Knowledge Graphs: Semantics, Machine Learning, and Languages
M. Acosta et al. (Eds.)

doi:10.3233/SSW230012

TRANSRAZ Data Model: Towards a Geosocial Representation of Historical Cities

Oleksandra BRUNS [a,b], Tabea TIETZ [a,b], Sandra GÖLLER [a] and Harald SACK [a,b]
[a] *FIZ Karlsruhe – Leibniz Institute for Information Infrastructure*
firstname.lastname@fiz-karlsruhe.de
[b] *Karlsruhe Institute of Technology (AIFB)*

Abstract.
Preserving historical city architectures and making them (publicly) available has emerged as an important field of the cultural heritage and digital humanities research domain. In this context, the TRANSRAZ project is creating an interactive 3D environment of the historical city of Nuremberg which spans over different periods of time. Next to the exploration of the city's historical architecture, TRANSRAZ is also integrating information about its inhabitants, organizations, and important events, which are extracted from historical documents semi-automatically. Knowledge Graphs have proven useful and valuable to integrate and enrich these heterogeneous data. However, this task also comes with versatile data modeling challenges. This paper contributes the TRANSRAZ data model, which integrates agents, architectural objects, events, and historical documents into the 3D research environment by means of ontologies. Goal is to explore Nuremberg's multifaceted past in different time layers in the context of its architectural, social, economical, and cultural developments.

Keywords. cultural heritage, digital humanities, city exploration, knowledge graphs, archival documents, architecture

1. Introduction

City preservation plays an important role in the cultural heritage and digital humanities research domains. Exploring the historical development of city architectures along with people living in it, their progress in technology, their craftsmanship as well as arts and culture is highly relevant for historians, architects, sociologists and the general public. Modernization, economic and industrial development, environmental progress, and the occurrence of natural disasters and wars result in the ongoing transformation of the city. Without a digital preservation of culturally relevant locations and the curation of their historical resources, this important heritage will be lost forever. In recent years, a number of projects have dealt with the digital preservation of culturally meaningful cities and places[1] [22]. In line with these efforts is the project TRANSRAZ[2], a successor of the

[1] https://www.timemachine.eu/
[2] https://www.fiz-karlsruhe.de/en/forschung/transraz

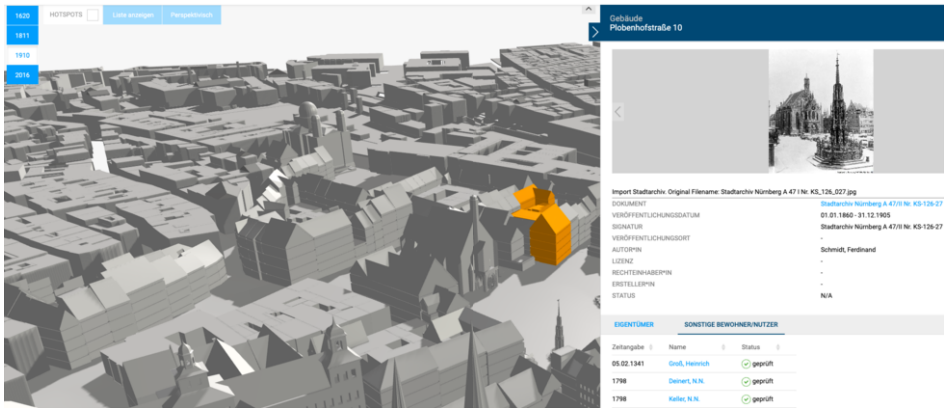

Figure 1. 3D visualisation of the building located on Plobenhofstraße 10 in 1910, and additional information about the building in the TRANSRAZ VRE.

project TOPORAZ [24]. Within TRANSRAZ heterogeneous historical data collections about persons, organizations and events are connected to an architectural 3D virtual research environment (VRE) using Knowledge Graphs (KGs) to enable the exploration of the historic city of Nuremberg in different time periods ranging from the Middle Ages to the 21st century. Nuremberg was one of the great European metropolises in the Middle Ages and beyond. It was the birthplace of renaissance artist Albrecht Dürer, who worked there all his life. The city developed into the epitome of German and European history and culture. Then, during the Second World War, the city was largely destroyed and only few buildings could be reconstructed. Therefore, a systematic and scientific reconstruction of the city in different time periods is necessary to research the history of Nuremberg in the context of its architectural, social, economical, and cultural developments.

In the TRANSRAZ VRE, KGs are utilized to connect historical documents retrieved by domain experts from archives and museums along with the entities extracted from the documents with the architectural 3D model (see Figure 1). These entities include among others persons, events, organizations and occupations. KGs furthermore allow to enrich these with additional information from external resources like Wikidata[3] and the German Authority Files[4] (GND). The VRE will allow to research residential and working areas, the establishment of educational institutions over time, the development of industry and the distribution of wealth in the city[5]. Furthermore, the connection of historical sources with the VRE allows for a scientifically accurate exploration. KGs have proven useful and valuable to integrate and enrich these heterogeneous data. However, this task also comes with a number of data modeling challenges, which include mapping a relational data model to existing ontologies, the efficient representation and management of data provenance, the change of entities over time, and the connection to the 3D environment.

The contribution of this paper is the TRANSRAZ data model, which includes ontologies to represent historical persons, their occupation and addresses, events, document

[3] https://www.wikidata.org/wiki/Wikidata:Main_Page
[4] https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html
[5] https://www.toporaz.de/toporaz-current/explore

annotations, and buildings in a 3D research environment. The TRANSRAZ data model consists of four main building blocks which have to be integrated into the VRE.

1. **Architectural Objects** are modeled as 3D elements within the VRE. The building parts are created as separate elements and have to be identifiable uniquely.
2. **Historical documents** are retrieved from archives and museums and contain relevant historical information about cities inhabitants, events etc. By means of information extraction and semi-automated annotations, the documents are connected to the 3D architecture with respect to their specific location and time mentioned.
3. **Agents** can be inhabitants of the historical city of Nuremberg and organizations. Entities, along with (family) relations, dates, addresses, and occupations are extracted from the existing TOPORAZ database as well as the historical document collections and connected to the 3D environment.
4. **Events** are extracted from the existing TOPORAZ database as well as the historical document collections. They include important city events and festivals as well as changes of state, e.g. architectural changes.

Goal is to represent these four main building blocks in a semantically meaningful and efficient way to enable a scientifically correct and intuitive exploration of Nuremberg in the context of its inhabitants and important events on the foundation of historical documents.The generalization of this contribution is ensured as it can be applied to any project related to the semantic representation of historical entities in a location based exploration environment. This work is reusable and reproducible, the data model is publicly available on the Web[6] including its documentation.
The remainder of the paper is structured as follows. In section 2 works related to modeling architectural data, the representation of heterogeneous cultural heritage data and the alignment of archival documents are discussed. In section 4 the design methodology along with data ressources and competency questions is discribed. Section 3 contains the main contribution of the paper, followed by evaluation use cases to explore the TRANSRAZ data. Section 5 concludes this paper.

2. Related Work

Three-dimensional models have traditionally been the purview for manufactures of navigation systems, planners, pollution researchers and virtual tourists [2]. However, in recent years the interest of connecting modern digital technologies with historical data from the past has raised immensely. Historians, museologists, architects and the general public are keen to research and curate heritage data, and basically to walk through the history hidden inside historical materials. Adding new dimensions to the past is a goal of the Time Machine projects[7]. Discovery, extraction, connection, reuse of historical data, and subsequently, the 3D reconstruction of European cities, e.g. Amsterdam[8] or Venice[9], is an ongoing goal to provide users with the ability to travel back in time. Meanwhile, a 4D browser for researching and communicating the history of the city of Dresden is

[6] https://ise-fizkarlsruhe.github.io/Transraz/datamodel
[7] https://www.timemachine.eu/about-us/
[8] https://www.amsterdamtimemachine.nl/
[9] https://www.epfl.ch/research/domains/venice-time-machine/

described in [22]. Apart from urban spaces, several projects aim at a detailed reconstruction of distinguished architectural objects. For example, Florence4D[10] is an initiative to revive prominent buildings of the Renaissance Florence. Similarly, the project "Digital Reconstruction of the Breslau Synagogue"[11] aims at reconstructing the largest synagogue in Breslau. A semantic information model will be presented, connecting objects with the resources used for the reconstruction. Additionally, the synagogue will be put into the social context by linking it to persons, corporate bodies, historical events, etc. Furthermore, there are several projects that utilize open source HBIM (Historic Building Information Modeling) software[12][7], including the HBIM Wiki Project[13], which aims to create an open database of historical buildings, the OpenHeritage project[14] that documents and conserves historic buildings across Europe, the H-BIM platform[15] and the Arches-HBIM project[16], which integrate HBIM with heritage management systems. In the context of these efforts, TRANSRAZ aims at not only 3D reconstructing of the entire urban environment of Nuremberg in 1910, but at enhancing the VRE via connecting its architectural objects to knowledge from external resources. To enable the enrichment, interoperability and reuse, tangible (e.g., historical buildings, places, documents) and intangible (e.g., festive events, traditions, life lines of persons) cultural heritage is meaningfully represented my means of ontologies.

Cultural heritage objects vary significantly in media type and attributes, however, they still share semantic similarities and benefit from cross-connections. The last decades have shown that semantic interlinking and representation of heterogeneous cultural heritage data is of interest for many research projects. There have been attempts to semantically model the data of Korean [21], Italian [6] and Finnish [18] national heritage. As part of the Europeana Project, the European Data Model (EDM) [9] was developed to provide a shared ontology infrastructure for cultural data from European GLAM institutions. All these data models are similar in nature of modeling, since the main reference for the models is the Conceptual Model CIDOC-CRM [8], a domain ontology of cultural heritage. Recently German National Library (DDB) has launched an initiative to develop the DDB Knowledge Graph and, thus, to enhance the frontend of the DDB. This is due to the complicated modeling of the EDM that negatively influences the usability and exploration of the data [25]. Despite CIDOC-CRM being also not fully sufficient for the user-oriented purposes of TRANSRAZ data model, for reasons of interoperability, the TRANSRAZ data model follows the best practices and provides a mapping to CIDOC-CRM.

A challenge of the modeling of TRANSRAZ space is the heterogeneity of the cultural heritage data that is obtained from a great amount of various resources, and that has to be represented in a direct connection with each other. This challenge can be addressed by exploiting existing ontologies and vocabularies. Related work towards modeling specific building blocks and design choices is discussed in section 3.2.

[10] https://florence4d.org/s/florence4d/page/home
[11] https://architekturinstitut.hs-mainz.de/projekte/digitale-rekonstruktion-der-breslauer-synagoge/
[12] https://github.com/UNIFE/hbim-suite
[13] http://www.hbimwiki.org/
[14] https://openheritage.eu/
[15] http://www.h-bim.com/
[16] https://www.arches-hbim.org/

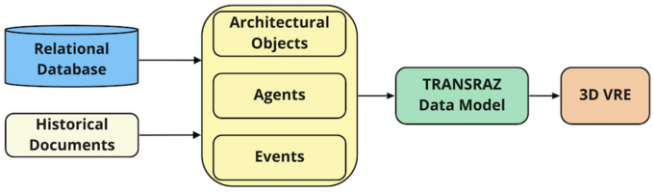

Figure 2. Workflow

3. TRANSRAZ Data Model

This section presents the main contribution of this work, the TRANSRAZ data model, which integrates agents, architectural objects, events, and historical documents into a 3D VRE by means of ontologies. This task presents versatile data modeling challenges: (i) Within TOPORAZ [24], a relational database model and a first version of the VRE were created. A challenge in this ongoing project TRANSRAZ is the transformation of the existing model to a semantic data model with mappings to existing ontologies relevant for the cultural heritage domain, as e.g. CIDOC-CRM. (ii) The historical documents retrieved from archives and museums are analyzed, semi-automatically annotated and connected to the VRE. To allow for a scientifically correct exploration of the city, the document annotation process has to be made completely transparent, i.e. data provenance plays an important role in the modeling decisions. (iii) Within the VRE, users will be able to explore different time periods. That means entities change their names, houses are destroyed and rebuilt, organizations change their ownership etc., which has to be considered in the modeling process.

The overall workflow is shown in figure 2. Architectural objects, agents, and events are represented based on ontologies on the foundation of the existing TOPORAZ relational database and the analyzed historical documents obtained by domain experts. They are then integrated into the TRANSRAZ data model and finally the 3D VRE.

In the following, the building blocks of the TRANSRAZ data model, the modeling requirements for each building block as well as the resulting ontologies are described. Additionally, the evaluation of the data model with the CQs is provided.

3.1. From Data Sources to Building Blocks

Due to the complexity, diversity and heterogeneity of the data sources and the knowledge contained within, a modular modeling approach to construct the data model was adopted. Therefore, the modules are represented by four main building blocks that correspond to the different aspects of the data model. Accordingly, the TRANSRAZ data model is organized in the following blocks:

* **Historical Documents** are used for the provenance annotation of statements, and also as a basis for automatic semantic annotation with references to other building blocks in the data model and classes in the ontologies.
* **Architectural Objects** represents all structures in the urban space, such as streets, districts, buildings, facades, roofs, rooms, etc.
* **Agents** refers to persons, clubs, organisations, associations, etc., who lived, worked or owned properties in the city or are socially connected to them.

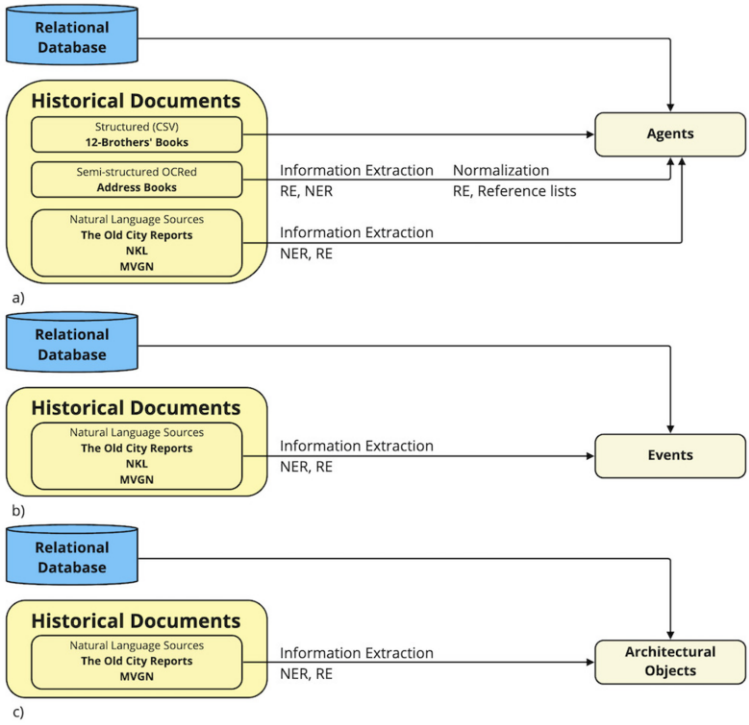

Figure 3. From data sources to building blocks

* **Events** represents events that can be associated with a point in time. On the one hand, this includes social, cultural, political, etc. activities of Nuremberg, such as Oktoberfest 1900. On the other hand, events that represent some duration and/or transformation, as e.g., deconstruction of a building, renaming of a street, etc.

Goals, challenges, and the proposed workflow towards extracting entities from historical resources, linking and disambiguating them are described in [26,5]. Subsequently, in [4], the first results to automatically obtain and model the information in the address books have been presented. Figure 3 schematically shows the workflow to populate the building blocks. However, the detailed discussion and results are out of the scope of this paper and will be addressed in the future work.

3.2. Ontologies of the TRANSRAZ data model

Following the guidelines of the ontology development, the TRANSRAZ data model aims at integrating, reusing, extending or mapping to established vocabularies and ontologies for the representation of the required classes and properties. This section presents concrete ontology engineering requirements that are derived from the CQs discussed in Section 4.2, taking into account the heterogeneity of data and concepts in the desired data model. Additionally, best practices for modeling of each building block are analyzed and compared against the requirements, and the design choices are presented.

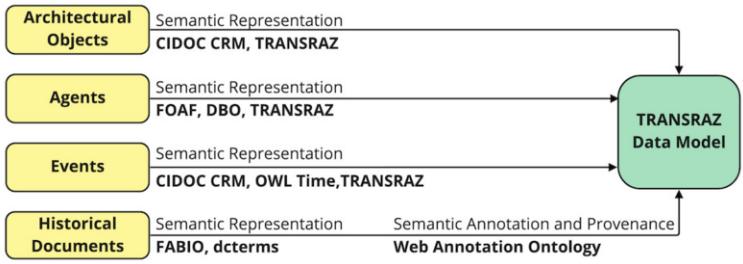

Figure 4. Ontologies of the TRANSRAZ Data Model

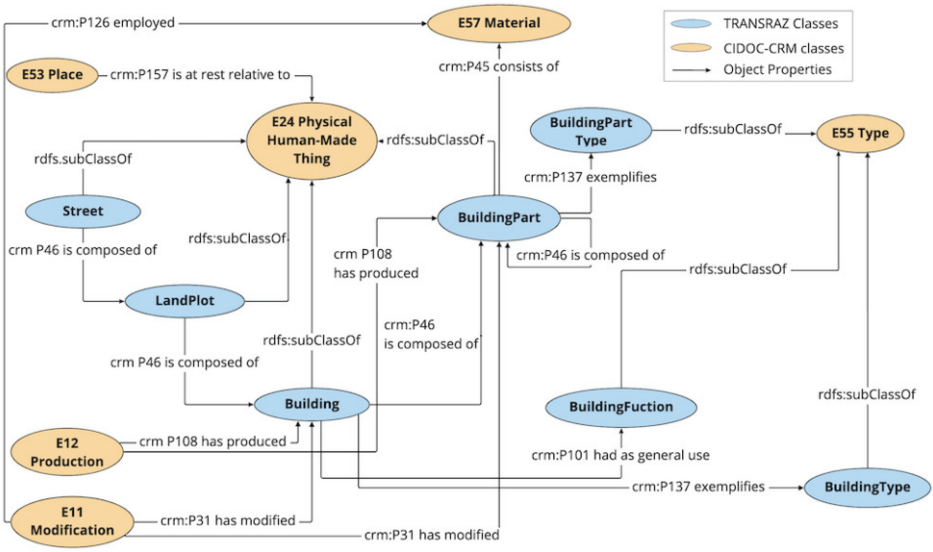

Figure 5. A part of the architectural object modeling within the TRANSRAZ Data Model

3.2.1. Modeling of City Architecture

The TOPORAZ relational database is the rich resource of the urban entities of the Main Market of Nuremberg. However, the representation of these entities and their relations is diverse and incomplete. In order to provide a comprehensive semantic representation of the tables in the TOPORAZ data model and further sources described in section 4.1, the following requirements have to be addressed:

REQ1: Architectural objects of a city are represented hierarchically. For example, a city consists of districts, a building consists of floors. This representation may include dozens of levels, describing every small part of a building (CQ4.1-2).

REQ2: Attributes of architectural objects rely on historical resources, thus, may be uncertain or incomplete. For example, for some buildings information about exact height, roof shapes, wall openings (windows, doors) and stylistic elements of the facade are provided, while for other buildings only the location area is known.

REQ3: Architectural objects, their functions, names are highly dynamic and change over time, for example, buildings are destroyed, streets change their names, a house that used to be a pharmacy can become a store, etc (CQ3.2, CQ4.2-3).
REQ4: Due to information dynamics, uncertainty and incompleteness of the data, facts require provenance annotation (CQ6.1-2, CQ6.4).
REQ5: Architectural objects of old cities, such as Nuremberg, are also considered cultural heritage that can further be explored via interconnecting with other cultural heritage objects (CQ3.2, CQ4.3, CQ5.1).

Research has been conducted towards the semantic representation of urban spaces [1,10,20], however, these ontologies are different from the architectural object block of the TRANSRAZ data model in their main goal – they are used to describe a modern city architecture. Moreover, these ontologies lack classes and properties to represent parts of the buildings (REQ1), and thus, are insufficient for modeling the level of detail required for TRANSRAZ buildings. Finally, the ontologies are not flexible and do not consider dynamics (REQ3), because they address urban spaces in real time only.

Due to the lack of ontologies to describe historical cities, a derivation of CIDOC-CRM is provided with CIDOC-CRM being the core ontology of the block. CIDOC-CRM is event-based, which allows the representation of change (REQ3) through events. Also, it provides a vocabulary to represent interconnections among objects and events in a flexible way (REQ2). And finally, it is frequently used by GLAM institutions (galleries, libraries, archives, and museums) to represent diverse cultural heritage, which puts architectural heritage into context with the related cultural objects (REQ5).

Since CIDOC-CRM does not provide an expressive semantic for modeling the structures of buildings and their parts, the model has been enhanced with the addition of self-defined classes, as e.g. *transraz:BuildingPart*. For readability, figure 5 reports only part of the elements of the TRANSRAZ building block, the full modeling is provided on GitHub[17]. The information is organized around three main classes:

* Class *crm: E24 Physical Human Made-Man Thing* allows to define subclasses of the architectural hierarchy and describe their components.
* Class *crm:E55 Type* connects architectural objects and their parts with functions, e.g. library, bakery or sleeping room, or their types, e.g. front house, park, balcony. Both functions and types are completed by a set of controlled vocabularies developed by the domain experts during the TOPORAZ project.
* Class *crm:E5 Event* and its subclasses, e.g. *crm:E12 Production* and *crm:E11 Modification*, provide a description of building processes. For instance, when a building was built or transformed, who did it and what materials were used.

TRANSRAZ classes are defined as subclasses of CIDOC-CRM and, thus, inherit all properties to specify the time of an event, the hierarchical nature of architectural objects, their types, functions, etc. Moreover, all architectural objects are associated with geographical coordinates and addresses that work as a bridge to connect them to the agents.

3.2.2. Modeling of Historical Agents.

One of the main goals of the TRANSRAZ project is to depict historical Nuremberg not only geographically, but also in its social context. Thus, the semantic representation of

[17] https://github.com/ISE-FIZKarlsruhe/Transraz/

its citizens and companies and their social networks is necessary. After the analysis of the complex and heterogeneous resources that contain information about persons and organizations of Nuremberg (see figure 3(a)), the following modeling requirements have been identified:

REQ1: Agents of a city are interconnected via different relations, which are of great interest to historic and genealogical research and have to be semantically represented (CQ1.1, CQ2.1, CQ2.3, CQ5.2).

REQ2: The descriptions of agents are often extracted from the historical resources, thus, may be uncertain, incomplete or even false. Thus, provenance annotation is required (CQ6.1-4).

REQ3: Attributes of agents, e.g. their addresses, names, occupations, etc. may change overtime (CQ1.1, CQ2.1-2, CQ3.1, CQ5.2).

There exist several ontologies that specialize in description of agents from different perspectives. However, they are mostly focused on describing either their existence in the virtual world, e.g. their activities on the Web [12], or basic information required for postal delivery [19]. In this work, agents serve as linking points among the building blocks, hence, among different domain-specific ontologies. The DBpedia Ontology (DBO) benefits from linking to the most common ontologies, as e.g. SCHEMA.ORG [18], Wikidata [19], FOAF [20]. Thus, the properties and classes from different vocabularies are interconnected and could simply be reused. Additionally, DBO contains a rich semantic representation and hierarchy of social relations, e.g. *dbo:mother, dbo:sister, dbo:friend* (REQ1). Mapping to CIDOC-CRM classes and properties, e.g. to CRM's *E21 Person*, enables the direct use of the CIDOC-CRM's events, that allow for the representation of change (REQ 3) (see figure 6).

3.2.3. Modeling of Events and Changes.

In historical research, when working with cultural data it is important to observe the development of entities and to keep track of changes that occur over time. That is why a proper semantic representation of temporal happenings is essential for the TRANSRAZ data model. Based on the data and data source analyses, the following requirements for modeling events have been established:

REQ1: Events in the TRANSRAZ model represent both social activities and changes of state. For example, the premier of the "Zapfenstreich" in 1907 and the change of address for a person or deconstruction of a building (CQ1.1, CQ2.4, CQ3.1-2, CQ4.2-3).

REQ2: Events of state change cover both semantic (occupation changes its function over time) and terminological (street is renamed) change of entities (CQ1.1, CQ2.2, CQ4.3).

REQ3: Events can be continuous and discreet, and thus, associated with time points or time intervals.

[18] https://schema.org/
[19] https://www.wikidata.org/wiki/
[20] http://xmlns.com/foaf/spec/

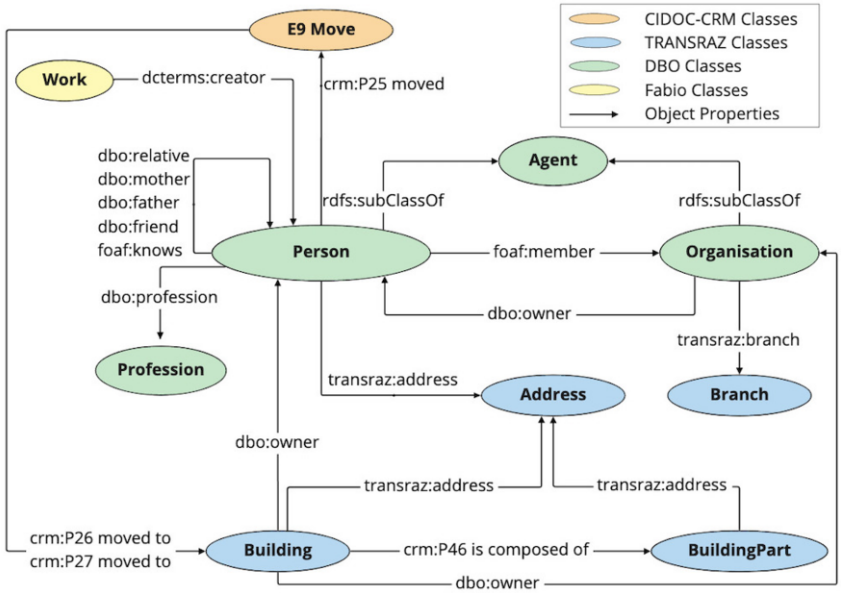

Figure 6. A part of the modeling of the TRANSRAZ agents block

REQ4: Events and their attributes, e.g. temporal information, are often extracted from the historical resources, thus, may be uncertain, incomplete or even false. Thus, provenance annotation is required (CQ6.1, CQ6.2, CQ6.3, CQ6.4).

REQ5: Events can be interconnected via temporal relations, as e.g., *overlaps, before* (CQ1.1, CQ2.1-2, CQ3.1-2, CQ4.2).

The entity-based representation of temporal relations requires either additional extensions to RDF [14,27,15], or a great amount of additional statements [17], while an event-based approach addresses this gap and allows for direct representation. In [11], a temporal event-centric Knowledge Graph is presented. In the work, subclasses of Wikidata and DBpedia classes *wdt:Q1656682* (Event) and *dbo:Event* are identified. However, these ontologies model events that aim at representing certain activities, e.g. lecture, concert, festival, and do not cover events that represent temporal states and change (REQ1, REQ2). The EventKG enhances the Simple Event Model (SEM) [28] and provides a vocabulary for modeling temporal relations of the types entity-entity (e.g. marriage), entity-event (e.g. Margarete Haagen played the Girl in the "Zapfenstreich"), but do not directly cover temporal relations between events, e.g. does reconstruction of the "Frauenkirche" overlap with the reconstruction of the St. Lorenz Church? (REQ5).

In contrast, CIDOC-CRM provides semantics for changes of state in cultural, social and physical systems, e.g. class *crm:E11 Modification* and its subclasses comprise a change of an object, class *crm:E8 Acquisition* comprises transfers of legal ownership from one agent to another (REQ1). Moreover, CIDOC-CRM's events are represented through the time spans of their validity in form of abstract temporal extents (REQ3), e.g. duration of the Ming Dynasty, and are connected to each other via temporal relations (REQ5), e.g. *crm:P114 is equal in time to, crm:P117 occurs during*, etc.

Since CIDOC-CRM does not fully suffice for the purposes of TRANSRAZ (REQ1 and REQ2), the model has been extended by adding further properties and classes. The

class *transraz:Event* is defined as subclass of *crm:E2 Temporal Entity* to represent cultural, political, military and further events that can be associated with points in time. *transraz:Event* inherits all properties of *crm:E2 Temporal Entity* to connect to other happenings, and additionally presents properties to connect events with agents, e.g. *transraz:participant* (REQ1). Moreover, the class *crm: E5 Event* is extended with two additional subclasses. The class *transraz:Relabeling* comprises changes of names of the instances of agents and objects. And the class *transraz:Refunction* addresses the changes on the function level, e.g. a building changes it function from pharmacy to shop (REQ2).

In order to address the uncertainty and incompleteness of time units in historical data, time expressions of different levels of granularity have to be considered. Thus, the new property *transraz:time* connects events with *time:TemporalEntity* from OWL-Time [16]. In contrast to *crm:E52 Time-Span*, it provides a more granular representation of time indications, it presents classes and relations for time instants and intervals, together with information about durations, and temporal position including date-time information. Additionally, the uncertainty and incompleteness of time intervals is addressed via providing temporal relations of the types interval-interval and interval-instant. Thus, the temporal relations can be modeled even if the exact event time is unknown, e.g. *interval1 time:overlaps interval2* (REQ4 and REQ5).

3.2.4. Representation and Annotation of Documents.

In the TRANSRAZ data model, the use of documents is two-fold:

REQ1: Historical documents are themselves considered cultural heritage objects and require a semantic representation to enable the findability and interconnection of resources (CQ6.1, CQ6.3).

REQ2: The provenience is required to keep track of the origin of statements and entities, the sources, they were derived from, are used as contextual evidence (CQ6.1-4).

Following the best practices, historical documents and their hierarchy are modeled with the FRBR-Aligned Bibliographic Ontology (FaBiO) [23]. FaBio provides rich semantics for recording and publishing bibliographic records. It enables the representation of archival records (*fabio:ArchivalDocumentSet* and *fabio:ArchivalDocument*), journals (*fabio:Journal*), their issues (fabio:JournalIssue) and articles (*JournalArticle*), books *fabio:Book* and their chapters (*fabio:BookChapter*). Partly making use of the DCMI Metadata Terms vocabulary[21] historical documents are annotated with the relevant metadata, e.g. *dcterms:creator, dcterms:publisher*. The Web Annotation Ontology[22] (OA) is used for providing annotations to the resources. Annotations (*oa:Annotation*) connect information to the source that the information was extracted from, for example a person mentioned in an address book or a fact derived from the NKL. Moreover, the property *oa:hasSelector* allows for linking an annotation to a specific part of a document including the text coordinates the annotation was made in. In this way, the provenance of a data excerpt can always be retrieved and proven. The task of representing knowledge from the rich, complex and heterogeneous resources, and, furthermore, enabling the usability and exploration of the information, presents diverse modeling challenges. To address them, in

[21] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
[22] https://www.w3.org/ns/oa

the TRANSRAZ data model a modular modeling approach is developed and presented, that enables to benefit from the connection of data coming from different domains via domain-specific ontologies.

3.3. Evaluation of the TRANSRAZ Data Model

The competency questions introduced in Section 4.2 were developed to scope the data model, guide the modeling process, and, provide a way to evaluate the data model. The CQs were iteratively developed together with the domain experts – art historians and experts in digital art history. An evaluation of the data model is performed by examining if the CQs could be transformed into SPARQL queries and by verifying the appropriateness of the data model in delivering correct answers to the CQs[23][24]. The evaluation shows that the integration of the four main building blocks of the TRANSRAZ data model enables multifaceted queries. The use cases furthermore show that an interconnection between the building blocks is necessary and enables the most valuable findings. The presented contribution will not remain static and will be further enhanced and extended upon the integration of further resources in collaboration with domain experts.

4. Design Methodology and Use Cases

This section presents the methodology of data model development, discusses data sources and knowledge that serve as a basis of the data model and provides user-oriented interrogatives developed in the first stage of the the data model design.

4.1. Data and Data Sources as Basis of the TRANSRAZ Data Model

The TRANSRAZ data model aims at providing a semantic representation of the knowledge stored in historical resources, and, thus, is developed following a bottom-up approach. In this section, the selection of the currently processed data sources is introduced. The extraction and connection of the knowledge contained in these sources is analyzed and used to obtain modeling requirements of the data model.

TOPORAZ relational database is an essential source of the TRANSRAZ data model presented in this work. The database was developed as an effort of the predecending TOPORAZ project[25] and is hosted by FIZ Karlsruhe. Based on historical resources, such as city plans, cadastral plans, photographs, maps etc., domain experts collected and provided a small biography for every building in Nuremberg's main market square for four different time periods – 1620, 1811, 1910, and 2016. The data was then structured and manually inserted into subject-based tables. In particular, the database contains entities for architectural objects like streets and buildings, their residents and owners, constructing events (e.g. destruction and creation) and connecting time identification (e.g., "before 1890"), etc. in form of around 50 classes. In the ongoing TRANSRAZ project the area of the coverage has been extended to the reconstruction of 3000 buildings, which is 30 times more than the area of TOPORAZ. In contrast to TOPORAZ, contextual infor-

[23] CQs and SPARQL queries: https://ise-fizkarlsruhe.github.io/Transraz/usecases
[24] The SPARQL Endpoint: https://www.toporaz.de/sparql
[25] https://www.toporaz.de/toporaz-current/

mation about Nuremberg's inhabitants, events and buildings are extracted and integrated into the TRANSRAZ VRE automatically. Furthermore, more diverse data sources are analyzed and connected, which increases the heterogeneity and complexity of the task.

Address Books of Nuremberg is one of the first printed sources on the residents of Nuremberg starting with 1792. The address books are physically stored in the Nuremberg City Archives[26] and in The Germanisches Nationalmuseum[27], and are also provided digitally in a scanned image form. The books contain important information about persons, e.g. their addresses, professions, places of ownership and work, etc., and about companies, e.g. their names, industries, addresses, etc. Due to the complicated nature of the books, such as bad paper quality, distortion of pages, poor inking, as well as challenging linguistic features: Gothic fonts, ligatures, archaic terms, old spelling variants, abbreviations and typos, the extraction and structuring of the knowledge hidden in the books is a challenging task.

Nuremberg Artists Lexicon. The "Nürnberger Künstlerlexicon" (NKL) [13] is a collection of bibliographical articles about artists of Nuremberg based on various archival records ranging from the 12th century to the mid 20th century. The articles provide both personal information of artists such as addresses, professions, birth and death places and dates, family relations, places and periods of study, and information about their artworks and their public life. The articles of NKL are based on administrative records, the text is saturated with temporal units to describe the events, e.g., date of marriages and artworks creation, periods of study and work, change in the ownership of properties.

Journal of the Association for History of the City of Nuremberg. The "Mitteilungen des Vereins für Geschichte der Stadt Nürnberg" (MVGN)[28] is a journal focused on the history of Nuremberg. It has been publishing scholarly articles since 1879 and includes an annual issue with up to 40 reviews on significant historical events, prominent individuals and organizations of Nuremberg, and important buildings like St. Sebaldus Church. The articles are characterized by their complex sentence structures, filled with coordinating conjunctions and descriptive phrases.

Books of Nuremberg's Twelve Brothers. The "Nürnberg Zwölfbruderbücher"[29] are medieval books that feature portraits and biographical information about retired Nuremberg craftsmen residing in an old people's home. These books have been digitized, transcribed, and indexed[30]. They provide details such as the craftsmen's names, professions, birth and death dates and places, registration dates in the retirement home, and length of stay. Descriptions of the individuals' portraits are also included.

Nuremberg Old Town Reports. The "Nürnberger Altstadtberrichte" are reports that were first published in 1976 by the Altstadtfreunde Nürnberg Association[31]. They document significant construction projects through images, photographs, and text, while also featuring essays on the history, art, and culture of Nuremberg's old town. The reports extensively describe historical buildings, including their addresses, functions, and construction events, but also cover historical events, festivals, and individuals or companies associated with the architectural objects.

[26] https://www.nuernberg.de/internet/stadtarchiv_e/
[27] https://www.gnm.de/en/museum/
[28] https://www.bayerische-landesbibliothek-online.de/mvgn
[29] https://hausbuecher.nuernberg.de/index.php?do=page&mo=2
[30] https://hausbuecher.nuernberg.de/index.php?do=page&mo=5
[31] https://www.altstadtfreunde-nuernberg.de/de/home

4.2. Competency Questions of the TRANSRAZ Data Model

The development of the TRANSRAZ data model followed a user-centered design and evaluation methodology ([3]). Users of the data model are on the one hand the general public, generally interested in he city's history, architecture and its inhabitants. On the other hand, users are researchers in the fields of history, social science, digital humanities, and architecture. With the goal to scope the desired data model and provide future users with the ability to gain answers through exploring the data model and its associated knowledge, a set of competency questions (CQs) in collaboration with domain experts was developed incrementally and iteratively, and resulted in 6 categories:

1. **Ancestor Search.** Users with ancestors in Nuremberg have the possibility to explore, where their relatives lived, how they lived and what their neighborhood looked like.

   CQ1.1: What information is available for a specific person or family living in the city? This involves events related to the person, family relations, their occupation.
   CQ1.2: Who lived in a certain house or at a certain address?
   CQ1.3: What did a certain neighborhood look like at a certain year? What kind of industries were based there?

2. **Wealth Distribution.** Researchers who are interested in the distribution of wealth throughout the city can explore the location of industries, infrastructure and living areas of Nuremberg's inhabitants.

   CQ2.1: Which properties and real estates belonged to a person or family and how did they distribute throughout the city?
   CQ2.2: Who owned properties and real estates in a certain street or district?
   CQ2.3: Where did people of certain occupations (e.g. bakers, teachers) live?
   CQ2.4: Where did specific industries and branches (e.g. schools, breweries) establish?

3. **Change and Events.** Researchers interested in the transformation of the city and changes of entities can explore the movement of a person within the city over the years, the destruction and construction of architectural objects and social events taking place.

   CQ3.1: Where did one person live in the city throughout time?
   CQ3.2: Where were specific infrastructures like hospitals and schools located in a certain time? And how their establishment developed throughout time?

4. **Architecture and Infrastructure.** Users intending to learn about the city architecture and infrastructure development can query architectural objects and the development ob buildings and their functions over time.

   CQ4.1: In what area of the city were the highest buildings (by number of of floors)?
   CQ4.2: What different roof types were built throughout the years in the city?
   CQ4.3: What building function did a building have over the years (e.g. pharmacy, hotel) and how was this documented in the historical sources?

5. **Social and Cultural.** To learn about the social life of the historical city over the years, users are able to explore touristic and culturally important areas.

   CQ5.1: What were the main cultural areas in the city, e.g. cultural and art institutions?
   CQ5.2: Which parts of the city did most artists live in?

6. **Historical Sources and Provenance.** Researchers can verify findings presented in the VRE by directly exploring the connected historical source documents. Furthermore, document annotations and their origins can be reproduced.

   CQ6.1: What historical literature is available for a certain street, building. person?
   CQ6.2: Where can information about a certain street, building, person be found in the historical resource, e.g. page number or coordinates?
   CQ6.3: What types of documents are available that mention a certain concept, e.g. certificates, archival records?
   CQ6.4: How and when were the annotations of a certain concept created? Manually or automatically?

5. Conclusion and Future Work

Within the TRANSRAZ project the TRANSRAZ data model has been created, which enables semantic representation of agents, events, architectural objects, and historical documents of historical cities. This model enables scientifically accurate exploration of the Nuremberg's history in a 3D virtual environment, and can be applied to other historical cities with similar aims. It is available publicly, it is planned to extend it by adding new vocabularies and controlled vocabularies for consistency in historical terms. Moreover, the TRANSRAZ Knowledge Graph integrated into TOPORAZ VRE will be extended with addition of further data resources. Exploring the knowledge extracted from the historical resources and their interconnection with a 3D VRE will allow users to obtain an impression of the hidden past and, hence, better understand the formation of today's world, its society and oneself.

References

[1] Auer, S., Lehmann, J., Hellmann, S.: Linkedgeodata: Adding a spatial dimension to the web of data. In: Int. Semantic Web Conference. pp. 731–746. Springer (2009)
[2] Behley, J., Steinhage, V.: Generation of 3D city models using domain-specific information fusion. In: Int. Conference on Computer Vision Systems. pp. 164–173. Springer (2009)
[3] Bravo, M., Luis, H.R., Reyes, J.: Methodology for ontology design and construction. Contaduría y Administración **64**, 134 (03 2019).
[4] Bruns, O., Tietz, T., Chaabane, M.B., Portz, M., Xiong, F., Sack, H.: The Nuremberg Address Knowledge Graph. In: 18th Extended Semantic Web Conference (ESWC), Poster and Demo Track. Springer (2021)
[5] Bruns, O., Tietz, T., Vafaie, M., Dessí, D., Sack, H.: Towards a representation of temporal data in archival records: Use cases and requirements. In: Proc. of the Int. Workshop on Archives and Linked Data. pp. 128–134 (2021)

[6] Carriero, V.A., Gangemi, A., Mancinelli, M.L., Marinucci, L., Nuzzolese, A.G., Presutti, V., Veninata, C.: ArCo: The Italian cultural heritage knowledge graph. In: Int. Semantic Web Conference. pp. 36–52. Springer (2019)
[7] Diara, F., Rinaudo, F.: Open Source HBIM for Cultural Heritage: a project proposal. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences **42**(2) (2018)
[8] Doerr, M.: The CIDOC CRM, an ontological approach to schema heterogeneity. In: Dagstuhl Seminar Proceedings. Schloss Dagstuhl-Leibniz-Zentrum für Informatik (2005)
[9] Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C., Van de Sompel, H.: The europeana data model (edm). In: World Library and Information Congress: 76th IFLA general conference and assembly. vol. 10, p. 15 (2010)
[10] Dsouza, A., Tempelmeier, N., Yu, R., Gottschalk, S., Demidova, E.: WorldKG: A World-Scale Geographic Knowledge Graph. In: Proc. of the 30th ACM Int. Conference on Information & Knowledge Management, pages=4475–4484, year=2021
[11] Gottschalk, S., Demidova, E.: EventKG: a multilingual event-centric temporal knowledge graph. In: European Semantic Web Conference. pp. 272–287. Springer (2018)
[12] Graves, M., Constabaris, A., Brickley, D.: FOAF: Connecting people on the semantic web. Cataloging & classification quarterly **43**(3-4), 191–202 (2007)
[13] Grieb, M.H.: Nürnberger Künstlerlexikon: Bildende Künstler, Kunsthandwerker, Gelehrte, Sammler, Kulturschaffende und Mäzene vom 12. bis zur Mitte des 20. Jahrhunderts. Walter de Gruyter (2011)
[14] Gutierrez, C., Hurtado, C., Vaisman, A.: Temporal rdf. In: European Semantic Web Conference. pp. 93–107. Springer (2005)
[15] Hartig, O.: Foundations of RDF* and SPARQL*:(An alternative approach to statement-level metadata in RDF). In: AMW 2017 11th Alberto Mendelzon Int. Workshop on Foundations of Data Management and the Web, Montevideo, Uruguay, June 7-9, 2017. vol. 1912. Juan Reutter, Divesh Srivastava (2017)
[16] Hobbs, J.R., Pan, F.: Time ontology in OWL. W3C working draft **27**(133), 3–36 (2006)
[17] Honneth, A.: Reification: A new look at an old idea. Oxford University Press (2008)
[18] Hyvönen, E., Mäkelä, E., Kauppinen, T., Alm, O., Kurki, J., Ruotsalo, T., Seppälä, K., Takala, J., Puputti, K., Kuittinen, H., et al.: CultureSampo—Finnish culture on the Semantic Web 2.0. Thematic perspectives for the end-user. In: Proceedings, Museums and the Web. pp. 15–18 (2009)
[19] Iannella, R., McKinney, J.: vCard Ontology-for describing People and Organizations. W3C Group Note NOTE-vcard-rdf-20140522 (2014)
[20] Karalis, N., Mandilaras, G., Koubarakis, M.: Extending the YAGO2 knowledge graph with precise geospatial knowledge. In: Int. Semantic Web Conference. pp. 181–197. Springer (2019)
[21] Kim, S., Ahn, J., Suh, J., Kim, H., Kim, J.: Towards a semantic data infrastructure for heterogeneous Cultural Heritage data-challenges of Korean Cultural Heritage Data Model (KCHDM). In: 2015 Digital Heritage. vol. 2, pp. 275–282. IEEE (2015)
[22] Kröber, C., Messemer, H., Bruschke, J., Maiwald, F., Niebling, F., Münster, S.: Ein 4D-Browser für historische Photographien: NutzerInnen zwischen medialer Vermittlung und Forschung, pp. S. 25–27. arthistoricum.net, Heidelberg (2020).
[23] Peroni, S., Shotton, D.: FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. Journal of Web Semantics **17**, 33–43 (12 2012)
[24] Razum, M., Göller, S., Sack, H., Tietz, T., Vsesviatska, O., Weilandt, G., Grellert, M., Scharm, T.: TOPORAZ: Ein digitales Raum-Zeit-Modell für vernetzte Forschung am Beispiel Nürnberg. Information-Wissenschaft & Praxis **71**(4), 185–194 (2020)
[25] Tan, M.A., Tietz, T., Bruns, O., Oppenlaender, J., Dessí, D., Sack, H.: DDB-KG: The German Bibliographic Heritage in a Knowledge Graph. In: The 6th Int. Workshop on Computational History (HistoInformatics2021-JCDL2021). Ed.: Y. Sumikawa (2021)
[26] Tietz, T., Bruns, O., Göller, S., Razum, M., Dessì, D., Sack, H.: Knowledge graph enabled curation and exploration of Nuremberg's city heritage. In: Proc. of the Conference on Digital Curation Technologies (Qurator 2021) (2021)
[27] Udrea, O., Recupero, D.R., Subrahmanian, V.: Annotated rdf. ACM Transactions on Computational Logic (TOCL) **11**(2), 1–41 (2010)
[28] Van Hage, W.R., Malaisé, V., Segers, R., Hollink, L., Schreiber, G.: Design and use of the Simple Event Model (SEM). Journal of Web Semantics **9**(2), 128–136 (2011)