Modeling as a Scholarly Process: The Impact of Modeling Decisions on Data-Driven Research Practices

Aline J. E. Deicke
Academy of Sciences and Literature | Mainz
Mainz, Germany

Abstract

In the digital humanities, the creation of a data model usually represents a pivotal stage in the research process. Data modeling not only serves to describe the domain in question and to guarantee interoperability, but it also helps to structure the available information and to gain new insights and perspectives into the subject at hand. At the same time, the concomitant processes of classification may have a considerable impact on subsequent studies. In the present paper, this tension is explored through the analysis of two exemplary use cases, each of which employs a different data model, whose impact on the research process is in turn examined: the modeling of the entity of a 'burial,' and the modeling of the spatial relations between objects in a grave. The conclusion will examine questions arising from this inquiry with a special focus on scholars working with research-centered databases.

1 Introduction

As the digital humanities continue to grow and develop, it has become increasingly clear that the research process starts long before the actual analysis whenever digital methods or digital data are involved. Any engagement

In: Tara Andrews, Franziska Diehr, Thomas Efer, Andreas Kuczera and Joris van Zundert (eds.): Graph Technologies in the Humanities - Proceedings 2020, published at http://ceur-ws.org.

with the data at hand must be preceded by a thorough consideration of the characteristics of the phenomenon to be studied, and the structuring of its information domain – or in other words, by the creation of a data model. This holds true not only for research groups working on large-scale projects such as digital editions or corpora, but also for individual scholars focusing on specific data-driven questions. In terms of data structure and data models, these two scenarios are characterized by very distinct requirements and limitations: while the objectives and architecture of larger projects are often conducive to the use of standardized formats, schemata, and ontologies, the application of the same principles to studies conducted by individual researchers or small research groups can prove to be quite a challenge.

For research projects rooted in the humanities, data modeling can serve a variety of purposes which often – but not necessarily – result in the functional implementation of the respective models (Beynon et al., 2006, p. 146). First and foremost, data modeling creates a formal model that represents one or more objects, concrete or abstract (Jannidis, 2017, p. 100), and describes "some segment of the world in such a way [as] to make some aspects computable" (Flanders and Jannidis, 2015, p. 3), which in turn enables complex logical operations on data, as well as communication about this information between computational systems. According to Jannidis (2017, p. 100) and Rehbein (2017, p. 162), if the model is well-formed and conforms to generally accepted standards, it secures a higher quality of data, and permits the exchange or merging of individual datasets, thereby guaranteeing interoperability. As such, it constitutes a vital prerequisite for the sustainable management of research data according to FAIR principles (Wilkinson et al., 2016).

In addition, modeling one's data also creates a basis for human communication in that it describes the domain in question, structures the available information, and offers new insights and perspectives regarding the subject at hand. Ciula and Eide go as far as to prioritize this function of modeling as "a creative process of thinking and reasoning where meaning is made and negotiated" over its more prosaic uses in database implementation (Ciula and Eide, 2017, p. i34).

However, the processes of classification that accompany such efforts can also shape subsequent studies to a considerable extent. For example, choices of ontologies and decisions on how to apply them can open up new paths of inquiry, but, at the same time, can also close off others well before the collection, analysis, and interpretation of data has even begun. The assignment of classes and properties, especially in the early stages of research, when some implications or theoretical perspectives might not yet have been thoroughly explored, can canonize certain points of view with regards to the data, when a more flexible understanding would have been beneficial (cf. Bowker and Star, 2008).

Obviously, such challenges are nothing new in humanities research (Flanders and Jannidis, 2019, p. 3). The ordering, categorization, and hierarchical modeling of information entities and the relations between them has long since been one of the primary tasks of scholars, for example in archaeology, where typology is used to create elaborate taxonomies of objects and their stylistic and functional development, and philosophy, which is where such ontological work originates (Arp et al., 2015, p. xxi). Yet, as Flanders and Jannidis note, "we inherit from the humanistic tradition a set of modeling practices and concepts that, while foundational, are often unsystematic, poorly understood by non-specialists, and invisible through their very familiarity" (Flanders and Jannidis, 2019, p. 5). The challenge in formalizing these practices and concepts with the goal of making some aspects computable lies in the need to reflect on this heritage in order to avoid the codification of its negative by-products into the work that is being done using digital methods. When conducted in a thorough manner, the practice of data modeling supports these efforts: not only does it contribute to our ability to present data in a machine-readable fashion and to formally structure information, but it also helps to make implicit biases or assumptions explicit.

The humanities already possess the tools required to meet this challenge – most notably in the form of the critical reflection of research practices that is part and parcel of postmodern and post-processual theory, and that can be readily transferred into the domain of the digital humanities. Indeed, there have been many instances of importation and adaptation of critical perspectives over the last decades. For example, Michael Shanks proposes a "symmetrical archaeology" (Shanks, 2007) that acknowledges how the study of the past is always a recreation of the past, and therefore shaped by modern biases, conventions, and habits. To create models of this "present past" means to reflect on these underlying assumptions, and – ideally – to challenge them. In a similar fashion, Susan Pollock and Reinhard Bernbeck question the validity of practices of classification that invariably categorize material remains by modern, as opposed to (pre-)historic, standards (Pollock and Bernbeck, 2010). In the digital humanities, Elena Pierazzo explores the "inevitable subjectivities" of the modeling process and their consequences for subsequent analysis (Pierazzo, 2019). This fundamental subjectivity is also emphasized by Arp et al.: drawing on theories such as epistemological representationalism and epistemological idealism, including the idea that we perceive 'reality' only through the filter of our own thoughts or concepts, they propose that ontologies cannot describe those realities, but only "conceptual items" (Arp et al., 2015, pp. 7-8). Going even further, Ciula and Eide emphasize the active element of creation, negotiation, and manipulation of external representations that is an integral part of the modeling process (Ciula and Eide, 2017, p. i34).

As far as the practice of data modeling is concerned, these examples remind us that researchers must be aware – in creating as well as in working with data models – that there cannot be one true data model representing the entirety of a certain real-world phenomenon. Instead, modeling practices reflect conventions, traditions, and individual habits of researchers, as well as the particular goal of the respective modeling project.

These considerations seem especially relevant as the way that knowledge is represented and structured has a profound and immediate relevance for society. As Hui states, digital objects can be understood as "externalized memories that condition our retrieval of the past and our anticipation of the future" (Hui, 2012, p. 390). Meanwhile, Arp et al. see ontologies as "publicly available representations of scientific information about reality" (Arp et al., 2015, p. 24). Through their implementation in openly accessible databases, data models shape how information is constructed, perceived, and processed in the public mind. Furthermore, in the face of the ever-growing field of Artificial Intelligence, carefully curated knowledge bases are becoming more important than ever (Rehbein, 2017, p. 162) (Schelbert, 2019, p. 146), since the biases unconsciously encoded into these datasets have a direct bearing on the choices made by AI entities. These arguments, of course, also extend to algorithms and their critical reflection.

The challenge of the subjectivity of researchers and research perspectives applies especially to those types of databases that Flanders and Jannidis call "research-driven" (Flanders and Jannidis, 2015, pp. 4-5) (Jannidis, 2017, p. 102), and that Ciula and Eide designate as modeled "for understanding" (Ciula and Eide, 2017, p. i36). In contrast to large-scale, curation-driven projects that focus on providing end users with generalized, standards-conforming data, research-driven data models are used to explore and express specific ideas or research interests, usually in the context of work conducted by individual scholars on highly specialized objects of inquiry. Often, they will be tailored to specific analytical methods, such as the quantitative methods that are used in network analysis.

Consequently, interoperability might be of secondary concern. Instead, different ways of modeling the same 'real-world' phenomenon reflect different research questions, or theoretical frameworks employed to approach these questions. In such cases, data models have a profound effect on research processes as a formal, structured way to think about the phenomena studied and the information from which these data models are built. Furthermore, data modeling shapes the scholarly process by focusing on specific aspects which are analyzed in greater detail, while others potentially remain unexplored.

This tendency can also be observed in the context of graph databases, particularly in combination with network research: on the one hand, modeling data in a graph promotes 'thinking in relations' even at the very beginning of the scholarly process and long before these relations are realized in the database itself, because the task of transforming unstructured data into nodes and edges can shift a researcher's focus towards connections that were hitherto unnoticed or disregarded; on the other hand, once defined, these modeling decisions can complicate certain avenues of analysis, or prevent them from being pursued altogether.

In this paper, two use cases are presented to demonstrate how the scholarly process starts with and impacts data modeling, and how differently constructed models facilitate different perspectives on the same subject. While both originate in a study about elite networks of the Late Bronze Age (Deicke, 2021, 2020), the first considers modeling the entity of 'burial,' which leads to the various theoretical concepts tied to this term and the various meanings it can hold depending on the specific context, whereas the second looks at different possibilities of modeling spatial relations between objects in a grave, which are reciprocally linked to the questions one can ask from the subject material.

2 Modeling as a Scholarly Process

2.1 Case Study: Identity Construction in Funeral Status Networks of the Late Bronze Age

The case study presented in this paper deals with elite burials of the late Urnfield period (c. ninth century BC), which constitutes the last period of the Central European Bronze Age in an area that reaches from Eastern France to the entrance of the Carpathian Basin and from the Alps to the German *Mittelgebirge*. Grave goods and other characteristics, such as grave architecture, are analyzed in the form of a network consisting of burials and the grave goods and features that they share (Figure 1). As the construction of burials and the selection of grave goods were most likely intentional acts that served to represent intersecting identities, statuses, and sources of power of the deceased or their successors, the patterns evident in the network reveal which strategies the elite(s) pursued to gain prominence at the beginning of the transition from Bronze to Iron Age, a time of marked socio-political unrest.

The burial of a member of the elite also constituted a focal point for the construction and negotiation of communal as well as individual identities in a manner analogous to Giddens' theory of the duality of structure, which states that "the continuous recreation and re-articulation of identities through burial rite [...] feeds back into the structuring of societies" (Giddens, 1979, p. 69) (Lucy, 2005, p. 105).

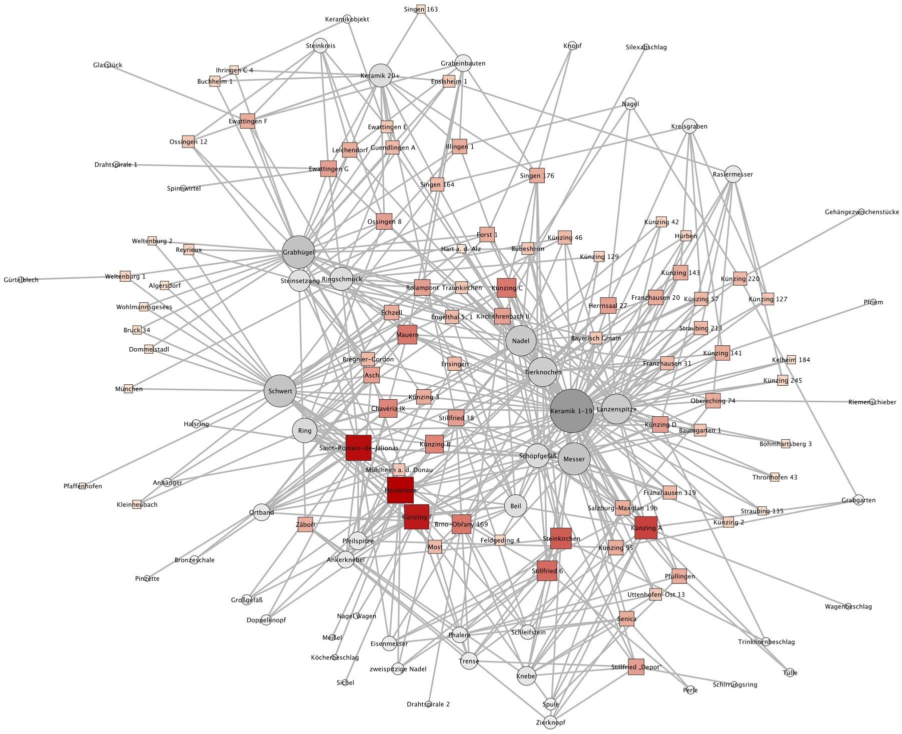

Figure 1: Two-mode network of burials (squares) and associated grave goods and features (circles). Node size and colour: degree centrality. Layout: Stress minimization (incl. overlap removal and manual adjustments).

For the purposes of this study, 82 previously published burials were selected on the basis of the presence of objects that most likely represented prestige goods in the late Urnfield culture, as well as additional signifiers of high status that can be assumed to have been characteristic of the funeral presentation of elites, such as the presence of an abundance of ceramic vessels and elaborate grave architecture (Deicke, 2021, pp. 13-20).

The data was stored in a graph database[1] and exported into network analysis software[2] in various configurations in order to analyze it from different

[1]https://neo4j.com/release-notes/neo4j-3-4-10

[2]http://visone.info

network perspectives. The underlying data model as well as the ones presented in this paper are mainly based on the CIDOC CRM (Doerr et al., 2020), the standard ontology in the field of cultural heritage (cf. Deicke, 2016).

2.2 What is a Burial?

As discussed above, the question of how to model the burials themselves was of central importance to the study – the goal being not only to be able to store, analyze, and retrieve data quickly and effortlessly in and from the graph database, but also to facilitate the export of specific one- and two-mode networks that figured prominently in the course of the network research. In this regard, the structure of the model directly affected which inferences could be drawn from the subsequent analysis. To illustrate this point, two models are presented below. The first one is based on the research question outlined above, namely: how is elite identity constructed through elements of the burial?

As can be seen in Figure 2, in this case, burial as a general term is characterized as *E19 Physical Object*, which, according to the specification, is comprised of "aggregates of objects made for functional purposes" (Doerr et al., 2020, p. 14). It is further classified according to *E41 Appellation* and *E22 Type*, it forms part of a specific *E27 Site*, and it contains human remains (*E20 Biological Object* and *E21 Person*). The elements relevant for status display – features such as ditches or tumuli, biological remains such as animal bones or meat offerings, and artifacts such as grave goods – are expressed through the classes *E25 Man-Made Feature*, *E20 Biological Object*, and *E22 Man-Made Object*, respectively. As these nodes represent the individual objects themselves (e.g. a particular sword of type Mörigen with a specific appearance and object biography (cf. Kopytoff, 2009) (Quillfeldt, 1995, pp. 242-243)), they are also assigned a general *E22 Type*, such as 'sword.' This broad classification ties the individual objects to the socio-cultural significance vested in them – for swords, the expression of military power is commonly assumed – while simultaneously ensuring the degree of generalization that is necessary for the construction of a network made up of *E19 Physical Objects* and the *E22 Types* found among them.

While this model serves its purpose and is supported by the specification of the CRM, it does present some problems, especially when it comes to the assignment of the class *E19 Physical Object* to the burial itself. By designating the individual physical structures of the burial classified as *E25 Man-Made Feature* as distinct elements, the burial node no longer embodies an actual physical object, but rather an abstract container that is defined by the association of the different types of entities described above. At the same time, it can be argued that a burial in general should be seen as a conceptual entity comprised not only of material remains, but also of (mainly ritual) actions – a point to which I will return in my discussion of the second model. The CIDOC CRM offers the class *E28 Conceptual Object* for "non-material products of our minds and other human produced data that have become objects of a discourse about their identity, circumstances of creation or historical implication" (Doerr et al., 2020, p. 18). Yet its hierarchical inheritance pattern does not allow this class to be used in conjunction with the *P46 forms part of* property necessary to connect the burial directly to its elements. Therefore, even if the burial itself is seen as an abstract entity defined purely by its associated features, the model based on the *E19 Physical Object* class pushes the inquiry into a reductionist direction that might not be intended by the researcher, and that does not adequately represent the research process. As such, the model presents an – if relatively benign – example of how one's choice of ontology can lead to very specific analytical approaches.

By way of contrast, the second research question that I wish to explore in this paper focuses less on the material background and more on the concept of social practice as an integral building block of societies: which social processes constitute a burial?

Analogous to the first example, the data model constructed in response to this research question is based on the CIDOC CRM. As the CRM is explicitly described as an "event-centric model" (Doerr et al., 2020, p. xix), an approach based on social practice rather than entities and their attributes appears to be well suited to the spirit of the ontology. Yet, the processes inscribed in the CRM to date mostly concern events from the realm of cultural heritage management and museum documentation (Doerr et al., 2020, p. i). In order to help bridge this gap, classes and properties of the compatible model CRMsoc were included, whose first version was published in May 2019 (Alamercery et al., 2019). The document, which currently exists in draft form, proposes a "domain ontology [...] that can be used to (re-)encode data that document social phenomena and constructs that are typically recorded by humanities and social science scholars" (Alamercery et al., 2019, p. 2).

Even at a cursory glance, this model appears to be much more complex than the first (Figure 3). Adding a meta level in the form of actual persons and their activities to the objects that are typically the focus of archaeological research expands each relation by two or three additional steps. Here, too, the burial is understood as an *E7 Activity*; *E20 Biological Objects*, *E22 Man-Made Objects*, and *E25 Man-Made Features* are now connected to it through their use in the burial rite. Added to this model are the *E39 Actor*, representing the successor of the deceased, and the *E74 Group*, which – as the burial community – provides the social framework for all actions taking place in the context of the burial. As such, the model directs the researcher's attention not to the objects themselves, but rather to the events of production and modification that made them a part of the burial, their purpose, and the relationships encoded within them. While it is certainly possible to derive status-related inferences from the interplay of these actions, the focus shifts away from the intra- and supra-regional significance of the *social persona* of the deceased (Binford, 1971, p. 17) (Saxe, 1970, pp. 5-7) as constructed specifically for presentation in the grave, which exists as a relatively stable result of the *rite de passage* that is the funeral (van Gennep, 2005, pp. 142-159). Instead, questions concerning the influence of different actors and groups on this presentation, their agency and role in the various processes of creation, modification, and production surrounding the funeral, and the configuration of the specific burial community itself assume greater prominence. What is required is thus a source record that is comprised not only of the traditional description of artifacts, but that also encompasses an exact documentation of their discovery, data from scientific analysis (e.g. traces of usage or material composition), and an extensive overview of studies on comparable cases, which operate under the same framework of a processual – as opposed to merely descriptive – approach.

2.3 Spatial Relations within a Grave

Another aspect that emphasizes function and social practice over typological and chronological attributes is the arrangement of artifacts in the grave. As Martina Löw states, space represents a social construct that materializes and impacts perceptions and interactions of communities (Löw, 2017, pp. 166-172). Similar to the selection of grave goods, the placement of these goods carries a variety of meanings which, while not always reconstructable, can still be observed.

Within this context, different conceptions and encodings of space answer different research questions. A geospatial coordinate system, for example, seems most useful for the exact documentation of the location of an object, as in the case of an excavation. This attributive absolute understanding of space as topology can be easily expressed through the CRM (Figure 4): graves, cemeteries, or other types of locations of finds – in this case expressed through a generic *E27 Site* node – can be assigned an *E53 Place* through the property *P53 has former or current location*. The *E53 Place* class can then be enriched by adding an *E44 Place Appellation*, coordinates in the form of an *E94 Space Primitive*, and – if necessary – an *E55 Type*. Hierarchies of these places can be constructed through the property *P89 falls within*. In theory, the same approach could be taken for individual elements of a burial, such as grave goods or features.

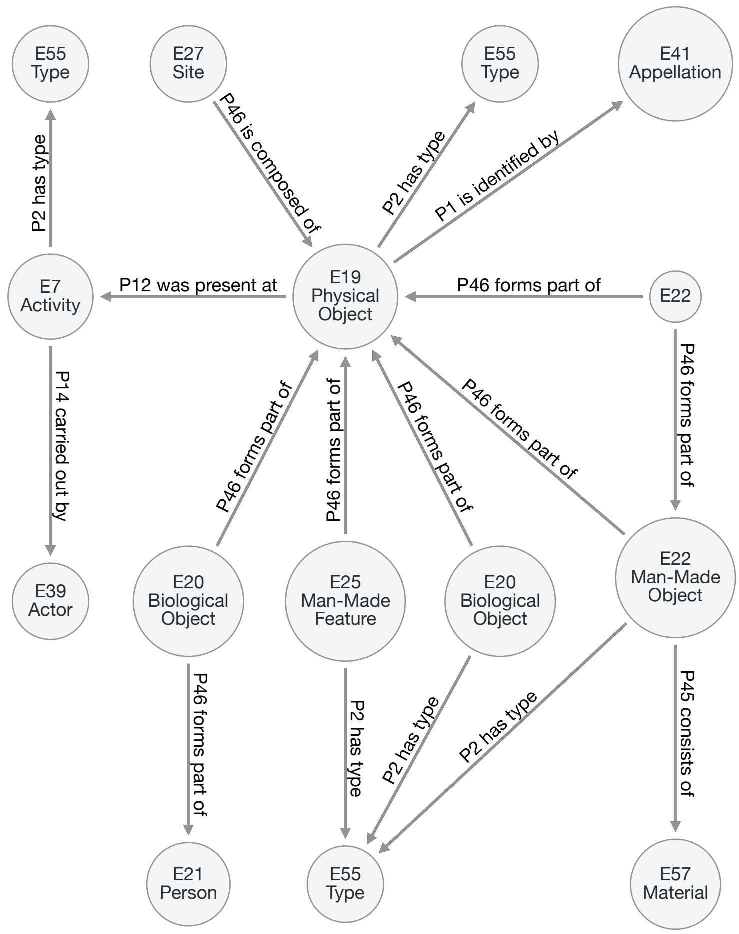

Figure 2: Conceptual data model of a burial consisting of artifacts, animal remains, architectural features, and human remains, based on the CIDOC CRM

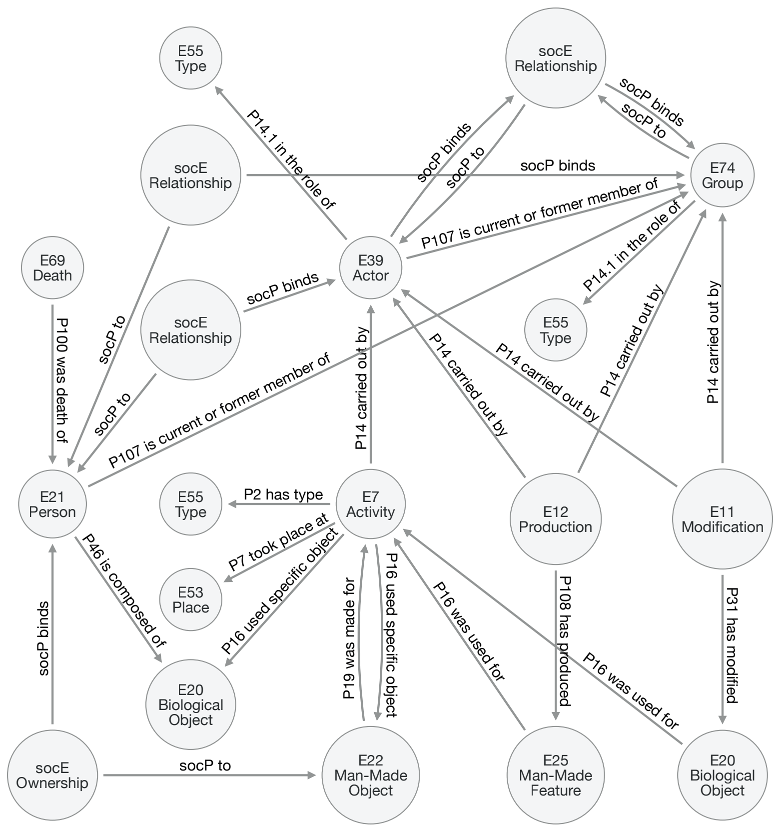

Figure 3: Conceptual data model of a burial consisting of artifacts, animal remains, architectural features, and human remains that focuses on underlying social processes, based on the CIDOC CRM and CRMsoc

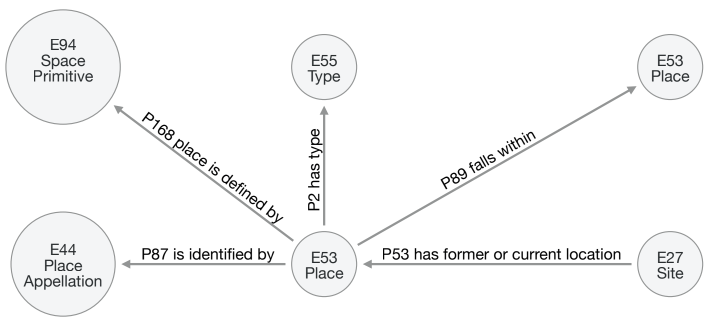

Figure 4: Conceptual data model of spatial attributes of a site, based on the CIDOC CRM

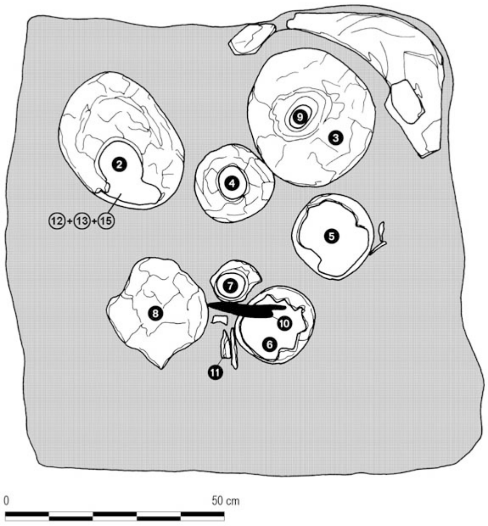

Figure 5: Franzhausen-Kokoron, Grave 119, plan of the burial (Lochner and Hellerschmid, 2016b, pl. 71)

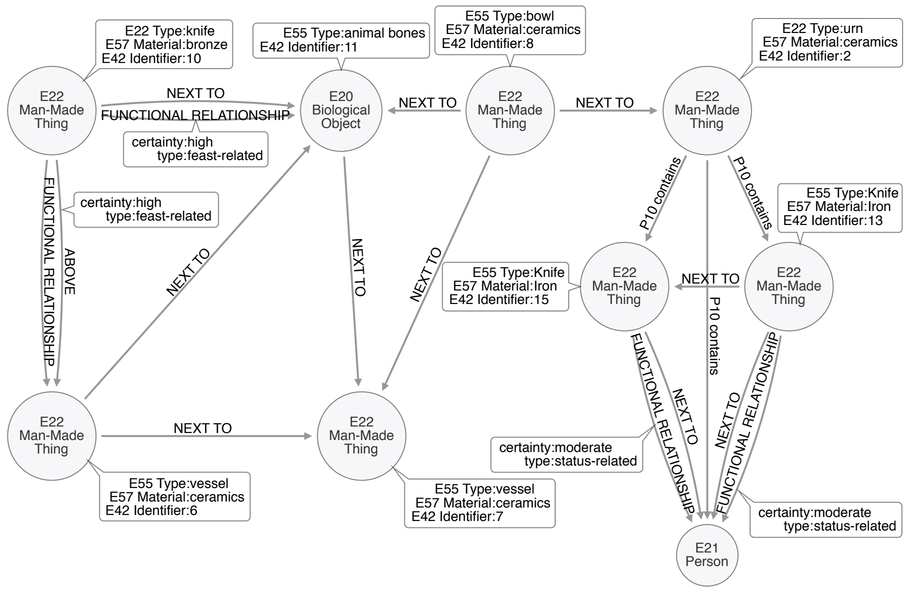

Figure 6: Draft of a conceptual model of spatial relations between a selection of finds from Grave 119 in Franzhausen-Kokoron, based on the CIDOC CRM and extended through additional *E42 identifiers* corresponding to Figure 5

Apart from documentary purposes, the assignments of geometric information also serves as a general frame of reference that can be translated into other, less normative conceptions of space. For example, the location of objects in relation to the architectural features of a burial pit or chamber can answer questions regarding the standardization of burial activities, including deposition patterns, the construction of spaces within the grave as zones that give material form to different identity aspects of the deceased, and their relationship to each other, or even the role of space as an actor in the perception and negotiation of the burial process.

In this paper, however, the focus lies on yet another aspect: the placement of objects in relation to each other. On a very basic level, the analysis of these kinds of spatial patterns allows inferences about the function of objects that might otherwise remain obscure, and, in a broader sense, about intra- and supra-regional meanings of individual artifacts or groups of artifacts in connection or opposition to each other. Grave 119 of the cemetery of Franzhausen (Nußdorf ob der Traisen, Lower Austria) provides a comparatively simple example, particularly when it comes to the bronze and iron knives it contained.

As can be seen in Figure 5, the assorted ceramic and metal grave goods found in Grave 119 are systematically distributed over the area of the rectangular burial pit. The metal objects (no. 9, 10, 12, 13, 15) are divided up and placed into several of the ceramic vessels. Of particular interest here are one bronze (no. 10) and two iron knives (no. 13, 15), whose contexts are very different: while the bronze knife lies across the remains of a vessel that unfortunately cannot be reconstructed and is accompanied by animal bones (most likely a meat offering or remains of a funeral feast), the iron knives were found inside of an urn, alongside the ashes of a cremated body (Lochner and Hellerschmid, 2016a, Grave 119). Similar associations of bronze knives with food offerings and iron ones with the body of the deceased can be observed in contemporaneous burials as far away as Eastern France, for example in the tumulus of Saint-Romain-de-Jalionas (Dép. Isère) (Brun, 1987, pp. 216-217) (Deicke, 2021, pp. 152-153).

While these findings might seem trivial at first glance, the different treatment of the same type of object depending on its material ties into a wider cultural context: the transition from Bronze to Iron Age, and the increasingly widespread adoption of iron as a production material. Especially with regard to the way in which objective sources of power (Lehman, 1969, p. 454-455) are materialized as part of an elite funerary identity, the deposition of the iron knife *not* in a utilitarian context, but as part of the personal accoutrements of the deceased, hints at the important role of this new technology in personal strategies of preservation, consolidation, and attainment of power as the Bronze Age drew to its close.

It can be assumed that many more insights into the function and meaning of objects, as well as the social construction of funeral space, could be inferred from the respective arrangements of grave goods. In particular, incorporating spatial relations into the data model of a corresponding knowledge graph that extends one of the models shown above could allow automatic or semi-automatic queries that would in turn point researchers to other potentially fruitful constellations. At the moment, such socio-theoretical understandings of space have not been integrated into the CIDOC CRM. While compatible models such as CRMarchaeo (Doerr et al., 2019) and CRMgeo (Hiebel et al., 2015) do expand the classes and properties of the core ontology related to the documentation of space, they do not (yet) support the detailed modeling of the spatial relations between artifacts and features as described above.

In the absence of suitable standards, Figure 6 represents an initial attempt to construct a simple model of the use case outlined above. It incorporates only the section of the inventory connected to the three knives and their possible functions, focusing on two types of relationships or properties: those that describe the placement of an object in relation to another, and those that provide an interpretation of this connection. Regarding the latter, type and certainty are expressed through attributes. As the model is merely a tentative first step towards a possible structure for such data, some caveats apply: for example, the plainly labelled relations of 'next to' and 'above' are in need of a more formal treatment, preferably in the form of a controlled vocabulary or even a hierarchical thesaurus. Likewise, the question of which connections should be incorporated into the model requires further clarification: should the spatial relations between all objects be codified? Or does it suffice to provide an explication of the ones deemed to be 'important'?

In contrast to the CRM, space or spaces as such do not appear as nodes of their own in this concept, at least not at this point in the modeling process. Drawing on the two processes involved in the creation of social space proposed by Löw, *spacing* and *synthesis* (Löw, 2017, pp. 158-161), the edges explicating the positioning between objects could be used to indicate the former, while the ties indicating functional relationships could act as a starting point for the latter. Additional aspects of socially constructed spaces could then be inferred from the graph created through these connections, and included as additional classes, and consequently, nodes.

3 Conclusion

These two short examples show how the practice of data modeling constitutes an explicit part of the research process that influences the conclusions that can be drawn from the material at hand. At the same time, the theoretical frameworks or premises that are present in the primary design of a given case study exert a strong influence on the resulting model. For this reason, the goal of data modeling in the humanities, especially in the context of research-driven projects like the ones discussed above, cannot be to 'streamline' data structures. Instead, what is desirable is for the diversity that is to be found in study design to be mirrored in the respective modeling approaches. It is incumbent upon humanities researchers to address crucial questions such as the following: how can the task of modeling be further integrated into the scholarly process? How can one efficiently and clearly disclose the choices that lead to particular models? And which possible implications and biases could arise from these data structures when analyzing specific data-driven research questions?

Beyond these lines of inquiry, which mainly concern communication between humans, there also remains a question about the communication between machines, or in other words, interoperability. As Arp et al. note,

> [o]ne central goal of the annotation of data using ontologies is to enable what is called 'semantic interoperability' between heterogeneous computer systems, defined as the ability of two or more such systems to exchange information in such a way that the meaning of the information generated by any one system can be automatically interpreted by each receiving system accurately enough to produce results useful to its end users. (Arp et al., 2015, p. 38).

Today's digital humanities emphasize the linking of information that enables such exchanges, as the steady growth of fields such as knowledge graphs, the Semantic Web, and Linked Open Data shows. Yet, especially in the case of research-driven databases, the question of how to navigate expectations between the poles of universal interoperability and case-specific models, and how to integrate such models in the wider network of the knowledge domain, can be difficult to answer. Nevertheless, such considerations are of fundamental importance, particularly in cases where researchers are publishing previously unknown material as part of their projects (e.g. recently excavated archaeological finds or newly edited historical sources), or where they aim to integrate data from other publicly accessible databases.

On a very fundamental level, a partial solution to these challenges lies in the thorough publication and documentation of datasets. In the case of research-driven projects (e.g. in the form of quantitative analysis), databases should customarily be published along with the findings. For these as well as for curation-driven databases, data models should be readily accessible and thoroughly described. In the accompanying documentation, the underlying logic and purpose should be made explicit – in general, researchers should aim to develop a critical approach to practices of modeling.

Concerning the interoperability of databases, it seems an easy solution to call for more generic data models that can be easily mapped onto each other, or to dismiss the contribution of research-driven data collections to the overall record in general – especially since, at present, the online publication of these collections can be difficult to manage for individual researchers or small teams without access to the necessary infrastructure.

Indeed, the task of mapping two data models or schemata onto each other – in the sense of "a sufficient specification to [allow the] transformation of each instance of schema 1 into an instance of schema 2 with the same meaning"[3] – involves numerous challenges. On a structural as well as a semantic level, data models can differ significantly from one another. Some factors have already been discussed, yet more can arise, concerning data types, naming conventions, the level of detail or completeness, inconsistencies, as well as fundamental discrepancies in the design of the underlying ontologies (cf. Dröge, 2010, p. 144, tab. 1). However, numerous tools, methods, and use cases for mapping different models or schemata onto each other already exist (several examples are listed and described on the CIDOC CRM website[4]) and can either be used directly or serve as a guideline for comparable processes.

For research-driven projects, these approaches might not be immediately usable. Yet it is precisely in these cases that the exact goal of the mapping is particularly relevant. As pointed out above, the objective must be to produce a system that is "*useful* to its end users" (Arp et al., 2015, p. 38; emphasis by author). Two cases can easily be imagined: the integration of collected data into more general, curation-driven databases (for example into union catalogues such as Kalliope[5]); or the combination of two databases that have been created with similar research interests in mind.

In the first case, the data model might matter less than the careful enrichment of individual entries with standardized metadata, such as identifiers from the Getty Thesaurus of Geographic Names[6] or the integrated author-

---

[3] http://www.cidoc-crm.org/short-intro-mappings

[4] http://www.cidoc-crm.org/mapping-methods-technology, http://www.cidoc-crm.org/mapping-tools, http://www.cidoc-crm.org/reports_mappings

[5] https://kalliope-verbund.info/en/index.html

[6] http://www.getty.edu/research/tools/vocabularies/tgn/index.html)

ity file (GND) of the German National Library.[7] Even if certain standard schemata are required, the identification of basic entities such as persons and places is likely to pose only a moderate challenge, particularly if the respective target database is well documented.[8]

In the second case, researchers must first identify to what degree the two databases in question overlap in terms of their theoretical and structural scope. A complete mapping of concepts and relations might not be a sensible approach if, for example, certain terms have different meanings or ranges. Particularly in the case of graph databases, the question of which entities are presented as nodes and which as attributes, and whether events or circumstances are modeled as nodes or edges, can also distinguish individual models. If approaches vary considerably, the exploitation of ontological hierarchies or the use of upper ontologies[9] can build bridges to connect diverse datasets. As the top level classes of the CIDOC CRM form one such upper ontology (Doerr et al., 2020, pp. xviii-xx), its use allows basic mappings to other models that use the CRM, while facilitating connections to models based on other top-level ontologies, such as the Basic Formal Ontology (BFO; Smith, 2015). While this method certainly leads to substantial simplifications, it also removes ambiguities and provides a starting point from which complexity can be re-introduced into a merged data collection.

In conclusion, data modeling presents a necessary and fundamental part of the scholarly process, and can provide insights into research topics beyond the immediate goal of database implementation. Data modeling also confronts researchers – creators as well as users of research data – with its own set of challenges: this paper has shown how data models and the (implicit) premises and theories built into them *a priori* can shape the study of phenomena in the humanities; how they influence researchers in terms of the analytical approaches, interpretations, and conclusions that are available to them on the basis of specific modeling choices; and how they can obstruct or block entire paths of inquiry, whether it be on account of technical barriers or one-directional thought processes. To address these issues in a critical and deliberate manner, the publishing process must include clear and expli-

[7] `https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html`

[8] For example, the web service correspSearch of the Berlin-Brandenburg Academy of Sciences and Humanities (`https://correspsearch.net/index.xql?l=en`, that collects metadata from scholarly editions of letters gives detailed information on which formats and authority files need to be provided to add information to the service (`https://correspsearch.net/index.xql?id=participate_steps&l=en`, and which can then be exported even from highly specialized databases with a reasonable amount of effort.

[9] Many thanks to Cristina Vertan (University of Hamburg) for her helpful suggestion regarding upper ontologies in the discussion that followed the presentation in Vienna on which this article is based.

cit documentation which outlines the reasons for selecting specific models, discusses the implications of these choices for subsequent research activities, and thoroughly describes the data models, schemata, and authority files used. If, however, mapping approaches carefully consider the scope and goal of the datasets in question, they can facilitate interoperability even in the case of research-driven databases, and enable both individual researchers and large-scale projects to share their data, analytical approaches, and research results with a much broader audience.

References

Alamercery, V., Beretta, F., Bruseker, G., Doerr, M., et al. (2019). *Definition of the CRMsoc. An Extension of CIDOC CRM To Support Social Documentation. Version 1.0.* ICOM/CIDOC Documentation Standards Group/CRM Special Interest Group, http://www.cidoc-crm.org/crmsoc/sites/default/files/CRMsoc_20190326.pdf.

Arp, R., Smith, B., and Spear, A. D. (2015). *Building Ontologies With Basic Formal Ontology*. MIT Press, Cambridge, MA.

Beynon, M., Russ, S., and McCarty, W. (2006). Human Computing – Modelling With Meaning. *Literary and Linguistic Computing*, 21(2):141–157, DOI: 10.1093/llc/fql015.

Binford, L. R. (1971). Mortuary Practices: Their Study and Their Potential. *Memoirs of the Society for American Archaeology*, 25:6–29, http://www.jstor.org/stable/25146709.

Bowker, G. C. and Star, S. L. (2008). *Sorting Things Out: Classification and Its Consequences*. MIT Press, Cambridge, MA, 8th edition.

Brun, P. (1987). *Princes et princesses de la celtique: le premier âge du Fer en Europe 850-450 av. J.-C.* Collection des Hespérides. Errance, Paris.

Ciula, A. and Eide, Ø. (2017). Modelling in Digital Humanities: Signs in Context. *Digital Scholarship in the Humanities*, 32(1):i33–i46, DOI: 10.1093/llc/fqw045.

Deicke, A. (2016). Cidoc CRM-Based Modeling of Archaeological Catalogue Data. In De Luca, E. W. and Bianchini, P., editors, *DHC 2016. Digital Humanities and Digital Curation. Proceedings of the First Workshop on Digital Humanities and Digital Curation*, number 1764 in CEUR Workshop Proceedings. http://ceur-ws.org/Vol-1764/.

Deicke, A. (2020). Entangled Identities: Processes of Status Construction in Late Urnfield Burials. In Donnellan, L., editor, *Archaeological Networks and Social Interaction*, Routledge Studies in Archaeology, pages 38–63. Routledge, Abingdon/New York, NY.

Deicke, A. (2021). *Zwischen Individuum und communitas. Identitätskonstruktion späturnenfelderzeitlicher Eliten im Spiegel funeraler Statusnetzwerke*. Number 358 in Universitätsforschungen zur prähistorischen Archäologie. Dr. Rudolf Habelt, Bonn.

Doerr, M., Bekiari, C., Bruseker, G., Christian-Emil, O., et al. (2020). Definition of the CIDOC Conceptual Reference Model. Version 7.0. http://www.cidoc-crm.org/sites/default/files/CIDOC%20CRM_v.7.0_%2020-6-2020.pdf.

Doerr, M., Felicetti, A., Hermon, S., Hiebel, G., et al. (2019). Definition of the CRMarchaeo. An Extension of CIDOC CRM. Version 1.4.8. http://www.cidoc-crm.org/sites/default/files/CIDOC%20CRM_v.7.0_%2020-6-2020.pdf.

Dröge, E. (2010). Leitfaden für das Verbinden von Ontologien. *Information — Wissenschaft und Praxis*, 61(2):143–147, https://www.phil-fak.uni-duesseldorf.de/jp/student-research-projects/015/.

Flanders, J. and Jannidis, F. (2015). Knowledge Organization and Data Modeling in the Humanities. https://nbn-resolving.org/urn:nbn:de:bvb:20-opus-111270.

Flanders, J. and Jannidis, F. (2019). Data Modeling in a Digital Humanities Context. In Flanders, J. and Jannidis, F., editors, *The Shape of Data in Digital Humanities. Modeling Texts and Text-Based Resources*, Digital Research in the Arts and Humanities, pages 3–25. Routledge, London/New York, NY.

Giddens, A. (1979). *Central Problems in Social Theory: Action, Structure and Contradiction in Social Analysis*. Macmillan, London.

Hiebel, G., Doerr, M., Eide, Ø., and Theodoridou, M. (2015). CRMgeo: a Spatiotemporal Model. An Extension of CIDOC-CRM to Link the CIDOC CRM to GeoSPARQL Through a Spatiotemporal Refinement. Version 1.2. http://www.cidoc-crm.org/crmgeo/sites/default/files/CRMgeo1_2.pdf.

Hui, Y. (2012). What is a Digital Object? *Metaphilosophy*, 43(4):380–395, DOI: 10.1111/j.1467-9973.2012.01761.x.

Jannidis, F. (2017). Grundlagen der Datenmodellierung. In Jannidis, F., Kohle, H., and Rehbein, M., editors, *Digital Humanities: Eine Einführung*, pages 99–108. J.B. Metzler, Stuttgart.

Kopytoff, I. (2009). The Cultural Biography of Things. Commodization as Process. In Appadurai, A., editor, *The Social Life of Things. Commodities in Cultural Perspective*, pages 64–91. Cambridge University Press, Cambridge, MA, 7th edition, DOI: 10.1017/CBO9780511819582.

Lehman, E. W. (1969). Toward A Macrosociology of Power. *American Sociological Review*, 34(4):453–465, DOI: 10.2307/2091956.

Lochner, M. and Hellerschmid, I. (2016a). *Dokumentation Franzhausen-Kokoron: Ein Gräberfeld der jüngeren Urnenfelderkultur aus Zentraleuropa. Erweiterte interaktive Datenbank mit Illustrationen und Fundbeschreibungen. Version 03/epub*. Verlag der österreichischen Akademie der Wissenschaften, Wien, DOI: 10.1553/KatalogUFK.

Lochner, M. and Hellerschmid, I. (2016b). *Fundmaterial und Befunde des urnenfelderzeitlichen Gräberfeldes von Franzhausen-Kokoron – Tafelformat*. Verlag der österreichischen Akademie der Wissenschaften, Wien, DOI: 10.1553/Dokumentation_UFK.

Lucy, S. (2005). Ethnic and Cultural Identities. In Díaz-Andreu, M., Lucy, S., Babić, S., and Edwards, D. N., editors, *The Archaeology of Identity Approaches to Gender, Age, Status, Ethnicity and Religion*, pages 86–109. Routledge, London/New York.

Löw, M. (2017). *Raumsoziologie*. Suhrkamp, Frankfurt am Main, 9th edition.

Pierazzo, E. (2019). How Subjective Is Your Model? In Flanders, J. and Jannidis, F., editors, *The Shape of Data in Digital Humanities. Modeling Texts and Text-Based Resources*, Digital Research in the Arts and Humanities, pages 117–132. Routledge, London/New York, NY.

Pollock, S. and Bernbeck, R. (2010). An Archaeology of Categorization and Categories in Archaeology. *Paléorient*, 36(1):37–47.

Quillfeldt, I. v. (1995). *Die Vollgriffschwerter in Süddeutschland*. Number IV 11 in Prähistorische Bronzefunde. Steiner, Stuttgart.

Rehbein, M. (2017). Ontologien. In Jannidis, F., Kohle, H., and Rehbein, M., editors, *Digital Humanities: Eine Einführung*, pages 162–176. J.B. Metzler, Stuttgart.

Saxe, A. A. (1970). *Social Dimensions of Mortuary Practices*. PhD thesis, University of Michigan, Ann Arbor,MI.

Schelbert, G. (2019). Ein Modell ist ein Modell ist ein Modell – Brückenschläge in der Digitalität: Architekturmodell, Datenmodellierung, Digital Humanities, Kulturerbe-Dokumentation, Modelltheorie, Theoretisches Modell, Wissenschaftstheorie, Wissensmanagment. In Kuroczyński, P., Pfarr-Harfst, M., and Münster, S., editors, *Der Modelle Tugend 2.0. Digitale 3D-Rekonstruktion als virtueller Raum der architekturhistorischen Forschung*, number 2 in Computing in Art and Architecture, pages 137–153. arthistoricum.net, Heidelberg, DOI: 10.11588/arthistoricum.515.c7449.

Shanks, M. (2007). Symmetrical Archaeology. *World Archaeology*, 39(4):589–596, https://www.jstor.org/stable/40026151.

Smith, B. (2015). Basic Formal Ontology 2.0. Specification and User's Guide. https://raw.githubusercontent.com/BFO-ontology/BFO/master/docs/bfo2-reference/BFO2-Reference.pdf.

van Gennep, A. (2005). *Übergangsriten (les rites de passage)*. Campus, Frankfurt am Main, 3rd edition.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., et al. (2016). The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data*, 3(1):160018, DOI: 10.1038/sdata.2016.18.