Metz)
- **norm.**: Normand
- **orl.**: Orléanais
- **ouest**: Ouest
- **pic.**: Picard
- **pic. sept.**: Picard septentrional (Flandres)
- **pic. mérid.**: Picard méridional (Soissonnais)
- **wall.**: Wallon
- **liég.**: Liégeois

- **creation date** (list[date], opt., uniq.): A list of two or one dates; the first date is either the earliest or the only date associated with the *Witness*'s production, and, in the case of a range, the second date is the latest date associated with the production.
- **creation date text** (text, opt., uniq.): The date associated with the *Witness*'s production as it is written in a catalogue or other scholarly source.
- **note** (text, opt., uniq.): Notes about the *Witness* that will not be standardized and used in computational methods but may serve to protect the integrity of the record.

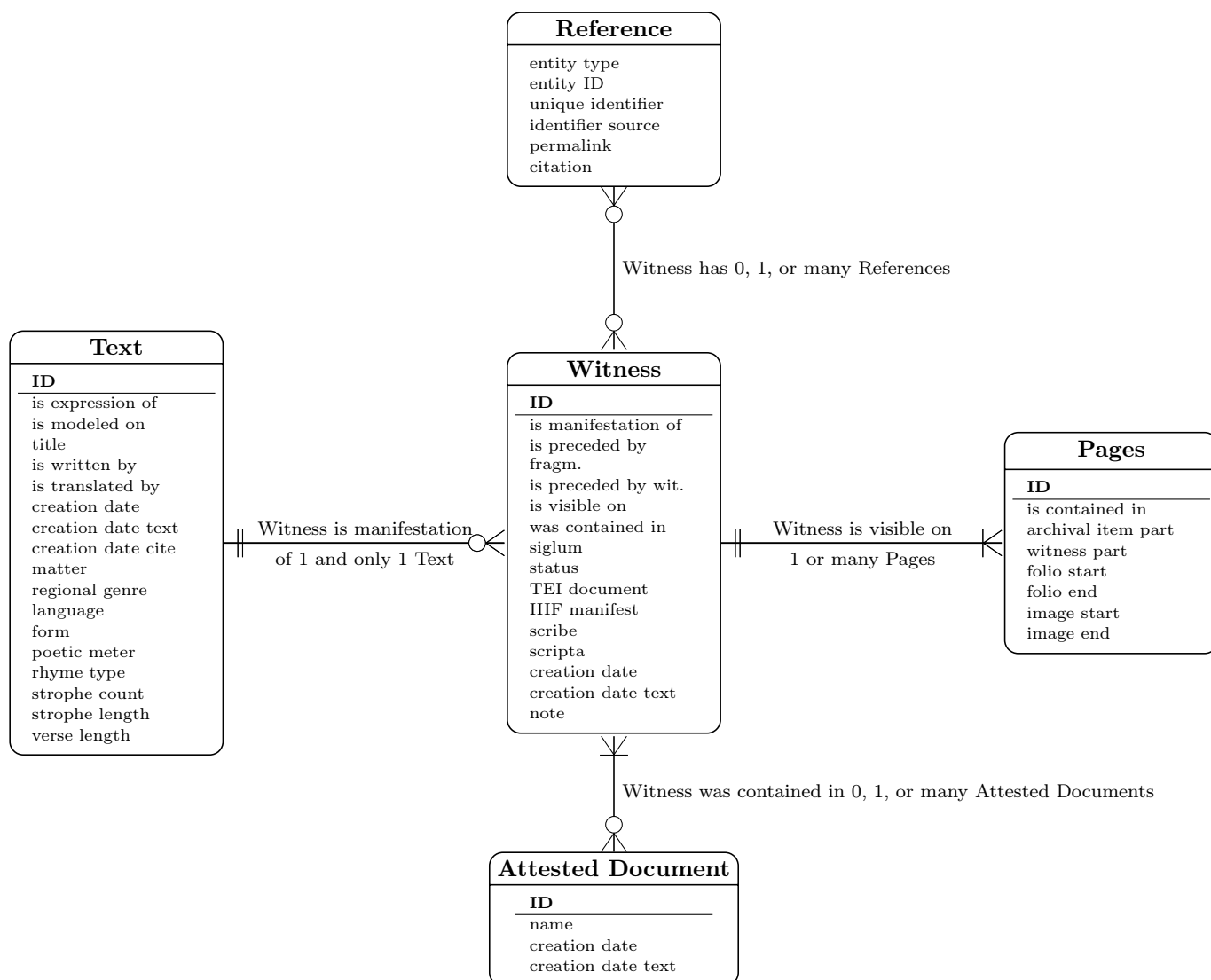


Figure 4.5: *Witness* entity relationships.

## 4.6 Attested Document

Definition: Hypothetical text object, conceived of and produced as one unit, in which *Witness* were transmitted; it no longer exists today.

Attributes:

- **name** (text, req., uniq.): Name given to the *Attested Witness* used to help identify it.
- **creation date** (list[date], opt., uniq.): A list of two or one dates; the first date is either the earliest or the only date associated with the *Attested Witness*'s production, and, in the case of a range, the second date is the latest date associated with the production.
- **creation date text** (text, opt., uniq.): The date associated with the *Attested Witness*'s production as it is written in a catalogue or other scholarly source.

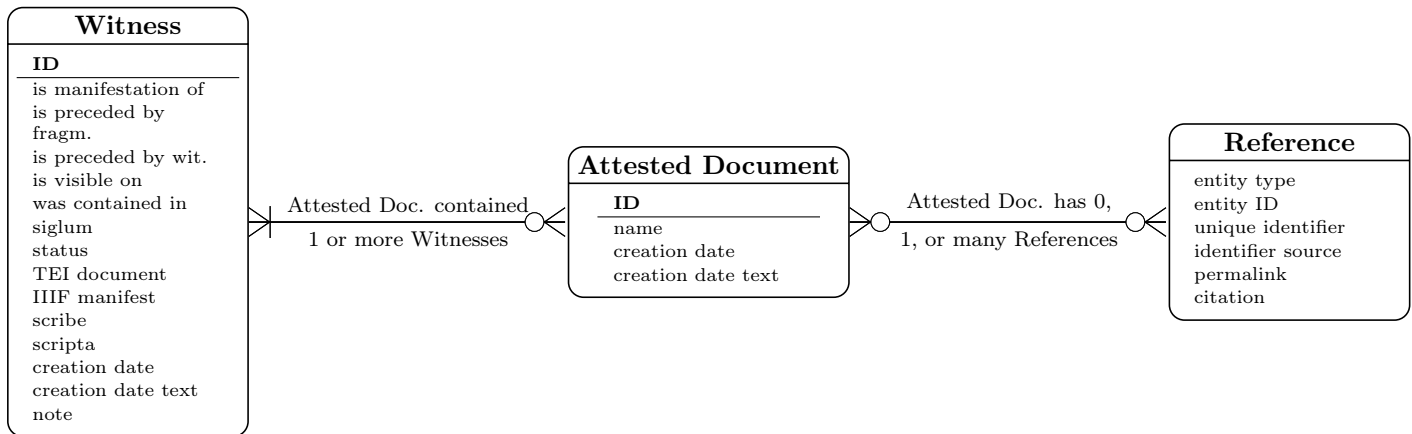


Figure 4.6: *Attested Document* entity relationships.

## 4.7 Pages

Definition: An uninterrupted series of leafs or pages in an *Archival Item* on which is inscribed the *Witness* of a *Text*.

Attributes:

- **is contained in** (foreign key [**Archival Item**], req., uniq.): The physical archival item in which the pages are currently bound or otherwise stored.
- **archival item part** (text, opt., uniq.): If the item's archival description and/or catalogue entry has subdivided the item into parts (i.e. Part II), the part containing the *Pages*; this information is used to enrich bibliographic references.
- **witness part** (integer, opt., uniq.): If the *Witness* is composed of discrete parts (i.e. multiple volumes), the *Pages*' order in the sequence; this information is used to correctly order multiple *Pages* entities descending from a *Witness*.
- **folio start** (text, req., uniq.): The folio number inscribed on the first page.
- **folio end** (text, req., uniq.): The folio number inscribed on the last page.
- **image start** (integer, opt., uniq.): If the pages are digitized, the number of the first digital image.
- **image end** (integer, opt., uniq.): If the pages are digitized, the number of the last digital image.

## 4.8 Archival Item

Definition: An extant physical document conserved in an archival institution.

Attributes:

- **is conserved in** (foreign key [**Repository**], req., uniq.): The archival institution that currently holds the item.
- **collection** (text, opt., uniq.): The name of a collection in the archival institution.
- **shelfmark** (text, req., uniq.): The shelfmark of the item in the archival institution.
- **belatedly compiled** (boolean, req., uniq.): Whether the item is a collection of manuscripts that pre-date the item's production and were extracted from their original contexts, belatedly compiled together (i.e 19th-century collection of Medieval fragments).

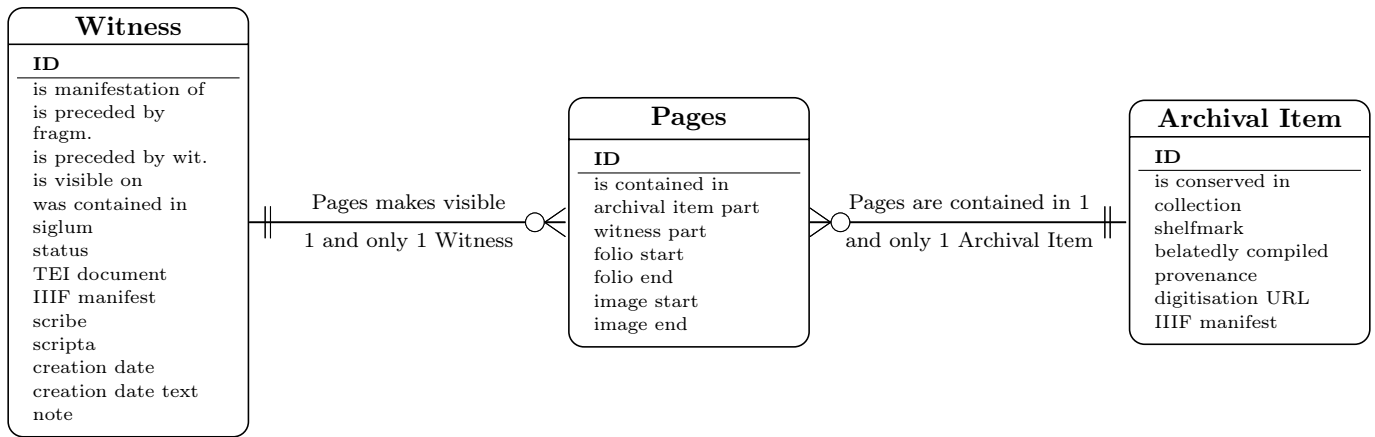


Figure 4.7: *Pages* entity relationships.

- **provenance** (text, opt., uniq.): Notes about the provenance of the *Archival Item*.
- **digitisation URL** (text, opt., uniq.): If digitized, the URL to the digital document.
- **IIIF manifest** (text, opt., uniq.): If a IIIF manifest is available of the digitized document, a link to the manifest's JSON.

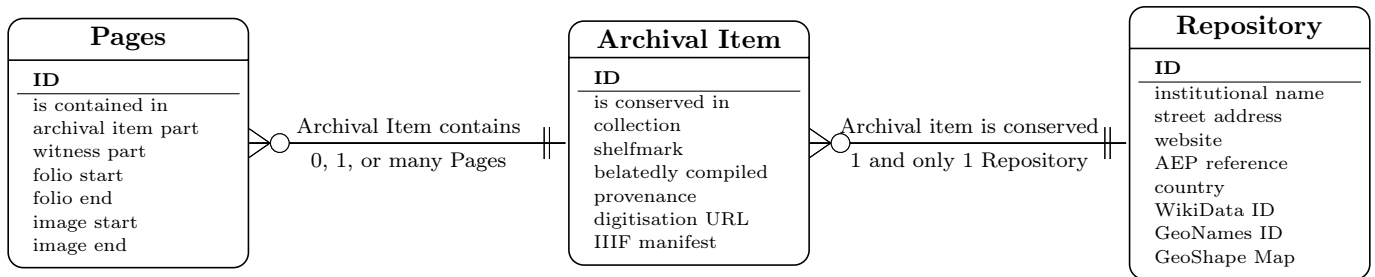


Figure 4.8: *Archival Item* entity relationships.

## 4.9 Repository

Definition: An archival institution.

Attributes:

- **institutional name** (text, req., uniq.): The full, official name of the archival institution in the original language.
- **street address** (text, opt., uniq.): The institution's principal address, which could be used to localize it.
- **website** (text, opt., uniq.): A link to the homepage of the institution's website.
- **AEP reference** [Archives Portal Europe reference] (text, opt., uniq.): URL to the archival institution's information on the Archives Portal Europe website.
- **country** (terms, opt., uniq.): Country in which the institution is located.
- **WikiData ID** (text, opt., uniq.): WikiData ID of the country in which the institution is located.
- **GeoNames ID** (integer, opt., uniq.): GeoNames ID of the country in which the institution is located.
- **GeoShape Map** (text, opt., uniq.): Link to a GeoShape map on Wikimedia.

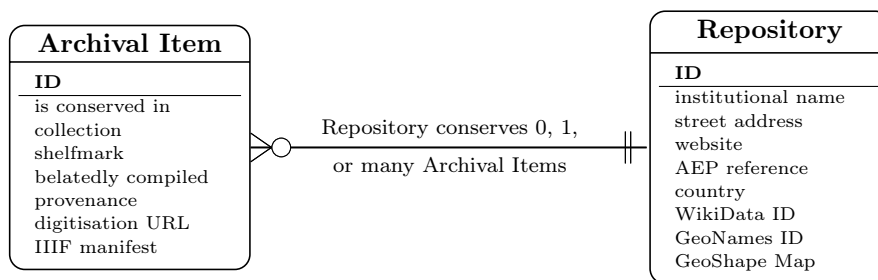


Figure 4.9: *Repository* entity relationships.