AARON PATTEE

Graph Databases for the Organisation and Analysis of Digital Heritage

Introduction

The purpose of this article is to encourage researchers in the *Digital Humanities* (DH) to make use of Labelled Property Graph Databases (LPGs) in order to organise and curate their datasets. Establishing and curating an organised database for all data flowing into a project is essential for provding an impirical basis for conclusions drawn at the intersection of different datasets. Although the traditional *Relational Database Management Systems* (RDBMS) and *Resource Description Framework* (RDF) Triple Stores also provide stable platforms for organising data, they do not offer the same interoperability, malleability, and efficiency required for many DH projects as property LPGs do. Such projects are built upon an ever-tilting lever in which the humanities-based and computer science-based analyses are constantly encountering one another on both the theoretical and practical level – allowing one of the two to easily gain the foreground at the cost of the other. The fulcrum of this balancing act must therefore be adjusted to the two fields of study and provide a base in which the weight of both is stabilised. The image of the lever-system can, of course, be applied to virtually all interdisciplinary studies and their associated methodologies. However, the ever-growing sector of DH projects linking seemingly unrelated disciplines to one another that have rarely interacted so far (Blanke et al. 2017; Mash et al. 2016; Kuczera 2017), is tilling new fields of cooperation. This requires the development and application of innovative tools to cultivate their common ground. With regard to databases, it is precisely the fulcrum's ability to adjust to the questions posed by different disciplines that exemplifies a malleable platform. It must be underlined that such databases are an efficient way for conducting research not simply due to the speed with which the data are imported or queried, but also because they facilitate the manner in which data can be indexed and maintained during the development or adaption of the data structure (Robinson/Webber/Eifrem 2015, 102).

The construction of a database incorporating undigitised historical data in addition to 3D data can be a time-consuming and labor-intensive process. However, the benefits of a well-curated database outweigh the perceived disadvantages. As data is gathered and the database grows, it becomes more capable of finding trends and connections between data that are not immediately obvious or even hidden. A malleable data structure allows researchers in the humanities and computer sciences to find patterns or test new connections without fear of losing data or the

© 2024 Aaron Pattee (CC BY-SA 4.0). Published in: Carsten Wergin and Stefanie Affeldt (Eds.),
Digitising Heritage. Transoceanic Connections between Australia and Europe. Heidelberg 2024, 123–138.
DOI: https://doi.org/10.17885/heiup.1305.c18423

foreboding risk of putting in hours of work for little change. It can be shared and explored as a team and changes are easily adapted. This paper discusses the key advantages of property LPGs, when compared to their counterparts, focusing upon how to structure a database from the perspective of a humanities researcher, with a recent dissertation serving as the case study. The development of such a database and, more importantly, drawing new conclusions from the data, requires interdisciplinarity and communication, though an adventurous spirit is never harmful in the process.

Databases and Digital Humanities

Whether in the *Digital Humanities* (DH), natural sciences, or any other field, all projects accumulate data from research that require a method in which to mediate and store the data.[1] Ideally, this method also includes a retrieval system in order to locate data that have been saved. However, not all databases are equal in this manner, and some are better suited for certain tasks than others. The two types of databases discussed in this paper are *Relational Database Management Systems* (RDBMS) and *Labelled Property Graph Databases (LPGs)*. The difference between the two may not be immediately obvious. Although one includes graphs and the other does not, they are often discussed as similar methods for data solutions (Tomasi 2018, 11). When examined more closely, LPGs offer a wider range of applications and are more easily managed, making them better suited for DH projects. This will be discussed in the remaining sections of the paper. Both types of databases store data and can be paired with retrieval systems, can host large amounts of data, and can be used as foundations for software and websites equipped with intuitive interfaces (Winberg/Zubac 2019). Despite these similarities in application, the underlying structures warrant a closer look, with a specific emphasis upon which database type is most conducive to interdisciplinary communication. In many cases, DH projects consist of a team of researchers from different fields, each bearing the standard of their respective disciplines. More often than not, the spectre of discord hovers above the teams as researchers from the humanities and computer sciences have different investigative approaches, disparate glossaries of terms, and expertise in disciplines that rarely come into contact with one another – not to mention the diversity of analytical methods and procedures within the humanities which are anything but a monolith. Thus, clear and open communication is vital, buttressed by a database system that does not entirely exclude the

1 For the creation of a knowledge base that interweaves a broad variety of data on Indigenous communities, locations, and items in a digital knowledge base, see Paul Turnbull: *Restoring Dignity*, pp. 29–45 in this volume.

humanities researchers becoming involved in its structure, without a certificate or extensive experience in database development.

The database construction is typically a task rendered unto the computer scientists whilst the researchers from the humanities concern themselves more with the collection and evaluation of the data. With proper communication and financial support, this process can prove successful as it often has (Bol 2018, 7). Nevertheless, a certain requirement for emerging projects and junior researchers in DH is to not only familiarise oneself with new technologies, but to learn how to apply them at a rudimentary level. This is not a call for everyone to learn every topic, theory, or method and obtain an expertise in every field. Instead, it is a call for an improved communication between the researchers by recommending a database system that is more suited for interdisciplinary dialogue. LPGs implement a graphical model for connecting data through the use of *nodes* (entities) and *edges* (relationships) (Robinson/Webber/Eifrem 2015, 25). Although it is a relatively new technology, some of the best interactive visualisations for this sort of model were produced in the Middle Ages. The 13th-century genealogy of *Kuenringer* family tree in the *Stifterbuch des Klosters Zwettl (*Schubert *2018, 247)* and the 14th-century *Genealogical Roll of the Kings of England (ibid., 63)*, as well as a host of other medieval genealogies, make use of nodes and edges to portray lineages featuring names, relations, and even portraits. Conceptualising ideas as graphs rather than purely as lists, allows one to visualise the development of the database much in the same way that sociograms allow one to visualise social networks. GDBs do precisely this, while attaching more levels of complexity.

Why Opt for a Graphical Model?

A graphical model combines efficiency with malleability by grouping data into nodes and edges. The graphical model allows for new connections to be made between previous data without altering the previous data lists, more akin to how humans interact when meeting new people who perhaps have mutual friends or associates. In essence, the graphical model establishes a digital sociogram that can be modified and adapted as the project expands. Thus, graphical models make use of methods that are already familiar to researchers in the humanities. The visualisation components of a graph, however, provide a simpler way of seeing the structure of the data.

Although the graphical model unites both mind-sets from the humanities and from computer science, not all graph databases (GDBs) operate in quite the same way. A key example is found in the difference between RDF triple stores and LPGs. Both RDFs and LPGs implement the graphical model and are directed graphs (Robinson/Webber/Eifrem 2015, 5), meaning that all edges include a distinct direction connecting nodes to one another. These directions are essential for modelling and querying the connections between data in the database, though only LPGs allow one to query regardless of the direction (ibid., 67). With a RDF triple store, queries rigidly follow the direction of the relationship with defined labels, which means that if two nodes are connected, only the direction of the connection will be followed. In stark contrast, a property graph database can follow relationships between two nodes in either direction, even if a direction is defined. Thus, only one relationship is necessary between two nodes rather than two relationships in order to query in both directions. This substantially reduces the number of relationships between nodes, thereby reducing the data to the most necessary amount. Furthermore, LPGs allow for data to be stored as properties of the relationships – something a RDF cannot. These edge-properties make a tremendous difference in the size of the data structure, because data can be stored in a variety of combinations, reducing the overall number of edges necessary to model nuances in the connections between nodes. This important aspect is second only to the significance of choosing a graphical model, as it demarcates a key difference between the two types of GDBs. Fig. 1 represents an example from the CITADEL doctoral project. It shows the modelling of a father-son relationship for which the type of relation and the certainty of the relation were easily modelled and visualised.

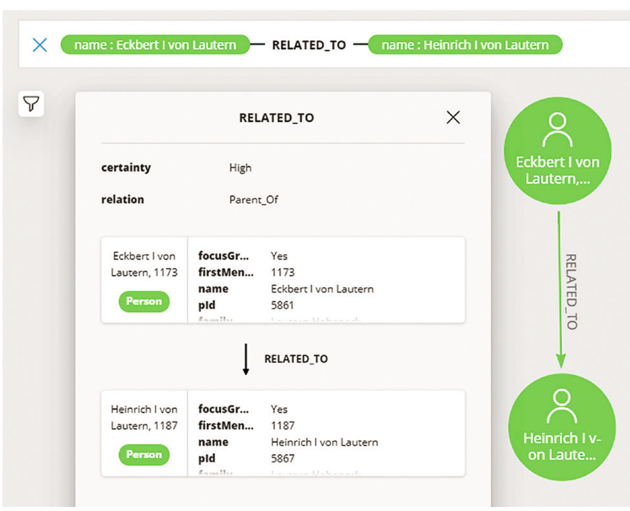

Fig. 1: An example of a relationship with an ascribed property, visualised in Neo4j Bloom.

Which Graphical Model is Suited to my Project?

Selecting a specific graphical model depends, of course, upon the nature of the inquiry in addition to which mid-ranged questions arise during the development of the project. Mid-ranged questions are the analyses that arise as the project develops (Hardesty/Little 2009, 69–70), constituting a crucial aspect of the overall project and one of the most important areas for communication between researchers.

These questions are the driving force behind the development of a data structure as more elements are taken into account that were either unkown or unclear at the project's inception. Mid-ranged questions often take into account various aspects from different disciplines, necessitating interdisciplinary dialogue. For example, the discussion of modelling a person's social standing in the High Middle Ages requires a discussion between Historians and Sociologists.

Approaching a topic from only one angle, will almost certainly leave out important insights garnered from other scientific perspectives. Inherent to the development of these questions, is the addition of more data. This category of investigation can necessitate a restructuring of the database and can happen at almost any time during a project's development. As projects begin large, become narrower, and again swell shortly before they are succinctly summarised for a final time. Adapting a data structure to sudden changes or adding new amounts of information previously unknown is common and essential. If the goal were to digitise a library collection or archive, an RDF triple store would be as well suited as a property graph database (Kaaij 2019).

If the objective is to embark upon a project that does not have an immediately known final result, such as an archaeological, architectural, geo-spatial, or historical investigation, then a LPG is better suited (Kuczera/Wübbena/Kollatz 2019). The reason for this has to do with the key difference between the two graphically-based databases: LPGs incorporate properties into nodes *as well as* edges. Distributing data across both core components creates a more flexible database that benefits both types of researchers. After selecting a LPG, the next step is to begin the process of organising the data in order to move it into the database.

When compiling data from various text resources such as geographical data or information derived from 3D models, the first challenge is to be able to control the sheer amount of data formats flooding into the project. Seeking a universal system for organising and simultaneously visualising all of the data of DH projects is the fever dream of most researchers, but it bears more similarities with the Quest for El Dorado, than an attainable result within a realistic timetable. Therefore, a good first step is to identify a data format that can be adapted to virtually any data. None is more befitting to the construction of database than the *Comma Separated Value* (CSV) format. *Microsoft Excel* is well suited for this task as it can export data in a variety of formats.

The following section illustrates this point by outlining the CITADEL project. The project combined historical, architectural, and geo-spatial data that was imported into a single LPG from which queries could be scripted using data from all three inquiries in order to draw new conclusions at their intersections.

Case Study: The CITADEL Dissertation

The overall goal of the CITADEL project was to establish a novel approach using integrated digital methodologies in order to pose new questions about the architectural development of four German castles at the turn of the 13th century. The combination of several methodologies has produced new results rather than focusing on the merits of each methodology individually (Pattee 2023). The investigation of the sites emphasised the role and the signalling of status using architecture, in which builders sought to indicate (or signal) their access to resources, knowledge, or kin groups by constructing monumental buildings. These buildings – all of which were castellated structures – included architectural elements that were both symbolic and utilitarian, yet unmistakably associated with the reign of the Hohenstaufen dynasty at the turn of the 13th century. However, the scant remains of three of the sites, and the conditions of all four as archaeological ruins, do not leave much to be studied, absent historical context. Although context was key to understanding the sites and their function, the context was mired by the peculiar nature of the historical figures who inhabited and constructed the sites, namely the mysterious *ministeriales* of the High Middle Ages. This elusive group of medieval governmental administrators muddled the line between noble and non-noble, as they were considered to be wholly neither, nor were they wholly respected. Though the latter was the effect of the former, it weighed just as heavily with regard to their perceived position in society (Bosl 1950; Bosl 1975; Hechberger 2010; Keupp 2002). As their numbers waxed and waned throughout the 11th to the 13th centuries – based largely upon the politics of the reigning German Emperors – their enfeoffed castellated homes established them as a political, military, and social force to be reckoned with. The process whereby they achieved social status, attained administrator positions and the favor of the kings was largely affected by their ability to not only excel as administrators of royal palaces and lands, and as famed military leaders in Italy, but also to construct constant reminders of their own importance. These constructions took the form of castles, palaces, and commandries that peppered the landscape of the hilly German Palatinate (Keddigkeit et al. 2007)[2] – as well as many other regions including Swabia, Saxony, and Bavaria. These buildings embodied who they were, which aspirations that had, and how they wished to be seen. The effectivity of the signal depended upon their actual position in society and how outsiders interpreted the architecture of their castles. This portrays a constant feedback loop between status and architecture. As castles atop hills, nestled in valleys, and cresting over glistening lakes were magnified by their extraordinary positions within the landscape, geo-spatial data were a necessity alongside the already abundant historical and architectural data.

2 Keddigkeit et al. (2007), *Pfälzisches Burgenlexikon Band I: A–E*, Kaiserslautern. This is the first volume of five documenting the castles of the German Palatinate.

Where to Begin When Developing a Database?

Everything began with spreadsheets. From architectural 'Roombooks' to coordinate points of geo-referenced maps, to details of historical individuals and charters, spreadsheets formed the first layer of documentation for the project database. As spreadsheets are easily accessible and essentially universal, they are a practical first tool to employ. They are particularly useful when exported as CSV files, which are compatible with essentially every database software as well as *Geographical Information Systems* (GIS). Although spreadsheets are an excellent first step and may even be a catalyst for new mid-ranged questions, they are not a result in and of themselves. Instead, they constitute the pathway whereby results can later be achieved in a database in which network analyses are conducted. This is accomplished by reaching into data from different investigations to bring forth new interpretations at the intersection of these data. Before such interpretations can be achieved, one must first begin with accumulating the data in a consistent manner.

As the objective of the CITADEL project was to research the castles, the sites themselves composed the primary source of information and the first order of business (Grossmann 2010, 44). In order to capture their condition in the most precise manner, both *Terrestrial Laser Scans* (TLS) and *Structure from Motion* (SfM) photogrammetry were employed for the generation of 3D models that were precisely measured and included high-resolution image textures. These provided the foundation for the architectural analyses that consisted of a traditional stone-by-stone construction research, albeit without having to conduct hand-drawn architectural illustrations. These detailed investigations were annotated on what were essentially orthophotos of each wall of each castle, in which building phases, ornamentation, and architectural elements were documented. The results were then recorded in a large *Excel* spreadsheet that organised the information according to their location, various building phases, and building types, as well as the construction elements, substances, and components. These terms each represented a node-group, or set of similar data types that create the nodes of the database. Each node-group received its own properties consisting of unique identification numbers, and in some cases, an amount. The data were based upon their positions within the 'Roombooks' of the sites, which were nothing more than a numerical listing of each wall of each site in an order according to how one walks through the respective site. The spreadsheet was relatively larger consisting of 595 rows and 18 columns in which all of the architectural were recorded. Such accumulations of data represent an authority file. They are only interesting insofar as they can be broken down into more manageable units. Whether at first glance or after repeated glances, a screen of data does not reveal much, despite organising the data into a scheme that relates to the project question. Nevertheless, such files are a vital first step, as they can always be modified and apportioned – though it is important to keep a backup!

The geo-spatial analyses began in a similar manner in which dozens of early modern maps were first geo-referenced – a process in which the spatial content of the maps was implicitly or explicitly directed to positions on the earth's surface (Conolly/Lake 2006, 17) – in order to create diagrams of environmental data such as lakes, rivers, and political territories (Pattee et al. 2018). These diagrams were vector data consisting of distinct points referencing buildings, or polygons representing areas/lakes. The point data had distinct coordinates that were exported as a CSV, consisting of 786 locations with an average of nine properties including the name, a unique ID, the coordinates, and further details regarding their modern position. This spreadsheet of the location data was also an authority file consisting of data drawn mainly from the historical maps, while also including data from the architectural analyses and, most importantly, from the historical investigation. However, in order to determine why and where the ministeriales built certain architectural elements, it was first necessary to model their social network.

Modelling their network provided an empirical basis in order to reconstruct their position in society leading to more refined interpretations of where their aspirations were to lead. This required a close reading of hundreds of transcribed charters, in addition to dozens of secondary historical sources regarding the castles, the ministeriales, and contemporary events at the turn of the 13th century. The diversity of the charter data consisted of 515 charters sourced from 23 different analogue volumes and 166 charters from the *Regesta Imperii Online* (RI Online) digitised resource.[3] As the online sources accounted for only 24 % of the total amount, it was not possible to introduce an computational reading all of the charters, such as *Optical Character Recognition* (OCR). Furthermore, the application of OCR was out of scope for the project as the objective was to provide context for the overall goal of researching the backgrounds for constructing the castles, and not to devise a method for reading the various forms of texts. These were recorded in a variety of fonts, including several forms of German *Fraktur*; languages, consisting mainly of Middle High German and Latin, in addition to the High German translations; and German dialects, whose modern linguistic derivatives most closely resemble *Pfälzisch* and *Schwäbisch*. The published format of the charters also varied, as some were much more complete in that they included clearly marked witness lists, as well as place names, dates, and transcription numbers. Others were recorded in telegram form, at times providing almost no information regarding the individuals involved in the charters, save their initials. Determining to whom the initials belonged, it was therefore necessary to read the charters on the pages before and after the one in question, a process that sometimes included up to 15 additional close readings of charters not meant for the project corpus. Although this may seem excessive and possibly misallocation of time, it was necessary because a

3 An additional 22 charters were also collected, but did not have a proper citation and were therefore not used in the analyses.

name that was wrongly interpreted drawn from initials could create a completely altered social network. In order to maintain a highly detailed and accurate model of the social networks, attention had to be paid to even the smallest details. Such instances emphasise the integration of expert knowledge for data curation. It is worth mentioning that not all of the charters were accumulated prior to the initiation of the close readings. Instead, at the time of their inception, the corpus consisted of slightly more than 100 charters – or merely 14 % of the final project text corpus.

The process of parsing the data necessary for the network modelling was rather straight forward at the beginning. It consisted of another large authority file of only the individuals mentioned in the charters and another for the charters themselves. The spreadsheet for the individuals was the largest, consisting of 1590 rows across 16 columns. However, not all of the information in the spreadsheet was used for the properties of the individuals. Instead, an average of five properties were attributed to each person, including a unique ID, the date of their first mention, whether or not they belonged to a family residing in one of the castles (the focus group), their heritage (regarding their familial association with the nobles, ministeriales, or unknown), and various other properties referencing their social positions. The immediate effect of this process was that it resulted in a decent overview of the individuals mentioned and their general position in society. However, there were a number of properties that had to be adapted almost immediately as people tend to move and attain new roles in society over time. As the members of the focus group were almost all ministeriales, modelling their position in the social network of the area was complicated by the fact that they often changed roles and even their names, depending upon where they were commissioned. Furthermore, their roles were actually a conglomerate of two types of positions in society, namely their *status* position referring to more static positions (such as knight), and *administrator* positions referring to more dynamic positions (such as sheriffs); and sometimes they occupied both types at once. Both social positions had direct effects upon their access to resources and more elite social circles along the hierarchical chain of medieval society. Thus, modelling their position in society was the best determinate of understanding their capacity and reasons for constructing buildings in a manner that reflected how they perceived themselves.

The spreadsheet for the charters consisted of 706 rows across eight columns, with properties including a unique ID, the date of the charter, a brief description, the location where the charter was issued, and a reference for the source of the charter. Although the charters themselves were simpler to record, they still contained an intrinsic ambiguity, namely which description to use when at times multiple events took place. This was a core problem at the beginning and was the point of departure for actually implementing a LPG for modelling the network. For instance, charters issued at the royal palaces rarely concerned only one topic and describing such a charter by the first topic discussed – or at random – would result in a highly inaccurate modelling of the social network at the date of the charter's issue. Some individuals, though mentioned in the same charter, were not affiliated with one another if their topic was altogether different. In short, multiple people were present for multiple reasons and for the sake of scribal brevity, their proceedings were bundled together into a single charter rather than two or more. Modelling these proceedings, while taking into account the individuals and their roles in society, complicated the structure of the CSVs, requiring additional spreadsheets cataloguing the events in the charters. These joined tables together, in the same way that JOIN tables combine spreadsheets in a RDBMS. However, these sorts of tables should be limited as the greater amount of JOIN tables leads to redundancies and complications in the data. Increasing the quantity of these tables can lead to inefficiencies in performance and reduce the malleability of the database (Bruschke/Wacker 2014, 3).

Labelling and Adapting the Database to New Data

As previously mentioned, all nodes, i.e. the individuals, charters, locations, and building analyses, consisted of a number of properties including unique ID numbers. These were attributed to each node during the development of the project and bore no relevance to the transcription numbers of the charters or any other number drawn from the data. Instead, they were entirely for internal use and allowed data to be connected based upon distinct numbers rather than names. Additionally, the IDs of the node groups were given numbers of the same amount of digits. For example, the Persons node-type (all of the individuals) received a number four digits long, beginning with 5000, whereas the Charters node- type were given a number five digits long beginning with 10.000. The purpose of this system was to not only provide IDs, but also provide quick information regarding what the numbers represented, absent any text. Thus, a four-digit number beginning with five or six, found within a swarm of other IDs, always represented a Person node, and a five-digit number beginning with 10 always represented a Charter node. This labelling system was very effective as it removed the prospect of long computer-generated IDs that only lead to confusion when searching for quick information. Once the core data had been labelled, the process of combining the data in order to construct a network was beyond the capabilities of Excel. This required a LPG, which in this case was Neo4j, as mentioned earlier. A key feature of LPGs such as Neo4j is that they implement a scripting language for connecting data, rather than having to rely upon JOIN tables in order to build connections. The specific scripting language that was used for this project was *Cypher*, which is essentially a pattern-matching language based upon *Structured Query Languages* (SQL), and is the most widely employed within the realm of LPGs (Robinson/Webber/Eifrem 2015, 24–25).

Although learning the basics of Cypher constituted another challenge, it was not so much an obstacle as it was a relief, because it essentially removed the prospect of having to create new JOIN tables. Combining tables is achieved by only a few lines of script that can be implemented in a manner of milliseconds allowing one to visualise the changes immediately thereafter. Furthermore, once the general form of the data was established, future additions such as persons, charters, or locations can be added with scripting in which they are directly connected with the other nodes and edges according to the project schematic. As new data enter the project, they are simply added in by script and are immediately connected to the rest of the data following the graph schematic. This sort of importation could be seen as a disadvantage because the new data are no longer in the original authority file spreadsheet. However, Neo4j and other property GDBs allow one to export CSVs from the database for each node type, thus offering a quick solution to this problem. Additionally, the queries themselves can be viewed as a graph, that is, as nodes and edges, as a table, or as a Cypher script, and all results can be exported in a variety of file formats including CSV and JSON.

Results

The results of the project consisted of 7,719 nodes in 14 node-types, and 14,687 edges in 26 edge-types. These included architectural, historical, and geo-spatial data that were combined in a single graph schematic allowing one to query across multiple data types while taking into account multiple mid-ranged questions. The castles and their building phases can be neatly visualised using the database allowing for quick interpretations as shown in Fig. 2. The final schematic of the project is shown in Fig. 3 in which all node-types and edge-types are depicted.

As mentioned in the first sections of this paper, the ability to balance the interests and analytical approaches of multiple researchers from different disciplines is an essential component in any DH project. As each researcher has their own research emphasis, modelling their mid-ranged questions in a single database is essential. Although the CITADEL project did not represent a project of multiple researchers, as it was a single doctoral project, it did include a host of advisors from the disciplines of Art History, Archaeology, Computer Science, Geoinformatics, and History. All of these advisors raised important questions pertaining to architectural function, site significance, social network analyses, the impact of landscape, and the perception of rank. These mid-ranged questions contributed to the overall graph schematic in which modelling pathways implicitly took each question into account. Thus, the final queries determining who built the sites and why the sites were built combined each analysis. The social network analyses regarding the combination of status and administrator positions as well as the specific appearances of individuals in events sourced from charters, filtered out who would have been

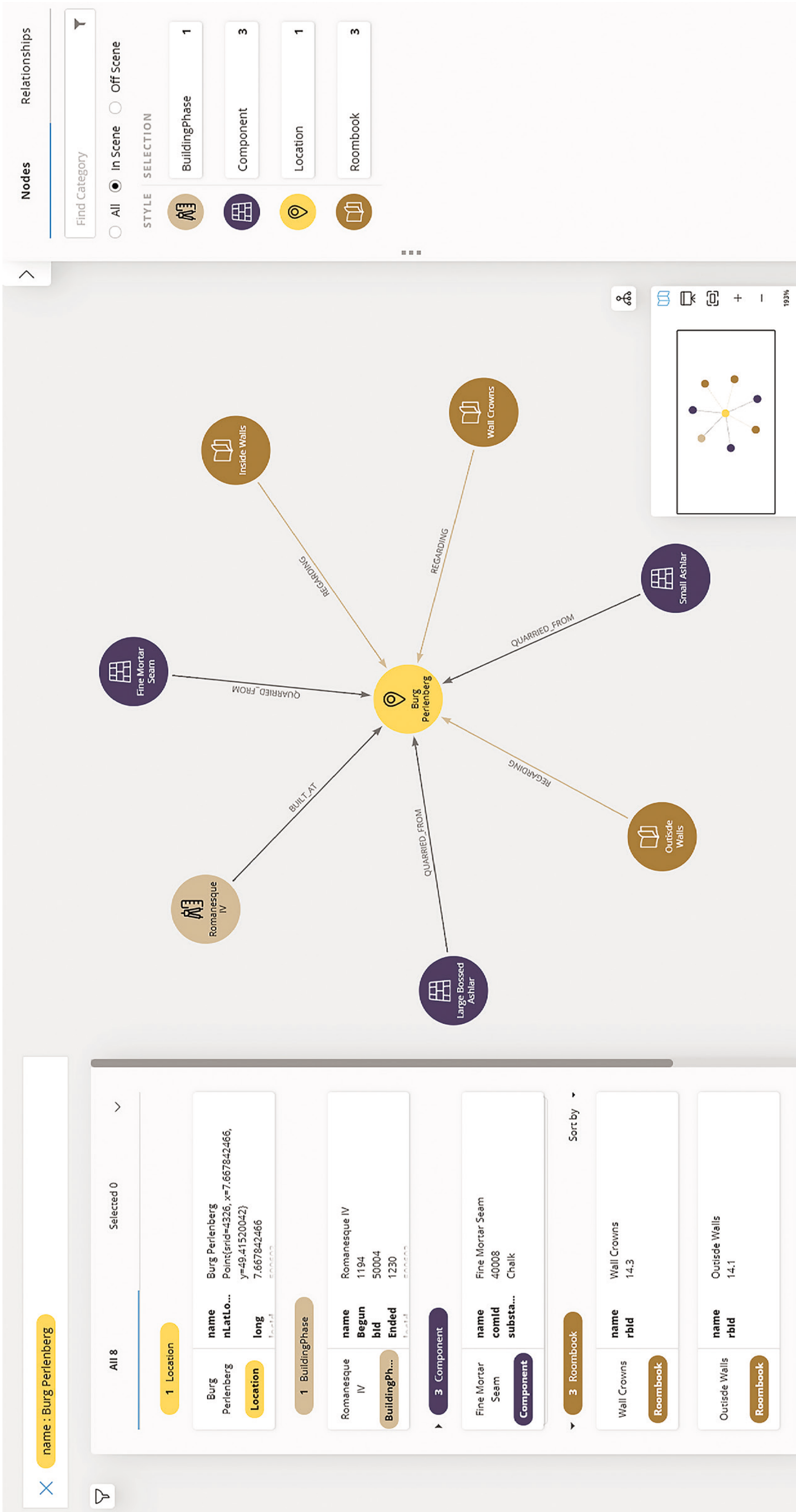

Fig. 2: Visualisation of Castle Perlenberg's Construction Research.

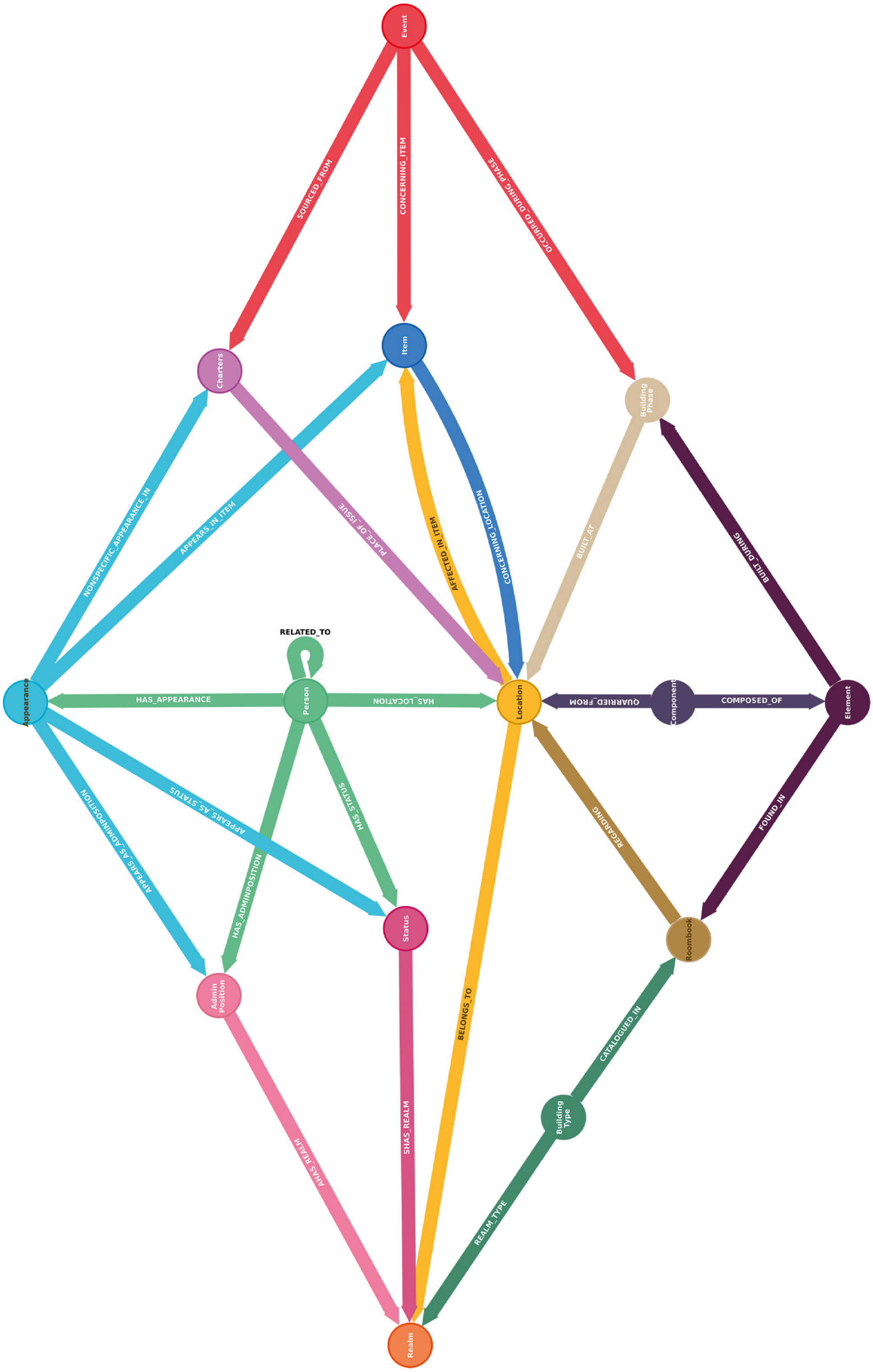

Fig. 3: Final Graph Schematic of the CITADEL Project.

in a position to gain permission and support for construction projects. The spatial organisation of the data and the locations in which charters were issued, provided information regarding site significance and recurrence, as well as supply points of resources for construction and transportation routes. The overall layout of the castles including the distribution of architectural elements, the use of specific materials, and the building phases identified particular points in time when construction could have occurred. When all three were combined in a single query, the result was a list of potential builders that also revealed an interesting pattern for some of the sites, namely that the builders of Castle Hohenecken, for example, all had the same first name across three generations.

Conclusions

These results were made possible by the application of a LPG, without which the same results would have either not have been possible or would have necessitated a much longer investigation. As it was a single doctoral project, there was a time constraint on the overall length for the production of the research. Nevertheless, the LPG comprised the fruit of a little less than two years. This is perhaps the clearest indication of the time efficiency to be gained when employing such a database. The malleability of the database was also a clear benefit as the graph schematic changed no fewer than 11 times over the course of its development, regularly adapting new information and pathways, as well as reallocating properties along edges. The two fields of Humanities and Computer Science find common ground in this process, where the manifold disciplines of the former meet the modern emphasis upon the latter. As communication is a key determinate in any DH project, it must be underlined that there were a host of advisors and mentors who both expedited and invigorated the development of the database. The use of such a database was uniquely suited to such discussions as the visualisations of the graphical model could be quickly examined and understood by all parties, regardless of their disciplinary background. It would also be remiss to not include a brief description of which further steps can be undertaken having established a LPG. A common application would be as the database for a website or software, or even for a *Conceptual Reference Model* (CRM) such as the CIDOC CRM ontology (Bruschke/Wacker 2014, 3). Furthermore, GDBs (including RDF triple stores) have a unique ability to 'speak' between databases in order to link information from previous databases to one another, or to a new one entirely. The interoperability, malleability, and efficiency of a LPG provides an excellent data solution for DH projects, in which data from various sources, disciplines, and analyses can be combined for well-structured queries and data organisation.

Bibliography

Blanke, Tobias/Bryant, Michael/Frankl, Michal/Kristel, Conny/Speck, Reto/Vanden Daelen, Veerle/Van Horik, René (2017), "The European Holocaust Research Infrastructure Portal", in: *Journal on Computing and Cultural Heritage* 10 (1), 1–18, https://doi.org/10.1145/3004457.

Bol, Peter (2018), "How the Digital Is Changing Research and Teaching on Asia", in: *ASIANetwork Exchange: A Journal for Asian Studies in the Liberal Arts* 25 (2), 7–28, https://doi.org/10.16995/ane.278.

Bosl, Karl (1975), "Die Adelige Unfreiheit", in: Friedrich Ludwig Wagner *(ed.), Ministerialität im Pfälzer Raum: Referate und Aussprachen der Arbeitstagung vom 12. bis 14. Oktober 1972 in Kaiserslautern*, 64, Speyer, 9–23.

Bosl, Karl (1950), *Die Reichsministerialität der Salier und Staufer: Ein Beitrag Zur Geschichte des Hochmittelalterlichen Deutschen Volkes, Staates und Reiches* (Schriften der Monumenta Germaniae Historica 10), Stuttgart.

Bruschke, Jonas/Wacker, Markus (2014), "Application of a Graph Database and Graphical User Interface for the CIDOC CRM", held at: *CIDOC 2014 Annual Conference* (Dresden 2014), http://www.drematrix.de/media/uploads/files/pdf/cidoc2014_final.pdf (accessed 7/9/2023).

Conolly, James/Lake, Mark (2006), *Geographical Information Systems in Archaeology*, Cambridge.

Grossmann, G. Ulrich (2010), *Einführung in die Historische und Kunsthistorische Bauforschung*, Darmstadt.

Hardesty, Donald L.,/Little, Barbara J. (2009), *Assessing Site Significance: A Guide for Archaeologists and Historians*, Plymouth.

Hechberger, Werner (2010), *Adel, Ministerialität und Rittertum im Mittelalter,* München.

Kaaij, Martijn van der (2019), "Automatic Metadata (Entity) Extraction and Workflow Efficiency: Real Life Solutions", in: *Archiving Conference* 2019 (1), 90–93. https://doi.org/10.2352/issn.2168-3204.2019.1.0.20.

Keddigkeit, Jürgen/Thon, Alexander/Scherer, Karl/Übel, Rolf/Burkhart, Ulrich (eds.) (2007), *Pfälzisches Burgenlexikon Band I: A–E*, Kaiserslautern.

Keupp, Jan Ulrich (2002), *Dienst und Verdienst: Die Ministerialen Friedrich Barbarossas und Heinrichs VI*, Stuttgart.

Kuczera, Andreas (2017), "Graphentechnologien in den Digitalen Geisteswissenschaften", in: *ABI Technik* 37 (3), 179–196, https://doi.org/10.1515/abitech-2017-0042.

Kuczera, Andreas/Wübbena, Thorsten/Kollatz, Thomas (2019), "Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Zur Einführung in diesen Band", in: *Zeitschrift für digitale Geisteswissenschaften* Sonderband 4, https://doi.org/10.17175/sb004_013.

Mash, Cole/Crompton, Constance/INKE Research Group/White, Travis (2016), "Unknown But Not Unknowable: The Network of Identified and Unidentified Hands in the Social Edition of the Devonshire Manuscript", in: *Scholarly and Research Communication* 7 (2), https://doi.org/10.22230/src.2016v7n2/3a260.

Pattee, Aaron (2023), *CITADEL: Computational Investigation of the Topographical and Architectural Designs in an Evolving Landscape*, Dissertation an der Philosophischen Fakultät der Universität Heidelberg zur Erlangung des akademischen Grades des Doktors der Philosophie, Heidelberg: Universität Heidelberg, https://doi.org/10.11588/artdok.00007958.

Pattee, Aaron/Volkmann, Armin/Anders, Katharina/Mara, Hubert (2018), "Analysing the Medieval Landscape of the German Palatinate", in: *GI_Forum* 1, 39–49, https://doi.org/10.1553/giscience2018_02_s39.

Robinson, Ian/Webber, Jim/Eifrem, Emil (2015), *Graph Databases: New Opportunities for Connected Data*, Sebastopol.

Schubert, Alexander (ed.) (2018), *Richard Löwenherz: König – Ritter – Gefangener*, Regensburg.

Tomasi, Francesca (2018), “Modelling in the Digital Humanities: Conceptual Data Models and Knowledge Organization in the Cultural Heritage Domain”, in: *Historical Social Research* Supplement 31, 170–179, https://doi.org/10.12759/hsr.suppl.31.2018.170-179.

Winberg, Victor/Zubac, Jan (2019), “A Comparison of Relational and Graph Databases for CRM Systems”, Master’s Thesis, Lund University.

Figure Credits

Fig. 1–3 Aaron Pattee.