alj

The GESAH Graphic Arts Ontology (GESAH GAO) and its reference application – concept and implementation

Tatiana Walther, Birte Rubach, Georgy Litvinov and Christian Hauschke

The GESAH Graphic Arts Ontology (GESAH GAO) was developed to describe works of art on paper such as prints and drawings. The design was guided by the content-related, specialized needs of the domain experts, as well as the requirement to record the data in a structured manner in order to enable subsequent use and to ensure technical connectivity to current and future developments. The ontology models cultural objects by means of activities related to them (creation, production, inscription, preservation, exhibition etc.), people/organizations and their roles, subjects depicted and other relevant concepts. It incorporates concepts from upper ontologies such as the Basic Formal Ontology (BFO) and reuses classes and properties from ontologies such as Friend of a Friend (FOAF), VIVO, and Simple Knowledge Organization System (SKOS). With TIB SAH digital, we show the reference implementation of the GESAH GAO. It was constructed using the open-source knowledge graph suite Vitro, which provides custom entry forms for cataloguing and public access to the digitized collection.

Introduction

The digital representation of cultural heritage collections provides new access for researchers and the general public alike. This is because digitized images can be viewed at any time and from any location via mobile devices. The use of structured metadata enables users to rapidly discover whole collections or individual objects in detail and in depth. This is in stark contrast to the conventional perception and use of collections. The precarious and valuable holdings require special protection for good reasons, but this means that the originals are difficult and sometimes impossible for the public to access. Moreover, only a select few individuals are granted the opportunity to access and explore them. The digitization of the collections is one aspect of the process; the findability of the digital objects is another. The establishment and application of standards is driving semantic networking in the digital world forward and is also indispensable for the provision and dissemination of knowledge about our cultural heritage. The emergence of the Semantic Web and Linked Open Data offers unique opportunities to achieve this comprehensive interconnectivity. The utilization of identifiers to reference all entities (persons, processes, objects, etc.) within a corpus and the application of standardized ontologies to represent the links between these entities and their properties, enables the creation of knowledge graphs that provide their users with a multitude of possibilities for querying, reusing, and the discovering of known or previously unknown information and cultural goods.

In this paper, we introduce the GESAH Graphic Arts Ontology (GESAH GAO), including its content, metrics, and implementation. Created during the

DFG-funded GESAH project (2019–2023),[1] the GESAH GAO was designed in the Web Ontology Language (OWL)[2] to contribute to the knowledge graph with the digital collection of single sheet prints and drawings from the Albrecht Haupt Collection. The OWL is one of the technologies recommended by the W3C for handling data on the Web.[3] It incorporates elements from the Resource Description Framework (RDF) and RDF Schema (RDFS). However, in comparison with these standards, OWL offers more functions in terms of expressivity. It allows defining more complex classes (e.g. owl:unionOf) and types of properties (e.g. owl:FunctionalProperty). The axioms can be interpreted by reasoners, enabling the deduction of additional information from the data described by the ontology. Moreover, it provides methods for matching or mismatching classes (owl:equivalentClass/owl:disjointWith), properties (owl:equivalentProperty) and individuals (owl:sameAs/owl:differentFrom) from different ontologies and datasets, thereby enhancing the interoperability of data described by OWL.[4]

The Albrecht Haupt Collection

The Albrecht Haupt Collection is a historic collection of national and international importance in the fields of architecture, art history, garden history and others. It is owned by the TIB – Leibniz Information Center for Science and Technology (TIB) in Hanover, Germany.[5] In 1901, architect and university lecturer Albrecht Haupt (1852–1932) bequeathed his extensive private collection of books, prints, drawings and travel sketches to what was then the library of Hanover College of Technology. With a passion for travel and collecting, Haupt devoted more than 25 years of his life compiling this collection.[6] In the GESAH project, our initial objective was to conduct an investigation into only one subset of the collection such as the individual graphic sheets (Grafische Einzelblätter der Sammlung des Architekten Albrecht Haupt), which had rarely been studied prior to the start of the GESAH project.

With just over 6,200 sheets, this relatively compact collection of prints and drawings dating from the sixteenth to the nineteenth centuries is characterized by its remarkable thematic range. While architecture and figurative drawings predominate within the approximately 1,000 hand drawings, the collection of prints presents itself as a panopticon of all genres and themes. In addition to architectural prints, the figurative prints of the great and small early European masters are represented as well as extensive documentation of wall and ceiling paintings of the Baroque era or neoclassical interiors. The collection also includes ornaments, portraits, vedute and, from the field of applied graphics, bookplates, seals, calligraphies, initials and much more. The drawings and prints also offer a comprehensive insight into European architecture and art as well as arts and crafts in chronological and geographical terms. They date from the end of the fifteenth century to the end of the nineteenth century and were created in Germany, Italy, France, the Netherlands, Flanders, Switzerland, Great Britain, Spain and – in a few cases – Russia. The collection includes a wide variety of drawing and printing techniques of different times. In addition to the obvious pleasure Albrecht Haupt derived from his collection, he employed it in two distinct ways: On the one hand, he utilized the material for his studies in preparation for his own creations in architecture, interior design or in restoration projects he carried out. On the other hand, he also utilized it for teaching purposes. Haupt was convinced that an understanding of the past was not only the best foundation for creative invention, but also essential. This position was highly controversial at the time and was seen differently by his contemporaries with more modern views (to make a long story short). Nevertheless, thanks to Haupt's conservatism, we are able to appreciate the works of art today and still learn from and about them.

The Ontology

The GESAH GAO was developed with the objective of adequately cataloguing the Albrecht Haupt Collection and publishing it on the World Wide Web. It enables the description of the complex process of creating a work of art and events related to it in the past or in the future.

1. For more information on the project that received follow-up funding, see the homepage of "GESAH+", accessed 4 July 2024, https://projects.tib.eu/haupt and "DFG-Gepris", accessed 4 July 2024, https://gepris.dfg.de/gepris/projekt/424739761?context=projekt&task=showDetail&id=424739761&.
2. "W3C Recommendation 11 December 2012, OWL 2 Web Ontology Language Document Overview (Second Edition)", accessed 4 July 2024, https://www.w3.org/TR/owl2-overview.
3. "Web Ontology Language (OWL)", accessed 4 July 2024, https://www.w3.org/2001/sw/wiki/OWL.
4. "OWL Features", accessed 4 July 2024, https://www.w3.org/TR/owl-features/.
5. Markus Jager and Simon Paulus, eds., Architekturzeichnungen der Sammlung Albrecht Haupt, (Petersberg: Imhof, 2023).
6. Hedda Saemann and Susanne Nicolai, "Albrecht Haupt als Sammler und Architekt," in Herrenhausen und Europa. Ein Gartennetzwerk. Druckgrafik aus der Sammlung Albrecht Haupt (exhibition catalogue Hanover), ed. Andreas Urban (Hanover 2019): 22–27, https://doi.org/10.15488/5593.

Related Works

One of the established reference models in Digital Humanities for describing works of art is CIDOC Conceptual Reference Model (CRM).[7] CIDOC CRM (and its owl-variant the Erlangen CRM[8]) were too extensive to be used in a Vitro-based application.[9] They serve to describe processes and manifestations from the beginning of the existence of an entity to its end, which are researched in very different disciplines, including archeology, musicology, theatre, art and architecture, and others. As a result, they are very large and complex.

Another data model is the Europeana Data Model,[10] which does not offer the required complexity for the event centric approach and primarily employs Dublin Core properties to describe agents participating in events, including properties such as e.g. dc:creator and dc:contributor. Dublin Core[11] is not based on OWL and therefore a direct reuse of the components of the Europeana Data Model would not match the requirements of a Vitro-based application.

Ontology scope

At the beginning of the ontology development process, competency questions were created. These questions shape the scope of the ontology and facilitate its evaluation. The following questions illustrate various approaches to the corpus. The ontology development is considered successful if the questions can be answered by data described by the ontology.

Selection of competency questions:

1. Which German prints from before 1600 are part of the collection?
2. Which Italian architectural drawings can be found in the collection?
3. Which prints were issued in Antwerp?
4. Which drawings were formerly attributed to Parmigianino?
5. Which drawings and prints have a collector's mark?
6. Which drawings and prints were kept in the Collection of Hippolyte Destailleur (Paris)?
7. Which drawings were kept under the former/current inventory no. (e.g. Albrecht Haupt's folder numbered e.g. 'XXXIIIa')?
8. Which drawings and prints depict the subject 'Adam and Eve'?

Therefore, the ontology is designed to serve the inventory, classification, and structured description of works on paper. It should provide a means of describing changing attributions to artists, complex events (such as the multi-phase creation of a work of art or its provenance history), as well as of recording inscriptions, collector's marks and other features in a structured manner. The ontology should facilitate easy access to the material for the public, enable scholarly indexing by the specialist community and curatorial documentation by the preserving institution. To define structure, fields, and relations our domain experts made extensive use of Knaus et al. 2019 (LIDO-Handbuch für Graphik)[12] and the so-called "Gemeinsamer Feldkatalog".[13]

Ontology metrics

The ontology was developed using the open-source ontology editor Protégé.[14] Its design was conducted in an iterative process involving continuous input from domain experts, and it was constructed in OWL DL.[15] The ontology comprises 273 declaration axioms and contains 80 named classes, 131 properties, 47 data properties, and 21 annotation properties. It includes incorporated classes and properties from other ontologies and vocabularies. Complex class expressions such as domain unions were used for the domain definition of several object and data type properties.

The Basic Formal Ontology (BFO)[16] was employed as an upper-level ontology, with several of its classes incorporated to represent fundamental concepts such as entity, role, quality, and so forth. The Friend of a Friend Ontology (FOAF)[17] was reused to describe persons, organizations, and groups of people. Additionally, the class skos:Concept from the Simple Knowledge Organization System (SKOS) vocabulary[18] is widely utilized in the GESAH Ontology as well. Particular basic classes like vivo:Relationship, vivo:DateTimeValue, vivo:DateTimeInterval

7. "CIDOC CRM", accessed 4 July 2024, https://www.cidoc-crm.org/.
8. "Erlangen CRM", accessed 4 July 2024, http://erlangen-crm.org/.
9. Cf. paragraph on "Implementation" below.

10. Cf. "Europeana Data Model Primer", 2013, p. 17, accessed 4 July 2024, https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf.
11. "Dublin Core", accessed 4 July 2024, https://www.w3.org/wiki/Good_Ontologies.

12. Gudrun Knaus, Regine Stein and Angela Kailus, *LIDO-Handbuch für die Erfassung und Publikation von Metadaten zu kulturellen Objekten, Band 1: Graphik*, (Heidelberg: arthistoricum.net, 2019), https://doi.org/10.11588/arthistoricum.382.544. Cf. https://cidoc.mini.icom.museum/working-groups/lido/lido-overview/lido-schema/.
13. The "Gemeinsamer Feldkatalog" was a list of fields that should be filled in when indexing art on paper. The list was agreed on by the community and made available via *AK Graphik vernetzt* on *arthistoricum.net* (https://www.arthistoricum.net/netzwerke/graphik-vernetzt). After the publishing of the more extensive LIDO manual (cf. note 12), the

including related properties such as the object property vivo:dateTimeInterval were integrated from the VIVO Ontology.[19] This allowed us to reuse the existing structure of forms and display configurations based on these classes for our purposes.

It should be noted that a significant share of the relationships between cultural objects have equivalences in the Lightweight Information Describing Objects (LIDO) Terminology,[20] although these are not declared as equivalent via axioms. All possible references to corresponding entities in LIDO, CIDOC CRM, Art & Architecture Thesaurus[21], ICONCLASS[22] and other vocabularies were integrated via rdfs:seeAlso.

Application of the GESAH GAO on the Albrecht Haupt Collection

The GESAH GAO can be employed to describe works on paper. These are not necessarily limited to drawings and prints; due to the composition of the Albrecht Haupt Collection, other classes such as photomechanical prints, printed (text) materials, manuscripts, musical notations and book illumination are also covered.[23] We differentiate between the work (cultural object) and its digital representation (digital image), which are described separately and linked to each other.

A cultural object is described using various relationships and properties. Firstly, it is identified by its title and inventory number. If available, catalogue raisonné entries and/or authority data of works of art assist with unambiguous identification. The works are classified by object type, geographical and stylistic categorization and assignment to an object category.

Other features that serve to precisely describe the object are the state of the printing plate or the edition, dimensions, watermarks, collector's marks and the state of preservation. Any documentation of the work is included in the fields on secondary literature and exhibitions. Eventually, the current holding institution and location are listed, as well as the licenses under which the digitized image and the metadata are made available.

Titles, dimensions, watermarks, collector's marks, catalogue raisonné entries, exhibitions and licenses are complex description objects consisting of context nodes. This structure allows us not only to store the metadata as texts or simple relationships, but also to classify them.

Events in the history of a work are modelled as activities, from its creation, through changes of ownership (provenance), to current preservation measures. For instance, a print is first conceived as a design during the act of creation (fig. 1) before it is engraved into the printing plate during the production activity, and subsequently issued by a publisher (edition activity). These activities also encompass the used techniques and materials. We understand an activity as a specific action occurring at a particular time and place, carried out by a person or organization occupying a specific role at that moment, the type of which is to be determined. The nature of the attribution to this person or organization can be expressed by the attribution type, e.g. *traditional*, *attributed to* or *rejected*. This is important in cases where there is uncertainty as to the identity of the artist or in cases where the objective is to register both former and current states of research. Therefore, several roles can be linked to the same creation or production activity with different attribution types.

A distinct feature of the GESAH GAO is the fact that inscriptions on the works themselves as well as on their mounts and backing boards can also be recorded in a structured form, namely as an activity (fig. 2). This means that the inscriptions can be assigned to individuals and located in time. Furthermore, they can also be differentiated according to their position and classified in terms of content by assigning attributes.

The inscriptions on the works provide valuable information about the history of the collection. Old inventory numbers of previous owners, lot numbers from auctions, attributions, dates or identifications of the pictorial subjects were noted on the sheets. These details reveal much about the provenance of the works and also about the history of their research. Handwritings can be attributed to individuals (fig. 3). Not only Albrecht Haupt himself and the long-time curator Lieselotte Vossnack, but also some prominent visitors to the collection, such as Heinrich Geissler (1927–1990), noted and even signed attributions made on the mounts.[24] In combination with archive material, the structured digital recording of

catalogue was removed from the site.

14. "Protégé", accessed 4 July 2024, https://protege.stanford.edu/.

15. "OWL DL", accessed 4 July 2024, https://www.w3.org/TR/owl-features/.

16. "Basic Formal Ontology (BFO)", accessed 4 July 2024, https://basic-formal-ontology.org/. Cf. Robert Arp, Barry Smith and Andrew D. Spear, *Building Ontologies with Basic Formal Ontology* (Cambridge, Mass.: MIT Press 2015).

17. "Friend of a Friend Ontology (FOAF)", accessed 4 July 2024, http://xmlns.com/foaf/spec/.

18. "Simple Knowledge Organization System (SKOS)", accessed 4 July 2024, https://www.w3.org/2004/02/skos/.

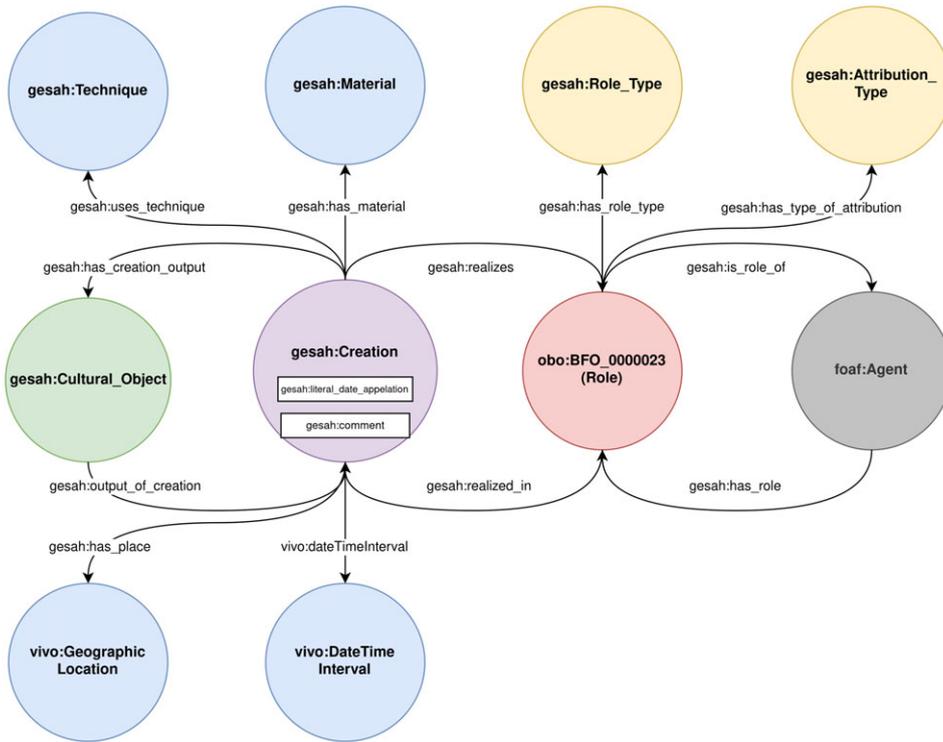19. "VIVO Ontology", accessed 4 July 2024, https://github.com/vivo-ontologies/vivo-ontology. Cf. Jon Corson-Rikert, Stella Mitchell, Brian Lowe, Nicholas Rejack, Ying Ding and Chun Guo, "The VIVO Ontology", in *VIVO,* eds. Katy Börner, Michael Conlon, Jon Corson-Rikert and Ying Ding (Heidelberg: Springer International Publishing, 2012), 15–33, https://d10.1007/978-3-031-79435-3_2.

20. "LIDO Terminology", accessed 4 July 2024, https://xtree-public.digicult-verbund.de/vocnet/?uriVocItem=http://terminology.lido-schema.org/&startNode=lido00409&lang=en&d=n#.

21. "Art & Architecture Thesaurus", accessed 4 July, https://www.getty.edu/research/tools/vocabularies/aat/.

22. "ICONCLASS", accessed 4 July, https://iconclass.org/.

23. An unique case of a three-dimensional object is a seal, for which an object class has been created, but no genre-specific description modalities, cf. "Siegel", accessed 4 July 2024, https://sah.tib.eu/display/n78456.

Fig. 1. Visualization of the activity *gesah:Creation* with object properties and data properties.

24. Cf. examples in the collection with Geissler's attributions https://sah.tib.eu/gesah/display/n5672, https://sah.tib.eu/gesah/display/n4425 (and comment by A. Haupt), https://sah.tib.eu/gesah/display/n5813, https://sah.tib.eu/gesah/display/n13913, https://sah.tib.eu/gesah/display/n766; for examples of detailed recording cf. https://sah.tib.eu/individual/n3105 or https://sah.tib.eu/individual/n433.
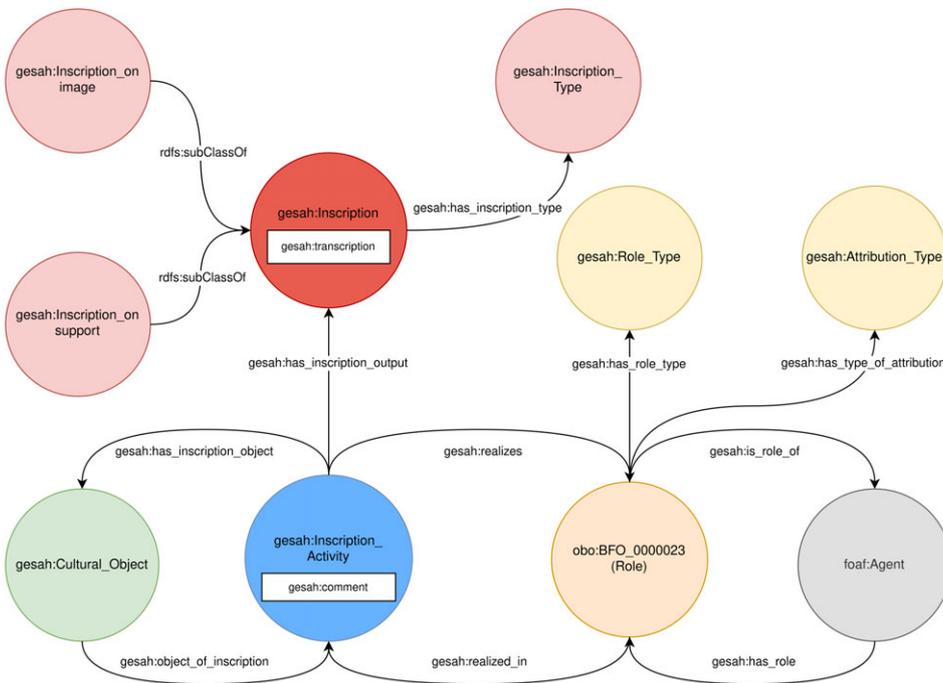
Fig. 2. Visualization of the activity *gesah:Inscription* with object properties and data properties.

this information allows for historical reconstructions of the perception and scientific study of the collection.

The subject of the image is recorded using keywords. In the case of people, places, events and real objects, they are recorded separately and stored in standardized data. For architectural representations, the method of representation is also recorded, such as whether it is an orthogonal or a perspective representation, a view or a plan.

Fig. 3. Oriental King, accompanied by soldiers, pen and wash, heightened with white on paper, inv. kl D Z 63: 7 (https://sah.tib.eu/display/n766) with attribution by Heinrich Geissler to Johann Wolfgang Baumgartner and traditional attribution to Johann Evangelist Holzer by an unknown hand on the drawing itself at lower right, by Albrecht Haupt on the support at left and Lieselotte Vossnack at lower right.

The digital images constitute a distinct class. They are assigned their own titles and are grouped in barcode collections. They are given additional information such as old and new folder numbers (reference to the scanned object in the collection), date of scanning and license. The property is_digital_representation_of is used to establish the relationship between the image and the cultural object data record.

Technical implementation in TIB SAH digital

The GESAH GAO was implemented and utilized in TIB SAH digital.[25] The open-source software Vitro was selected as a software for the implementation. Vitro is a web application with integrated ontology, instance editor and customizable public browsing. Vitro is most prominently used as the core of the community-based research information system VIVO,[26] but it is used in any other areas such as for a philosophical lexicon[27], as a linked-data based cataloguing tool[28] and more.[29] Vitro requires formalized ontologies and vocabularies according to the Web Ontology Language specification. Additionally, it is capable of handling the Simple Knowledge Organisation System.[30]

In this environment, domain experts catalogue drawings, prints and other works manually with the help of specially developed entry forms; they facilitate and speed up the indexing process and also prevent incorrect entries (figs. 4–5). In addition, the public has access to the entire collection, including works that have only been partially catalogued. The collection search function enables users to locate objects by employing both full-text search and predefined filters which are organized into groups for ease of use. Preview images, brief information and a sorting function facilitate the rapid identification of the object searched for or provide a convenient overview.

Reuse

The GESAH GAO has been published on GitHub under the Attribution 4.0 International license and is freely available for reuse.[31] The latest release is version v2024-11-25, with the next release being in planning. The ontology can also be

25. "TIB SAH digital", accessed 4 July 2024, https://sah.tib.eu/.

26. Michael Conlon, Andrew Woods, Graham Triggs, Ralph O'Flinn, Muhammad Javed, Jim Blake, Benjamin Gross et al., "VIVO: A System for Research Discovery", *Journal of Open Source Software. The Open Journal*, 26 July 2019, https://doi.org/10.21105/joss.01182.

27. Georgy Litvinov, "Customized Implementation of VIVO for Working with Philosophical Texts", *VIVO 2021 Conference*, 27 April 2021, accessed 4 July 2024, https://openreview.net/forum?id=1VP6e2gxSjs.

28. Huda Khan, Lynette Rayle and Rebecca Younes, "VitroLib: From an ontology and instance editor to a linked data cataloging editor", *Proceedings of the International Conference on Dublin Core and Metadata Applications*, 2017,
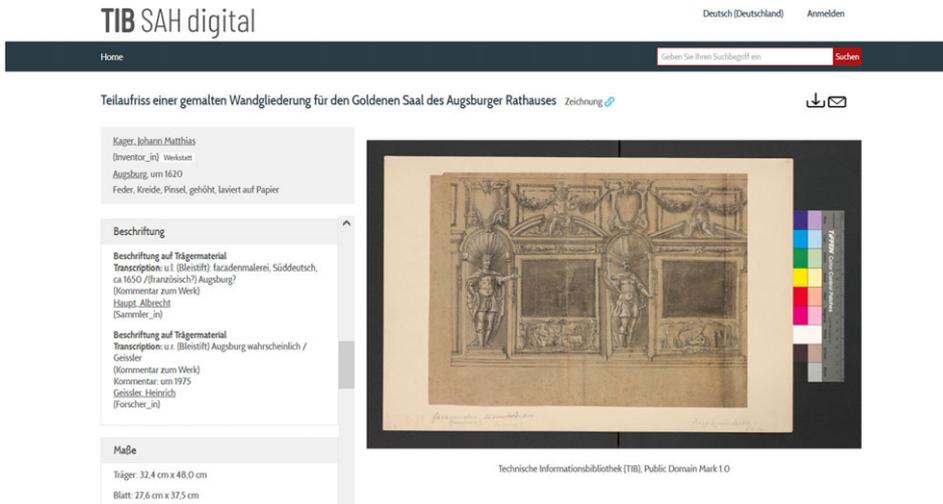
Fig. 4. Screenshot of TIB SAH digital, the profile of the cultural object https://sah.tib.eu/individual/n5672 with detail of an inscription.
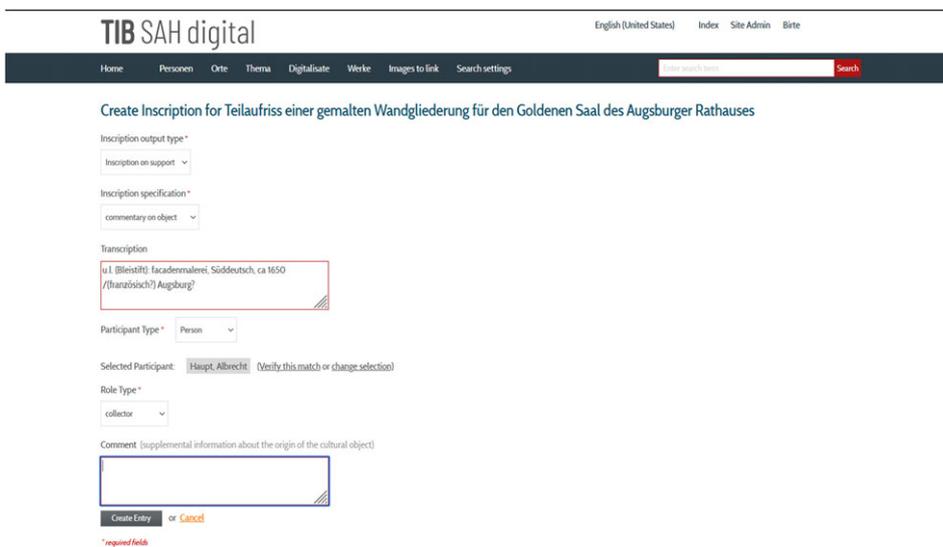

Fig. 5. Screenshot of *TIB SAH digital* with the custom entry form for inscriptions.

found on the TIB Terminology Service.[32] A data export was made possible via the LIDO exchange format to the Graphikportal, hosted by Deutsches Dokumentationszentrum für Kunstgeschichte – Bildarchiv Foto Marburg, which acts as a subject-specific aggregator.[33] This was done in order to foster the dissemination of the collection, the underlying ontology and its software implementation. A mapping to LIDO via XSLT has been created and is released on GitHub,[34] as well as the public endpoint for harvesting.[35]

Conclusion

It can be argued whether the creation of a local ontology should be favoured over the reuse of an existing one. With the GESAH GAO, we have a data model which is tailored to the needs of domain experts and fits the GESAH collection. Although the ontology was created to be used for this specific collection, it can also be reused by organizations with similar collections.

As the ontology is employed in its implementation by TIB SAH digital and is actively utilized, it is undergoing constant further development. Currently the ontology is being expanded in order to include further holdings such as bound graphic collections and 6000 travel sketches by Albrecht Haupt.[36] In addition to the necessary extensions in terms of content, genuine ontological improvements are

https://dl.acm.org/doi/abs/10.5555/3183831.3183854.

29. Christian Hauschke, Tatiana Walther and Graham Triggs, "Vitro – ein universell einsetzbarer Editor für Ontologien und Instanzen", *Informationspraxis*, 4, no. 1 (2018), https://doi.org/10.11588/IP.2018.1.49357.

30. Alistair Miles et al., "SKOS Core: Simple knowledge organisation for the Web", *International Conference on Dublin Core and Metadata Applications*, 3–10, accessed 4 July 2024, http://dcpapers.dublincore.org/pubs/article/download/798/794.

31. GESAH GAO on GitHub: https://github.com/tibonto/gesah.

32. GESAH GAO on TIB Terminology service: https://terminology.tib.eu/ts/ontologies/gesah.

33. "Graphikportal", accessed 4 July 2024, https://www.graphikportal.org/.

34. GESAH to LIDO mapping on GitHub: https://github.com/TIBHannover/gesah-installer/blob/main/gesah-installer/home/src/main/resources/rdf/dynapiAbox/everytime/endpoint_procedure_cultural_object_lido_xslt.n3.

35. Public endpoint for harvesting: https://sah.tib.eu/oai/provider.

36. In order to integrate bound graphic volumes we created a new

also being sought. There are several controlled vocabularies, containing individuals for assigning as role types, attribution types, materials, techniques etc. These controlled vocabularies are going to be published as well. Furthermore, it is conceivable to map the GESAH GAO to CIDOC CRM and Linked Art[37] in the future, thus enabling our flow into the Knowledge Graph of the worldwide Cultural Heritage and to be used in completely different contexts.

Acknowledgments

Funding

Tatiana Walther
Linked Data Librarian, Lab Group Open Research Information
German National Library of Science and Technology (TIB)
Welfengarten 1 B
30167 Hannover
GERMANY
Email: tatiana.walther@tib.eu

Birte Rubach
Art Historian, Research Assistant
Technische Informationsbibliothek (TIB)
Welfengarten 1 B
30167 Hannover
GERMANY
Email: birte.rubach@tib.eu

Georgy Litvinov
Software Developer, Lab Group Open Research Information
Technische Informationsbibliothek (TIB)
Welfengarten 1 B
30167 Hannover
GERMANY
Email: georgy.litvinov@tib.eu

Christian Hauschke
Lead, Lab Group Open Research Information
Technische Informationsbibliothek (TIB)
Welfengarten 1 B
30167 Hannover
GERMANY
Email: christian.hauschke@tib.eu

object class "Composite volumes" and use the object property "has_part"/"is_part_of" to describe the cultural objects belonging to that volume, furthermore with the data property "has_order_number" we can assign a sequence to the objects within the volume. While migrating already existing metadata of travel sketches into our environment we corrected the modelling of the former concept "depicted person" and "depicted place".
37. "Linked Art", accessed 4 July 2024, https://linked.art/.