Breaking Down Barriers: The Transition of ODIS from a Relational to a Triple Store Database

Bert Aernouts[1], Mehmet Celik[2], and Joris Colla[1]

[1]KADOC-KU Leuven/ODIS
[2]LIBIS-KU Leuven

In this paper, we discuss the ongoing transformation of the ODIS database – a key digital infrastructure for historical research on civil society in Flanders and Brussels – from a relational to a triple store architecture. Originally developed in the early 2000s and overhauled between 2009 and 2013, ODIS is now undergoing a comprehensive renewal to align with linked data principles and the FAIR data guidelines. The transition involves the implementation of a Virtuoso triple store, a redesigned RDF-based data model, and a new user interface. Central to this effort is the development of SOLIS (Smart Ontology Layer for Interoperable Systems), a tool that bridges domain expertise and technical implementation by generating SHACL-based APIs from spreadsheet-defined ontologies. The paper discusses the rationale behind adopting a triple store, the challenges of semantic modelling, multilingual data handling, and the integration of external authorities such as OpenStreetMap. It also highlights the benefits of semantic richness, interoperability, and enhanced data visualisation through network and geographical tools. This case study offers insights and best practices for digital humanities projects seeking to modernise their data infrastructures and foster interdisciplinary, data-driven research.

Keywords: civil society, contextual disclosure of heritage, digital humanities, heritage and research database, linked open data

1 Introduction

As research trends in the humanities change over the years, tools evolve to follow and sometimes even shape these new approaches. ODIS, a contextual relational database on the history of civil society that is used by a growing number of research and heritage organisations in Flanders and Brussels, was developed between 2000 and 2003. From 2009 to 2013, the database underwent a major overhaul to better meet the needs of its users. With the broader adoption of linked data in cultural heritage and humanities research, and with researchers demanding more advanced tools for data visualisation and analysis, the database is currently undergoing another renewal process, propelled by an infrastructure grant from the Research Foundation Flanders (FWO). At the core of the 'ODIS renewal project 2022–2027' is the development of a new triple store database and a new user interface. This stems from a desire to create new ways to query, analyse, and visualise the ODIS datasets, as well as to develop an interoperability and discovery layer.

This paper focuses on the backend transition from a relational to a triple store database, discussing the decisions, methods, and workflows implemented to enable the desired functionalities for the new user interface. We first provide some background information on ODIS, its history, and the current architecture of the database. We then discuss the approaches to data(re)modelling, as well as the tools used and created. Our aim is to offer an account of our efforts and journey through this transition. We seek to share our experiences with the digital humanities community, advocate for a linked data approach, and present some best practices.[1]

2 ODIS: Past and Present

ODIS is a research-driven relational and contextual database on the history of civil society (ODIS, n.d.b). It was developed between 2000 and 2003, thanks to a grant from the Research Foundation Flanders (FWO – Max Wildiers Fund) (Heyrman and Weber, 2007). Four major cultural archives in Flanders, all seeking a system to store, interconnect, and disclose contextual metadata related to their heritage collections, formed the basis of the project: ADVN - archive for national movements, Amsab-Institute of Social History (Amsab-ISG), KADOC-KU Leuven, and Liberas (the former Liberal Archives). Other heritage and research organisations and projects in Flanders and Brussels soon joined the partnership, making ODIS one of the most widely used historical research instruments in Flanders.[2]As ODIS does not receive structural funding, its management and maintenance are covered by the annual operational usage fees paid by the partners.

In 2006, the management of the database was entrusted to a non-profit association under Belgian law.[3] The universities of Antwerp (UA), Brussels (VUB), Ghent (UGent), and Leuven (KU Leuven) are represented in its General Assembly. KADOC, the interfaculty Documentation and Research Centre on Religion, Culture and Society at KU Leuven, is responsible for the day-to-day management of ODIS, while LIBIS-KU Leuven is the main technical service provider. From 2009 to 2014, a research infrastructure grant from the Hercules Foundation of the Flemish authorities enabled the development of a new database (De Maeyer, 2015). The data structure of ODIS was expanded with new modules for the description of buildings, families, and events. Bilingualism (Dutch–English) became a key feature of the database, significantly increasing its usability, though also adding complexity to the data model. A mix of language-dependent and language-independent fields makes up the descriptions,

1 We would like to thank Peter Heyrman (KADOC-KU Leuven), Stephan Pauls (LIBIS-KU Leuven), Rachele Ricceri (LIBIS-KU Leuven), Winand Van Meerbeek (KADOC-KU Leuven), and Roxanne Wyns (LIBIS-KU Leuven) for their valuable input.

2 A complete overview of the ODIS partners, with links to their websites, is available on the ODIS website (ODIS, n.d.a).

3 The bylaws of the non-profit organisation (vzw) 'Onderzoekssteunpunt en Databank Intermediaire Structuren in Vlaanderen (19de-20ste eeuw)' were published in the Belgian Official Gazette (*Belgisch Staatsblad*), 13 November 2006, 884.703.544. Their most recent update was published on 8 January 2024 (noa, 2024).

resulting in shared data between the two language versions of a record, but also discrepancies. Translation projects aimed at closing this information gap have already yielded significant progress, though the issue remains on the agenda.

Stimulating cross-fertilisation between heritage and research has always been a major goal of ODIS. By offering researchers and heritage experts a shared platform, the database aims to foster interdisciplinary research into the history and heritage of civil society. First and foremost, it provides its partners with a stable, reliable, and user-friendly environment to centralise and store contextual information on civil society and its documentary heritage. This enables them to safeguard, supplement, update, and correlate their dispersed datasets and repertories, keeping them accessible in the long term and ensuring the reproducibility of research. Secondly, the system is used to validate data and publish them on the World Wide Web, via the ODIS public catalogue (OPAC) and specific OPACs on the websites of partners and projects. Finally, ODIS enables its partners to analyse their datasets by means of advanced search tools (Angelaki et al., 2019; Colla and Heyrman, 2019; Heyrman, 2025).

The broad scope of the ODIS partnership is reflected in the comprehensive content of the database. Many thematic fields are addressed at the international, national, regional, and local levels: politics, social movements and organisations, culture, art and architecture, religion, national movements, education, care provisions, migration, and more. In May 2021, the ODIS datasets on religion, culture, and society, managed by KADOC, were recognised as a core research facility of KU Leuven (KADOC-KU Leuven, 2024; KU Leuven, 2024). As of early 2024, ODIS contained 308,144 records on organisations, persons, families, buildings, events, archives, and publications (mainly periodicals). Nearly 48% of them are available for consultation under a Creative Commons licence (BY-NC-SA 4.0), facilitating their reuse in other initiatives and projects. In 2023, the ODIS OPAC was visited 91,642 times by 51,517 unique visitors, resulting in 604,982 page views. Visitors use the database in several ways: 1) as an encyclopaedia; 2) as a heuristic instrument; 3) as an authority database; and 4) as a digital humanities research tool.[4]

The current ODIS data model consists of eight interlinked modules (see Figure 1). The structure of the modules is based on international standards wherever possible: ISAAR(CPF) for organisations, persons, and families; the DOCOMOMO Guidelines for buildings; ISAD(G) for archives; ISBD for publications; and ISDIAH for repositories (Docomomo International, n.d.; International Council on Archives, 2024a,b,c; International Federation of Library Associations, 2011). All modules share a parallel structure. Each includes ‘authority entries’ for identification data (e.g. names, titles, dates and places of birth and death). Furthermore, data inputters can use both free text fields and repeatable fields/field groups. In the latter category, systematic data input is carried out based on validated thesauri and lists of choices. Finally, ODIS features relational field groups. These ensure the interconnection between the ODIS modules and enable the creation of clear and unambiguous links between records. By interconnecting records, the ODIS data model makes it possible, for example, to map networks and discover unexpected connections. With some research effort, the contextual information clusters in the database provide a multifaceted view of civil society in Flanders/Belgium, its key players and documentary heritage. Since only some authority entries are mandatory, authors and partners can determine the scope and depth of their datasets in alignment with future research objectives. They also

[4] More information on the content and use of ODIS can be found in the ODIS annual reports, available on the ODIS website (in Dutch) (ODIS, n.d.a).

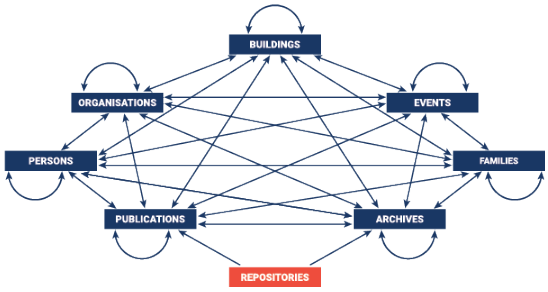


Figure 1: *Entity-relationship diagram of the ODIS database.*

independently determine whether they would like to publish their series in the OPAC.

Striving for versatility and cooperation, ODIS seeks to address and even anticipate the technical needs of its diverse partners and users. These users are a balanced mix of academic researchers, heritage experts, individuals involved in local historical research, genealogists, journalists, and other information seekers.

As several features of the database have become somewhat outdated and no longer provide the expected functionalities, nearly ten years after the so-called 'ODIS Hercules Project 2009–2014', another comprehensive renewal of ODIS became necessary. Supported by a research infrastructure grant from FWO and benefiting from a technical collaboration with LIBIS-KU Leuven, the database has embarked on a transformative journey since 2022.[5] The 'ODIS renewal project 2022–2027' aims to update the mission and goals of ODIS, foster data input and data quality management, strengthen participation, and implement a revamped communication strategy. However, its core ambition is the technical renewal of the database.

Although the project resources did not allow for extensive user research, ODIS had a good understanding of the technical needs of its backend and frontend users through a user survey conducted in 2020, and through several technical workshops at the annual ODIS community day. A key element is the development of new ways to query, analyse, and visualise the ODIS datasets, such as network and geographical visualisation. This will enable ODIS to provide its users with more hands-on research tools. Secondly, we aim to develop an interoperability and discovery layer, offering more and better connections with other catalogues, research instruments, platforms, and linked open data resources. This serves as the cornerstone of a sustainable open access policy, based on the so-called FAIR principles (Findable, Accessible, Interoperable, Reusable) (GO FAIR initiative, 2022). Endpoints (JSON:API and SPARQL endpoint) will enable end users – both ODIS partners and external users – to retrieve data from ODIS for reuse in other databases or within research tools. Rights and responsibilities of the data creators and users will be defined in a data management and sharing plan, which is currently being developed. The endpoints will also facilitate the semantic enrichment of the ODIS content through interconnections with complementary datasets.

To achieve these project goals, the development of a new database and user interface is underway. For the database, it was decided to switch from a relational Oracle database to a triple store database developed in Virtuoso. In the following sections of this contribution, we will discuss the challenges involved in the backend renewal of

[5] Research Foundation Flanders, I011122N. The supervisors of the renewal project are Prof. Dr. Kim Christiaens (KADOC-KU Leuven), Dr. Peter Heyrman (KADOC-KU Leuven), and Jo Rademakers (LIBIS-KU Leuven).

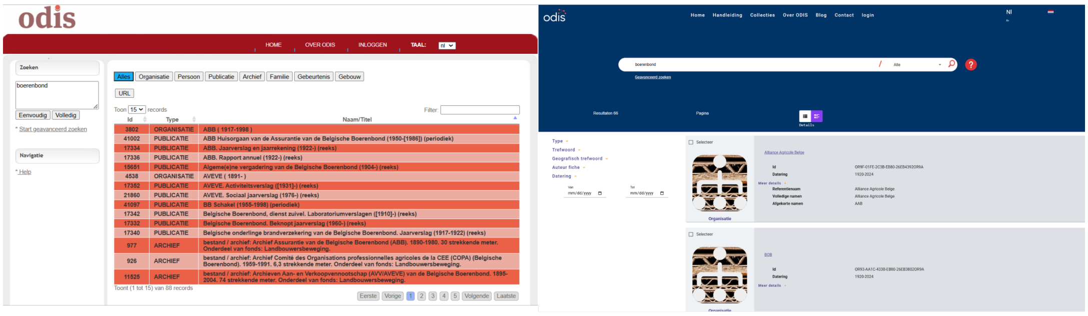


Figure 2: *Left: the current ODIS search interface. Right: the new ODIS search interface.*

ODIS in more detail.

3 A Database in Transition: *ODIS Yet to Come*

In the current phase of the project, a new frontend has been envisioned to maintain a strong connection with users' evolving needs. To implement this renewed vision within the existing infrastructure, a new backend had to be designed and developed. To better understand the choices made for the backend, allow us to outline a vision of *ODIS Yet to Come*.

The website will undergo a facelift, featuring a new logo, new brand colours, updated graphics, and more (see Figure 2). Users will be able to query the OPAC more easily through updated simple search functionalities, a more user-friendly advanced search, the ability to filter the result set using facets, and the option to represent the result list either in a table or in a more detailed manner. All of this should help accommodate researchers of all levels. Navigation will be more intuitive within the website and between the various records, allowing for seamless movement back and forth within the OPAC.

We also aim to enable a 're-presentation' of the rich data already stored within ODIS by offering new ways to explore them both visually and analytically. Not only will the records have a new layout, but there will also be an integrated geographical tool to plot results on a map, helping to contextualise the data spatially. To conclude our vision of *ODIS Yet to Come*, we must mention one final visualisation tool: the network graph. Inspired by, for instance, the *Gemeinsame Normdatei* (GND), managed by the German National Library, we will integrate a network representation within our records (GND-Explorer., n.d.). Users will be able to open the network from a record and further explore the multitude of relationships interconnecting records, hopefully discovering network clusters at a higher level. In short, the current OPAC will be updated to meet the latest standards in search functionalities, along with some useful extras.

Tackling the renewal of the entire infrastructure is no easy task, and neither is describing the process. To meet all the needs of the envisaged frontend, the entire backend architecture must be stable and versatile enough to support it. This can be broken down into a few components: a new database environment, a new data model to structure that environment, and a new API layer to facilitate the exchange of data between the database and the OPAC, as well as between the database and external authorities.

The first task we faced was selecting a new database platform that would meet our needs. Our decision was partially influenced by the type of data model we intended to use. The cultural heritage sector has increasingly shifted towards a linked data approach, exemplified by the International Council on Archives (ICA) launching its new description standard, Records in Contexts (RiC),[6] and the Flemish authorities with their Open Standards for Linking Organisations (OSLO) (International Council on Archives, 2024d; Vlaanderen, n.d.).[7] Given that our frontend applications strongly suggest a linked data framework, this direction appeared to be the most logical choice. Separate from the frontend requirements, the opportunities for the research data in ODIS to be disseminated more widely and at the same time being enriched by external authorities are too important to be overlooked. ODIS data were already interlinked by the many relationships between the different modules, such as between organisations and persons. These links, however, were not always as easy to perceive on a larger scale. This is where an RDF (Resource Description Framework) approach could allow the data to shine.

Although RDF models can be stored in relational databases – typically yielding high-quality results (Hernández et al., 2016) – we decided to embrace the triple store approach.[8] While many are familiar with the concept of relational databases, it is helpful to briefly describe what a triple store entails. A triple store is a database optimised for storing and querying RDF data. Conceptually, it functions as a key-value store, where the key is split into a subject (identifier) and predicate (property name). Unlike relational databases, triple stores lack predefined table structures or constraints, instead relying on standards like RDF and SHACL for data organisation and validation.

While triple stores offer unparalleled flexibility, their layered approach can make them difficult to adopt. Developers often perceive linked data technologies as overly complex, with limited immediate value. This situation underscores the importance of having an API that takes care of validation and database communication.

Several challenges are connected to the adoption of a triple store database, often cited by developers as reasons to refrain from embracing this novel approach. Firstly, it is well known that linked data has a steep learning curve, involving effort and time investment that are often unavailable within the constraints of a specific project. Secondly, some inherent challenges of adopting a triple store system include higher complexity when integrating with other platforms, such as frontends. Specific queries

---

[6] The International Council on Archives (ICA) launched its first drafts of a conceptual model and ontology for Records in Contexts in 2016 and most recently published version 1.0 in 2023.

[7] The Flemish authorities have been rolling out linked open data standards across several fields and have been in the process of working on a standard for cultural heritage since 2020. Their aim is to help facilitate data exchange within the heritage sector in Flanders, for example, between archives and museums, but also to connect with international players.

[8] Ravat et al. (2020) have shown that for querying vast datasets, relational databases still have the upper hand. However, we believe that a triple store approach is appropriate for our case since, despite being large in comparison to other historical research infrastructures, our dataset (308,144 records as of 1 January 2024) is still relatively small compared to those in most performance tests.

must be constructed to make relevant information human-readable. In terms of maintenance, we still observe more complex debugging, as well as performance issues for larger triple stores. Moreover, as ecosystems are not yet as expanded as RDBMSs, triple stores come with a tooling complexity that can be perceived as too high at first.

Yet, we believe that the path to linked data can and should be pursued using triple stores, and that these challenges are outweighed by numerous long-term advantages. Firstly, data models are more flexible and adaptable to the needs of evolving projects, accommodating the needs of different users. A new record structure or graph can thus be easily created using construct queries. Additionally, a significant added value of a triple store approach includes semantic richness and interoperability. These two reasons support our choices. Semantic richness is guaranteed by the possibility of referring to different schemas (e.g. schema.org), and to maintain consistent values. Agent entities are more easily integrable into the data model. Interoperability is a major asset of the new data model built using triple stores, as it is achieved both by smoothly exporting information to other triple stores and by acquiring data from external systems. The apparent tooling complexity can be overcome by implementing an appropriate API layer to facilitate data exchange. Finally, the relative performance issues that can be experienced when running large triple stores are mitigated by using Elasticsearch.

We were not the first humanities database infrastructure in Flanders to transition from a relational to a triple store database. Archiefbank, the database of Archiefpunt, made a similar transition between 2020 and 2022 (Archiefpunt, n.d.; FARO Vlaams steunpunt voor cultureel erfgoed vzw., 2020).[9] Like our colleagues at Archiefpunt, we opted for a Virtuoso Triple Store. Virtuoso is open-source software that provides all the functional necessities for our project: it is a well-maintained product, offers high-performance SPARQL querying for efficient handling of complex queries and optimised RDF graph traversal, supports federated SPARQL queries enabling remote RDF data source queries, is scalable for large datasets through distributed deployment, and includes reasoning and inference support. Additionally, it facilitates data integration across multiple formats, such as XML, CSV, and other databases, while ensuring ACID-compliant transactions.[10] From a financial perspective, it also offers excellent support at a relatively low cost. It is important, however, not to become dependent on one database platform. The way we have built our data model and surrounding infrastructure (as discussed in the next section) ensures that we can easily migrate between platforms if and when we choose to do so. The only requirement for the platforms is that they have an HTTP SPARQL endpoint.

4 Working out Solutions, Challenges Along the Way

The successful implementation of data-driven applications in the digital humanities hinges on the seamless integration of domain expertise with technological proficiency. While scholars possess invaluable knowledge of historical sources, cultural contexts, and research methodologies, they often lack the technical expertise to translate their

9 Archiefpunt vzw is a cultural heritage organisation that aims to raise awareness of private archives in Flanders and Brussels. Its Archiefbank database makes thousands of archival descriptions available to the public. In their 'DATA project', spanning 2020 to 2022, the infrastructure (backend and frontend) underwent a total facelift to better serve its stakeholders.

10 ACID (Atomicity, Consistency, Isolation, and Durability) is a set of rules expected from database platforms.

data models into accessible and interoperable APIs. Conversely, developers may possess the necessary programming skills but lack the nuanced understanding of the domain-specific requirements and constraints that underpin effective data modelling. This inherent disconnection between domain experts and developers presents a significant bottleneck in the development lifecycle, leading to:

1. **Increased development time:** The manual translation of data models into APIs is a time-consuming and error-prone process, requiring constant communication and iterative refinement between domain experts and developers.
2. **Reduced agility:** The rigid separation of roles hinders rapid prototyping and experimentation, both of which are crucial aspects of research and development in the digital humanities.
3. **Limited interoperability:** Inconsistent and poorly documented APIs impede data sharing and integration across different projects and research teams, hampering collaborative research efforts.

This suggests that tools designed to reduce the time taken to transition from a model to an API are highly welcome. The team working on the ODIS transition consists of various profiles, each with its own skill set. On the KADOC side, most members have limited technical skills, but they possess extensive domain knowledge of the existing data model, the data, and the functionality of the database and its interface. LIBIS, the technical partner in the project, has all the technical proficiency but lacks knowledge of the intricacies of ODIS and its data. At first glance, they complement each other beautifully, but this also brings its own set of challenges.

In short, we required a workflow that would allow us to:

1. describe a new RDF-based model,
2. enable people with domain knowledge to work on it independently while ensuring that the deliverables are suitable for the developers,
3. ideally generate the blueprint for an API.

4.0.1 *Spreadsheets as Description Tool*

Describing a model is a way to express what you know about a business domain. It can be used for documentation, data validation, or in our case, the creation of APIs. Although many methods are available for describing models, the easiest approach – requiring no prior knowledge of a specific methodology – is to use one or more **spreadsheets**. This allows for description of the model in a structured yet user-friendly manner and easy segmentation between namespaces, entities, properties, and links. A system of spreadsheets was set up using Google Sheets, where we began to outline the necessary entities and overall structure of the new model. The spreadsheets act as a bridge between the functional and technical sides of the project, enabling team members with more functional knowledge of the database to interact with and help shape the model in an accessible way. In doing so, attention must always be paid to the bilingual nature of the data in ODIS. Initially, ten workbooks were created: one for each module (namespace), a 'general' workbook for modelling entities shared between the various namespaces, and another one to model all code tables. The first tabs in every workbook, as shown in Figure 3, begin with an underscore and are specific to the system:

	A	B	C	D	E	F
4	baseDataTypes	Description	AllDatatypes	AllDatatypes	ProjectEntities	ProjectEntities
5	rdf:JSON		geo:wktLiteral	xsd:string	odis:Agent	odis:Agent
6	rdf:langString	Multi language string	odis:Agent		odis:Archief	
7	xsd:anyURI	A valid URI	odis:Archief		odis:Archiefcode	
8	xsd:boolean	boolean	odis:Archiefcode		odis:Archiefvormer	
9	xsd:date	date	odis:Archiefvormer		odis:Auteur	
10	xsd:dateTime	date + time	odis:Auteur		odis:Beroep	
11	xsd:double	converted to float. double is 64bit, float 32bit	odis:Beroep		odis:BeroepTitel	
12	xsd:duration	iso8601 duration	odis:BeroepTitel		odis:Beschermingsstatus	
13	xsd:float	a float	odis:Beschermingsstatus		odis:Beschrijvingsniveau	
14	xsd:gYear	an integer	odis:Beschrijvingsniveau		odis:Bevoegdheid	
15	xsd:integer		odis:Bevoegdheid		odis:Bewaarplaats	
16	xsd:string		odis:Bewaarplaats		odis:BewaarplaatsCode	
17	xsd:time		odis:BewaarplaatsCode		odis:BouwhistorischFeitSystematisch	
18	time:DateTimeInterval		odis:BouwhistorischFeitSystematisch		odis:Bouwstructuur	
19	schema:temporalCoverage		odis:Bouwstructuur		odis:Bron	
20	geo:wktLiteral		odis:Bron		odis:BronParallelId	
21	v:Geometry		odis:BronParallelId		odis:ChronologischeBouwgeschiedenis	
22			odis:ChronologischeBouwgeschiedenis		odis:Codetabel	
23			odis:Codetabel		odis:Contact	
24			odis:Contact		odis:Context	
25			odis:Context		odis:Copyright	
26			odis:Copyright		odis:DateringSystematisch	
27			odis:DateringSystematisch		odis:Diploma	
28			odis:Diploma		odis:Entiteit	
29			odis:Entiteit		odis:Ereteken	
30			odis:Ereteken		odis:Erkenning	
31			odis:Erkenning		odis:ExterneCatalogus	
32			odis:ExterneCatalogus		odis:Familie	
33			odis:Familie		odis:FamilieFunctie	
34			odis:FamilieFunctie		odis:FamilieWoonplaats	
35			odis:FamilieWoonplaats		odis:Formaat	
36			odis:Formaat		odis:Functie	

_README · _DATATYPES · _PREFIXES · _METADATA · _REFERENCES

Figure 3: *Example of the* _DATATYPES *tab.*

- **_README**: A brief explanation of the spreadsheet structure.
- **_DATATYPES**: A comprehensive list of data types across namespaces.
- **_PREFIXES**: Taxonomies and external references.
- **_METADATA**: Metadata such as the model's title and version.
- **_REFERENCES**: Links to additional spreadsheets within the model.

4.0.2 *Namespace-Specific Sheets*

The main difference between the global and namespace workbooks is the `_ENTITIES` sheet. Each namespace has a dedicated `_ENTITIES` worksheet, which describes entities, their plural forms, and relationships (e.g. inheritance (`subClassOf`) or external equivalence (`sameAs`)). For each entity, additional worksheets define its properties, including data types, constraints, and links to other entities. Every entity defined in the `_ENTITIES` sheet can have a separate tab that describes all the properties associated with that entity. When an entity has a `subClassOf` relationship, it will inherit all of the properties of its parent entity, as illustrated in Figure 4.

Let's have a look at a couple of entities in Figure 5. `Entiteit` does not have a `subClassOf` relationship, as it is the basis for all entities within the model. In Figure 6, we can see that it has three properties: id, parallel_id, and _audit. When we look at the entity `Codetabel`, we see it is a `subClassOf` of `Entiteit`, meaning it inherits all of its properties. The properties of `Codetabel`[11] thus include `id`, `parallel_id`, `_audit`, `waarde`, and `context`, with the latter two defined in the `Codetabel` tab.

[11] `Codetabel` is a separate entity that describes all code tables used throughout the database.

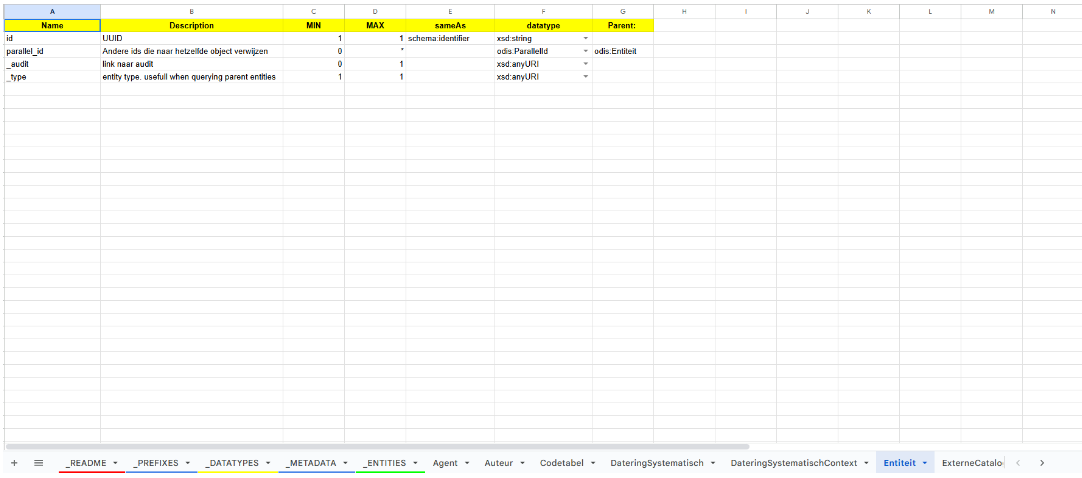
	A	B	C	D	E
1	**Name**	**NamePlural**	**Description**	**subClassOf**	**sameAs**
2	Agent	Agenten		odis:Entiteit	rico:Agent
3	Auteur	Auteurs		odis:Entiteit	
4	BeroepTitel	BeroepTitels		odis:Codetabel	
5	Beschrijvingsniveau	Beschrijvingsniveaus		odis:Codetabel	
6	Bevoegdheid	Bevoegdheden		odis:Codetabel	
7	BewaarplaatsCode	BewaarplaatsCodes		odis:Codetabel	
8	BouwhistorischFeitSystematisch	BouwhistorischFeitSystematisch		odis:Codetabel	
9	Bron	Bronnen		odis:Codetabel	
10	BronParallelId	BronParallelIds		odis:Codetabel	
11	Codetabel	Codetabellen	Basis Klasse voor code tabellen	odis:Entiteit	
12	Context	Contexten		odis:Codetabel	
13	Copyright	Copyright		odis:Codetabel	
14	DateringSystematisch	DateringenSystematisch		odis:Entiteit	
15	Entiteit	Entiteiten	Basis klasse voor alle entiteiten		schema:Thing
16	ExterneCatalogus	ExterneCatalogi		odis:Entiteit	
17	Functie	Functies		odis:Codetabel	
18	GeografischeTrefwoord	GeografischeTrefwoorden		odis:Codetabel	
19	Geslacht	Geslachten		odis:Codetabel	
20	Illustratie	Illustraties		odis:Entiteit	
21	Landcode	Landcodes		odis:Codetabel	
22	Naam	Namen		odis:Entiteit	
23	Naamsoort	Naamsoorten		odis:Codetabel	
24	Onderwijsinstelling	Onderwijsinstellingen		odis:Organisatie	
25	Onderwijsniveau	Onderwijsniveaus		odis:Codetabel	
26	OnlineBijlage	OnlineBijlagen		odis:Entiteit	
27	Orgaan	Organen		odis:Codetabel	
28	ParallelId	ParallelIds		odis:Entiteit	
29	Periodiciteit	Periodiciteiten		odis:Codetabel	
30	PersoonEreteken	PersoonEretekens		odis:Codetabel	
31	PersoonTitel	PersoonTitels		odis:Codetabel	
32	Publicatiesoort	Publicatiesoorten		odis:Codetabel	
33	Relatie	Relaties		odis:Codetabel	

+ ≡ _README _PREFIXES _DATATYPES _METADATA _ENTITIES Agent Aute

Figure 4: *Example of the* `_ENTITIES` *tab.*

Name	**NamePlural**	**Description**	**subClassOf**	**sameAs**
Agent	Agenten		odis:Entiteit	rico:Agent
Auteur	Auteurs		odis:Entiteit	
BeroepTitel	BeroepTitels		odis:Codetabel	
Beschrijvingsniveau	Beschrijvingsniveaus		odis:Codetabel	
Bevoegdheid	Bevoegdheden		odis:Codetabel	
BewaarplaatsCode	BewaarplaatsCodes		odis:Codetabel	
BouwhistorischFeitSystematisch	BouwhistorischFeitSystematisch		odis:Codetabel	
Bron	Bronnen		odis:Codetabel	
BronParallelId	BronParallelIds		odis:Codetabel	
Codetabel	Codetabellen	Basis Klasse voor code tabellen	odis:Entiteit	
Context	Contexten		odis:Codetabel	
Copyright	Copyright		odis:Codetabel	
DateringSystematisch	DateringenSystematisch		odis:Entiteit	
Entiteit	Entiteiten	Basis klasse voor alle entiteiten		schema:Thing
ExterneCatalogus	ExterneCatalogi		odis:Entiteit	

Figure 5: *Example of class inheritance.*



Name	Description	MIN	MAX	sameAs	datatype	Parent:
id	UUID	1	1	schema:identifier	xsd:string	
parallel_id	Andere ids die naar hetzelfde object verwijzen	0	*		odis:ParallelId	odis:Entiteit
_audit	link naar audit	0	1		xsd:anyURI	
_type	entity type. usefull when querying parent entities	1	1		xsd:anyURI	

Figure 6: *Example of class* `Entiteit` *worksheet.*

The `sameAs` column references a context outside this model. For `Entiteit`, this indicates that it shares the same context as `schema:Thing`. That approach ensures the interoperability of our model with external standards and systems. By defining parent and child entities, we dismantle the barriers that previously separated the eight modules in ODIS. In the old model, humans could identify similarities in field groups and assume their equivalence; however, this logic was not apparent to machines. By employing shared entities, we define and embed these similarities in aspects such as dates and names, as well as in our conceptualisation of these 'base entities' as agents.

4.1 SOLIS

The sheets provide a foundation, yet the transition from a conceptual data model to functional APIs is a resource-intensive process. Our research focuses on leveraging the expressive power of **SHACL** (Shapes Constraint Language), a W3C standard for defining and validating RDF data, to bridge the gap between data models and APIs (W3C, 2017). SHACL provides a declarative and human-readable syntax for expressing constraints and expectations on RDF data, making it an ideal candidate for capturing domain-specific knowledge and translating it into machine-processable specifications.

Building on the spreadsheets' foundation, we developed **SOLIS** (Smart Ontology Layer for Interoperable Systems). This engine can take the information in the sheets to generate ERDs and, more importantly, SHACL files. These files contain the logic and rules necessary to validate data ready for import into the triple store. SOLIS is also a novel library that automates the generation of RESTful APIs directly from SHACL-based data models. By leveraging the flexibility and scalability of **triple stores**, SOLIS enables developers to:

- **Rapidly prototype and iterate:** The automated API generation process significantly accelerates the development cycle, enabling researchers to quickly test and refine their data models and explore different API endpoints.

Code tables are controlled lists of values that aren't subject to constant change but provide a stable way of preserving controlled data and links. Since we are using an open API, we need to limit certain entities to read-only, and these are grouped under `Codetabel`.

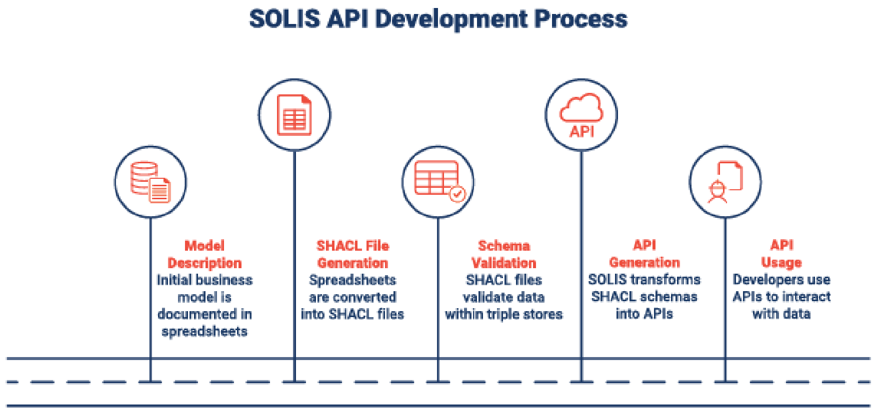


Figure 7: SOLIS pipeline diagram.

- **Focus on user experience:** By abstracting away the complexities of API development and RDF interpretation, SOLIS empowers developers to concentrate on building intuitive and user-friendly interfaces that cater to the specific needs of humanities scholars and researchers.

- **Enhance data interoperability:** The use of standardised formats and protocols ensures that APIs generated by SOLIS are easily discoverable, accessible, and compatible with a wide range of tools and platforms, fostering data sharing and collaboration within the digital humanities community.

SOLIS represents a significant step towards streamlining the development of data-driven applications in the digital humanities. By empowering domain experts to actively participate in the API development process, and by providing developers with a powerful and efficient toolkit, SOLIS has the potential to unlock new avenues for research and innovation in this exciting field. It explicitly addresses the challenges of the current situation by highlighting the inefficiencies caused by the disconnection between domain experts and developers, and it emphasises the significance and potential of SHACL for capturing domain knowledge and generating API specifications.

By leveraging SHACL, SOLIS transforms these constraints into functional API endpoints, bridging the gap between data modelling and application development. Based on the files, SOLIS generates two things: documentation about the model, and a JSON:API layer that will enable interoperability between the database, the new frontend, and third-party users (Aernouts and Celik, 2025; Celik, 2025; JSON:API, n.d.). This is crucial to our architecture, as it enforces the logic and rules for data validation entering the triple store, while also providing sense and structure to the data coming out of the database. The triple store serves as a backend storage: all logic is stored and enforced through the API, allowing for easy iterative work. Whenever any changes have to be made to the model during the modelling and remodelling process, a simple regeneration of the SHACL will resolve the issue efficiently.

4.2 Modelling Challenges

In reshaping the data model, there were two main goals: 1) it must facilitate an easy transition from the old model to the new; and 2) it must be compliant with RiC. The first requirement was driven by a desire to avoid alienating users familiar with the existing model, as well as by the need to transition as efficiently as possible. As mentioned earlier, the old model was based on several ICA description standards that were subsequently incorporated into RiC, making compliance easily achievable. Some logical changes were made in the new ODIS model, such as incorporating Agent entities and using generic entities to serve similar functions; for instance, the `odis:Naam` entity is utilised for both names and titles of publications or archives. These abstractions were necessary to rationalise the model. Obsolete fields – fields that were rarely used or often misused – were discarded where possible. This approach helps to maintain the quality of data input by limiting certain fields and better guiding users in their data input efforts.

The bilingual nature of the data presented another challenge. The method used to store language variants in the relational database is difficult to replicate. Records consisted of language-dependent and language-independent text fields, a specific solution for the 2009–2013 renewal that does not translate well to the new approach. By assigning the datatype rdf:langString to these properties in the new model, we were able to address the language issue. In the future, the database can potentially be expanded to accommodate more language variants, moving from bilingualism to multilingualism.

One of the requirements for the new frontend is an integrated geographical visualisation of the data. This necessitates, among other things, GIS data to enable plotting on a map. The current database already contains an extensive geographical thesaurus, built on a dataset from TomTom for Europe, which has been manually enriched with data for locations outside the European continent. While this provided a foundation for the geo tool, it was a self-contained thesaurus with no links to other authorities. This prompted the desire to use external authorities for our geographical thesaurus. The authority had to meet several requirements:

1. there should be metadata to reconstruct the hierarchy of place names,
2. it should grant free access to global data,
3. it should allow for multiple authority references such as Wikidata IDs,
4. given that ODIS works with historical subjects, historical place names should be available as aliases.

We found that OpenStreetMap met most of our needs and provided not only data but also a widespread community (OpenStreetMap, n.d.). It also offers us the opportunity to contribute to the community by documenting places used in ODIS that are not yet present in OpenStreetMap. At the time of writing, we have successfully mapped the old geodata to the OpenStreetMap model and are researching ways to incorporate their data, either by ingesting and regularly updating it or through a 'live call'.

One of the factors that ignited interest in linked data for this project was network visualisation. In the academic year 2023-2024, Yiyao Jin, an MSc student in Digital Humanities at KU Leuven (supervised by Prof. Dr. Katrien Verbert), worked on this topic for ODIS. She produced a first draft of what a network visualisation of our data could look like in her master's thesis (Jin, 2024). Her research was already very promising; using Gephi to represent a select dataset yielded some exciting results. Her findings will serve as a starting point for developing the visualisation tool in the frontend.

ODIS is a dynamic database, and part of the transition project involves migrating the legacy data to the new environment. To transfer the data from the old database to the new one, the existing structure had to be converted to the new triples format. Mappings were made to facilitate an easy transition, which served as a useful test to identify any omissions or logic errors in the new model. Once the conversion scripts were completed, the newly converted data were ingested in the triple store via a JSON:API. This functioned as a preliminary test to ensure that the necessary constraints and logic were being enforced. The process resulted in several iterations of the model, as well as updates to the SHACL and API. This highlighted the advantages of this approach, allowing for easy adjustments to the model in the spreadsheet and seamlessly triggering subsequent changes to the SHACL and API.

5 Concluding Remarks

In this paper, we shared insights from the ongoing technical renewal project for the contextual database ODIS. The journey to the new ODIS environment is an adventurous one; we are currently halfway through and on track to complete the project on time. We have established a new triple store database in Virtuoso, reworked the data model, and set up a user-friendly workflow for easy modelling and API generation with SOLIS. Furthermore, we have successfully converted the data and rigorously tested the setup. The shift to a triple store approach enables us to break down barriers not only between ODIS and other data repositories but also between the separate modules within the database. Additionally, it has enabled us to align the ODIS data model with Records in Contexts. Alongside redesigning the database structure, the public interface of ODIS will undergo a revamp.

The technical renewal programme aims to strengthen the flexibility and multifunctional use of ODIS, fostering its broad applicability in individual and collaborative scientific research, all within a multidisciplinary and international-comparative perspective. The system will help ODIS users compose and query consistent data packages. It will also enable them to visualise and analyse ODIS datasets in novel ways, supporting, for example, a more in-depth understanding of associational life, the prosopographical analysis of civil society organisations, the visualisation of biographical and associational itineraries, the spatial and territorial impressions of civil society, and more. In addition to opening new perspectives for data-driven research in the social sciences and humanities, we are confident that the investment programme will also encourage more research units to join the ODIS partnership and incorporate their own data series, further strengthening the thematic scope and research potential of the database.

We hope that our journey will inspire similar projects and demonstrate that it is possible to navigate the challenging middle ground between business and technical teams. We also hope to have offered convincing reasons for developers to take the leap into the RDF domain and overcome the obstacles. Our approaches can serve as a foundation for others to further develop their own workflows and streamline their transition processes.

References

Onderzoekssteunpunt en Databank Intermediaire Structuren in Vlaanderen (19de-20ste eeuw). *Belgisch Staatsblad*, 2024. URL `https://www.ejustice.just.fgov.be/tsv_pdf/2024/01/08/24004971.pdf`.

Bert Aernouts and Mehmet Celik. ODIS Ontologie, 2025. URL https://data.odis.be/_doc.

Georgia Angelaki, Karolina Badzmierowska, David Brown, Vera Chiquet, Joris Colla, Judith Finlay-McAlester, Klaudia Grabowska, Vanessa Hannesschläger, Natalie Harrower, Freja Howat-Maxted, Maria Ilvanidou, Wojciech Kordyzon, Magdalena Król, Antonio Gabriel Losada Gómez, Maciej Maryl, Sanita Reinsone, Natalia Suslova, Mark Sweetnam, Kamil Śliwowski, and Marcin Werla. How to Facilitate Cooperation between Humanities Researchers and Cultural Heritage Institutions. Guidelines. Technical report, Digital Humanities Centre at the Institute of Literary Research of the Polish Academy of Sciences, Warsaw, Poland, March 2019. URL https://zenodo.org/records/2587481.

Archiefpunt. Home - Archiefpunt, n.d. URL https://archiefpunt.be/.

Mehmet Celik. GitHub - mehmetc/solis: Turn a SHACL file into an API, April 2025. URL https://github.com/mehmetc/solis. original-date: 2021-09-17T12:17:43Z.

Joris Colla and Peter Heyrman. Het middenveld en zijn erfgoed in context: het veelzijdig gebruik van de onlinedatabank ODIS (2000-2017). In D Leyder and N Istasse, editors, *Ten dienste van de gebruiker. Over de zoekinstrumenten ontwikkeld door archivarissen en bibliothecarissen. Akten van het colloquium van 8 december 2017*, volume 105 of *Archief- en Bibliotheekwezen in België*, pages 99–111. Archief- en Bibliotheekwezen in België, 2019.

Jan De Maeyer. ODIS: Database Intermediary Structures Flanders (AKUL043). Final Report, June 2015. Technical report, 2015. URL https://www.odis.be/hercules/docs/FinalReport_AKUL043_ODIS.pdf.

Docomomo International. Docomomo International – Architecture Archive, n.d. URL https://docomomo.com/.

FARO Vlaams steunpunt voor cultureel erfgoed vzw. Digitale Architectuur Traject Archiefpunt (DATA). Verbetering van gebruiksvriendelijkheid en interconnectiviteit in Linked Open Data tijd, fase 1 | FARO Vlaams steunpunt voor cultureel erfgoed vzw, 2020. URL https://faro.be/project/digitale-architectuur-traject-archiefpunt-data-verbetering-van-gebruiksvriendelijkheid-en.

GND-Explorer. Homepage, n.d. URL https://explore.gnd.network/.

GO FAIR initiative. FAIR Principles - GO FAIR, 2022. URL https://www.go-fair.org/fair-principles/.

Daniel Hernández, Aidan Hogan, Cristian Riveros, Carlos Rojas, Enzo Zerega, Fabian Flöck, Elena Simperl, Marta Sabou, Alasdair Gray, Paul Groth, Freddy Lecue, Yolanda Gil, and Markus Krötzsch. Querying Wikidata: Comparing SPARQL, Relational and Graph Databases. In *The Semantic Web–ISWC 2016: 15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016, Proceedings, Part II 15*, volume 9982 of *Lecture Notes in Computer Science*, pages 88–103. Springer International Publishing AG, Switzerland, 2016. ISBN 0302-9743. doi: 10.1007/978-3-319-46547-0_10.

Peter Heyrman. Online tools and instruments supporting the study of transnational elites: the functionalities of the contextual database ODIS. In Andrea Ciampani and Thomas Kroll, editors, *Transnational Encounters?*, European Elites, International Associations and National States (1882-1914). De Gruyter Oldenbourg, 2025. ISBN 978-3-11-168059-0.

Peter Heyrman and Donald Weber. Intermediaire structuren in Vlaanderen: ODIS. In Marijke Hoflack, editor, *Het Max Wildiersfonds geëvalueerd: archief en wetenschappelijk onderzoek.*, Archiefkunde : verhandelingen aansluitend bij Bibliotheek- en archiefgids 9, pages 73–80. VVBAD Vlaamse vereniging voor bibliotheek-, archief- en documentatiewezen, Antwerpen, 2007. ISBN 978-90-72679-32-1.

International Council on Archives. ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families, 2nd Edition, 2024a. URL https://www.ica.org/resource/isaar-cpf-international-standard-archival-authority-record-for-corporate-bodies-persons-and-families-2nd-edition/.

International Council on Archives. ISAD(G): General International Standard Archival Description - Second edition, 2024b. URL https://www.ica.org/resource/isadg-general-international-standard-archival-description-second-edition/.

International Council on Archives. ISDIAH: International Standard for Describing Institutions with Archival Holdings, 2024c. URL https://www.ica.org/resource/isdiah-international-standard-for-describing-institutions-with-archival-holdings/.

International Council on Archives. Records in Contexts - Ontology, 2024d. URL https://www.ica.org/resource/records-in-contexts-ontology/.

International Federation of Library Associations. International Standard Bibliographic Description Consolidated Edition 2011, 2011. URL https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/isbd/isbd-cons_20110321.pdf.

Yiyao Jin. *The Network Visualization Approach for the ODIS Database*. PhD thesis, KU Leuven. Faculteit Wetenschappen, Leuven, 2024.

JSON:API. JSON:API — A specification for building APIs in JSON, n.d. URL https://jsonapi.org/.

KADOC-KU Leuven. KADOC - ODIS: KU Leuven Core Facility on Religion, Culture and Society, 2024. URL https://kadoc.kuleuven.be/english/3_research/31_ourresearch/ODIS.

KU Leuven. KU Leuven Core Facilities, 2024. URL https://research.kuleuven.be/en/research-areas/core-facilities.

ODIS. About ODIS, n.d.a. URL https://www.odis.be/hercules/_en_overODIS.php.

ODIS. Home, n.d.b. URL https://www.odis.be.

OpenStreetMap. OpenStreetMap, n.d. URL https://www.openstreetmap.org/.

Franck Ravat, Jiefu Song, Olivier Teste, and Cassia Trojahn. Efficient querying of multidimensional RDF data with aggregates: Comparing NoSQL, RDF and relational data stores. *International Journal of Information Management*, 54:102089, October 2020. ISSN 0268-4012. doi: 10.1016/j.ijinfomgt.2020.102089. URL https://www.sciencedirect.com/science/article/pii/S0268401219306097.

Vlaanderen. OSLO, n.d. URL https://www.vlaanderen.be/digitaal-vlaanderen/onze-diensten-en-platformen/oslo.

W3C. Shapes Constraint Language (SHACL), July 2017. URL https://www.w3.org/TR/shacl/.

Guohui Xiao, Diego Calvanese, Roman Kontchakov, Domenico Lembo, Antonella Poggi, Riccardo Rosati, and Michael Zakharyaschev. Ontology-Based Data Access: A Survey. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pages 5511–5519, Stockholm, Sweden, July 2018. International Joint Conferences on Artificial Intelligence Organization. ISBN 978-0-9992411-2-7. doi: 10.24963/ijcai.2018/777. URL https://www.ijcai.org/proceedings/2018/777.