

The Database of Byzantine Book Epigrams Project: Principles, Challenges, Opportunities

Rachele Ricceri¹, Klaas Bentein¹, Floris Bernard¹, Antoon Bronselaer¹, Els De Paermentier¹, Pieterjan De Potter¹, Guy De Tré¹, Ilse De Vos¹, Maxime Deforche¹, Kristoffel Demoen¹, Els Lefever¹, Anne-Sophie Rouckhout¹, Colin Swaelens¹

1 Ghent University, Belgium

Rachele Ricceri (corresponding author)

Ghent University, Faculty of Arts and Philosophy, Department of Literary Studies. Blandijnberg 2, 9000 Ghent, Belgium rachele.ricceri@ugent.be

Klaas Bentein

Ghent University, Faculty of Arts and Philosophy, Department of Linguistics. Blandijnberg 2, 9000 Ghent, Belgium klaas.bentein@ugent.be

Floris Bernard

Ghent University, Faculty of Arts and Philosophy, Department of Literary Studies. Blandijnberg 2, 9000 Ghent, Belgium floris.bernard@ugent.be

Antoon Bronselaer

Ghent University, Faculty of Engineering and Architecture, Department of Telecommunications and Information Processing. St-Pietersnieuwstraat 41, 9000 Ghent, Belgium antoon.bronselaer@ugent.be

Els De Paermentier

Ghent University, Faculty of Arts and Philosophy, Department of History. Sint-Pietersnieuwstraat 35, 9000 Ghent, Belgium els.depaermentier@ugent.be

Pieterjan De Potter

Ghent University, Ghent Centre for Digital Humanities. Sint-Pietersnieuwstraat 35, 9000 Ghent, Belgium pieterjan.depotter@ugent.be

Guy De Tré

Ghent University, Faculty of Engineering and Architecture, Department of Telecommunications and Information Processing. St-Pietersnieuwstraat 41, 9000 Ghent, Belgium guy.detre@ugent.be

Ilse De Vos

Ghent University, Faculty of Arts and Philosophy, Department of Linguistics. Blandijnberg 2, 9000 Ghent, Belgium i.devos@ugent.be

Maxime Deforche

Ghent University, Faculty of Engineering and Architecture, Department of Telecommunications and Information Processing. St-Pietersnieuwstraat 41, 9000 Ghent, Belgium maxime.deforche@ugent.be

Kristoffel Demoen

Ghent University, Faculty of Arts and Philosophy, Department of Literary Studies. Blandijnberg 2, 9000 Ghent, Belgium Kristoffel.Demoen@ugent.be

Els Lefever

Ghent University, Faculty of Arts and Philosophy, Department of Translation, Interpreting and Communication. Groot-Brittanniëlaan 45, 9000 Ghent, Belgium els.lefever@ugent.be

Anne-Sophie Rouckhout

Ghent University, Faculty of Arts and Philosophy, Department of Literary Studies. Blandijnberg 2, 9000 Ghent, Belgium annesophie.rouckhout@ugent.be

Colin Swaelens

Ghent University, Faculty of Arts and Philosophy, Department of Translation, Interpreting and Communication. Groot-Brittanniëlaan 45, 9000 Ghent, Belgium colin.swaelens@ugent.be

Abstract

This paper presents an overview of the history, conceptualization, and development of the Database of Byzantine Book Epigrams, an ongoing research project hosted at Ghent University. It also offers a glimpse into current and future research threads carried out within the project, with an eye on long-term sustainability.

The first part of the paper pinpoints the position of DBBE within the broad field of Digital Humanities and addresses the question of how and why Byzantine metrical paratexts have been collected in an

open-access online database. In the second part of the article, we describe the main features of the relational database currently available, both from the perspective of its users and from a technical point of view. The third section of the paper includes the description of four subprojects connected to DBBE, which at present involve the development of a graph database complementary to the relational one, the implementation of natural language pre-processing applied to the DBBE corpus, the linguistic analysis of formulaicity in book epigrams, and the exploration of the broad implications of the study of book epigrams for a better understanding of Byzantine book culture.

Keywords

Relational Database; Graph Database; Manuscript Studies; Byzantine Epigrams; Paratexts; Natural Language Processing; Formulaicity

I THE PROJECT

The present paper outlines the challenges and opportunities of the digitization and valorisation of a textual corpus that is non-uniform, highly formulaic and intimately interconnected with its material context. These characteristics call for innovative tools and techniques that enable flexible, interdisciplinary and “fuzzy” research. It is our conviction that this not only benefits the research of humanities scholars (in this case, historians, philologists, linguists and manuscript scholars), but also computer scientists, who are confronted with challenges that are both technical and conceptual. From the perspective of Digital Humanities, our *Database of Byzantine Book Epigrams* (DBBE) project may thus provide an inspiring example of evolving mutual exchange between Humanities and Computer sciences.

The first section of the paper includes an introduction to the corpus (I.1) and outlines the history and vision behind the project (I.2). In the second section we present an overview of the current database features (II.1), a case study on the challenges (and solutions) for the automatic detection of textual similarities (II.2), the technical description of the database architecture (II.3), and an overview of the approach we adopted to store and share data (II.4). The third part of the paper outlines the research that will be conducted within four subprojects, namely Natural Language Pre-processing (III.1), the graph approach inducing further technical development (III.2), linguistic analysis of the corpus (III.3) and research questions related to medieval Greek book culture (III.4).

I.1 Corpus

The textual corpus under view here consists of Byzantine book epigrams. The term “book epigrams” is used to designate poems that are not only written “in” manuscripts (just like most Byzantine poetry) but also “inscribed on” these manuscripts. With this, we mean that these epigrams tell the reader more about the very manuscript they are part of: they present its subject, tell us who authored it, physically wrote it, paid for it, read it, and how, when, where, and why. By this token, book epigrams are not essentially different from inscriptions that are attached to other objects in Byzantine culture (religious works of art, buildings, etc., see Rhoby 2009; 2010; 2014). Hence, book epigrams are firmly tied to, and only gain meaning when considered together with, their material physical context (for a thorough examination of book epigrams preserved in illuminated manuscripts, see Rhoby 2018).

This lends them a very strong paratextual character. The term “paratext”, as propagated by Genette (1987), refers to texts which present the author, genre, purpose, etc. of a “main text” to the reader; they are to be found on the threshold between the purely mental world of the text and the material world of the physical realization of this text. Genette self-avowedly talked strictly about the medium of printed books. Medievalists have been applying the term “paratext” also to a wide range of other “dependent” texts, such as summaries, commentaries, etc (see Andrist 2018). Our project, however, is concerned with poems that are literally on this threshold between the materiality of the book and the world of the main text, and thus provide a space of communication between on the one hand author, scribe, and/or patron, and on the other hand, the reader.

The interest of these book epigrams lies in the fact that they are interconnected in many ways. First, there is the material embedment. The epigrams are still to be found in their original context of use. Visually and physically, we encounter them in the same shape as the original readers encountered them. They are to be found in liminal places of the book: beginning or end, quire divisions, at the end of texts. Even when they only offer a one-line title of a work, it is meaningful that they do so in a metrical shape and often in a visually distinctive script. Hence, they are the place where codicology and palaeography interact with techniques and attitudes of reading.

Second, there is the social entanglement. These poems are the direct expression of all the actors and communities involved with book culture and text transmission. Their study allows us a direct insight into the history of reading culture, and the social and economic background of manuscript production.

Third, this is a corpus of living texts. Because they are written down by people who are often non-professional poets, they provide an interesting sample of less erudite literary devices and linguistic developments, especially compared with Byzantine poetry written by well-known authors, which tends to be antiquarian (although this feature is not absent in some book epigrams either). By the same token, book epigrams are also often formulaic. Verse lines, parts of verse lines, and poems or chunks of poems recur throughout the corpus, in various operations of adaptation, permutation, and variation.

I.2 Project's Vision and Positioning Within Digital Humanities

When our attention was first drawn to the corpus and the project was conceptualized (late 2000s), it was difficult to even gain access to book epigrams. They were mainly published in manuscript catalogues, that is, the descriptions of manuscript collections preserved in various libraries or monasteries. These catalogues vary greatly in quality and attention for philological detail. A significant part of the corpus was in fact still unedited and thus only available for someone studying the manuscripts themselves.

In its first phase (2010-2015, funded by the Hercules Foundation of the Flemish Government), the DBBE project chiefly intended to collect the textual material in one place and make the texts digitally available. The database was conceived as a relational database: data was connected to each other around the core of a textual record of an epigram.

The textual records of the separate epigrams received most attention. The scanning of catalogues and other publications was a core task of this data collection process. Our team also discovered many epigrams that were still hidden unedited in manuscripts; some significant new editions were published in scientific journals.¹ It became clear that the corpus presented challenges both in terms of quantity and quality. Book epigrams proved to be far more numerous than anticipated, which testifies to their importance in Byzantine culture. Moreover, the initial project was not adapted to the thickness and complexity of contextual data relevant to a more comprehensive understanding of book epigrams.

In a second phase (2015-2020, funded by the Special Research Fund of Ghent University), the focus shifted from collecting data to exploration and analysis. The contextual datasets (persons, place in manuscript, geographical origin) underwent a major revision that resulted in a more encompassing, rational and granular treatment of data. Text critical problems were

¹ <https://www.projectdbbe.ugent.be/publications-and-papers/> (all links cited throughout the paper have been last consulted on 28/02/2023).

tackled through the revision of “type” texts, i.e., provisional “editions” of epigrams that in many cases regrouped several “occurrences”, i.e., specific instances of book epigrams in manuscripts (more on types and occurrences below). The database itself underwent a substantial update of its technical infrastructure (see *infra*, II.3). Meanwhile, we developed a standardized way of referring to book epigrams by means of unique identifiers. The implementation of this identification system facilitates cross-referencing with parallel digital repositories.²

At this point, a mass of data had been collected and made accessible. The scholars connected to our team exploited and used the corpus and its potential for various textual queries successfully, focusing on interesting subsets of our data.³ The question remained, however, how to make sense of the patterns occurring in the dataset as a whole; how to detect, visualize and explore the interconnections between the multiple layers that are present in our data set. This called for both a more interdisciplinary approach and a more innovative way of analysing data.

The challenges currently faced by the DBBE team during the third phase of the project (2021-2026, again funded by the Special Research Fund of Ghent University) are illustrated in the third section of this paper and mirror an evolving scope. As DBBE is a long-term enterprise, which has undergone several adjustments, the vision underlying the project has gradually changed. The growth and maintenance of the current database has been put beside the growing necessity of dealing with research questions that require the combined expertise of philologists, linguists and computer scientists.

To enhance the visibility of the project and to increase the public awareness of the rich heritage of Byzantine book epigrams, we have strongly invested in outreach activities, such as an [online lecture series](#),⁴ a [blog](#),⁵ several initiatives on social media⁶ and a [workshop](#)⁷ aimed at pupils from Flemish secondary schools. These initiatives have proven to be successful for engaging a wide and diverse array of people, ranging from specialists in Byzantine philology to neophyte enthusiasts.

² E.g., <https://www.manuscripta-biblica.org/>; <https://septuaginta.uni-goettingen.de/catalogue>; <https://medieval.bodleian.ox.ac.uk/>.

³ <https://www.projectdbbe.ugent.be/about-the-project/#poetry-from-the-margins>.

⁴ <https://www.projectdbbe.ugent.be/lectures/>.

⁵ <https://www.projectdbbe.ugent.be/blog/>.

⁶ <https://www.facebook.com/byzantinebookepigrams>; <https://www.instagram.com/dbbe.ugent/>; <https://twitter.com/dbbeugent>.

⁷ <https://www.dcg.ugent.be/lesmateriaal/>.

II STATE OF THE ART

II.1 Database Exploration and Navigation

The first version of the project data platform was launched online in 2015, principally featuring objects (i.e., manuscripts) and epigrams. Technical development has not stopped ever since, and in 2019 we proudly announced the launch of the current version of the DBBE, including textual and metatextual data presented through a new web application and stored using a completely revised data model.

This section of the paper aims at illustrating the appearance of the front-end of the relational database and at clarifying the principles that underlie its contents. It is an update of existing publications (e.g., Bernard & Demoen 2012) and draws from current practices, as described on the project's [Help page](#).⁸

The book epigrams recorded in DBBE are arranged according to a fundamental distinction between Occurrences and Types. **Occurrences** are all the instances of epigrams, exactly as they are found in manuscripts, including all kinds of idiosyncrasies in terms of orthography, punctuation etc. The texts offered in Occurrence records are usually based on the diplomatic transcriptions made by DBBE team members on the basis of (reproductions of) manuscripts. In some cases, however, we still must rely on manuscript catalogues or other related publications, which do not always apply rigorous standards and very often normalise the manuscript text.

Each of these Occurrences is linked to one or more overarching texts, called **Types**. The concept of “type” has been introduced to point to a reconstructed text that groups one or more Occurrences with an identical or similar text. Type records provide normalised texts, “readable” adaptations of the evidence (i.e., the Occurrences) found in manuscripts. The text source of Type texts is always mentioned: this is either an external source, such as an existing critical edition or a manuscript catalogue, or “DBBE”, especially for previously unedited Types or for edited texts that have been substantially updated by the DBBE team. In some instances, however, the identification of a Type as a normalised, readable version of one or more Occurrences is unsatisfactory. A standardised Type text does no justice to the reality of

⁸ <https://www.dbbe.ugent.be/pages/help>.

textual transmission of book epigrams, which due to their very nature are open texts, copied and reworked in a similar, yet substantially different way (Demoen 2022: 1378-1379).

Since book epigrams are metrical inscriptions in and on books, each Occurrence is linked to the **Manuscript** in which it is found. Information provided on the manuscript detail pages is based on the available scholarship and is only rarely the result of our own investigations. In particular, we rely on the information available in the Pinakes database and provide standardised identifiers (Diktyon numbers) widely used to refer to Byzantine manuscripts.⁹ We focus on the paratexts themselves and any information on manuscripts is added simply in order to facilitate our users' understanding of the material context of these paratexts.

All three types of database pages mentioned above (i.e., Occurrences, Types and Manuscripts) can have **Persons** linked to them. The Persons include the Byzantine people involved in the production of book epigrams (scribes and patrons of manuscripts and poets of epigrams) as well as people who function as the subject of epigrams.

The last category is **Bibliography**, i.e., all books, book chapters, articles, blog posts, online sources and (PhD) theses linked to one or several of the Occurrence, Type, Manuscript or Person records in the database.

These five categories constitute the menu items on the database website, each of which leads to the search page for the respective category. Guidelines on how to search these categories have been meticulously described on the [Search tips and tricks](#) page;¹⁰ they will not be elaborated on here. Instead, in the next paragraphs, we will navigate through the different detail pages of the five categories and show how these pages and the data they contain are linked to each other.

As an illustration of Type detail pages, we will consider [Type 3818](#) (Figure 1),¹¹ which is a short poem summarising the content of the Byzantine novel *Hysmine and Hysminias* and a “typical” type: it does not pose specific philological challenges and can be perfectly presented within the current infrastructure of the DBBE.

⁹ See <https://pinakes.irht.cnrs.fr/> and <http://www.diktyon.org/en/>.

¹⁰ <https://www.dbbe.ugent.be/pages/search-tips-tricks#how-to-search>.

¹¹ <https://www.dbbe.ugent.be/types/3818>.

Type 3818

(formerly typ/1921)

1 Ἐρωτὶ βλαβθεῖς Ὑσμίνης καρδίαν,
2 ἔρωτα φελέγον προτρέπει πᾶσι νέους,
3 ἔρωσ γὰρ ὀθέτερον πτωσκει βέλους.

Title(s)	Στίχον πρὸς ἔρωτα
Text source	M. Marcovich, 2001, <i>Exaltatus Macrembolitae, De Hysminae et Hysminiae amoribus</i> (In: <i>Manchester-Lapsing 2001</i>)
Text status	Text completely known
Editorial status	Critical text
Genre(s)	→ Header-related epigram → Text-related epigram
Metric(s)	Epidicasyllable
Subject(s)	Hysminae and Hysminias
Tag(s)	→ (Contrib) advice to the reader → Reference to content of the book
Translation(s)	→ Hysminias, his heart pierced by Eros, bids all young people to flee Eros, for Eros wounds more sharply than a dart. Language: English Source(s): E. Joffe, 2012, Four Byzantine Novels, Liverpool 177 → <i>Cor utus Hysminias amare ut sapienter omni haveses, hortatur amorem, Nam telo aculeis vulnerat amor: Language: Latin Source(s): I. Hardt, 1810, Catalogus codicum manuscriptorum graecorum Bibliothecae Regiae Bavaricae, Cod. CCCCXVIII-CCCCXXII continens (vol. 4), Munich 251</i>

Comment	This epigram was also written in the now lost, 16th-century manuscript Turin, Biblioteca Nazionale Universitaria C. VI. 15 (Pasin 239).
Bibliography	<i>Primary</i> I. Hardt, 1810, <i>Catalogus codicum manuscriptorum graecorum Bibliothecae Regiae Bavaricae, Cod. CCCCXVIII-CCCCXXII continens (vol. 4), Munich 251</i>
Number of verses	3
Occurrence(s)	→ [21295] Ἐρωτὶ βλαβθεῖς Ὑσμίνης καρδίαν [15th c.] MUNICH - Bayerische Staatsbibliothek - gr. 405 [15th c.] (f. 1v) (3 verses) → [19754] Ἐρωτὶ βλαβθεῖς Ὑσμίνης καρδίαν [1501-1510] NAPLES - Biblioteca Nazionale - gr. III AA 3 [1501-1510] (f. 1r) (3 verses) → [32526] Ἐρωτὶ βλαβθεῖς Ὑσμίνης καρδίαν [16th c.] PARIS - Bibliothèque Nationale de France (BNF) - gr. 2914 [15th c.] (f. 1r) (3 verses) → [32518] Ἐρωτὶ βλαβθεῖς Ὑσμίνης καρδίαν [1364] PARIS - Bibliothèque Nationale de France (BNF) - gr. 2915 [1364] (f. 1v) (3 verses)
Related type(s)	→ [4733] Τρωθεῖς ἔρωτος ὄζειν ψυχρομόρφος (6 verses) (Same cycle) → [4735] Ὁ μὲν Σολομὼν ἀσμετοτοφαλλομορφος (16 verses) (Same cycle)
Acknowledgements	Contributor(s) Anne-Sophie Rouckhout The credits system has been implemented in 2019. Credits from before the new system was in use might be incomplete.
Identification	Vassis ICB 2005, 256: "In fabulam Eustathii Macrembolitae de Hysminae et Hysminiae amoribus"
Permalink	http://www.dbbbe.ugent.be/types/3818

Figure 1: DBBE Type 3818

A Type detail page (or “record”), in addition to the normalised text of the epigram or the actual critical text if it is available (as is the case in our example), also provides the following information (fields for which no data are entered do not appear in the record; this is the case for all records):

Title(s)	in Greek when available. We provide a Latin title as well, but only if the epigram does not have a Vassis (2005/2011) identifier. ¹²
Text status	indicates the completeness of the text we offer. There are two options: 1. Completely known: we know the text in its entirety; 2. Partially unknown: incomplete Types.
Editorial status	“Critical text” or “Not a critical text”
Genre(s)	we divide our corpus of epigrams into six genres, based on the main actors that play a role in the communicative situation typical for book epigrams, i.e., the author, the text, the scribe, the patron, the reader and images (Bernard & Demoen 2019: 416-423; Demoen 2022: 1374-1378). This classification is rather fluid: one epigram can often be connected to more than one of these categories.
Person(s)	the author of the epigram, i.e., the poet (if known)

¹² These reference works provide an alphabetical listing of the opening lines of some 20.000 Byzantine poems from the beginning of the 4th to the 15th century. Each opening line is accompanied by information on the author, subject, sources and length of the poem. In the DBBE, we refer to the page number and the title/subject in Latin (see ‘Identification’ further down in this list).

Metre(s)	metrical patterns typical of classical and Byzantine poetry
Subject(s)	the main subject(s) of the poem, which can be both Persons and non-Persons (such as characters/titles of literary works, cities, general concepts like “grammar”, “music” or “astrology”,...). These subjects facilitate thematic searches.
Tag(s)	the motifs recurring in the epigram. Like Subject(s), Tag(s) facilitate thematic searches.
Critical notes	although we do not claim our Types to be full critical editions, occasionally we do include critical notes.
Translation(s)	we rarely translate the Types ourselves, but whenever available we do provide existing translations, both in modern languages and in Latin.
Comment	other information that may be relevant to better contextualise the epigram
Bibliography	in addition to the text source, which is mentioned at the top of the page, we here provide primary and secondary sources
Number of verses	either the number of verses displayed in the Text field or a different value, as sometimes we know from an external source that a Type has more verses than we currently have access to
Occurrence(s)	a list of all Occurrences linked to the Type in question
Related types	other Types related to the Type in question, if any
Acknowledgements	we acknowledge all DBBE team members who worked on this record as well as anyone who has provided us with information ¹³
Identification	whenever possible, we refer to entries in other publications or on other websites where our Types can be found, such as Vassis (2005/2011) or Rhoby (2018) ¹⁴
DBBE identifier	all Types in the database have been assigned a unique DBBE identifier, which can be found at the top of each detail page and below as part of the permalink

Table 1: Information to be found on a Type detail page

¹³ There are several ways in which users of the DBBE can provide us with information: it is possible to give feedback on existing records via the feedback form on the bottom of every detail page, to inform us about new occurrences via an [online form](#) or to contact us on dbbe@ugent.be. Many valuable contributions have also been made thanks to our outreach initiatives (see above, I.2).

¹⁴ For a list of all external identifiers currently used in DBBE, see below (section II.4).

A very important section of any Type record is the list of Occurrences and of related Types. Looking at these lists, users immediately get a glimpse into the transmission of a particular epigram. For instance, all preserved Occurrences related to Type 3818 are written in manuscripts from the 14th century onwards, they all count three lines (which is a valuable piece of information, as we will see below, in section II.2) and one of them (preserved in a 15th-century Parisian manuscript) was possibly added at a later stage, as it is dated to the 16th century. All this information is conveniently epitomised on the Type record itself and is presented more diffusely in each of the pages linked to in the screenshot in Figure 2.

Occurrence(s)	<ul style="list-style-type: none"> → [21295]*Έρωτι βληθείς ὑσμινίας καρδίαν [15th c.] MUNICH - Bayerische Staatsbibliothek - gr. 405 [15th c.] (f. 1v) (3 verses) → [19754]*Έρωτι βληθείς ὑσμινίας καρδίαν [1501-1510] NAPLES - Biblioteca Nazionale - gr. III AA 3 [1501-1510] (f. 1r) (3 verses) → [32526]*Έρωτι βληθείς ὑσμινίας καρδίαν [16th c.] PARIS - Bibliothèque Nationale de France (BNF) - gr. 2914 [15th c.] (f. 1r) (3 verses) → [32518] ἔρωτι βληθείς ὑσμινί(α)ς καρδίαν [1364] PARIS - Bibliothèque Nationale de France (BNF) - gr. 2915 [1364] (f. 1v) (3 verses)
---------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figure 2: DBBE Type 3818 (detail)

The hyperlink behind the incipit of [Occurrence 32518](#),¹⁵ the oldest attestation of the epigram, leads us to its detail page (Figure 3).

Occurrence 32518

1. * Έρωτι βληθείς ὑσμινί(α)ς καρδίαν.
2. Έρωτι φεόγενι ηρωτρεπει πησι νεοας.
3. Έρωας γαρ οξίτερον επιρωσασα βέλλουσας.


Type(s)	[3818]Έρωτι βληθείς ὑσμινίας καρδίαν (3 verses)
Text source	DBBE (reproduction of a reproduction of the manuscript)
Text status	Text completely known
Date	1364
Manuscript	PARIS - Bibliothèque Nationale de France (BNF) - gr. 2915 [1364]
Place in Manuscript	f. 1v
Contextual information	The epigram is the first in a series of three poems, which precede Eustathios Makrembolites' <i>Hyssmine</i> and <i>Hyssminias</i> .
Media(s)	DedeeasyLibra
Genre(s)	→ Isthmic epigram → Isthmic epigram
Subject(s)	Hyssmine and Hyssminias
Image source(s)	https://piclica.bnf.fr/ark:/12148/btv1b10722532v/f616em

Bibliography	Primary M. Marcovich, 2001, <i>Eustathius Macrembolites, De Hyssmines et Hyssminiae amoribus libri XI</i> , München-Leipzig: XXII
Number of verses	3
Related occurrence(s)	→ [21295]*Έρωτι βληθείς ὑσμινίας καρδίαν [15th c.] MUNICH - Bayerische Staatsbibliothek - gr. 405 [15th c.] (f. 1v) (3 verses) (3S) → [19754]*Έρωτι βληθείς ὑσμινίας καρδίαν [1501-1510] NAPLES - Biblioteca Nazionale - gr. III AA 3 [1501-1510] (f. 1r) (3 verses) (3S) → [32518]*Έρωτι βληθείς ὑσμινίας καρδίαν [16th c.] PARIS - Bibliothèque Nationale de France (BNF) - gr. 2914 [15th c.] (f. 1r) (3 verses) (3S)
Acknowledgements	Creator(s) Anne-Sophie Roukhout Transcriber(s) Anne-Sophie Roukhout The credits system has been implemented in 2019. Credits from before the new system was in use might be incomplete.
Permalink	https://www.dbbe.ugent.be/occurrences/32518

Figure 3: DBBE Occurrence 32518

¹⁵ <https://dbbe.ugent.be/occurrences/32518>.

The text of Occurrence records is presented as faithfully as possible, mostly by means of manuscript transcriptions. The editorial conventions we use can be found on the [Search tips and tricks](#) page.¹⁶ The detail pages of Occurrences and Types have the fields on Metre(s), Genre(s), Subject(s), Comment, Bibliography, Number of verses and Acknowledgements in common; note, however, that for some of these information fields Occurrences can diverge from their Type. In addition, the following essential information is given on Occurrence records:

Title	if present in the manuscript, the title is transcribed and displayed before the verses of the epigram
Verse variants	if there is a little link icon next to a verse () , this verse may be found in other Occurrences as well. When you click on the verse in question, a so-called verse detail page opens (on which see infra, II.2).
Type	the relevant Type(s) to which the Occurrence has been linked
Text source	the source on which we relied for the text, i.e., either an external source or “DBBE” in case we have established the text ourselves based on “Autopsy of the manuscript” or “Inspection of a reproduction/microfilm of the manuscript” ¹⁷
Text status	indicates the completeness of the text we offer. There are three options: 1. Completely known: we know the text in its entirety; 2. Partially unknown: we only have access to a partial text; 3. Completely unknown: epigrams we know exist, but to which we have no access.
Date	the exact date or a date interval; the date of the Occurrence can differ from the composition date of its manuscript (see below)
Manuscript	the manuscript in which the Occurrence can be found. You can click on its hyperlink to open the Manuscript detail page.
Place in Manuscript	preferably a precise folio number; alternatively, a more general indication of where in the manuscript the Occurrence

¹⁶ <https://www.dbbe.ugent.be/pages/search-tips-tricks#the-occurrence-detail-pages>.

¹⁷ At the time of writing this article, 75% of the occurrences in the DBBE have been transcribed by DBBE team members.

	can be found
Palaeographical information	indications on script, ink colour, or other palaeographical features
Contextual information	surrounding text(s), whenever relevant
Person(s)	the patron(s) and / or scribe(s) of the epigram. Often (but certainly not always) the patron and / or scribe of an Occurrence will be the same as those of its manuscript
Image source(s)	link(s) to an image of the Occurrence in an online, often freely accessible, repository ¹⁸
Related occurrence(s)	any Occurrence that has at least one verse in common with this Occurrence
DBBE identifier	all Occurrences have been assigned a unique DBBE identifier, which can be found at the top of each detail page and below as part of the permalink

Table 2: Information to be found on an Occurrence detail page

The hyperlink in the Manuscript field on Occurrence detail pages leads us to the Manuscript detail page (Figure 4). DBBE Occurrence 32518 belongs to [manuscript Par. gr. 2915](#),¹⁹ written in 1364 and containing four book epigrams in total.

¹⁸ Besides downloaded reproductions from online repositories, we have more images at our disposal, which unfortunately cannot be published in the database due to copyright restrictions.

¹⁹ <https://dbbe.ugent.be/manuscripts/32517>.

Manuscript

PARIS - Bibliothèque Nationale de France (BNF) - gr. 2915

Content	→ Epistolographica > John Chrysostom (saint) (person page) → Erotica > Eustathios Makrembolites (person page)
Person(s)	Scribe Markos Mousouros (1470-1517) - RGK: I.265, II.359, III.433 - VGH: 289.E, 290.A
Date	1364
Bibliography	M. Marcovich, 2001, Eustathius Macrembolites, De Hysmines et Hysminiae amoribus libri XI, München-Leipzig: XI
Identification	Diktyon (Pinakes), 52554
Occurrences	→ f. Iv [32518] ἔρωτι βληθεὶς ὑσμινί(α)ς καρδίαν [1364] → f. Iv [32519] τρωθεὶς ἔρωτος ὀξέσι ψυχοφθόροις [1364] → f. Iv [32520] ὁ μὲν σολομῶν ἀσματοψαλμογράφος [1364] → f. 102v [32522] μάρκος ἐγὼ τάδε πάντα κελαινοτάτιο ρεέθρου [1476-1525]
Acknowledgements	Creator(s) Anne-Sophie Rouckhout <div style="background-color: #e0e0e0; padding: 5px; text-align: center;">The credits system has been implemented in 2019. Credits from before the new system was in use might be incomplete.</div>

Figure 4: DBBE Manuscript 32517

A Manuscript record includes:

Name of the manuscript	we follow the system used by the Pinakes database , i.e., city, library, collection and shelfmark
Content	a general idea of what the manuscript contains, based loosely on the generic epithets used by the Thesaurus Linguae Graecae . ²⁰ We are using a parent-child system: well-known authors and texts have been given a separate entry within the overarching categories. Both the generic categories and their “children” have been provided with a hyperlink, which leads to the Manuscript search page listing all manuscripts with this specific content. If an author is included in the manuscript content, an additional hyperlink “person page” next to it links to the Person detail page of the Person in question.
Person(s)	the scribe(s) and patron(s) of the manuscript, but in some cases also later annotators of the manuscript
Date	the exact date or a time interval for the composition of the

²⁰ <http://stephanus.tlg.uci.edu/>.

	manuscript. We mostly rely on external sources, such as Pinakes or secondary literature, to which we refer for more information about unsure, debated or complex dating.
Origin	location of composition, whenever available and as detailed as possible. We also use a parent-child system to organise the origins.
Bibliography	a list of catalogues and/or other publications that offer information on the manuscript
Identification	the Diktyon number which directs to the corresponding pages in Pinakes and, if applicable, in Paratexts of the Bible (Manuscripta Biblica)
Comment	other information that can be relevant
Occurrences	a list of Occurrences found in the manuscript, ordered by folio number. Each incipit links to the Occurrence detail page.
Acknowledgements	we acknowledge all DBBE team members who worked on this record as well as anyone who has provided us with information.

Table 3: Information to be found on a Manuscript detail page

A major effort since the launch of the second version of DBBE in 2019 has involved the fine-tuning of the prosopographical metadata by revising the Person records, with the aim to enable scholars to find links between people, textual content and contextual metadata. Persons can function as manuscript content (Manuscript detail page), subject of an epigram (Type and Occurrence detail pages), poet of an epigram (Type detail page) and scribe or patron of one or more manuscripts or epigrams (Manuscript and Occurrence detail pages). Each reference to a Person as content, scribe, patron, poet and subject links to the Person detail page.

The 14th-century Parisian manuscript that contains one of the four attestations of the book epigram on the Byzantine novel *Hysmine and Hysminias* has one person linked to it as a scribe, named Markos Mousouros (Figure 5). He is not the main scribe, as the manuscript was written in 1364 and Markos lived in the 15th-16th century, but he added a metrical ex-libris on f. 102v and is therefore included as scribe of both the manuscript and the occurrence he wrote. Clicking on his name will lead you to his [Person page](https://dbbe.ugent.be/persons/212).²¹

²¹ <https://dbbe.ugent.be/persons/212>.

Person
Markos Mousouros (1470-1517)

Name	Markos Mousouros
Date	1470-1517
Identification	<ul style="list-style-type: none"> → RGK_I.265 → RGK_IL359 → RGK_III.433 → VGH_289.E.290.A
Manuscript(s)	<p>As Scribe:</p> <ul style="list-style-type: none"> → LONDON - British Library - Burney 96 [1476-1500] → PARIS - Bibliothèque Nationale de France (BNF) - gr.2915 [1364] → VATICAN CITY - Biblioteca Apostolica Vaticana - Palat. gr.261 [15th c.] → VATICAN CITY - Biblioteca Apostolica Vaticana - Palat. gr.275 [15th c.] → VATICAN CITY - Biblioteca Apostolica Vaticana - Palat. gr.287 [14th-15th c.] <p>As Related:</p> <ul style="list-style-type: none"> → MILAN - Biblioteca Ambrosiana A.119 sup. [1276-1325] → MILAN - Biblioteca Ambrosiana L.41 sup. [15th c.] → VENICE - Biblioteca Nazionale Marciana - gr. app. XI 12 [1476-1500]
Occurrences(s)	<p>As Scribe:</p> <ul style="list-style-type: none"> → [2092] εἰ καὶ σοὶ σφίγγε θυμὸς πάρος ἕκασ βέβηλος [1476-1500] → [1768] Κόσμησι τῶδε κάρλων, ὃν ἄμμετι μεθίδωνι [16th c.] → [1768] κάρλων κτημὸ τῶδ' ἔστιν· θεὸν λαοπένο σπάλλαια [15th c.] <p>display 5 more occurrences (8 in total)</p> <p>As Subject:</p> <ul style="list-style-type: none"> → [3078] κάρλων τῶδ' οὐκ ἔστ', ἀλλὰ μωσοῦρου κτήρ [18th c.] → [3252] μάρκος γῶν τῶδε πάντα κελαινεύουσι μετ'ἄρου [1476-1525]
Type(s)	<p>As Poet:</p> <ul style="list-style-type: none"> → [4468] εἰ καὶ σοὶ σφίγγε θυμὸς πάρος ἕκασ βέβηλος → [3078] κάρλων τῶδ' οὐκ ἔστ', ἀλλὰ μωσοῦρου κτήρ → [2525] Κόσμησι τῶδε κάρλων, ὃν ἄμμετι μεθίδωνι <p>display 4 more types (7 in total)</p> <p>As Subject:</p> <ul style="list-style-type: none"> → [3078] κάρλων τῶδ' οὐκ ἔστ', ἀλλὰ μωσοῦρου κτήρ → [3252] μάρκος γῶν τῶδε πάντα κελαινεύουσι μετ'ἄρου
Bibliography	<ul style="list-style-type: none"> → L. Ferret, 2014, L'Italia degli Ormanzi I. Marco Musura. Turinout: Matematicis Genealogy Project (last accessed: 2019-07-06) → https://www.genealogy.math.ndsu.nodak.edu/~lshp/16-15250/ → Pinakes Bibliothé de recherche et d'habitats des textes (last accessed: 2018-02-18) → https://pinakes.fr/notes/epigramme-possesseur-autre/1596/ → Album de copistas de manuscritos griegos en España (last accessed: 2019-07-06) → http://webs.ucm.es/info/copi/stas/copista.html?num=66

Figure 5: DBBE Person 212

Apart from listing all clickable database records to which the person is linked, subdivided into Manuscripts, Types and Occurrences (and in each category according to their role), Person detail pages also mention:

Name	the full name of the person and the provenance of the person, if available
(Self-)designation	any Greek word describing the person in question. This information is either inferred from the text of the epigrams related to the Person or drawn from existing prosopographical sources.
Date	either an exact ‘born’ and ‘died’ date or date intervals to account for inexact dates. In addition, in order to be as objective as possible, we also provide ‘attested’ dates and intervals whenever known: these are based on references to the Person in primary sources.
Office	office(s) held by the person in question, transliterated instead of translated
Identification	references to external identifiers or prosopographical databases ²²
Bibliography	bibliographical references

Table 4: Information to be found on a Person detail page

As the screenshots above illustrate, database records have been consistently linked to each

²² For a list of all external identifiers currently used in DBBE, see below (section II.4).

other, in order to facilitate navigation through the database (see *infra*, II.3). Therefore, each record contains several hyperlinks that link either to another database record or to an overview of all records linked to a specific piece of information. Links to Occurrences, Types, Manuscripts, Persons and Bibliography direct users to the detail pages of these records. The hyperlinks of genres, metres, subjects (apart from persons), tags and contents enable the user to search for all types, occurrences or manuscripts with that label: they link to the Search page of Types, Occurrences or Manuscripts, where the search can be further refined.

The effort described has resulted so far in an ever-growing and ever-evolving corpus that at present counts over 4800 Type records, linked to more than 12000 Occurrence records (mostly based on autoptic transcriptions), referring to some 4300 manuscript records, in which in one way or another more than 2000 Persons were involved.

II.2 Textual Similarity: Challenges and Solutions

The safe journey through which database users are guided can sometimes be surprisingly perilous. Book epigrams are prone to variations and modifications, as they are often copied and reworked in different times and contexts. Occurrences [19576](https://www.dbbe.ugent.be/occurrences/19576),²³ [24352](https://www.dbbe.ugent.be/occurrences/24352)²⁴ and [22293](https://www.dbbe.ugent.be/occurrences/22293)²⁵ constitute a meaningful case study to illustrate how the presentation and visualisation of similar texts has been implemented in the current version of DBBE. These three Occurrences have three twelve-syllable verses in common: τὸν εὔλαλον τέττιγα τῆς ἐκκλησίας, / τὸν μουσικὸν νοῦν, τὸν παρ' ἡμῖν Ὀρφέα / Δαυὶδ τὸν ἐσθλὸν ἐν προφήταις τὸν μέγαν, “the eloquent cicada of the church, / the musical mind, our Orpheus, / the splendid David, great among the prophets”. These poetic lines briefly point to the content of the three manuscripts that contain the epigrams, all of which transmit the text of the biblical book of the Psalms, whose composition was traditionally attributed to the prophet David, here compared to the Greek lyric poet Orpheus.

The syntactic subjects of those three lines are expressed in the second part of the epigrams and point to the scribes of the three manuscripts in which the Occurrences are copied, Anthimos, Gregorios, and Euthymios, with reference to the physical act of book production, resulting in different closings of the three Occurrences, counting either one or two lines.

²³ <https://www.dbbe.ugent.be/occurrences/19576>.

²⁴ <https://www.dbbe.ugent.be/occurrences/24352>.

²⁵ <https://www.dbbe.ugent.be/occurrences/22293>.

Occurrence 19576 thus counts five lines in total (ll. 4-5: Τέτευχε. Θερμῶς. Ἄνθιμος καλλιγράφος / εἰς κλέος, εἰς καύχημα τῶν ὀρθοδόξων, “the scribe Anthimos has ardently composed for the fame and pride of the faithful”), whereas Occurrence 24352 only four (l. 4: γρηγόριος τετευχεν ὁ πρέσβυς πόθῳ, “the priest Gregorios has composed out of desire”). These two Occurrences share the same three-line incipit, referring to the content of the book. This may be enough to identify them as belonging to the same “Type” of text, but they substantially differ as far as the specific information about the manuscript production is concerned. Occurrence 22293, moreover, presents yet a different textual scenario. The three verses in question (incipit τὸν εὐλαλον τέπτιγα) are not found at the beginning of the poem, but as lines 7-9, following another book epigram on the Psalter (incipit Τοῦ Πνεύματος τὰ θεῖα τόξα καὶ βέλη, “the divine bows and arrows of the Spirit”), which frequently occurs in Byzantine manuscripts. The closing of the epigram, just as in Occurrence 24352, consists of only one verse, where the name of the zealous scribe is mentioned (l. 10: Γέγραφεν εὐθύμιος εἰς αὐτοῦ κλέος, “Euthymios has written for his glory”). The layout of the manuscript folio suggests that Occurrence 22293 was indeed perceived by the scribe as one continuous poem of ten verses, therefore the epigram is recorded as such in DBBE.²⁶

At the time of writing this article, Occurrences 19576, 24352 and 22293 are all linked to Type [3654](#),²⁷ a five-line epigram based on the text of Occurrence 19576, which has been taken over from external sources (Hutter 1977: 27; Parpulov 2014: 217). The variant readings, or rather the different versions of the texts, are synthetically sketched in the Critical Notes field of the Type record (Figure 6), in order to give the user an idea of the complexity of the textual tradition.

Critical Notes	4 τέτευχε θερμῶς Ἄνθιμος καλλιγράφος] γέγραφεν Εὐθύμιος εἰς αὐτοῦ κλέος Par. gr. 164, Γρηγόριος τέτευχεν ὁ πρέσβυς πόθῳ Turin. B.VII.30 5 om. Par. gr. 164 and Turin. B.VII.30
----------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figure 6: Type 3654 (detail)

Moreover, the list of Occurrences linked to the Type record provides a synoptic view of the transmission of the epigrams in the three manuscripts. The incipit and the number of verses included in the single epigrams, as displayed in Figure 7, do not match with the Type in two out of three Occurrences.

²⁶ A picture of the Occurrence is available at <https://gallica.bnf.fr/ark:/12148/btv1b550060713/f17.item>.

²⁷ <https://www.dbbe.ugent.be/types/3654>.

Occurrence(s)	<ul style="list-style-type: none"> → [19576] Τὸν εὐλαλον τέτιγα τῆς ἐκκλησίας [951] OXFORD - Bodleian Library - Auct. D. 4.1 [951] (f. 35r) (5 verses) → [22293] Τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλη [1070] PARIS - Bibliothèque Nationale de France (BNF) - gr. 164 [1070] (f. 5r) (10 verses) → [24352] τὸν εὐλαλον τέτιγα τῆς ἐκκλησίας [9th c.] TURIN - Biblioteca Nazionale Universitaria B. VII. 30 (Pasini 342) [9th c.] (f. 139r) (4 verses)
---------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Figure 7: Type 3654 (detail)

The alternative readings provided in Occurrences 22293 and 24352, which do not belong to the Types text field, are not searchable through the Type Search Page. A better visualisation of this kind of textual variation is one of the challenges that is being dealt with in the current, third phase of the project (see *infra*, III.2).

As said, the text of Occurrence 22293 is the faithful transcription from a manuscript where the scribe actually merged two epigrams into one. This Occurrence is therefore linked both to Type 3654 and to Type [3878](#) (incipit Τοῦ Πνεύματος τὰ θεῖα τόξα καὶ βέλη), to which the first six lines of the poem correspond. An Occurrence record is thus designed to link a specific poem to one or more Types in a flexible way, in order to show how texts are freely combined. In addition, a more refined system is in place to visualise portions of text on the level of the verse. Each line of any Occurrence is ideally linked to a so-called **verse variant page**, where parallels and deviations of single verses are clearly shown next to each other. The first verse of Occurrence 22293 (Τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλη) is for instance to be found twenty-six times in DBBE, with a minimal degree of variation (Figure 8).²⁸

²⁸ https://www.dbbe.ugent.be/verse_variants/15714.

Τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλιη	1/10	[22293] Τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλιη	f. 5r	1070	PARIS - Bibliothèque Nationale de France (BNF) - gr. 164
+τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλιη	1/7	[21705] τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλιη	f. 3v	12th-13th c.	ROME - Biblioteca Casanatense 1908
※ τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλιη	1/7	[24334] τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλιη	f. 2v	12th-13th c.	SINAI - Mone tes Hagias Aikaterines - gr. 22
Τοῦ πνεύματος τὰ θεῖα, τόξα καὶ τὰ βέλιη:	1/5	[21817] Τοῦ πνεύματος τὰ θεῖα, τόξα καὶ τὰ βέλιη	f. 508r	1452	SINAI - Mone tes Hagias Aikaterines - gr. 27
Τοῦ πνεύματος τὰ θεῖα τόξα κ(αὶ) βέλιη:	1/7	[24359] Τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλιη	f. 1v	14th c.	SOFIA - Centar I Dujcev D. gr. 389
τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλιη:	1/7	[25079] τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλιη	f. 14r	10th c.	VATICAN CITY - Biblioteca Apostolica Vaticana - Barb. gr. 340
τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλιη	1/6	[24327] τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλιη	f. 37v	11th c.	VATICAN CITY - Biblioteca Apostolica Vaticana - Ottob. gr. 398
τοῦ πνεύματος τὰ θεῖα τόξα κ(αὶ) βέλιη	16/22	[17485] ὁ πάντα κόσμον ἐν λόγῳ κτίσας πάλιν	f. 17r-17v	1058	VATICAN CITY - Biblioteca Apostolica Vaticana - Vat. gr. 752

Figure 8: Verse variant page 15714 (detail)

Verse variant pages have been conceptualised as a pragmatic solution to better group individual verses collected in DBBE, which appear in Occurrences that are not necessarily linked to the same Type. Verse variant pages are a suitable device to dig into textual transmission in a swift and user-friendly way.

First and foremost, such overviews are a helpful tool to collate verses, whose various attestations are conveniently displayed one after another in the first column, which is useful from a philological point of view.

Secondly, these pages also help to visualise, up to a certain extent, the textual fluidity typical of book epigrams. In the second and third column of the grid appear respectively the place occupied by each particular verse in each Occurrence and the full incipit of the poem. The reference to the number of verses and to the relative positioning of the verse in question within the Occurrence is an indication that similar or identical verses may appear in shorter or longer texts. The presence of a different incipit is a firm hint to the fact that the transmission of that verse is not a linear one. Another incipit testifies to a form of textual contamination at some point of the tradition.

Finally, verse variant pages provide users with a glimpse into the manuscript evidence by listing—in the fifth and sixth column of the grid—references to the shelfmarks of the manuscripts where the verses are copied and to the dating of the relevant Occurrences. This information is valuable to address at least two kinds of research questions. On the one hand it helps to put manuscripts in contact and to research their relationships (see *infra*, III.2), contributing to the identification of clusters of manuscripts that share similar paratexts and,

perhaps, were related to each other. On the other hand, it gives an additional input to the philological study of the corpus by means of a compact and clear visualisation of the relative chronology of the epigrams.

Verse variant pages have been created by manually linking verses, based on suggestions created by a specifically designed tool (see *infra*, II.3). The creation of the links is ultimately a human decision, which can result in inconsistent choices. Moreover, in the verses collected in these overview records no distinction is made between orthographical variation and actual variant readings.

+ δαυιτική πέφυκε βίβλος ἄσμάτων·	1/6	[24487] δαυιτική πέφυκε βίβλος ἄσμάτων	f. 3v	12th-14th c.	ROME - Biblioteca Casanatense 1908
+ εὐαγγελικός πέφυκα δέλτος ἄσμάτων	1/4	[23732] εὐαγγελικός πέφυκα δέλτος ἄσμάτων	f. 104r	1067	SINAI - Mone tes Hagias Aikaterines - gr. 172
Εἰρμῶν μέλος πέφυκε δέλτος ἄσμάτων	1/6	[22096] Εἰρμῶν μέλος πέφυκε δέλτος ἄσμάτων	(gen.) end of ms.	1266	SINAI - Mone tes Hagias Aikaterines - NE gr. M 176
+ δαβιτικῶν πέφυκα δέλτος ἄσμάτων(ν)	1/4	[22093] + δαβιτικῶν πέφυκα δέλτος ἄσμάτων(ν)	f. 32v	10th c.	SINAI - Mone tes Hagias Aikaterines - NE gr. ΜΓ 19
Δαυιτική πέφυκα δέλτος ἄσμάτ(ων)	1/6	[24360] Δαυιτική πέφυκα δέλτος ἄσμάτ(ων)	f. 1v	14th c.	SOFIA - Centar I Dujcev D. gr. 389
+ δαυιτική πέφυκα δέλτος ἄσμάτ(ων)	1/6	[33537] δαυιτική πέφυκα δέλτος ἄσμάτ(ων)	f. 139r	1317-1320	VATICAN CITY - Biblioteca Apostolica Vaticana - Palat. gr. 367
δαυιτη κή πέφυκα δέλ τος ἄσμάτων·	6/22	[17485] ὁ πάντα κόσμον ἐν λόγῳ κτίσας πάλιν	f. 17r-17v	1058	VATICAN CITY - Biblioteca Apostolica Vaticana - Vat. gr. 752

Figure 9: Verse variant page 14628 (detail)

The verse variant page²⁹ corresponding to the first line of Type [4568](#)³⁰ (Δαυιτική πέφυκα δέλτος ἄσμάτων, “I am the book of David’s songs”, also referring to the book of Psalms) includes twenty Occurrences. As visualised in Figure 9, the text of this verse significantly varies from manuscript to manuscript. In fact, the decision to group these verses together is based on the intuitive observation that their text is similar enough to belong together.

From a philological viewpoint, the visualisation of this example of horizontal transmission (Stussi 2006: 14) proves extremely fruitful. The very fact that these verses have been grouped together by means of the verse variants system currently in place allows for exploring the non-conventional tradition of a paratext that mostly accompanied the text of the Psalms but

²⁹ https://www.dbbe.ugent.be/verse_variants/14628.

³⁰ <https://www.dbbe.ugent.be/types/4568>.

was occasionally modified and copied along the text of the Gospels (Occurrence [23732](https://www.dbbe.ugent.be/occurrences/23732)³¹) or a liturgical book (Occurrence [22096](https://www.dbbe.ugent.be/occurrences/22096)³²).

However fortunate the verse linking through verse variant pages can be, nonetheless, a more refined digital approach to the concept of similarity within the corpus and its objective measurement will only prove beneficial (see *infra*, III.2).

II.3 Technical Description

The data platform released in 2015 made it possible to publish a sizable amount of research data that had previously only been available to a select group of researchers. Upon extensive usage of the platform, we realised that additional features, such as the possibility to link some types of records and the ability to find records based on certain fields, that could improve the platform's usability, would have been desirable. The way the data model was set up, however, made it impossible to implement some of the required features. As major changes were going to be made to the data model, it was decided to create a new web application as well, including both a public interface with search and detail pages and an editing interface. The new web application (launched in 2019) was developed by the [Ghent Centre for Digital Humanities](https://www.dbbe.ugent.be/), which brings together Digital Humanities efforts in order to make the implementation and maintenance of projects more sustainable. The database is hosted on a Ghent University server, which offers a long-term guarantee on data safety and sustainability. The current data model was primarily conceived to facilitate the linking of records. In this data model, almost every class hierarchically inherits from a base entity class, facilitating the creation of links between these classes with much fewer join tables. A special link table "factoid" is introduced to generically model relations between objects and a related date, interval or location where relevant. Figure 10 depicts an example of a factoid. In the previous data model, a specific join table was used to define relations between Occurrences and Types. In the current data model, these relations can be defined using the generic join table factoid.³³

³¹ <https://www.dbbe.ugent.be/occurrences/23732>.

³² <https://www.dbbe.ugent.be/occurrences/22096>.

³³ The following relations can currently be modelled as factoids: "completed at", "reconstruction of", "located at", "written", "related to", "subject of", "origination", "based on", "appears immediately after", "died", "born", and "attested". This system is also used to model different kinds of relations between two Types: "Is part of", "Variant (permutation of words)", "Variant (other metre)", "Variant (other subject)", "Consists of", "Same cycle", "Variant (other wordings)", "Unknown". Additional "factoid types" can be added if it is ever needed to model additional relationships between objects.

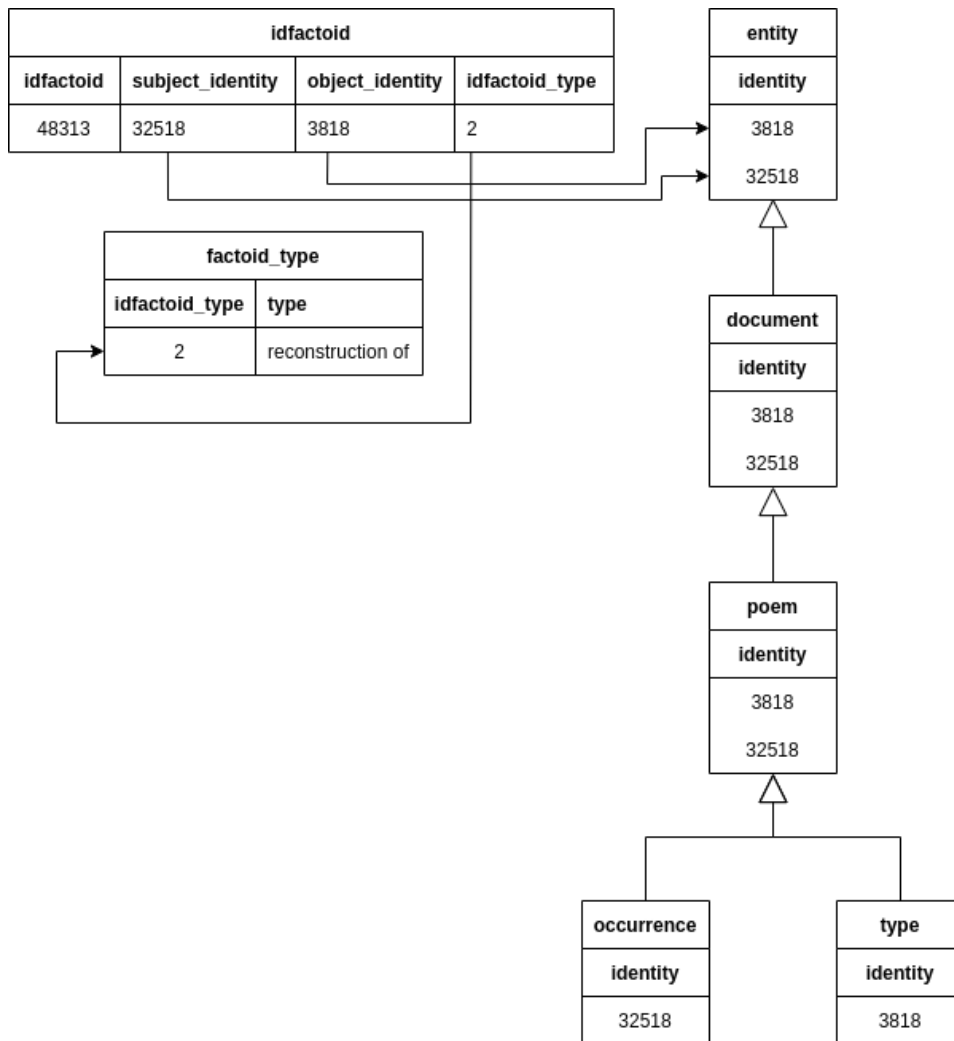


Figure 10: Simplified current data model representing a relationship between a Type and an Occurrence defined in the generic factoid join table. This specific example represents the fact that Type 3818 is a reconstruction of Occurrence 32518.

Records from different tables representing the same data were detected and migrated towards a single record in a single table. As an example, we will consider a person that is both the subject and patron of an Occurrence. As demonstrated in Figure 11, in the previous data model, a single person was represented by two different records, which could (and do in this specific example) contain different information. In the current data model, the data about this single person are stored in a single record, as visualized in Figure 12. This approach makes the data model more transparent, prevents data duplication and reduces the chances of data inconsistencies.

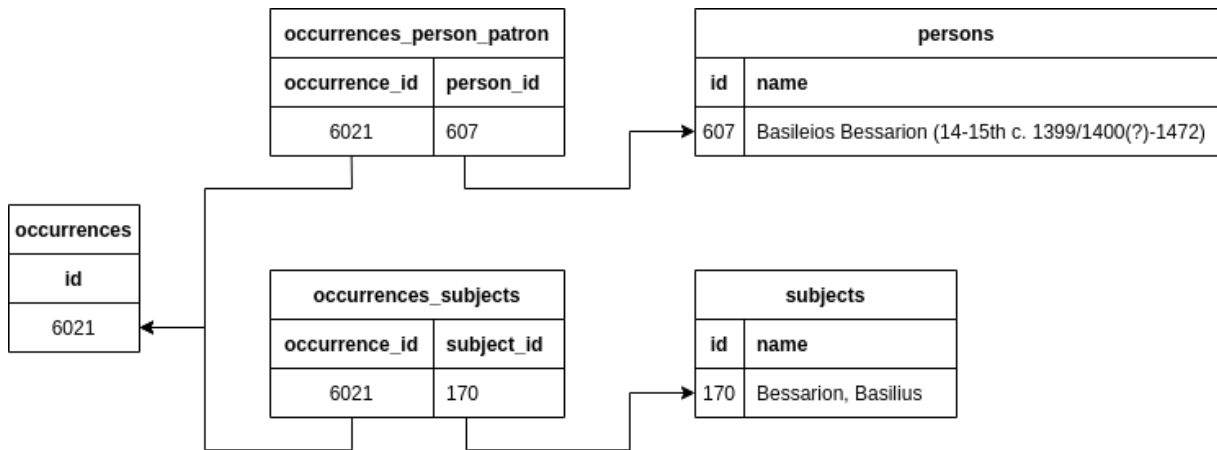


Figure 11: Simplified previous data model representing a subject and patron relationship between an Occurrence and a Person

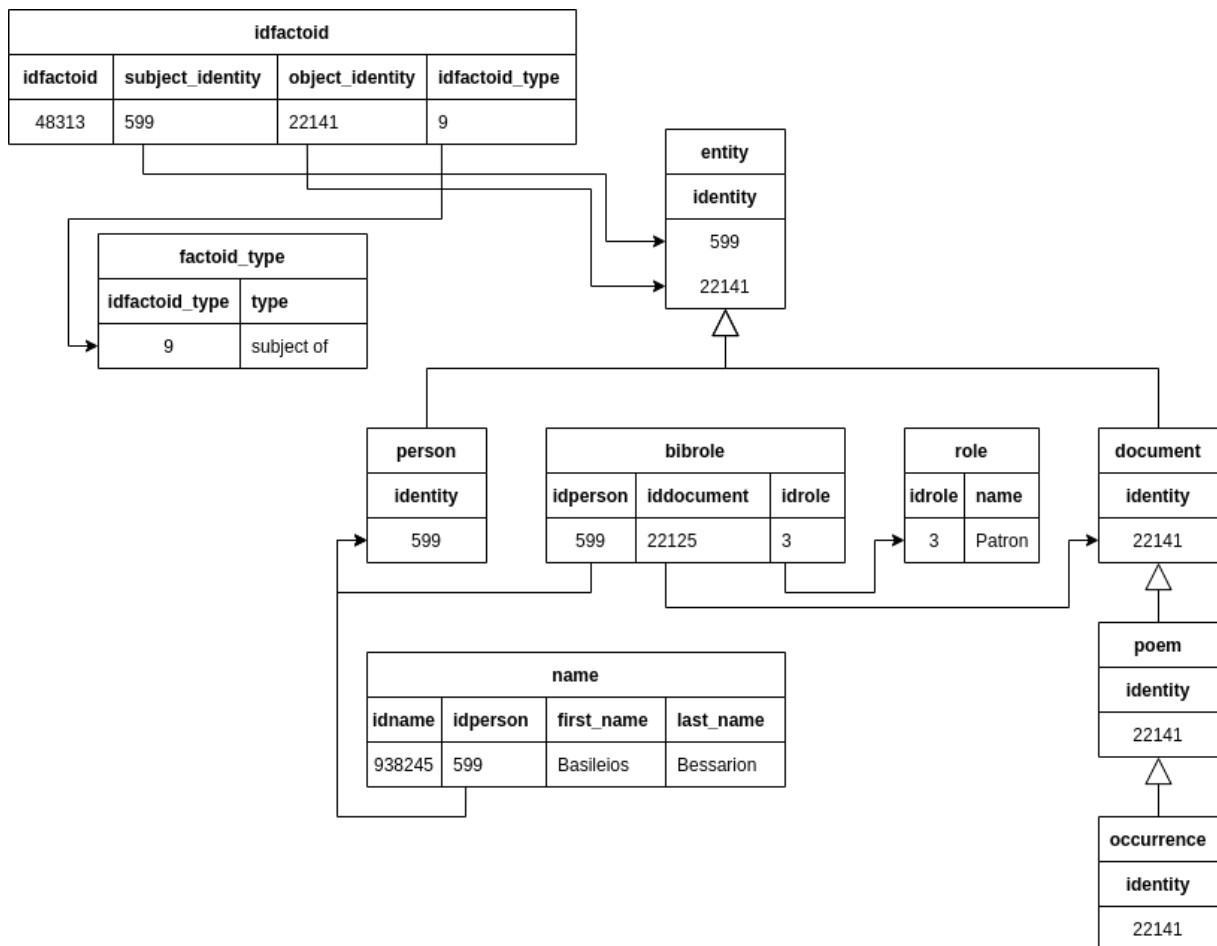


Figure 12: Simplified current data model representing a subject and patron relationship between an Occurrence and a Person

Some objects, such as regions and manuscript contents, can be placed in a hierarchical structure (the so-called “parent-child system”). For example, the region of Apulia is part of

the region of Southern Italy, which is part of the region of Italy. In the current data model, regions and genres are stored in such a way that any of the hierarchical levels can be linked to other objects. This allows filtering based on these different levels, where underlying levels are included in the results as well (a search for the region Southern Italy includes the results linked to the region Apulia, see *infra*). Figure 13 depicts the data model that makes it possible to have any number of hierarchical levels. A region and its parents can be retrieved using a recursive SQL query, as listed in Figure 14.

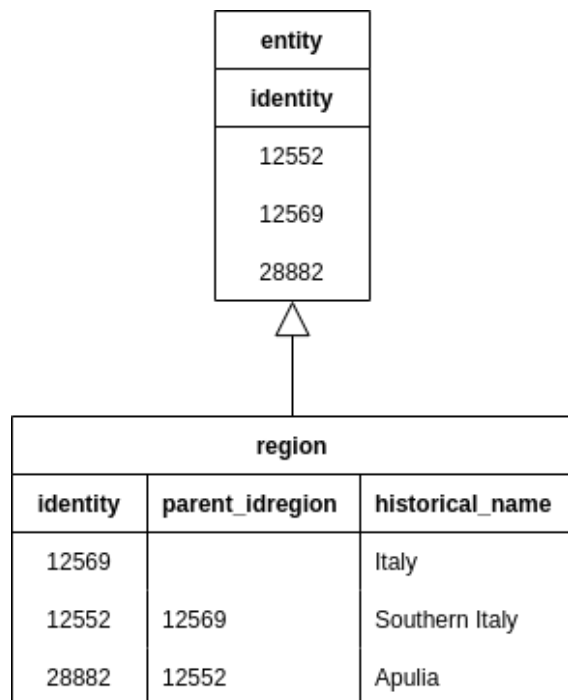


Figure 13: Simplified current data model representing Apulia and its parent regions

```

WITH RECURSIVE rec (id, ids, historical_names, depth) AS (
  SELECT
    r.identity,
    ARRAY[r.identity],
    ARRAY[COALESCE(r.historical_name, '')],
    1
  FROM data.region r

  UNION ALL

  SELECT
    r.identity,
    array_append(rec.ids, r.identity),
    array_append(rec.historical_names, COALESCE(r.historical_name, '')),
    rec.depth + 1

  FROM rec
  INNER JOIN data.region r
  ON rec.id = r.parent_idregion
)
SELECT
array_to_json(ids) AS ids,
array_to_json(historical_names) AS historical_names
FROM rec
INNER JOIN (
  SELECT id, MAX(depth) AS maxdepth
  FROM rec
  GROUP BY id
) rm
ON rec.id = rm.id AND rec.depth = rm.maxdepth
WHERE rec.id = 28882

```

Figure 14: Simplified recursive query used to retrieve information about a region and its parent regions using the current data model

Data types “fuzzydate” and “fuzzyinterval” were introduced to enable the description of uncertainty in dates. In the old data model, uncertain dates were described by a start and end date in some places and by a textual description in other places. In the current data model, “fuzzydate” is defined by the earliest and latest date possible. A “fuzzyinterval” is defined by the earliest and latest date possible for the start of the interval and the earliest and latest date possible for the end of the interval. Using a “fuzzydate” instead of a textual description enables date-based searching and sorting. For example, it is possible to describe the date of birth of [John XI Bekkos](https://www.dbbe.ugent.be/persons/31646)³⁴ between 1230 and 1240 by using (1230-01-01,1240-12-31) as “fuzzydate” definition. Likewise, his date of death, March 1297 can be described with “fuzzydate” (1297-03-01,1297-03-31).

³⁴ <https://www.dbbe.ugent.be/persons/31646>.

As described in detail in section II.2, verses are grouped together in groups of verse variants and these groupings can be used to discover interesting links between Occurrences that might not be discovered by solely relying on the relations between Occurrences and Types. Figure 15 visualizes some of the verses and the occurrences they are part of for verse variant 15714.

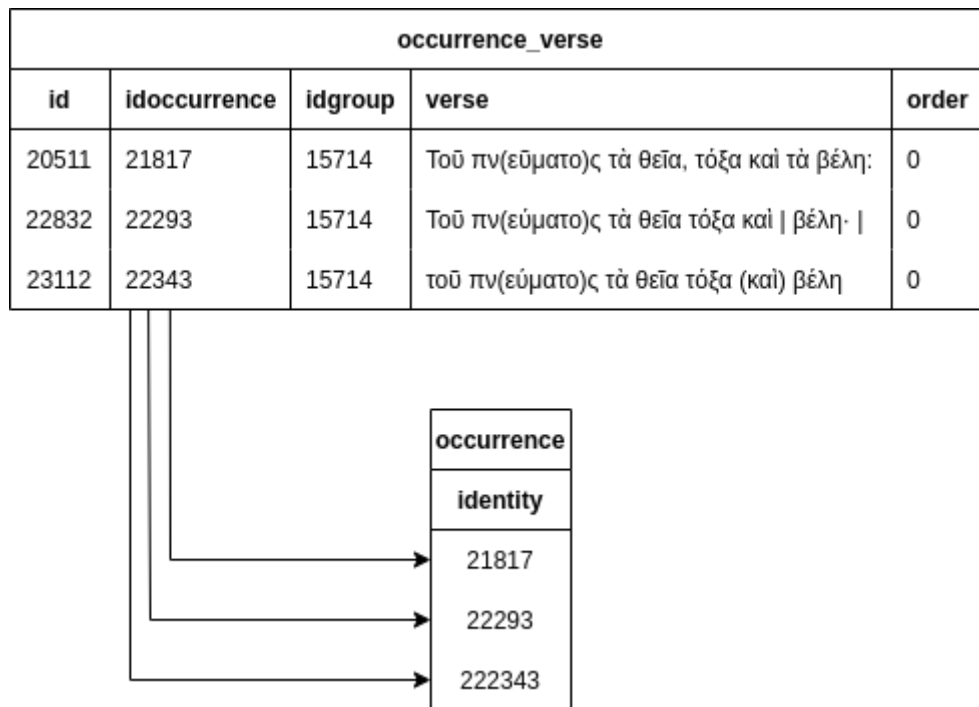


Figure 15: Simplified current data model representing verse variant 15714 (only 3 verses shown)

In the new web application that was needed to take full advantage of the changes in the data model, the front-end framework Vue.js is used to improve the user experience of the data platform. This makes it easier to display both the search filter configuration and the corresponding search results on a single page and makes it possible to immediately update the results after filtering. The search results contain links to detail pages including more information on certain objects; links to related objects; and relevant search pages with predefined filters. In the editing environment, the use of a front-end framework allows to provide in-place validation of the entered data and enables an uncluttered representation of the data to be edited. It also allows the creation of tools to speed up some specific tasks, such as the linking of verses in verse groups when editing Occurrences. Detail pages are completely rendered on the Symfony back-end to maintain findability by web crawlers. The initial request for a search or edit page results in a minimal HTML page that loads the JavaScript with the Vue.js application that renders most of the web page. The information flow between

the different components for detail page requests and the initial edit and search page requests is visualised in Figure 16. On edit and search pages, user actions within the same page lead to data requests to the Symfony back-end initiated by the Vue.js application, resulting in updates of the web page without a complete page reload. Figure 17 depicts the general architecture for these interactions on edit and search pages.

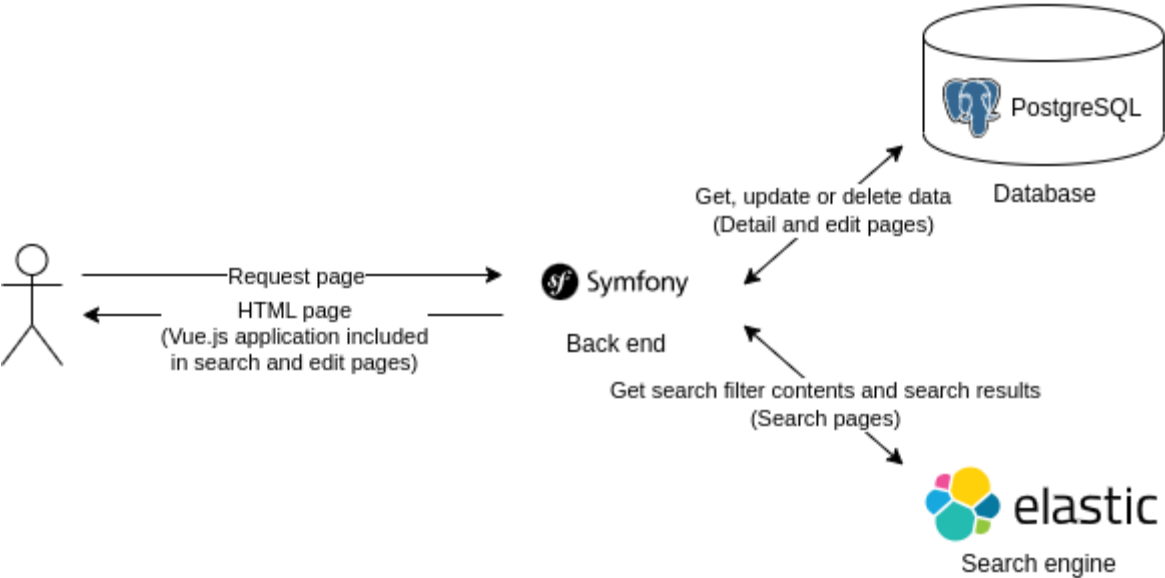


Figure 16: Overall architecture of the current version of the online data platform (initial page loads)

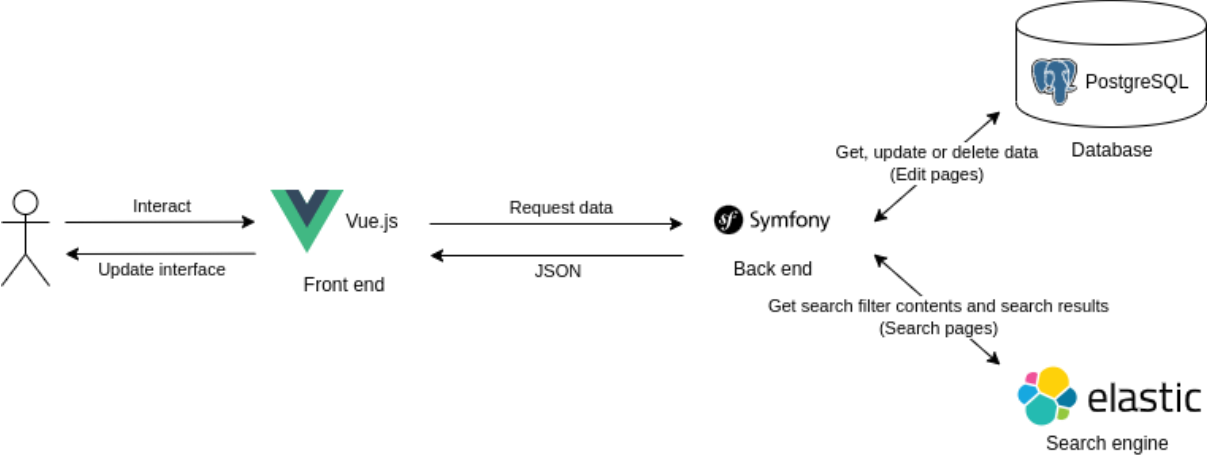


Figure 17: Overall architecture of the current version of the online platform (interactions after the initial page load on edit and search pages)

The engine behind the search pages is Elasticsearch. Its aggregation feature is used to populate the dropdown menus that can be used for faceted navigation. For each object that has hierarchical values in a field (see supra; e.g., manuscript contents), both the value itself and all its parents are added to the Elasticsearch document, making sure that if there is a search query containing a parent field, all objects that contain children of this parent field can be found.

Greek texts from Occurrences and Types are pre-processed before indexation: special characters (round, square, and angular brackets, vertical pipe, and plus) are removed; accents are removed using the ICU Analysis plugin; and all characters are converted to lowercase. The same operation happens with search strings entered by end users, leading to more expected search results. Elasticsearch is also extensively used to help find verse variants. Two queries are executed to create a list with suggested verse variants: a query to retrieve 10 groups containing the most similar verses and a query to retrieve 25 ungrouped verses that are most similar. In the editing environment, the editor can link any of the results returned by these two queries together in an existing or new verse variant group, as illustrated in Figure 18. An additional tool was developed right after the introduction of verse variants, where verse variant groups were suggested for all the verses in the complete corpus, and a single click per group sufficed to create this specific group.

Search linkable verses

Q

Verse	Line	Occurrence	Occ. loc.	Manuscript	(Un)link
Τοῦ πν(εύματος)ς τὰ θεῖα, τόξα καὶ τὰ βέλη:	1	[21817] Τοῦ πν(εύματος)ς τὰ θεῖα, τόξα καὶ τὰ βέλη	f. 508r	SINAI - Mone tes Hagias Alkaterines - gr. 27	-
Τοῦ πν(εύματος)ς τὰ θεῖα τόξα καὶ βέλη,	1	[19854] Τοῦ πν(εύματος)ς τὰ θεῖα τόξα καὶ βέλη	f. 1r	OXFORD - Bodleian Library - Barocci 160	
Τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλη:	1	[20932] Τοῦ πνεύματος τὰ θεῖα τόξα καὶ βέλη	f. 5r	MILAN - Biblioteca Ambrosiana B 106 sup.	
Τοῦ πν(εύματος)ς τὰ θεῖα τόξα κ(αὶ) βέλη·	1	[21149] Τοῦ πν(εύματος)ς τὰ θεῖα τόξα κ(αὶ) βέλη	f. 8v	MILAN - Biblioteca Ambrosiana † 24 sup.	
+τοῦ πν(εύματος)ς τὰ θεῖα τόξα καὶ βέλη	1	[21705] τοῦ πν(εύματος)ς τὰ θεῖα τόξα καὶ βέλη	f. 3v	ROME - Biblioteca Casanatense 1908	
ὁ προφη τὰναξ θείου πν(εύματος)ς πνέων	2	[24555] δα(ν)δ προφήτου, καὶ βασιλέως μέλος	f. 1r	UPPSALA - Universitetsbiblioteket - gr. 24	+
τοῦ γὰρ δυνατοῦ τὰ τεθηγμένα βέλη	5	[17799] ἡ βίβλος αὐτῆ τοῦ κρατίστου δεσπότη	f. 3r-3v	VATICAN CITY - Biblioteca Apostolica Vaticana - Vat. gr. 666	+
τοῦ γὰρ δυνατοῦ τὰ τεθηγμένα βέλη	5	[24641] ἡ βίβλος αὐτῆ τοῦ κρατίστου δεσπότη	f. 1r-1v	MOSCOW - Gosudarstvennyj Istoričeskij Muzej (GIM) - Sinod. gr. 387 (Vlad. 224)	

Figure 18: Part of the editor interface facilitating the linking of verses

II.4 An Open Access Project

The DBBE project has been developed as an open access project from the start. In the current phase of the project, we will go even further by adhering to the FAIR principles (European Commission 2016; Wilkinson 2016), which prescribe best-practices to make data easy to Find, easy to Access, Interoperable and easy to Reuse. Even if we did not specifically focus on these principles in the past, their application is already found in the current platform.

First, data and metadata gathered in DBBE are easily **findable**, as records have been assigned unique identifiers and associated to permalinks. DBBE data can therefore be referred to in an easy and stable way. Furthermore, the URLs resulting from filtering on the search pages can be used to enable the sharing of search queries. A transparent referencing system has also been implemented to identify external resources, both online projects and publications. Table 5 contains a list of all external identifiers currently used in DBBE. Most of these identifiers can be used to search for resources.

Record category	External identifier	Can be used to search
Type	Anthologia Graeca	yes
Type	BEiÜ IV	yes
Type	Paratexts of the Bible (Pinakes)	yes
Type	Vassis ICB 2005	yes
Type	Vassis ICB 2011	yes
Manuscript	Diktyon (Pinakes)	yes
Manuscript	Paratexts of the Bible (Manuscripta Biblica)	no
Person	PBE	yes
Person	PBW	yes
Person	PLP	yes
Person	PMBZ	yes
Person	RGK	yes
Person	VGH	yes
Book, Article	Olivier 1995	no

Book, Article	Olivier 2018	no
Book, Book Chapter, Article, PhD	Vassis ICB 2005 / 2011	no

Table 5: Overview of external identifiers used in DBBE

Secondly, **accessibility** has guided the development of the project. DBBE has been an open access data platform since the first release in 2015. In the second major release of the platform (launched in 2019), several features have been added to enhance collaboration with external researchers and projects. A feedback component has been integrated into all detail pages to facilitate the feedback process. This has ensured that DBBE data are continuously updated and improved also thanks to external input. While team members have access to the backend interface and can edit data, authorisation protocol put in place also allows external users to be granted read-only access to the backend platform.

Interoperability is a key element to ensure that data are integrated with other datasets. This aspect of the data management has so far been fulfilled by means of mutual collaboration agreements with other relevant projects that have resulted in reciprocal references. Among the online databases that refer to DBBE it is worth mentioning the Pinakes database, where around 320 DBBE Type records are listed.³⁵ The online platform [Manuscripta Biblica](#), developed in the framework of the Paratext project, also provides links to DBBE, which is listed as an Inventory from which information is drawn.³⁶ At the time of writing, the online manuscript catalogue of the [Göttinger Septuaginta Project](#) links to 94 DBBE manuscript records, cited among the online sources relevant to retrieve information on manuscripts.³⁷ Other examples of online manuscript catalogues that include references to DBBE in their description are the [Bibliotheca Palatina - digital](#) project³⁸ and the [Medieval Manuscripts in Oxford Libraries](#) catalogue.³⁹

The implementation of automatized systems to link data, by means of APIs, is one of the challenges that lie ahead the further development of the project, as collaboration with other

³⁵ Cross-references to DBBE are recorded as *Identifiants*, such as in <https://pinakes.irht.cnrs.fr/notices/oeuvre/19744/> and <https://pinakes.irht.cnrs.fr/notices/oeuvre/15671/>.

³⁶ See e.g., <https://www.manuscripta-biblica.org/paratext/?id=19741&type=Oeuvres>; <https://www.manuscripta-biblica.org/paratext/?id=15369&type=Oeuvres>.

³⁷ See e.g., https://septuaginta.uni-goettingen.de/catalogue/Ra_13; https://septuaginta.uni-goettingen.de/catalogue/Ra_189. We are grateful to Maria Tomadaki who provided us with these figures.

³⁸ See e.g., https://digi.ub.uni-heidelberg.de/de/bpd/virtuelle_bibliothek/codpalgraec/beschreibungen/bav_pal_gr_139.html.

³⁹ At the time of writing of this paper, 186 DBBE records have been mentioned as online resources: <https://medieval.bodleian.ox.ac.uk/?utf8=%E2%9C%93&q=dbbe>.

projects has so far been realized by manual work, which can result in inconsistencies and lack of mutual synchronization.

Finally, in order to increase the **reusability**, reproducibility and sustainability of the data gathered for the DBBE project, the underlying datasets have been published at <https://doi.org/10.5281/zenodo.7682523> under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.⁴⁰ The source code for the data platform has been published at <https://github.com/GhentCDH/dbbe> under the MIT license. Moreover, the daily work of the team has been thoroughly documented in a dedicated private GitHub repository, also including a Wiki where good practices have been collected.

III CURRENT AND FUTURE RESEARCH OPPORTUNITIES

In its current form, the database is very well suited to retrieve material and to perform queries that have already a specific research question in mind. Examples are: what epigrams can be found in Vat. gr. 1650? In which epigrams is the word εὐσέβεια (“piety”) to be found? What are epigrams typically found in Psalters of the 11th century? This is the type of queries for which a traditional relational database is eminently suited, and it will indeed fulfil the needs of a substantial part of the scholarly community.

However, the corpus of book epigrams brings forward challenges and opportunities that exceed the limits of a traditional relational database. Specifically, relational data stores require a predefined database scheme according to which the data have to be structured, which implies complex transformations of data to fixed tabular formats and potential information loss. Moreover, relational data stores are not efficient in handling complex and numerous relationships among data elements. The multiple interconnectedness of these paratexts and their material carriers, outlined above, as well as the large variety in contextual metadata call for technologies that enable more complex research questions. Instead of generating simple lists of results, it would be interesting to analyse the corpus according to network patterns, nodes, and scales of similarities.

Such an interest is obvious when considering the formulaic nature of the corpus. More complex digital data mining will allow us to perform in-depth analyses to detect connections between patterns of a textual kind on the one hand, and patterns of contextual data (chronological, geographical, codicological, etc.) on the other hand, or even a combination of both. The peculiar nature of the Occurrence records, where the inconsistent orthography used

⁴⁰ <https://creativecommons.org/licenses/by/4.0/>.

by medieval scribes is retained, is an innovative and challenging field of investigation to further develop lemmatisation techniques.

Besides, the methodology of the dichotomy between Occurrence and Type has its limitations. Instead of groups of quasi-identical Occurrences (which would then be categorised under one Type), the corpus displays a nuanced range of similarities between epigrams where the distinction between mere textual variance and a different “text” is hard to make. The danger with the current practice is that an impression of objectivity is created, whereas the choice of creating yet another “type” is sometimes in fact rather arbitrary. More granular and flexible technologies can thus allow us to reflect more objectively textual variance and to leave behind more traditional philological solutions.

The recognition of the various interconnections of our corpus between text, context and materiality, and at the same time of the complex textual patterns that recur through multiple variants of epigrams, inspired the project to mount an interdisciplinary research team where linguists, literary scholars, historians, palaeographers, and engineers collaborate.

Under the aegis of DBBE, four subprojects (including the preparation of as many doctoral dissertations) are therefore currently or shortly being carried out, in order to pursue different yet complementary scientific objectives. These subprojects are subsequently referred to as NLP subproject, Graph subproject, Language subproject, and Book Culture subproject. In the next pages we will show how ongoing and future research will further enhance the contribution DBBE has made to the field of Digital Humanities. Cooperation within the interdisciplinary team will be highlighted as a key element to ensure the advancement of the project.

III.1 NLP Subproject

First, to cope with the problems the inconsistent orthography in the Occurrences causes, a pipeline for linguistic annotation will be developed. Such a pre-processing pipeline consists of three parts: a tokenizer to split the words and punctuation, a part-of-speech tagger to perform morphological analysis and a lemmatiser to provide every word with its lemma. Such pipelines already exist for ancient Greek (Crane 1991, Keersmaekers et al. 2019) but those are dictionary-based approaches that cannot deal with inconsistent orthography or out-of-vocabulary words. For example, a dictionary-based approach will be able to analyse βιβλίου as a genitive from βιβλίον (“book”), but the variant βοιβληου, as it occurs in [Occurrence 32232](#), will not be recognised as such. To deal with this inconvenience, we will develop a

machine learning approach for the morphological analysis and lemmatisation of our corpus. The underlying language model will convert every word within our database into a vector. Since such a vector is based on the context in which that specific word appears, the vectors of both βιβλίου and βιβλίου should be nearly identical and thus recognise both as a form of βιβλίον. The development of this pre-processing pipeline also enables more complex queries of the DBBE, e.g., queries on syntactic patterns or specific morphological forms. Once the pipeline is developed and all epigrams in the database are linguistically pre-processed, the representation of the intertextual relations can be tackled. To find related verses and epigrams, several similarity detection techniques need to be developed, which can be divided into two categories, namely orthographic and semantic similarity measures. The orthographic similarity techniques, on the one hand, measure how many characters or words differ between verses to compute how similar those are. Semantic similarity, on the other hand, measures the distance between the word vectors mentioned in the previous paragraph to determine how similar two verses are. For example, the only lexical difference between the first verse of Occurrences [29724](#) and [18306](#) is the alternation of βλέπειν (“to look”) and ἰδεῖν (“to see”). The meaning of those verbs is not that dissimilar, neither are the contexts in which they appear, so the vectors of both words will be quite similar. Even more adventurously, semantic similarity can be explored over language barriers, as book epigrams circulated not only in Greek but also in numerous other languages (e.g., Arabic, Quiring-Zoche 2013, or Latin, Bouveret 1965-1982). We will perform a pilot study on building cross-lingual vector representations, in order to find similar epigrams in other languages as well. These cross-lingual embeddings rely on the inherent similarities in language structure and composition to project monolingual embeddings in the same vector space based on a small bilingual seed dictionary for the given language pair. Preliminary research has applied cross-lingual embeddings to evaluate sentence similarity (Bjerva & Östling 2017).

III.2 Graph Subproject

Secondly, in order to overcome the limits of relational databases, as outlined above, the Graph Subproject focuses on developing a novel graph database. Unlike the tabular structure of relational data stores, graph databases (Angles & Gutierrez 2008) store data in nodes, which can be connected by edges. Both nodes and their connections can have their own property structure. Hence, a fixed database structure is no longer a requirement and nodes can be traversed efficiently via the edges. The database will consist of nodes representing (parts of)

instances of Occurrences as well as nodes containing contextual and linguistic information. Three categories of relationships, connecting these nodes, will be introduced: (1) edges representing the structure of the Occurrences by connecting textual nodes, (2) links between textual nodes and corresponding contextual or linguistic nodes and (3) connections representing variations of texts or metadata by means of similarities. This new graph database will coexist with the current relational DBBE in a polyglot database system (Khine & Wang 2019). We will construct a pipeline to automate the data extraction, transformation and loading (ETL) process from the relational DBBE to the graph database as well as a synchronisation procedure to guarantee consistency between the two databases. As a result, the graph database will provide users with an innovative visual instrument that facilitates exploration and analysis of related textual and contextual data in the form of nodes.

Taking advantage of this new graph-like structure we will develop three (sets of) advanced facilities supporting the handling of more advanced research questions. First, we will create advanced similarity measures between subgraphs, based on the orthographic and semantic similarity measures between verses (see III.1) as well as on additional textual and contextual data. These similarity measures will allow us to measure the similarity between e.g., (subsets of) Occurrences or even entire Manuscript records. Second, we will develop relevance measures for subgraphs. These measures will make it possible to automatically detect the most relevant words or (half)verses in a subset of Occurrences. Third, we will construct a set of pattern recognition techniques, finding subgraphs of connected nodes (both textual and contextual) that often occur together and reveal hidden knowledge that cannot be detected by conventional querying techniques. For example, a pattern might reveal that a group of epigrams containing a specific set of words are all connected to the same person, origin, period and/or subject.

The development of this subproject is informed by the idea of open data. We will investigate and develop intuitive and practical interfaces to make data publicly available. As a guideline, we will use the FAIR principles (see above, II.4). We will make sure that the new interfaces require little technical knowledge to use, but at the same time offer sufficient flexibility to query this novel database system.

III.3 Language Subproject

Above we have commented on the “formulaic” as well as “fluid” nature of book epigrams. The main aim of the Language subproject is to gain a better understanding of the formulaic

patterns occurring in the corpus, by combining the novel digital possibilities offered by the NLP and Graph subprojects with modern linguistic insights.

Ancient/Byzantine Greek formulaicity has received attention from two different perspectives: on the one hand, scholars have investigated formulaic aspects of oral poetry, Homer in particular (e.g., Bakker 2005); on the other hand, the formulaicity of non-literary genres such as letters and petitions, particularly in the Post-classical and Byzantine periods, has also received in-depth study (e.g., Nachtergaele 2015). Byzantine book epigrams constitute an interesting go-between, because they are poetic on the one hand, but are much shorter and more ephemeral than the oral poems that have been handed down to us. Even though the formulaic nature of the epigrams has been observed on various occasions (e.g., Bernard & Demoen 2019), it has never been studied in any detail. With our linguistic subproject, we want to have a closer look at two sets of questions.

First, we want to get a better grasp of what the building blocks of Byzantine book epigrams were, or put differently, which kinds of formulaic sequences can be encountered in this corpus. We understand formulaic sequences in a broad sense, as ‘the usual phrasings of a speech community’ (Buerki 2016: 15), including idioms, collocations, formulas, proverbs, etc. Rather than applying such a predetermined typology, we want to describe the differences between the formulaic sequences retrieved in our corpus in terms of relevant parameters such as *length*, *lexical specificity*, *combinability*, and *prosody*. In order to do so, we want to make use of the technology that the DBBE-project has at its disposal, such as the verse variants page mentioned above, but at the same time we want to explore the novel NLP and graph-based technologies the other subprojects are developing.

The second question that we want to address is the relationship between formulaicity and creativity. As we already mentioned, scribes did not always stick to fixed patterns, but combined (parts of) epigrams with each other, introduced novel lexical and morphological features, changed the order of words, etc. Our intention is to use graph-based technology to detect and visualize outliers, in combination with close reading of the texts, and to create a typology of the sorts of variation we find. At the same time, we also want to study the social motivations behind such alterations: previous research (e.g., Wray 2002) has argued that formulaic language has a fixed number of functions, such as a reduction of effort for the speaker/writers, the marking of structure in conversation or discourse, and the manipulation of the hearer/reader, including how the hearer/reader perceives the speaker/writer’s identity. We want to explore in particular the interpersonal function of formulaic creativity in book epigrams, by looking closer into aspects of both the micro-context (is there a connection with

the manuscript itself, e.g. in terms of the text that is being handed down or the intended readership?) and the macro-context (are archaic forms substituted by more innovative ones, or vice versa; is there an overlap with other types of epigrams; can certain innovations be linked to the rhetorical, literary, or non-literary tradition?). In this regard, we intend to closely collaborate with the book culture subproject, which will similarly be exploring socio-cultural questions.

III.4 Book Culture Subproject

The Book Culture subproject investigates the material entanglement of book epigrams in manuscripts. It focuses on book epigrams as the “nodes” between the material realization of texts and their intellectual or spiritual significance in society. Its main goal is to reveal attitudes in Byzantine culture towards book production and consumption, and to understand the social embedment of manuscripts. Book epigrams were a preferred forum for individuals, institutions, and communities to formulate their intentions when producing and/or consuming specific texts (often connected to donations). Book epigrams can thus contribute a valuable new dimension to the existing research into social contexts of Byzantine manuscripts and cultures of literacy. Moreover, the narrow relation that book epigrams (as paratexts) have with the main text in the manuscript is a source for valuable new insights into various reading strategies and levels of interpretation.

To grasp the complexity of this interconnectedness, this subproject interprets patterns of similarities between different sets of data (textual and contextual), for which a graph database is a promising new tool. In other words, instead of analysing one subset of data or metadata, this project analyses relationships between those sets, and typical patterns that emerge from bundles of relationships.

(1) A first set of relations is that between book epigrams and main texts (i.e., the main works gathered in a manuscript in which book epigrams occur). Typical patterns of discourse that book epigrams present about a certain group of main texts (be it Psalters or a canonical classical Greek author) may reveal attitudes towards the interpretation of texts and the authority of certain texts.

(2) A second set of relations is that between book epigrams and the materiality of a manuscript. Attitudes towards reading (especially in the case of foundational or contested texts) have their impact on the material outlook of the manuscript: its design and structure, the

types of script, the use of ink and decoration, all of which are data meticulously preserved in the database.

(3) A third set of relations is that between paratexts and historical contexts. How can we relate patterns of textual similarities (as established in the NLP and Language Subprojects) to concrete historical contexts, and identifiable social, intellectual and spiritual communities? This question pertains to the more concrete historical metadata registered in DBBE together with each record of metrical paratexts: regional provenance or specific monastery, networks of related scribes and/or patrons, etc.

Based upon the newly developed database tools to visualize and analyse big sets of data, this subproject constitutes the ideal testcase to apply the novel computational approach to an analytical model in which philological, palaeographical and codicological facts are interpreted within a wider cultural, spiritual, and historical framework. Since book epigrams often thematize the relationships between textual, contextual and extratextual data, they are an excellent research object for such an enterprise.

IV CONCLUSIONS

The experience of the DBBE project, whose achievements and challenges have been described in this paper, can constitute an excellent reference point to other projects in Digital Humanities. Being at its third round of funding, the project has a rather exceptionally long life. This has allowed us to adjust the scope of DBBE in order to answer new, sometimes unexpected research questions. Moreover, the research approaches adopted to collect and analyse textual and metatextual data have evolved throughout the years, along with the ongoing technical development.

Although our project has a clear scope and focuses on a specific corpus, the approach we have used can potentially be extended to other corpora as well. The DBBE data platform has been designed to accommodate textual and metatextual data concerning evidence from languages other than Greek. Moreover, the conceptual design and the code base of the data platform have inspired and contributed to other projects (even not textually oriented ones) developed by the Ghent Centre for Digital Humanities, and vice versa.

Like many other projects that have put a digital platform at the centre of their research activity, we are also faced with the crucial question of long-term sustainability. To which extent a digitized corpus, compiled according to current practices and standards, will still be useful, beneficial, and appealing within a few years, let alone decades? How to respond to the

development of novel digital approaches in a timely, yet not too time-consuming manner? The solutions we have proposed to tackle these challenges have often been guided by a pragmatic approach. Being faced with time and resource constraints, we have chosen to invest in the improvement of the textual and metatextual information provided in our database records, as well as in the design of state-of-the-art digital platforms. Yet, this choice has implied that, so far, we could not focus on other potentially interesting options, such as implementation of encoding standards (e.g., TEI), which would have required too considerable an investment when compared to the academic return we anticipated at the time the project started.

The investment in digital technologies carries by its very nature the risk of becoming outdated. From the perspective of scholars in the Humanities, this limitation has been and will be overcome by producing “traditional” publications aside digital ones. These two complementary approaches can easily become a strength. On the one hand, current and future investigations and analyses within manuscript studies would be simply impossible without digital corpora. On the other hand, a fair record of traditional publications is still a necessary channel to convey standardised and reproducible data.

AUTHORS’ INFORMATION

RR drafted the structure of the paper, coordinated the writing of the manuscript, and wrote sections II.2, parts of II.4, and IV. KB wrote section III.3. FB wrote section I, the introduction to section III, and section III.4. AB was co-supervisor of the design of the current DBBE. EDP revised the paper, especially sections I and III.4. PDP wrote section II.3, parts of section II.4, and has developed the current DBBE. GDT revised the paper with special attention for the introduction of section III and section III.2, and was co-supervisor of the design of the current DBBE. IDV wrote parts of sections III.1 and III.2. MD wrote section III.2 and parts of the introduction to section III. KD is the director of DBBE since the start of the project; he revised the paper, especially sections I, II.1-2, III (introduction) and IV. EL wrote parts of section III.1. AR wrote section II.1 and parts of section I.2. CS wrote parts of section III.1.

ACKNOWLEDGEMENTS

The project has been generously funded by the Hercules Foundation of the Flemish Government (2010-2015) and by the Special Research Fund of Ghent University (2015-2020);

2021-2026). The development and hosting of the DBBE database is supported by the Ghent Centre for Digital Humanities, a Core Facility at Ghent University.

We are grateful to the PASSIM project team members for the kind invitation to present the DBBE project at the inspiring sessions on “Life and Afterlife of Digital Projects in Medieval Manuscript Studies” at IMC (July 2022). We wish to express our gratitude to Gleb Schmidt for his wonderful guidance throughout the submission and review process of this paper.

REFERENCES

- Andrist P: Toward a Definition of Paratexts and Paratextuality: The Case of Ancient Greek Manuscripts. In *Bible as Notepad. Tracing Annotations and Annotation Practices in Late Antique and Medieval Biblical Manuscripts*. Edited by Lied LI, Maniaci M. Berlin, Boston: De Gruyter; 2018:130-150.
- Angles R, Gutierrez C: Survey of graph database models. *ACM Computing Surveys* 2008, 40 (1):1-39.
- Bakker EJ: *Pointing at the Past: From Formula to Performance in Homeric Poetics*. Hellenic Studies 12. Washington, D.C., Cambridge (MA): Harvard University Press; 2005.
- Bernard F, Demoen K: Byzantine book epigrams from manuscripts to a digital database. In *From Ancient manuscripts to the Digital Era: Readings and Literacies, Proceedings*. Edited by Clivaz C, Meizoz J, Vallotton F, Verheyden J. Lausanne: PPUR; 2012:431-440.
- Bernard F, Demoen K: Book epigrams. In *A companion to Byzantine poetry*. Edited by Hörandner W, Rhoby R, Zagklas N. Leiden: Brill; 2019:404-429.
- Bénédictins de Bouveret: *Colophons de manuscrits occidentaux des origines aux XVI^e siècle*. Fribourg: Editions universitaires; 1965-1982.
- Bjerva J, Östling R: Cross-lingual Learning of Semantic Textual Similarity with Multilingual Word Representations. In *Proceedings of the 21st Nordic Conference on Computational Linguistics*. Association for Computational Linguistics. Gothenburg; 2017:211-215.
- Buerki A: Formulaic Sequences: A Drop in the Ocean of Constructions or Something More Significant? *European Journal of English Studies* 2016, 20 (1):15-34.
- Crane G: Generating and parsing classical Greek. *Literary and Linguistic Computing* 1991, 6(4):243-245.
- Demoen K: Reading: book epigrams: introduction. In *The visual culture of later Byzantium (1081-c.1350): Vol. 3.II*. Edited by Spingou F. Cambridge: Cambridge University Press; 2022:1369-1380.
- European Commission: Guidelines on FAIR Data Management in Horizon 2020. [https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf].
- Genette G: *Seuils*. Paris: Seuil; 1987.
- Hutter I: *Corpus der byzantinischen Miniaturenhandschriften. Oxford Bodleian Library I*, vol. 1. Stuttgart: Anton Hiersemann; 1977.
- Keersmaekers A, Mercelis W, Swaelens C, Van Hal T: Creating, Enriching and Valorizing Treebanks of Ancient Greek. In *Proceedings of the 18th International Workshop on Treebanks and Linguistic Theories (TLT, SyntaxFest 2019)*. Association for Computational Linguistics; 2019:109-117.

- Khine PP, Wang Z: A Review of Polyglot Persistence in the Big Data World. *Information* 2019, 10(4):141.
- Nachtergaele D: The Formulaic Language of the Greek Private Papyrus Letters. *PhD thesis*. Ghent University; 2015.
- Parpulov G: *Toward a History of Byzantine Psalters*. Plovdiv; 2014.
- Quiring-Zoche R: The Colophon in Arabic Manuscripts. A Phenomenon without a Name. *Journal of Islamic Manuscripts* 2013, 4(1):49-81.
- Rhoby A: *Byzantinische Epigramme in inschriftlicher Überlieferung*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften; 2009-2018.
- Stussi A: *Fondamenti di critica testuale*. Bologna: Il Mulino; 2006.
- TLG: Thesaurus Linguae Graecae. [<http://stephanus.tlg.uci.edu>].
- Vassis I: *Initia Carminum Byzantinorum*. Berlin, New York: De Gruyter; 2005.
- Vassis I: Initia Carminum Byzantinorum Supplementum I, *Parekbolai* 2011, 1:187-285.
- Wilkinson M et al: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 2016, 3, 160018.
- Wray A: *Formulaic Language and the Lexicon*. Cambridge, New York: Cambridge University Press; 2002.