Atramhasis: An online SKOS vocabulary editor

Koen Van Daele ![ORCID]¹, **Tinne Cahy**², **Wim De Clercq**², **Jonas Geuens** ![ORCID]¹, **Bram Goessens** ![ORCID]¹, **Jonathan Piraux**², **Bart Saelen**², **Maarten Taeymans**², **and Cedrik Vanderhaegen**¹

1 Flanders Heritage Agency, Belgium **2** GEO Solutions, Belgium

Summary

Atramhasis is an online vocabulary editor that allows users to create, maintain and consult controlled vocabularies and thesauri (Harpring, 2010) according to the SKOS specification (Miles & Bechhofer, 2008). SKOS (Simple Knowledge Organisation System) is a W3C Recommendation that supports controlled vocabularies within the framework of the Semantic Web. SKOS provides a standard way to represent these vocabularies using RDF (Resource Description Framework) (Schreiber & Raimond, 2014), another W3C Recommendation that allows passing data between computer applications in an interoperable way.

SKOS vocabularies record controlled vocabularies as a set of concepts, collections and their relations. A concept is something a researcher wants to describe and define, and a collection is a grouping of a number of concepts. Flanders Heritage is active in the domain of cultural heritage and describes concepts such as Roman period (Slechten, 2004), oppida, and archaeological excavations. A typical collection would be settlements by function, a grouping of settlement types according to their function, as opposed to by size or form. Similar concepts and collections are grouped in a conceptscheme, e.g., the heritage types conceptscheme consists of concepts and collections that describe types of heritage, ranging from solitary trees over burial mounds to airfields and swimming pools. Concepts and collections can be labelled with preferred labels and alternative labels, more elaborately described with definitions and notes, and provided with references to source material. The concepts can be related, in general or in a hierarchical way (broader/narrower), to other concepts and collections in the same vocabulary or to concepts in other vocabularies such as the Art and Architecture Thesaurus (AAT) (Getty Vocabulary Program, n.d.).
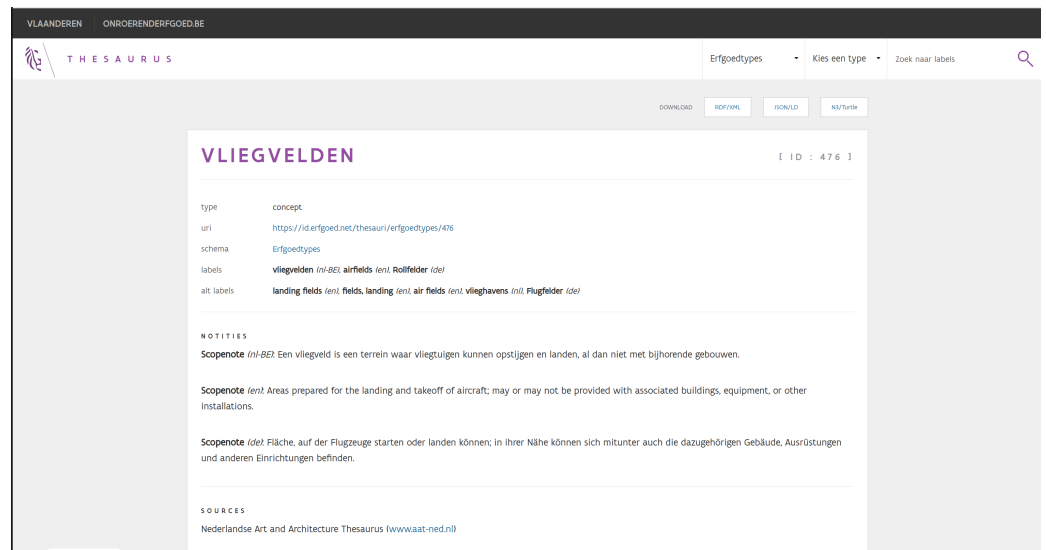
Figure 1: Airfields described as a SKOS concept.

Statement of need

Functional requirements

Atramhasis was written to be a user-friendly open source SKOS editor. First and foremost, we wanted a system that adheres to the SKOS standard, yet is useable for users without prior knowledge of SKOS or RDF. For a typical user, it had to feel as if they were consulting a normal website, as opposed to a RDF vocabulary, since the latter can feel rather daunting to non-technical users. This not only applies to users consulting the thesauri, but also to those editing them. Our thesaurus editors are not IT- or linked data specialists, but domain experts, researchers and other specialists in the field of cultural heritage. While a general understanding of thesauri is within their grasp, the technicalities of RDF are not. Thus, editors in Atramhasis do not write RDF statements, but edit data in a normal web admin interface, as seen in Figure 2. All mapping to RDF and SKOS is done behind the scenes, invisible to the editors.
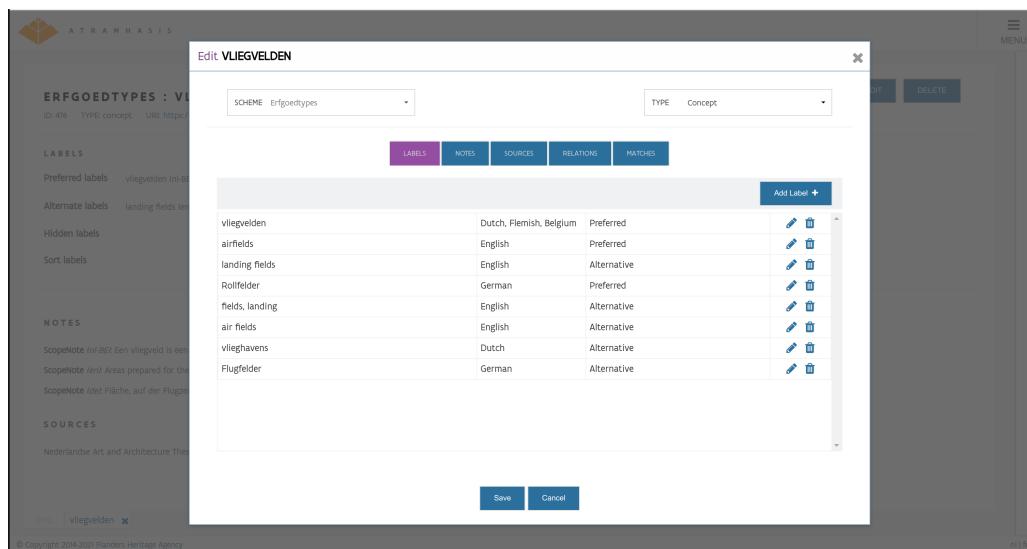
Figure 2: Editing the concept of airfields is simple and straightforward.

The system was conceived as Flanders Heritage's central platform for publication of internal and regional vocabularies dealing with cultural heritage (Mortier et al., 2017). The publication website allows humans to browse, search, and consult the vocabularies online in a user-friendly way. Search results can be downloaded in CSV format for further processing. Internal and external systems use the webservices provided by Atramhasis to consult or download vocabularies. Concept URIs are used in indexing data in systems such as the Inventory of Immovable Cultural Heritage (Haan & Vanmaele, 2021; Hooft, 2021; Van Daele et al., 2015) or the Flanders Heritage Image Database. This allows users to search those systems using the provided thesauri (Figure 3). For a typical end-user, the thesauri are presented as dropdown lists or specialised widgets that allow navigating the thesaurus from the top concepts along branches to the leafs. For most transactions between internal Flanders Heritage systems and the thesaurus system, simple JSON REST services are used. These are deliberately modelled on the implementation standards used in other Flanders Heritage sytems. This allows developers used to working in an enterprise IT context to feel comfortable and productive.
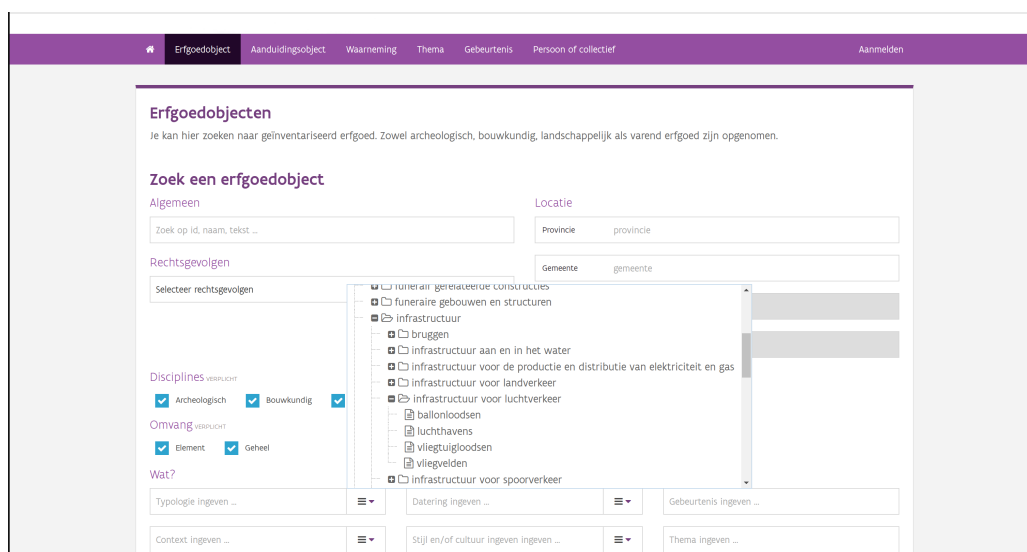

Figure 3: Searching for airfields in the Inventory of Immovable Cultural Heritage

Van Daele et al. (2023). Atramhasis: An online SKOS vocabulary editor. *Journal of Open Source Software*, *8*(83), 5040. https://doi.org/10.21105/joss.05040.

While interactions with internal system are done through plain JSON REST services, publishing of linked data for external consumption is also supported. Individual concepts and collections can be downloaded as RDF data in Turtle, RDF/XML, and JSON-LD format (Kellogg et al., 2020). Entire conceptschemes can be downloaded in Turtle or RDF/XML format. Finally, an integrated Linked Data Fragments (LDF) server is available, serving the thesauri through the Triple Pattern Fragments protocol (Verborgh et al., 2016), either from Turtle files or - for optimal performance - HDT files, a binary RDF representation (Fernández et al., 2013). An LDF server such as the Flanders Heritage Thesaurus LDF server can be browsed online for basic usage, but more importantly provides a full SPARQL interface to the Atramhasis thesauri (Figure 4) through an LDF client such as Comunica (Taelman et al., 2018). This makes all the versatility of SPARQL queries available without having to setup a triplestore, thus keeping the required technology stack small. Implementors who do need or want a full triplestore could easily add one and use the export capabilities provided by Atramhasis to populate the triplestore.

Figure 4: Querying the Flanders Heritage thesaurus of styles and cultures served by an Atramhasis server with a SPARQL query from a comunica client

An Atramhasis instance can be connected to external vocabularies and thesauri through an interface called a Skosprovider (Van Daele & Geuens, 2022), and any thesaurus providing

one can be used for linking external concepts. Out of the box skosproviders are available for Getty vocabularies such as the AAT (Getty Vocabulary Program, n.d.) and other Atramhasis instances, but any thesaurus adhering to the SKOS standard can be added with a little development work. Connecting an external thesaurus opens up the possibility of interlinking internal and external thesauri, importing concepts from such a thesaurus, and turning your vocabularies into true linked data.

Technical requirements

As a government agency, Flanders Heritage has its own corporate identity, part of the wider branding of the Flemsish Government. Therefore, Atramhasis comes with a default style but is easy to extend with a custom corporate identity. This can be seen by comparing the Flanders Heritage thesaurus with the default Atramhasis setup.

We needed software that was easy to integrate with our regular authentication and authorisation mechanism, a single sign-on environment used by most Flemish Government agencies. Therefore, Atramhasis does not come with a default authentication and authorization layer, but the underlying Pyramid framework provides hooks and integration points facilitating this. There are default libraries for this framework that can be configured according to a user's own corporate security needs.

Atramhasis uses SQLAlchemy, a database abstraction layer, so it can be run with different relational databases. We recommend PostgreSQL for an enterprise multi-user production environment such as the Flanders Heritage thesaurus, which has been serving 25.000 visitors annually. SQLite is very well suited for a single-user environment and rapid prototyping. By using this very simple file-based backend and not configuring any authentication, Atramhasis has been used as a quick SKOS editor by people not wanting to write SKOS by hand.

Since we already had multiple thesauri, a single instance of Atramhasis can host multiple conceptschemes (Figure 5). Creating a conceptscheme requires somewhat more work than creating a concept or collection. Generally it is best left to system admins and IT experts who can also set up a URI generation scheme and handler, decide on some special configuration settings, and know how the conceptscheme will be used in other applications.
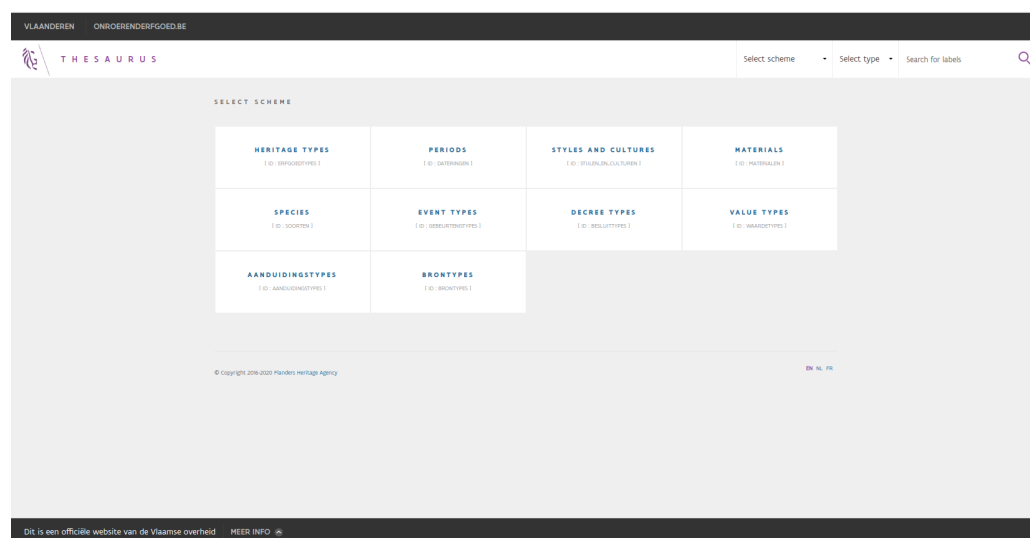

Figure 5: All conceptschemes in a single Atramhasis instance

Finally, we knew our thesauri were fairly small. The largest Flandes Heritage conceptscheme holds some 1.485 concepts. Bigger conceptschemes are certainly possible. No upper limit

has been reached so far and the software itself has no hardcoded limit. However, we do feel Atramhasis is not ideal for hosting very large thesauri such as the AAT (Getty Vocabulary Program, n.d.). Often such a thesaurus defines custom properties or has smaller subgroups to keep the thesarus navigable. At Flanders Heritage we have avoided creating heterogenous conceptschemes, opting for conceptschemes with a tight focus. For example, the Flanders Heritage thesaurus has different conceptschemes that each map to an AAT subgroup called a Facet (Styles and Periods, Activities, Materials, Objects). While the end result is very similar, Atramhasis does not currently support something like the facets the AAT employs to organise concepts in different subgroups within a single conceptscheme. So far, this has not proved to be an issue.

State of the field

Having decided on our functional and technical requirements, we surveyed vocabulary software available at the time (2014). Knowing we needed to integrate existing software in our normal technical environment, requiring a great degree of flexibility and customisation, we focussed our search on open source software.

Software like Protégé offers a lot of functionality for building RDF vocabularies, but is not suited for editing by non-technical users. Collaboration would have required all editors to be proficient in version control systems such as Git. A similar limitation is shared with a project such as SkoHub. Others projects such as Skosmos provide for publication of thesauri, but not editing. A true online editor such as TemaTres had a more user-friendly interface, but was difficult to evaluate properly since most of the documenation and code was in Spanish. Both TemaTres and OpenSkos are written in PHP. OpenSkos was also lacking good documentation so it was unclear how easy it would be to customise and adapt the software. iQvoc had exactly the kind of end-user experience we were looking for, but it runs on the idea that every conceptscheme requires a new instance of the application, which would have required a lot of work whenever a new scheme was needed. None of the available solutions had a ready-made integration with a single sign-on environment or made it easy to build one. Adding your own corporate identity would have been feasible in some ways, but often it would have to be done by forking the software as opposed to configuring it, complicating long term maintenance.

Conclusion

After careful consideration of our functional and technical requirements and the available open source software, we decided to write a simple but extensible editor in Python. We felt this was the best way to make sure we could support all our use cases in a sustainable way. So far this has proven to be the right decision.

Acknowledgements

References

Fernández, J. D., Martínez-Prieto, M. A., Gutiérrez, C., Polleres, A., & Arias, M. (2013). Binary RDF representation for publication and exchange (HDT). *Web Semantics: Science, Services*

and Agents on the World Wide Web, *19*, 22–41. https://doi.org/10.2139/ssrn.3198999

Getty Vocabulary Program. (n.d.). *Art and archtitecture thesaurus*. Getty Research Institute. https://www.getty.edu/research/tools/vocabularies/aat/about.html

Haan, A. de, & Vanmaele, N. (2021). *Inventariseren van landschappelijk erfgoed in vlaanderen. Historiek, methodologie, doelstellingen en resultaten* (No. 195; Onderzoeksrapporten Agentschap Onroerend Erfgoed). agentschap Onroerend Erfgoed. https://doi.org/10.55465/rsdd4339

Harpring, P. (2010). *Introduction to controlled vocabularies: Terminology for art, architecture, and other cultural works*. Getty Research Institute. ISBN: 978-1-60606-026-1

Hooft, E. (2021). *Inventariseren van bouwkundig erfgoed in vlaanderen. Historiek, methodologie, doelstellingen en resultaten* (No. 196; Onderzoeksrapporten Agentschap Onroerend Erfgoed). agentschap Onroerend Erfgoed. https://doi.org/10.55465/mmym4330

Kellogg, G., Champin, P.-A., & Longley, D. (Eds.). (2020). *JSON-LD 1.1*. World Wide Web Consortium. https://www.w3.org/TR/2020/REC-json-ld11-20200716/

Miles, A., & Bechhofer, S. (2008). *SKOS simple knowledge organization system reference*. World Wide Web Consortium, Working Draft WD-skos-reference-20080829. https://www.w3.org/TR/2009/REC-skos-reference-20090818/

Mortier, S., Van Daele, K., & Meganck, L. (2017). Calling it what it is. Thesauri in the Flanders Heritage Agency: History, Importance, Use and Technological Advances. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *IV-2/W2*, 151–156. https://doi.org/10.5194/isprs-annals-IV-2-W2-151-2017

Schreiber, G., & Raimond, Y. (Eds.). (2014). *RDF 1.1 Primer*. World Wide Web Consortium. https://www.w3.org/TR/rdf11-primer/

Slechten, K. (2004). Namen noemen. Het CAI-thesaurusproject. In E. Meylemans (Ed.), *CAI-I. De opbouw van een archeologisch beleidsinstrument* (pp. 49–54). Instituut voor het Archeologisch Patrimonium. https://oar.onroerenderfgoed.be/publicaties/IAPR/14/IAPR014-003.pdf

Taelman, R., Van Herwegen, J., Vander Sande, M., & Verborgh, R. (2018). Comunica: A modular SPARQL query engine for the web. In D. Vrandečić, K. Bontcheva, M. C. Suárez-Figueroa, V. Presutti, I. Celino, M. Sabou, L.-A. Kaffee, & E. Simperl (Eds.), *Proceedings of the 17th International Semantic Web Conference* (Vol. 11137, pp. 239–255). Springer. https://doi.org/10.1007/978-3-030-00668-6_15

Van Daele, K., & Geuens, J. (2022). *Skosprovider* (Version 1.2.0). Zenodo. https://doi.org/10.5281/zenodo.6984378

Van Daele, K., Meganck, L., & Mortier, S. (2015). Data-driven systems and system-driven data: The story of the Flanders Heritage Inventory (1995-2015). *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *II-5/W3*, 323–327. https://doi.org/10.5194/isprsannals-II-5-W3-323-2015

Verborgh, R., Vander Sande, M., Hartig, O., Van Herwegen, J., De Vocht, L., De Meester, B., Haesendonck, G., & Colpaert, P. (2016). Triple Pattern Fragments: A low-cost knowledge graph interface for the Web. *Journal of Web Semantics*, *37–38*, 184–206. https://doi.org/10.1016/j.websem.2016.03.003