Parsing the Papyri: The Potentials of Wikidata for Papyrologists

Jasmine A Sahu-Hough
Yale University, United States
jasmine.sahu-hough@yale.edu

***Abstract*— Papyrologists have developed many digital tools over the years but have not yet engaged robustly with Wikidata. In seeking to gather together all the material excavated in the Syrian city of Dura-Europos in this LOD environment, the International (Digital) Dura-Europos Project (IDEA) has offered a perfect opportunity to explore Wikidata's utility for representing papyri and parchment. This paper details the process of developing a data model for the Dura-Europos papyri and reflects upon both the challenges and potentials revealed by the exercise. It is concluded that if the papyrological community becomes more involved in developing the representation of data on Wikidata, the platform could offer an exciting way to open the field to broader conversations within the digital realm.**

***Index Terms*— Linked Data, Ontologies, Humanities, Data modeling**

I. Introduction

The International (Digital) Dura-Europos Archive Project (IDEA), formerly the Yale Dura-Europos Archive Project (YDEA), aims to collect the materials excavated from Dura-Europos under one digital roof, using the Linked Open Data (LOD) venue of Wikidata for the purpose [1]. The significant quantity of papyri and parchment texts included among these artefacts has offered the opportunity to explore how such objects may be modelled on this platform. Papyrologists have developed many digital tools over the years but few have been active on Wikidata so far. Although the (ongoing) process of creating a data-model has laid bare the challenges in mapping papyrological categories to Wikidata, it has also highlighted, as this paper aims to show, the numerous rewards to be gained by overcoming such issues.

*

Dura-Europos was an ancient Syrian city, founded around 300 BCE and destroyed in a siege in 256 CE. It was excavated in the 1920s and 1930s, first by Franz-Valéry-Marie Cumont, and subsequently by a joint team from the French Académie des Inscriptions et Belles-Lettres and Yale University [2]. The finds recovered from these excavations were distributed between institutions in Syria, France, and the USA through a partage agreement [3]. The dispersal of these materials has had a profound effect on research into Dura-Europos. Whilst their physical separation limits the ability of scholars to understand objects in relation to one another, the multiple languages in use by the institutions involved often curtails key-word searches [1]. We see this even in the name of the site itself. Researchers exploring the website of the Yale University Art Gallery must refer to it as 'Dura-Europos' in order to find artifacts recovered from the city in the museum's online catalogue. For scholars accessing the database of the Bibliothèque nationale de France, however, the spelling 'Doura Europos' (*sic*, without hyphen) must be used in order to return any hits at all. Linguistic barriers have also prevented local communities in Syria from accessing information about the site, leading to a sense of alienation [4]. Interviews conducted by Jen Baird and Adnan Al Mohamad with local residents at Salhiyeh (the modern town neighbouring Dura-Europos) have highlighted the need and desire for resources accessible in Arabic for a non-specialist audience [5]. The aim of the IDEA project is to digitally reassemble both the finds and archival materials, allowing them to "'speak' across institutional borders" [1]. The project selected Wikidata to serve as its LOD venue because it operates a low barrier-to-entry and supports contributions in multiple languages, thereby encouraging international contributions [1]. In addition, the widely recognized complexity of other ontologies, such as CIDOC-CRM [6] [7] [8], made a locally hosted database utilizing such vocabularies seem less suitable for promoting the kind of wide accessibility envisioned by the IDEA project. Eventually, the project aims to use Wikidata as a dynamic backend for an independent web-interface, which will provide access to the materials in a user-friendly way for a non-specialist audience.

The additional advantage of this choice was that it offered an opportunity to experiment with using Wikidata for representing parchment and papyri texts. More than 180 such items were recovered from the excavations. The majority were brought to Yale [9], making this collection a small but significant subgroup within the Dura-Europos materials. As with the numerous inscriptions recovered from the ancient city, moreover, the parchment and papyri present a modeling challenge because they can be understood as both artifacts and texts, and because there is a long history of digital representation to account for. Existing digital tools, such as Papyri.info and the Trismegistos databases, incorporate metadata but use a wide range of idiosyncratic terms (in a limited pool of languages) to refer to concepts such as text genre or date range. Only some of these, meanwhile, map intuitively onto the properties and items available on Wikidata. In the following sections, I will reflect on the process of developing a data-model by first sketching the landscape of digital papyrology and the structuring of information in Wikidata. I will then provide a summary of a selection of the modeling decisions taken and discuss how this framework may enhance the utility of Wikidata for the papyrological community in the future.

II. The Landscape of Digital Papyrology

A. Current Digital Tools

Two key fields in the Humanities are particularly concerned with papyri: Egyptology and Classics.[1] Parchment texts (the other main writing material used in pre-modern times before paper) also concern scholars, especially those working in Medieval Studies. Papyrological engagement with digital tools in these fields began as early as the 1960s, with an entire session on 'Computer uses in Papyrology' scheduled at the 12th International Congress in 1968 [10]. Since that time, each field has developed their own set of digital resources (some of which partially overlap with one another) ranging from online catalogues, text corpora, ontologies, and databases.[2] Papyrology is thus one of the best served areas of ancient world studies in the digital realm. For the purposes of this paper, three are especially important to discuss: Trismegistos, Papyri.info and ThOT (Thesauri and OnTology for documenting ancient Egyptian textual resources). On the side (primarily) of Classics, two key sites, Papyri.info and Trismegistos have absorbed or provide links to many of the previous offerings [11], making them the first port of call for anyone conducting research in the field [12]. ThOT, meanwhile, is a more recent development in Egyptology, although it has close connections with the more established TLA (Thesaurus Linguae Aegyptiae) [13].

Trismegistos, launched in 2006, serves primarily as a repository for metadata [10], although many of their earliest imports also included text transcriptions. At a basic level, Trismegistos allows the user to search for specific texts (each of which has a Trismegistos (TM) number as a unique identifier) and to retrieve metadata about them (such as dating and bibliography). More advanced features include the numerous embedded thematic databases, such as TM Archives (which groups texts according to the archives they are believed to have belonged to in antiquity) and TM People (which gathers attestations of names and specific individuals), that enable various quantitative analyses to be conducted [10]. Trismegistos also provides multiple ways to visualize such analyses, ranging from charts to maps [12]. Hosted at Leuven [10], the platform has required paid subscription to access the full range of its functions since January 2020 [12].

Papyri.info, developed in 2010 [12], offers the user a combination of basic metadata and digital text editions. Their papyrological navigator allows the user to conduct word searches across texts from a large number of collections in Europe and North America or to call up editions by narrowing a small range of metadata fields, such as language or date [14]. Originally funded by the Mellon Foundation, the site is now entirely reliant on volunteer labour both as editors and contributors, although two of the major associations of papyrologists, the Association Internationale de Papyrologues and the American Society of Papyrologists, have recently launched a joint call to establish an endowment for the resource [12].

The data model for ThOT was first completed in 2015-6 [15] and is described as "a multilingual repository of resources for metadata enrichment and data exchange" [16]. These resources consist of Simple Knowledge Organization System (SKOS)-compliant thesauri detailing an ideal framework for describing inscribed objects (not just papyri, but also epigraphic texts) [16]. ThOT operates similarly to CIDOC CRM and the Text Encoding Initiative (TEI) in providing other projects (hosted externally) with a framework for encoding their metadata such that it will be interoperable with different projects. It also hosts URIs for the concepts within its thesauri, thereby improving the usefulness of the project for Linked Data initiatives. Funded by initially by the Anneliese Maier Research Award, granted in 2015 at the 11th International Congress of Egyptologists by the Humbolt Foundation [15] [17], the project is housed at the Department of Egyptology at the University of Liege [17].

B. A Full Complement?

With such excellent tools available, the reader may well ask what need or room there is for any additional digital resources. It must be acknowledged, however, that Trismegistos, Papyri.info and ThOT have their limitations.

For Trismegistos and Papyri.info, the approach of each remains closely connected to the modes of representing papyri in print literature. In both, text exists separately from image, mirroring, as Lucia Vannini has pointed out, the relegation of plates to the end of edited volumes [12]. On Papyri.info, moreover, the text editions provided keep faithfully to the format and apparatuses that have been well-known in papyrology for generations [12]. While this has the advantage of being familiar to papyrologists, the approach does not take full advantage of the possibilities for semantic mark-up available in the TEI compatible EpiDoc XML files in which Papyri.info's editions are stored [12]. These editions present only static text, in contrast to the dynamic interaction fostered on the early imports on Trismegistos (where the user can select words and be transported to a database detailing the use of the same word in other texts). Neither make use of existing ontologies such as CIDOC CRM or FRBRoo (or indeed ThOT, although the fact that this data model has only recently been

[1] *Pace* scholars working on Arabic texts on papyrus. This is an exciting and expanding field that I have been unable to take into account for this article. Offerings for the field of Digital Papyrology in Arabic include *The Arabic Papyrological Database* [Online] Available: https://www.apd.gwi.uni-muenchen.de/apd/project.jsp

[2] Some especially notable recent digital offerings, which will not be discussed here, include the Greek Schools Project, Critical Editions for Digital Analysis and Research (CEDAR), PalPap, and the Demotic Palaeographical Database Project (DPDP). In particular, the Greek Schools Project, funded by the ERC, aims to offer a new kind of digital text edition which may potentially overcome some of the limitations of digital editions discussed here when the application is launched. A further exciting development in digital papyrology is the use of Optical Character Recognition (OCR) and Natural Language Processing (NLP) in AI tools to facilitate the transcription of texts. Applications such as Transkribus may well revolutionize the field in the future. I have refrained from discussing this area in the current article, however, since the focus here is on the representation of metadata rather than text transcription.

developed makes the omission more than understandable), meaning their categorisations of metadata lack interoperability with other cultural heritage projects, despite the existence of a specific CIDOC extension for representing ancient texts (CRMtex) [18]. Trismegistos and Papyri.info, therefore, largely float free from broader initiatives in digital humanities, outside of ancient world studies. Lastly, these resources perpetuate the modern linguistic biases that formed a traditional aspect of papyrology, since their outputs are primarily in English or German.

ThOT, meanwhile, fares better on the point of interoperability because many of its classes can be mapped to either CIDOC CRM or TEI [16]. The correspondence is not exact, however, and the very existence of this separate initiative bespeaks a perception of Egyptological material as particular and distinctive from other types of objects. For example, Stéphane Polis and Vincent Razanajao explain that their conceptualization of a *Document* diverges from the CIDOC CRM element E84 Information Carrier because the latter is designated as a subclass of E22 Man-Made Object, which does not match the relationship between a *Document* and an *Object* in ThOT [16]. The element *Witness* in ThOT, meanwhile, maps onto multiple elements in the TEI [16]. Such a perspective is, of course, understandable from the point of view of field specific scholars, whose deep knowledge of the subject makes them alert to details that differentiate their material from all others. From the perspective of linked data, however, this approach is limiting. The multi-lingual outlook of the thesauri offers an improvement over the linguistic landscape of the other digital tools discussed, but the terms are still limited to just four languages (German, English, French and Dutch), although they intend to offer Arabic as well [17].

In addition to these shortcomings, all three platforms raise concerns about access and sustainability. As noted above, Trismegistos no longer grants open access to all its services, meaning that researchers wishing to use the embedded databases and visualizations must rely on institutions to purchase subscriptions. Although the site still offers "details pages" for non-subscribers, all such services remain contingent on attracting enough subscribers to maintain the server costs, which, as the Trismegistos team notes, "is not self-evident" [19]. Papyri.info, on the other hand, fares better with regard to access – all of its data is freely available both through its own interfaces and in its github repository – but suffers when it comes to long-term stability. Its current infrastructure rests heavily on the voluntary activity of editors such as James Cowey [10], and while the plans to establish an endowment offer some hope for the future, it remains unclear how substantial or sustainable such a funding model will be. ThOT, meanwhile, is an open access project, but depends on grant funding and hosting by the University of Liège to maintain its server. Whilst the information within its data model can be easily downloaded and maintained in other places, one of the great advantages of the project at present is that it hosts URIs for all of its elements, which is crucial for its utility to linked data projects. If the University of Liège were no longer able to host the site, however, these URIs would be compromised.

III. LOD in Wikidata

Described as a "collaboratively edited knowledge graph" [7], Wikidata offers a LOD environment built on the principle of RDF (Resource Description Format) semantic triples – statements with a subject (item), predicate (property), and object (value), each element of which refers to a uniquely described concept, differentiated by its own identifying code (Q- or P-number). Qualifying properties (known as 'qualifiers') can be added to these statements to further refine the information represented. Linking each unique concept to a code means that the ontology of Wikidata can be described as "language independent" [20], which enables its structure to be better understood by computers. For human use, however, the labelling of concepts also operates in a multi-lingual setting, since most basic descriptors (such as 'human', 'book', 'name', 'country of origin' etc) have been translated into a very wide array of languages; indeed, Wikimedia supports 326 languages [21], even if not all of them are used as consistently in Wikidata as others [20]. As a result, even where an individual item (such as a specific papyrus fragment) is only labeled in one language, the statements that structure its overall description will incorporate a range of languages by default. A major advantage of this linguistic approach is that Wikidata's querying service, an implementation of SPARQL (SPARQL Protocol and RDF Query Language), can be deployed in the user's own language and produce the same results as a user operating in any other language. The power of this aspect of Wikidata can, of course, be enhanced by inputting multi-lingual data on an item level. In this way, the problem identified earlier concerning the name of Dura-Europos would be easily eliminated. Provided that the item page for the Syrian city, Dura-Europos, also included the French designation, 'Doura Europos' among its translated titles, a user querying in French would be brought to the same item page as a user querying in English.

The system offers almost unlimited possibilities for interconnection, not least because Wikidata represents only one prong of the multi-faceted Wikimedia Foundation, which also includes sites such as Wikipedia, and Wikimedia Commons [7]. Through such connections, items on Wikidata can be connected with images hosted on Wikimedia Commons, or encyclopedia entries expanding on the concepts described on Wikipedia, whilst the ideas represented on each of these sites can in turn be linked to the item page delineating those concepts on Wikidata. As this networked structure hints, a similar dynamic relationship can be established between Wikidata and external sites, with Wikidata serving as a backend database for a whole host of elaborate representations. A good example of this can be seen in sciencestories.io, a website application built using Wikidata APIs to gather information about female scientists into narrative units [22] [23]. As its creators reflect, the use of Wikidata for the backend means that the knowledge curated by the project is inherently dynamic, able to shift in accordance with the development of information on Wikidata itself [22].

As should be apparent from this description, anyone seeking to input information into Wikidata must take a dynamic approach to their material. In order to take advantage of the potential for connection, each category requires interrogation so that its expression may both link broadly and represent accurately the concept deployed.

IV. Creating a Data Model for Papyrology

A. *Papyri on Wikidata before IDEA*

Excluding the items added by the IDEA project, a Wikidata query for papyri yields only 610 results [24]. Compared to the 71,549 papyri yielded by a similar search on Trismegistos, the dearth of material on Wikidata becomes readily apparent [25]. If one examines the descriptions more closely, moreover, a key issue emerges – consistency. Some items, such as Papyrus Oxyrhynchus 27 [26], neglect to include even basic information such as an inventory number, whilst others, such as Papyrus 42 [27], offer a fuller set of statements but label themselves ambiguously ('Papyrus 42' refers to a Gregory-Aland number, an identifier only scholars of the New Testament are likely to use regularly). In this state, therefore, papyri on Wikidata afford little benefit to papyrologists or researchers of any kind since queries for any information beyond the material of the item could not guarantee to produce results. A standard data model must be developed before the potential of Wikidata for papyrology can even be tested.

B. *A Sketch of the Modeling Process*

The model being developed as part of the IDEA project has evolved into a complex, fine-grained schema. For the purposes of this paper, therefore, I will focus on the way a limited number of issues have been approached: the relationship of texts, fragments and witnesses, text genres, find-spot, modern acquisition history, editors, and identifiers. As the IDEA project as well as my own experimentation with Wikidata and papyrological data progresses, the model may change further. The following should be understood as a snapshot of the project's modeling decisions to date. Readers interested in exploring the full data model as it develops can find a read-only version at [28].

1) *Texts, Fragments and Witnesses*

Papyrus and parchment are frequently preserved in fragments, rather than intact sheets or entire rolls, and the writings they carry may be copies (witnesses) of works that also appear on other fragments. This situation presents an issue for how to model the relationship between these elements. Papyrology has traditionally been liable to ignore issues of fragmentation and to publish multiple fragments together, usually because the editors of these publications are able to argue that the fragments are pieces of one original whole. In this way, a group of fragments may be assigned a single publication number. The catalogues of papyri and parchment produced by the institutions that hold the fragmentary items are commonly found to perpetuate this issue. Among the Dura papyri, for example, P.CtYBR inv. DP 87 has multiple fragments that are all catalogued under a single inventory number [29]. The practice is not consistent, however. P.CtYBR inv. DP 82 [30] and P.CtYBR inv. DP 85 [31] are published together as P. Dura 42 [9], but have different inventory numbers in the catalogue. The problem in essence revolves around whether one privileges the text that is written on a papyrus or parchment, or the papyrus or parchment object itself.

Different digital projects have dealt with this issue in contrasting ways. As might be expected, based on the analysis above, Papyri.info and Trismegistos tend to perpetuate the publication approach. Trismegistos assigns a unique number to each record that they identify as constituting a single document. They argue that "all texts written on what was in antiquity a single writing surface belong together and form one document", with the only exception being where "there are good reasons to believe that the only (and unintended) relation between the two texts is the writing surface itself" [32]. The effect of this definition is that multiple fragments grouped together because they are believed to have formed one writing surface in ancient times are a single document, whilst one sheet of papyrus or parchment that might have an account written on one side, but a section of a novel on the other, will be treated as two different documents.

ThOT takes a different approach. The ThOT Data Model (TDM) separates four concepts: *Object*, *Document*, *Witness*, and *Text* [16]. An *Object* designates "a physically discrete material object", such as a single fragment of papyrus. A *Document* encompasses "an artefact reconstituted in its original entirety", such as the combination of papyrus fragments into the original sheet that would have existed in antiquity. A *Witness* covers "a single occurrence of a *Text*, in its material (and more broadly philological) dimensions", such as a copy of a few verses of Homer's epic poem, the *Iliad*, which might appear on a *Document*. Lastly, a *Text* represents "a textual composition as it can be reconstructed from …. *Witness(es)*", or in other words the abstract idea of Homer's epic poem, the *Iliad* [16]. As Polis and Razanajao detail, ThOT's approach follows (in spirit, if not in every element) the structure proposed by the TEI and CIDOC CRM [16]. Their attitude has the advantage of recognizing the materiality of the text and maintaining the integrity of fragments separate from scholarly assertions of their relationships to one another. In other words, keeping *Document* and *Object* separate acknowledges the possibility that the combination of fragments into a single text is an expression of scholarly opinion (regardless of how justified that opinion might be).

The best option regarding the input of papyri and parchment data into Wikidata would therefore seem to be to follow a similar principle to that adopted by ThOT. Each individual papyrus fragment would be its own item (ThOT's *Object*), to which data relating to its physical manifestation (e.g. dimensions or material) would be attached. These fragment items would then be linked to a separate item for the text represented on the fragments (ThOT's *Document*) through the properties 'part of' (P361), which would reside on the fragment/*Object* item page, and 'has part' (P527), which would attach to the *Document* item page. The relationship of the

Witness, meanwhile, would not be expressed as a separate item but would rather be captured as a relationship between the other items. It would be represented by attaching the property 'exemplar of' (P1574) to the *Document* item page and supplying the *Text*, which *would* need to exist as an item in its own right), as the value of this property. The precise portion of the *Text* that is witnessed by the *Document*, meanwhile, would be captured using the qualifier 'line(s)' (P7421) or 'section, verse, paragraph, or clause' (P958). Fig. 1 provides a graphic illustration of this relationship.

The IDEA Project's papyri on Wikidata are currently in the process of being adjusted to fit with this model. The initial data for the Dura papyri was scraped directly from the Beinecke Library catalogue and uploaded with a skeleton set of statements to Wikidata, meaning that the structuring of their

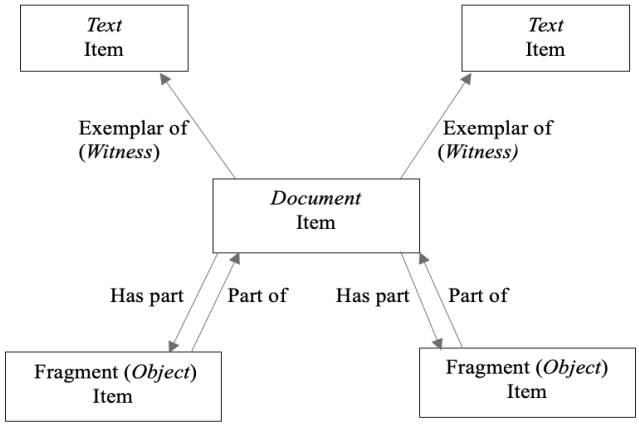

Fig. 1. Graphic representation of the relationship between *Object*, *Document*, *Witness* and *Text* as it might be realized on Wikidata.

data conformed to the (inconsistent) approach to fragments and text prevalent in the catalogue. While this process had the advantage of providing stub items on Wikidata on which subsequent work could be based, it is now necessary to disentangle the statements on these items to better reflect the way *Object*, *Document*, *Witness*, and *Text* relationships should ideally be represented.

Although this disentangling process does of course involve extra work (an inevitable part of the model development process in a relatively new digital project), the pitfalls associated with *not* carrying this work out can be demonstrated by an issue that is currently affecting the Dura papyri on Wikidata. As noted above, the scrape from Yale's Beinecke resulted in the Dura papyri items being based around the way they were inventoried in the online catalogue. The items were simultaneously linked to the TM number associated with their publication. Due to the fact that Trismegistos' definition of a 'document' does not exactly match with the inventorying practices at the library, however, this resulted in six papyri violating the TM text ID property's 'unique value' constraint which requires that only one item at a time should have this property [33]. In order to resolve the issue, the IDEA Project has entered a topic into the discussion page for the TM Text ID, a feature included in Wikidata to allow community members to clarify and/or build consensus around proposed changes to the ways properties may be used on the platform [34].

2) Texts Genres

The conceptualization of a 'genre' presents a crucial issue for modeling. Wikidata does not list many of the classifications of text type commonly used in papyrology as 'genres', reserving that label instead for categories of literary or audio-visual fiction such as 'horror' or 'romance'. We have therefore selected the property 'instance of' (P31) in order to represent papyrological text genres. This property would be added to the *Document* item since it is only through the process of reading and analysis that a papyrus' genre becomes apparent. Although superficially different from the concept expressed in papyrology, 'instance of' offers a huge range of suitable values which match the pre-existing text genres very well. The practice of separating papyri into 'documentary' and 'literary' categories can be encapsulated by 'document' (Q49848) and 'literary fragment' (Q1440453), the latter being selected over 'literary work' (Q7725634) because the texts studied by papyrologists are almost always fragments of works, not full texts. Genres such as 'letter', 'petition', 'receipt', meanwhile, can be readily found as items already on Wikidata; to date, I have not found any type to be missing from the lists used on Papyri.info in either English or German, although further work is required to ensure that every descriptor is exactly matched with a Wikidata item. The more specialized genres identified by ThOT, such as 'Necropolis Journal' (*thot-18752*) [35], are not currently available on Wikidata. In this case, however, it is simply a matter of creating a new item to match the genre identifications made within the discipline.

In order to improve the searchability of items (as well as the potential for interconnection within Wikidata), however, the values used for 'instance of' need to be multiplied. Thus, a papyrus list would need to be identified as both a 'document' and as a 'list' (Q12139612) enabling it to be found in queries looking specifically for letters or for documentary texts more generally. This layering of values is necessary because not every documentary papyrus genre is (currently) categorised as a subclass of 'document'; if they were, only the more precisely designated text-genre would be necessary. Moving forward, the IDEA project hopes to discuss the expansion of document subclasses with the broader Wikidata community.

Literary genres, meanwhile, should be expressed on the *Text* item, since they relate to the work as a whole, rather than the individual *Witness* found on a *Document*. Here the property of genre (P136) can be used, and any genres particular to the ancient world that are not currently represented on the platform simply added.

3) Findspot

Designating a findspot for papyri and parchment is often unnecessary, since the vast majority of collections consist of material acquired by purchase where the location of excavation is unknown. For the Dura-Europos manuscripts, however, findspots are usually recorded within the excavation reports. Within the project, the importance of modeling this information, regardless of the utility of the framework for other

papyrological projects on Wikidata, lies in the possibility of connecting papyri and parchment with other types of excavated objects found in the same location. Beyond the project, as modern excavations unearth more papyri, having a way to represent such information in Wikidata will undoubtedly take on greater importance.

The key property for expressing the idea of a findspot is 'location of discovery' (P189). It would need to be attached to the *Object* item, since it pertains to the spot where the material remains of the papyrus or parchment item was found. In some cases, representing the information provided by the Dura-Europos excavation report with this property was very straightforward. Many were found in 'room W13', located within the Temple of Azzanathkona. Both of these locations have been created as Wikidata items by the IDEA project, with 'room W13' being designated as a room within the temple. For these manuscripts, therefore, describing their findspot merely involved using 'room W13' (Q118390423) as the value for 'location of discovery'. When searching for such objects using Wikidata's querying service, the inquirer should in theory be able to find the items even when looking for them under the broader location designation. We tested out whether this theory worked by writing query [36], which sought to locate all papyri found in the Palmyrene Gate (Q98930725), the main entrance to the ancient city. According to the excavation report, four papyri were located in various locations within this building. Two were found in Room G of the South Tower, one in Room H of the North Tower, and one in the doorway between rooms F and G of the South Tower [9]. The query successfully returned all four papyri, demonstrating that the nesting of location data did allow the papyri to be discoverable even when researchers were not aware of the precise findspot.

Where, however, the excavation report provided not a single spot or area of discovery, but rather a triangulation of multiple location references (such as "in Wall Street back of Block L8" [9]), representing this descriptive configuration was much more difficult. Several solutions were proposed and will be detailed below, along with evaluation of their pros and cons.

The first suggestion, following the example of the text genres, was to layer multiple values. Thus, a papyrus found "in Wall Street back of Block L8" would have two values for 'location of discovery': 'Wall Street' (Q116621880) and 'Block L8' (Q116622036). The rationale for this solution was to ensure the item would be found through queries searching for either of the locations mentioned in the excavation report. Two key concerns arose over this modeling solution, however. First, it was feared that multiple values might imply uncertainty over the correct find spot, rather than a co-ordination of information. Secondly, this solution seemed to involve a significant loss of precision. The excavation report's phrase "in Wall Street back of Block L8" expresses that the papyrus was found in the section of Wall Street which abuts the city block L8 – an important qualification, since Wall Street runs the entire length of the westernmost side of the ancient city. In addition to these important caveats, it seemed likely that the map visualization tool offered through Wikidata's querying service (which was envisioned to be especially helpful for non-specialist users) might not be able to handle multiple values, leading to misleading results (e.g. the same papyrus being represented in several locations).

Nesting location information, so that it would be clear that these values were related to one another and still referred to only one find spot, seemed to be desirable. The next proposal, therefore, was to utilize the most specific, or most important element of the triangulation as the 'location of discovery' and to use qualifier properties to represent the additional information. Thus, for "in Wall Street back of Block L8", 'Wall Street' would be the value used for 'location of discovery', whilst 'Block L8' would be expressed using a qualifier.

Finding the right qualifier for this framework, however, has proved challenging. The property 'adjacent building' (P3032) seems to be the only one designed to express physical adjacency – it includes the term 'next to' among its aliases [37] – and yet the current description of the property on Wikidata stipulates that it is intended for describing a situation where a building is adjacent to an item, not any other type of geographical or location entity [38]. Although the excavation reports do record some find spots according to neighbouring buildings (such as the synagogue), the concepts of adjacency expressed are more usually between areas of the city. 'Significant place' (P7153) was put forward as an alternative property, to be used alongside the qualifier 'object has role' (P3831) and values such as 'adjacent' (Q5403187). A key drawback to this proposition, however, was that where it would be necessary to list multiple coincident locations (e.g. where the excavation reports recorded a find spot at the "back of block L7 and L8") the qualifier pair of 'significant place' and 'object has role' might produce confusion regarding which 'significant place' accorded with which 'object has role' qualifier. In addition (and more importantly), 'significant place' has only infrequently been used as a qualifier (only 26 times to date) and never alongside 'location of discovery'; using the property in this way, therefore, might reduce the discoverability of the find spot information for users familiar with Wikidata.

After conducting further research into the use of 'adjacent building' in Wikidata to date, it was found that numerous users had already utilized the property to express adjacency with entities which were not buildings (particularly churches like Église Saint-Leu de Bellebrune (Q41792531), which had a cemetery as the value). Moreover, on the discussion page for the property, other users had already been engaged in a debate over whether the application should be widened from 'building' to other location entities [38]. The IDEA project has accordingly raised our own particular case in order to persuade the Wikidata community that the property should be altered to make it appropriate for use with our items. When no objections were raised, we went ahead with using 'adjacent building' as a qualifier with 'location of discovery' and intend to request an alteration to the property name with the Wikidata admins. This mode of designating findspot was also subjected to a querying test, this time with query [39] checking whether an inquiry looking for papyri found next to block L8 would also pick up instances where there was uncertainty over whether the findspot was near L8 or L7. The excavation report indicated that there

should be 36 results [9], and this was the number that the query duly returned.

In addition to the issue of findspot triangulation, uncertainty surrounding findspot offered another complication for modeling location information. While the excavation reports provide a wealth of detail, some items seem to have been less well recorded than others, leading to assertions that a papyrus was "probably" found in a particular area, based on the general locations in which the majority of manuscripts were discovered. To express such tentative information, therefore, the qualifier 'sourcing circumstances' (P1480) was added to statements, with the value 'probably' (Q56644435), 'possibly' (Q30230067), or 'presumably (Q18122778), depending upon the language used within the report itself.

The complications outlined in this section would be unlikely to trouble papyri unearthed in modern excavations to the same degree. Not only have archaeological recording standards changed significantly from those practiced in the early 20th century, but highly precise measurement tools are now available to plot such information scientifically. Details such as GPS locations can be readily added to a 'location of discovery' statement using 'co-ordinate location' (P625) as a qualifier.

4) Acquisition History

Since the Dura-Europos materials travelled directly from their place of excavation to the collection in which they are now housed, this concept has not been utilized for the IDEA project. For papyri and parchment acquired through the antiquities market, however, acquisition history would be a very helpful concept to represent. Since the information concerns the physical papyrus item, once again we would expect it to be added to the *Object* item.

To represent an individual who sold or bought a parchment or papyrus item, the best mode of representation seemed to be to designate a 'significant person' (P3342), name the individual (who would themselves need to have an item page on Wikidata), then to qualify this statement with 'object has role', using the value 'vendor' (Q104099828) or 'buyer' (Q1308239). Although this model might seem cumbersome compared to the use of the property 'owned by'(P127), it also allows other types of relationships to be represented in a consistent manner, such as 'observer' (Q28973077) or 'archaeologist' (Q3621491) (if it is possible to determine the individual who excavated the item in question). The purchase itself, meanwhile, can be represented using the 'significant event' (P793) property with the value 'sale' (Q194189). This statement can then be qualified with the property 'point in time' (P585) to express the date at which the sale took place. Supplying such information would then allow a timeline of acquisition history to be created.

5) Editor

When papyri or parchment fragments are published, the individual who reads, transcribes, analyses and (often) translates their contents is known in papyrology as the editor. This title, of course, differs somewhat from the concept of 'editor' as it appears in most other contexts, since the term is usually associated with someone who corrects a text. Although Wikidata has a property for 'editor' (P98), therefore, it is significant that the English language description of the property designated this person as one "who checks and correct [*sic*] a work (such as a book, newspaper, academic journal, etc.) to comply with the rules of a genre" [40]. While editors of papyri will note features such as spelling mistakes in the specific *Witness* they are working on, it is not expected that they will *correct* these mistakes (unless providing a 'full transcript' alongside a 'diplomatic transcript' for a new literary fragment), since the purpose is to transmit the writings recorded on the papyrus faithfully for non-specialists [41].

In the French language description of the property, however, an editor is explained as a "responsable, rédacteur ou rédactrice d'une œuvre, tel un livre ou un journal scientifique. Concerne aussi celles et ceux qui établissent les textes (anciens ou manuscrits)" [42].[3] The broader application of 'editor' was therefore encapsulated in this description. We therefore raised the issue on the discussion page for the 'editor' property and received the suggestion that the English language description be changed to match the French language version [43]. The next move after having changed the description is to apply the property to the *Document* item for the required papyrus, since it relates to the process of reconstituting the papyrus from its physical fragments. Since some papyri are re-edited after their initial publication (often when some additional fragment has been found, or new technology permits better readings), we suggest that adding the qualifier 'point in time'(P585) will allow a first editor to be distinguished from any subsequent editors.

6) Identifiers

In keeping with the goals of linked data, Wikidata includes 'Identifiers' among its properties, thereby allowing users to connect items to external authorities and databases. By assigning as many relevant identifiers as possible to an item, the user is able to draw the entire digital landscape together. The ability to assign identifiers, moreover, ensures that the representation of the item on Wikidata incorporates other conceptualizations of the data, and so acknowledges the varied nature of scholarly opinion. The concept can be likened to the concordance of texts sometimes included in papyrological print publications, which aligns the various ways the papyrus being published has been referred to.

For papyri and parchment items, there are several identifiers that it is essential to include. The Trismegistos Text ID (P8532) is paramount among these. Despite the issues discussed above, this identifier is crucial for improving the query-ability of papyrological data, since the Trismegistos database has been the central hub for so many papyrological items for such a long time. We envision, therefore, that many researchers who might seek to locate texts on Wikidata may well rely upon this

[3] English translation: manager, editor or (female) editor of a work, such as a book or scientific journal. This also applies to those who establish texts (ancient or handwritten).

identifier to find the item they require. This identifier should be attached to the *Document* item, since that most closely aligns with the type of entity Trismegistos assigns individual unique identifiers to; doing so will also, as a bonus, lessen the likelihood of creating a constraint violation (see discussion above).

Another important identifier is the 'Handle ID' (P1184), a superset of the DOI [44]. This identifier can be used to provide a link to any online catalogue entry belonging to the institution where the papyrus or parchment item in question is currently housed. Since the catalogue is more closely linked to the physical form of the papyrus (although not always, see discussion above), it makes sense for this identifier to be linked to the *Object* item. Some collections, of course, are too small to host their own databases [45]. Where such repositories are available on the web, however, it is important to connect with them as an important authority for the information recorded on the Wikidata item page. It also provides a link to the institution that researchers would need to contact for additional inquiries about the papyrus item, although this function can also be fulfilled by including the property 'Collection' (P195) for every item where the institutional home of the item is known. In this latter case, however, the collection (whether it is a museum, library, or other institution) needs to be an item in its own right on Wikidata. The Handle ID, on the other hand, can provide a connection to the collection without needing to ensure the existence of any other item on the platform, since its role is to point outwards.

Beyond the TM number and Handle ID, there are numerous other identifiers which pertain to specific subsections within papyrology. For example, New Testament scholars would undoubtedly wish to include the Gregory Aland Number, which is captured by Gregory-Aland-Number (P1577) and would be added to the *Document* item. Of course, since papyrology is relatively new to Wikidata, researchers may sometimes find external repositories that have not yet been incorporated with appropriate identifiers into the platform. The process of proposing new properties and creating item pages for these databases, however, is relatively straight-forward and uncontroversial within the Wikidata community, as the IDEA project discovered in creating a new property for Mertens-Pack Number. Mertens-Pack Numbers (MP^3) are used for literary fragments and refer to the database listing Greek and Latin literary papyri which is administered by the Centre de Documentation de Papyrologie Littéraire at the University of Liège [46]. As a recognized authority for literary papyri, the proposal put forward by IDEA received immediate support and was created very quickly (just 15 days after the original proposal was posted) [47]. The timeline for creating this new property demonstrates (again) the flexibility of Wikidata. The Mertens-Pack Number identifier (P11399) would be used for the *Document* item.

V. The Challenges and Potential of Wikidata

Discussion of Wikidata among digital humanists has highlighted two key concerns: the nuance and granularity of categories for the representation of data, and control over the quality of information imparted [48] [49]. The process of developing a data-model for papyrology outlined above certainly demonstrates the validity of these concerns. First, the difficulties in selecting appropriate properties and arguably cumbersome nature of some of the solutions, supports the notion that such a general database environment risks over-simplifying or misrepresenting specialist information. Indeed, Rossenova, Duschesne, and Blümel argue that deploying the underlying system upon which Wikidata depends, Wikibase, independently of the LOD environment may offer the best way for humanities projects to retain control of the terms of data representation [49]. Kesäniemi, Koho, and Hyvönen, meanwhile, advocate using Wikibase as a tool for writing and maintaining a CIDOC CRM-based knowledge graph, thereby making use of subject-specific extensions available in the CIDOC family of data models but bypassing the notorious complexity of these ontologies by deploying the user-friendly interface offered by Wikibase [6]. The concerns about quality, meanwhile, are validated by the issues with consistency identified in the representation of papyri on Wikidata prior to the contributions of the IDEA project.

Despite these challenges, however, my exploration of Wikidata for this project has encouraged me to see the potentials of this platform, if the initial teething problems can be overcome. For instance, the need to map categories to the ontologies of Wikidata, rather than perpetuating those already used in a specific discipline can be considered a strength rather than a weakness. The process demands reflection on the nature of the categories required, encouraging us to ask what their purposes are for research analysis. We need to anticipate what questions might be asked in order to select the right descriptors to yield fruitful results in querying, whilst at the same time allowing for enough flexibility so that inquiries beyond our imaginings might be conducted in the future. The expansiveness of the categories available in Wikidata make it clear that such investigations could very readily range beyond the confines of papyrology. What part, for instance, might papyri play in studies seeking instances of 'letters' diachronically? Linking the metadata associated with parchment and papyri sources with broader concepts opens the material of the field to outside perspectives in a way the current digital tools, grounded as they are in papyrological traditions, would be unlikely to do. Fig. 2, created using basic searches on SPARQL, provides a demonstration of the kind of connections that might be produced from this approach. While the links extracted here might seem bizarre, and some might question the utility of drawing such items together, the exercise demonstrates how Wikidata is able to push knowledge beyond the boundaries conventionally imposed by subject specialists.

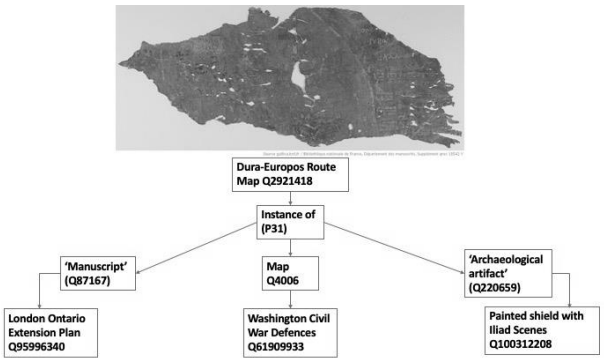

Fig. 2. Graphic representation of the connections that could be made between the item for the Dura-Europos Route Map parchment and other items through their 'instance of' designations.

These specific connections might not bear fruit, but others, outside of what we might imagine now, might prove explosive. We are dealing with the frontiers of the imaginable. Even a database utilising CIDOC CRM, despite the stated intention of this framework to facilitate interoperability across collections [50], would not achieve the same level of cross-disciplinary discoverability for the simple fact that it would necessarily be hosted locally, rather than in the broad LOD environment provided by Wikidata. For all that the ThOT framework is highly useful, as demonstrated by the way it has informed this data model, it requires implementation in this sort of linked open data environment for its potentials to be usefully realized.

For the more prosaic-minded, the potential of Wikidata to create connections between related subspecialities might seem more appealing. The division of labour within the fields of Classics and Egyptology mean that different types of objects are frequently handled by different specialists and may be artificially separated after excavation. This situation, indeed, prevails in the print documentation of the Dura-Europos material, since the different type of objects excavated are dealt with in different excavation reports. Through Wikidata, however, these materials can be easily digitally reassembled. Fig. 3 offers a sample of items connected by their discovery in the Palmyrene gate. Although each link provided here would need to be investigated further in order to determine what meaning it yields, this exercise serves to demonstrate the usefulness of Wikidata as a tool for supplying the fuel for future research.

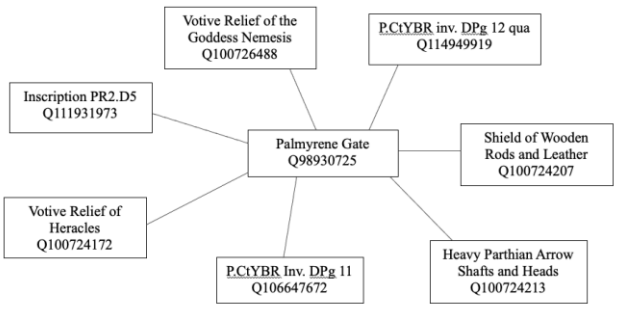

Fig. 3. Graphic representation of the connections that could be made through the location of discovery statement.

For both nuance of data representation and quality control, moreover, the greater involvement of the papyrological community would undoubtedly mitigate such concerns. With a data model established and more individuals contributing to the system, the quality control mechanisms deployed elsewhere in Wikidata should reduce the likelihood of wayward data representation. Not only does Wikidata automatically record all revisions made to item pages (similarly to the logging of activity also utilised on Papyri.info), but active community members are known to engage in discussion about and correction of perceived errors [48]. IDEA's participation in discussions about the application of the 'adjacent building' property demonstrates the flexibility of Wikidata if a user-community is active and pushing its structure in order to improve the platform's descriptive capabilities. In addition to this, if the Wikidata community were to take the step of integrating CIDOC CRM, FRBRoo and ThOT vocabularies into its framework, as the use of Wikibase for writing CIDOC CRM compatible RDF [6] and the overall similarity in their information structures [51] suggests would be possible, the platform could offer the opportunity for papyrologists to finally incorporate these standards into their data representation.

If such enhancements to the quality and nuance of Wikidata's representations take place, the other advantages of this platform could be expected to come to the fore. The multi-lingual aspect noted above would open papyrological resources to a wider range of people, beyond those possessing competence in both ancient and modern languages. The interconnection with other elements of the Wikimedia Foundation could be utilized to reconnect text and image, with Wikimedia Commons being mobilized to host papyri images, and Wikidata's image annotation tool offering the potential to add mark-up to these depictions – indeed, the IDEA project is experimenting with the possibilities offered by such interconnected features in the future. Finally, this same connectedness with the wider Wikiverse could offer security for the data represented on the platform, since unlike discipline-specific projects such as Papyri.info, Trismegistos, and ThOT, the funding for Wikidata relies upon a far broader base of support. With rewards such as these, therefore, working on the challenges presented by Wikidata ontologies seems well worth the effort.

ACKNOWLEDGMENT

I'd like to thank Dr Anne Chen and Professor Holly Rushmeier for giving me the opportunity to participate in the IDEA Project, and for introducing me to Wikidata in the first place. My profound thanks also go to Dr Mike Sampson, Dr James Cowey, Tom Gheldof, and Kat Thornton for consulting with me on the data modeling process.

REFERENCES

[1] IDEA Project Website About Page. [Online]. Available: https://duraeuroposarchive.org/about/

[2] L. R. Brody. “Yale University and Dura-Europos: From excavation to exhibition” in *Dura-Europos: Crossroads of antiquity*, Boston, McMullen Museum of Art, Boston College, 2011, pp. 17-32.
[3] J. A. Baird, *Dura-Europos*, London, UK, Bloomsbury Academic, 2018.
[4] J. A. Baird. “Unclassified: Towards and archaeological counter-archive” Guest Lecture Feb 8 2022, Yale University, Unpublished.
[5] J. A. Baird, A. Al Mohamed, “The Ruins that Remain: Remembering Dura-Europos in Salhiyeh” in *Dura-Europos Past, Present, and Future*, New Haven, Yale University Press, to be published.
[6] J. Kesäniemi, M. Koho, E. Hyvönen. “Using Wikibase for Managing Cultural Heritage Linked Open Data Based on CIDOC CRM” in *New Trends in Database and Information Systems. ADBIS 2022. Communications in Computer and Information Science, vol 1652*, New York, Springer, 2022.
[7] R. Heberlein. (2019) On the flipside: Wikidata for cultural heritage metadata through the example of numismatic description. Presented at IFLA WLIC 2019 – Athens, Greece – Libraries: dialogue for change. [Online]. Available: https://library.ifla.org/id/eprint/2492
[8] S. Middle. “Investigating Linked Data Usability for Ancient World Research.” PhD thesis The Open University. 2022 [Online]. Available: http://dx.doi.org/doi:10.21954/ou.ro.00014b1f
[9] C. B. Welles, R. O. Fink, and J. F. Gilliam, *The excavations at Dura-Europos conducted by Yale University and the French Academy of Inscriptions and Letters. Final report V, part 1: The parchments and papyri*, New Haven, Yale University Press, 1959.
[10] N. Reggiani, *Digital papyrology I: Methods, tools and trends*, Berlin, Germany, De Gruyter, 2017.
[11] Trismegistos "Partners" [Online]. Available : https://www.trismegistos.org/partners.php
[12] L. Vannini. “Online availability, impact and sustainability of digital papyrological resources” *DCO*, vol. 8, pp.23-45, 2022.
[13] ThOT “Partners & contributors” [Online]. Available : https://thot.philo.ulg.ac.be/project.html#partners
[14] Papyri.info website. [Online]. Available: https://papyri.info/search
[15] ThOT Sign List “About” [Online]. Available : https://thotsignlist.org/About
[16] S. Polis, V. Razanajao, “Ancient Egyptian Texts in Context. Towards a Conceptual Data Model (The ThOT Data Model – TDM)” *BICS*, vol. 59, no. 2, pp.24-41, 2016.
[17] ThOT “Introduction” [Online]. Available: http://thot.philo.ulg.ac.be/project.html
[18] A. Felicetti, F. Murano. “Scripta manent: a CIDOC CRM semiotic reading of ancient texts” *Int J Digit Libr*, vol. 18. pp.263-270, 2017.
[19] Trismegistos “What Changes on 1 January 2020?” [Online]. Available: https://www.trismegistos.org/keeptrismegistosalive_whatchanges.php.
[20] L. Kaffee, A. Piscopo, P. Vougiouklis, E. Simperl, L. Carr, L. Pintscher, “A Glimpse into Babel: An Analysis of Multilinguality in Wikidata” in *OpenSym '17*, Galway, Ireland, 2017, pp.1-5.
[21] Meta-Wiki “List of Wikipedias” [Online]. Available: https://meta.wikimedia.org/wiki/List_of_Wikipedias
[22] K. Thorton, K. Seals-Nutt “Science Stories: Using IIF and Wikidata to Create a Linked-Data Application” *ISWC (P&D/Industry/BlueSky)*, vol. 2, pp.1-4, 2018.
[23] Science Stories Website [Online]. Available: https://www.sciencestories.io/welcome
[24] Search for Papyri on Wikidata via SPARQL query service [Online]. Available: https://w.wiki/6ftf
[25] Search for Papyri on Trismegistos [Online]. Available: https://www.trismegistos.org/index_disambiguation_form.php?searchterm=papyrus|Material@&searchterm_date=&publ_date=&search_criteria_or=&strict_search=1&preset_criteria=
[26] Papyrus Oxyrhynchus 27 on Wikidata [Online]. Available: https://www.wikidata.org/wiki/Q656908
[27] Papyrus 42 on Wikidata [Online]. Available : https://www.wikidata.org/wiki/Q677742
[28] J. A. Sahu-Hough "Papyri Data Model" IDEA Shared Drive [Online]. Available : https://docs.google.com/document/d/1k9JH5gtVYN8Y22KGBdbX9_KAjag4FVQtJxBH650PXIw/edit?usp=sharing
[29] Yale Digital Collections, “P.CtYBR inv. DP 87” [Online]. Available: https://hdl.handle.net/10079/digcoll/2771545
[30] Yale Digital Collections, “P.CtYBR inv. DP 82” [Online]. Available: https://hdl.handle.net/10079/digcoll/2771932
[31] Yale Digital Collections, “P.CtYBR inv. DP 85” [Online]. Available: https://hdl.handle.net/10079/digcoll/2771766
[32] Trismegistos, “Stable Identifiers” [Online]. Available: https://www.trismegistos.org/about_how_to_cite.php
[33] Wikidata, “Database reports/Constraint violations/P8532” [Online]. Available: https://www.wikidata.org/wiki/Wikidata:Database_reports/Constraint_violations/P8532#%22Unique_value%22_violations
[34] JASHough (username) “Add exceptions to ‘distinct-value constraint’ or remove constraint” Wikidata [Online]. Available: https://www.wikidata.org/wiki/Property_talk:P8532#Add_exceptions_to_'distinct-value_constraint'_or_remove_constraint
[35] ThOT “Necropolis Journal” [Online]. Available: http://thot.philo.ulg.ac.be/concept/thot-18752
[36] Search for all papyri discovered in the Palmyrene Gate on Wikidata via SPARQL query service [Online]. Available: https://w.wiki/9JrS
[37] Adjacent Building Item on Wikidata [Online]. Available : https://www.wikidata.org/wiki/Property:P3032
[38] Ipoellet (user name) “Other Adjacent Things” [Online]. Available : https://www.wikidata.org/wiki/Property_talk:P3032
[39] Search for all papyri discovered adjacent to Block L8 on Wikidata via SPARQL query service [Online]. Available: https://w.wiki/9D$k
[40] Wikidata, “Editor (P98): Difference between revisions” [Online]. Available: https://www.wikidata.org/w/index.php?title=Property:P98&diff=prev&oldid=2067206797
[41] P. Schubert, “Editing a Papyrus” in *The Oxford Handbook of Papyrology*, Oxford, 20, pp.197-215.
[42] Wikidata “Editor” [Online]. Available: https://www.wikidata.org/wiki/Property:P98
[43] JASHough (username), “Change description to widen applicability” Wikidata, [Online]. Available: https://www.wikidata.org/wiki/Property_talk:P98#Change_description_to_widen_applicability
[44] Wikidata “Property talk: P1184” [Online]. Available: https://www.wikidata.org/wiki/Property_talk:P1184
[45] R. Ast, S.P. Davis, “The Advanced Papyrological Information System” *Storicamente*, vol. 4, pp.1-6, 2008.
[46] CeDoPaL “Welcome” [Online]. Available: http://web.philo.ulg.ac.be/cedopal/en/
[47] Wikidata “Property proposal/Mertens-Pack Number MP3” [Online]. Available: https://www.wikidata.org/wiki/Wikidata:Property_proposal/Mertens-Pack_Number_(MP3)
[48] K. Thornton, E. Cochrane, T. Ledoux, B. Caron, C. Wilson. “Modeling the Domain of Digital Preservation in Wikidata”, *iPRES*, 2017.
[49] L. Rossenova, P. Duschesne, I. Blümel. “Wikidata and Wikibase as complementary research data management services for cultural heritage data”, *The 3rd Wikidata Workshop, Workshop for the scientific Wikidata Community @ ISWC 2022. CEUR Workshop Proceedings*, 2022, pp.1-15.
[50] N. Crofts, M. Doerr, T. Gill, S. Stead, M. Stiff. “Definition of the CIDOC Conceptual Reference Model: Version 5.0.4” *ICOM/CIDOC CRM Special Interest Group, Nov 2011* [Online]. Available: https://cidoc-crm.org/sites/default/files/Documents/cidoc_crm_version_5.0.4.pdf
[51] P. Le Boeuf. “Modeling Rare and Unique Documents: Using FRBRoo/CIDOC CRM” *Journal of Archival Organization*, vol. 10, no. 2, pp.96-106, 2012 .