ZfdG

Zeitschrift für digitale Geisteswissenschaften

Fachartikel aus:
Zeitschrift für digitale Geisteswissenschaften, Heft 11 (2026)

Titel:
Cross-Linking of Vocabularies for Art, Architecture, and Material Culture: Methods, Strategies, Practice, and the AI Perspective

Autor*in:
Ksenia Stanicka-Brzezicka

Kontakt: ksenia.stanicka@uni-marburg.de
Institution: Philipps-Universität Marburg
GND: 132184338 ORCID: 0000-0003-2390-8903
Contribution (CRediT): Conceptualization | Writing – original draft

Weitere Beteiligte:
Sławomir Brzezicki (Herder-Institut für historische Ostmitteleuropaforschung Institut der Leibniz-Gemeinschaft – Data curation | Validation)

DOI des Beitrags:
10.17175/2026_005

Nachweis im OPAC der Herzog August Bibliothek:
1967117071

Erstveröffentlichung:
23.04.2026

Letzte Überprüfung aller Verweise:
07.04.2026

Format:
PDF ohne Paginierung, Lesefassung

GND-Verschlagwortung:
Thesaurus | Künstliche Intelligenz | Kulturerbe | Mapping | Linked Data | Interoperabilität

Empfohlene Zitierweise:
Ksenia Stanicka-Brzezicka: Cross-Linking of Vocabularies for Art, Architecture, and Material Culture: Methods, Strategies, Practice, and the AI Perspective. In: Zeitschrift für digitale Geisteswissenschaften 11 (2026). 23.04.2026. HTML / XML / PDF. DOI: 10.17175/2026_005

Änderungen in Version 1.1 (22.05.2026):
Ergänzung eines Förderhinweises

Ksenia Stanicka-Brzezicka

Cross-Linking of Vocabularies for Art, Architecture, and Material Culture: Methods, Strategies, Practice, and the AI Perspective

Abstract

This article explores methods and challenges of mapping domain-specific vocabularies to international reference thesauri, focusing on the ›Material Culture Thesaurus‹ (MCT) and the ›Getty Art & Architecture Thesaurus‹ (AAT). It addresses semantic, linguistic, and technical issues essential for integrating cultural data into digital research infrastructures. The article discusses mapping strategies, semantic frameworks such as *SKOS*, and editorial workflows for maintaining conceptual integrity. It also reflects on the potential and limits of AI-supported approaches in vocabulary work, advocating for hybrid workflows that combine machine-assisted tools with humanistic expertise.

Der Beitrag untersucht Methoden und Herausforderungen bei der Zuordnung fachspezifischer Vokabulare zu internationalen Referenzthesauri am Beispiel des ›Material Culture Thesaurus‹ (MCT) und des ›Getty Art & Architecture Thesaurus‹ (AAT). Im Zentrum stehen semantische, sprachliche und technische Aspekte der Integration kultureller Daten in digitale Forschungsinfrastrukturen. Diskutiert werden Mapping-Strategien, semantische Modelle wie *SKOS* sowie editorische Verfahren zur Sicherung begrifflicher Konsistenz. Abschließend reflektiert der Text über Potenziale und Grenzen KI-gestützter Verfahren und plädiert für hybride Workflows, die maschinelle Hilfsmittel mit geisteswissenschaftlicher Expertise verbinden.

1. Introduction

1.1 Relevance of Controlled Vocabularies in the Humanities

Specialized controlled vocabularies – understood here as an umbrella term, ranging from simple concept lists to thesauri and ontologies – are essential for capturing the complexity and disciplinary specificity of knowledge in the humanities. They support core tasks such as documentation, research, data enrichment, and information retrieval. Alongside metadata standards and ontologies, controlled vocabularies play a key role in improving the quality, interoperability, and reusability of research and collection data. When interlinked and published in open, machine-actionable formats, they enable broader visibility, semantic integration, and meaningful reuse of data across institutional, disciplinary, and national boundaries. In an increasingly interconnected digital research environment – shaped by developments such as open data platforms, knowledge graph technologies, and cloud-based infrastructures – controlled vocabularies are fundamental building blocks for making cultural and scholarly data findable, accessible, and contextually meaningful on a global scale.[1] [1]

The relevance of controlled vocabularies is also demonstrated by their prominent role in current research infrastructure initiatives, in Germany particularly within several National Research Data Infrastructure (NFDI) consortia, such as NFDI4Culture, NFDI4Objects, and NFDI4Memory, where extensive efforts are being made to advance the integration and use of vocabularies and ontologies. This includes the development of services and tools that facilitate the discovery, alignment, and reuse of knowledge organization systems. Notable examples include [2]

- the Basic Register of Thesauri, Ontologies & Classifications (BARTOC): a global registry for controlled vocabularies,

[1] Cf. Harpring 2010b.

- Cocoda: a web-based application for creating and managing mappings between different thesauri and classification systems,
- and DANTE: the data hub for authority data and terminologies, an online service that facilitates the easy integration of vocabularies into local digital environments.

These tools support both conceptual interoperability and cross-domain knowledge integration, reflecting the growing need for semantic infrastructure in humanities research. Moreover, efforts are increasingly aimed at aligning controlled vocabularies with FAIR[2] and CARE[3] principles, ensuring that they are not only technically interoperable but also culturally and ethically appropriate for diverse research communities. [3]

1.2 The Challenges: Integrating, Interoperability and Reusability of Specialized Vocabularies

At the same time, vocabularies pose complex challenges for information systems – such as data management platforms or databases – due to their extensive scope, variety of formats, and intricate hierarchical structures. For users, the main challenges lie in selecting the appropriate vocabulary and context-relevant terms, as well as in their technical implementation. [4]

Also, integrating subject-specific vocabularies into broader semantic infrastructures represents a critical conceptual and technical challenge in the digital documentation of art, architecture, and material culture. While these vocabularies capture the richness and disciplinary specificity of their domains, they are often developed in specific projects, shaped by national research contexts, and expressed in various languages. This fragmentation, in turn, hampers interoperability and reuse, especially in cross-project scenarios and AI-based translation systems, where such vocabularies could serve as valuable training data.[4] [5]

1.3 Objectives and Central Research Question

This article addresses the conceptual, technical, and practical considerations involved in mapping domain-specific vocabularies to international reference thesauri, with particular emphasis on the Getty Art & Architecture Thesaurus (AAT). It aims to offer methodological insights and guidelines for enhancing the semantic integration of humanities research data, while preserving the essential granularity and contextual specificity of specialized terminologies. Drawing on international standards such as ISO 25964 and semantic frameworks like SKOS, this article explores the formal models and challenges involved in achieving meaningful vocabulary alignment. It should be emphasized that this article focuses on specialized vocabularies, which distinguishes it from the tools mentioned above that operate with different types of vocabularies on a more general level and across various disciplines. [6]

The article also engages with the increasingly relevant role of generative AI models and machine translation systems, though it intentionally does not place this issue at the center. Instead, it argues that before AI can be meaningfully integrated into vocabulary development, there is a need to first clarify and systematize the goals and processes of thesaurus construction – many of which are still evolving or remain inconsistently defined. [7]

At the center of this exploration is the case study of mapping the ›Material Culture Thesaurus‹ (MCT) to the AAT. The Material Culture Thesaurus is a specialized vocabulary for architecture, art and cultural heritage. It is being developed by the Herder Institute for Historical Research on East Central Europe – Institute of the Leibniz Association in Marburg (HI), together with the German Documentation Centre for Art History Foto [8]

[2] FAIR Principles: Findable, Accessible, Interoperable, Reusable. Cf. Kailus 2023.
[3] CARE Principles: Collective Benefit, Authority to Control, Responsibility, Ethics. Cf. Carroll et al. 2020.
[4] Cf. Doerr 2001.

Marburg at the University of Marburg (DDK). The MCT, settled within the interdisciplinary field of material culture studies, offers a highly nuanced and detailed vocabulary capturing the diversity of objects, practices, and contexts central to the study of materiality. The mapping process, conducted between 2021 and 2023 as a subproject within NFDI4Culture[5], with technical support from digiCult[6], aimed to align the MCT with the AAT's overarching semantic model to enhance interoperability while maintaining the disciplinary depth and regionally rooted nuances of the original vocabulary.

From a methodological perspective, the MCT served also as a practical ›laboratory‹ for testing and refining strategies for mapping and integrating domain-specific terminology into the AAT. Crucially, this work was carried out using authentic examples drawn from real-world use cases, ensuring that the findings reflect practical challenges and solutions. This article describes the process, highlighting the issues encountered and the strategies adopted to address them. Above all, it situates these concrete examples within the broader context of multilingual vocabulary management and interoperability. [9]

2. Specialized Vocabularies between Precision and Interoperability

2.1 Cultural and Disciplinary Contexts as Obstacles to Alignment

In disciplines such as art history, archaeology, and material culture studies, terminology is shaped by culturally, historically, and linguistically specific contexts. These systems often evolve within national or regional scholarly traditions and are deeply embedded in local epistemologies. As such, they frequently diverge from internationally standardized reference vocabularies, posing both methodological and conceptual challenges for their integration into broader semantic frameworks. Technical expertise alone is insufficient – effective alignment also requires a nuanced, domain-specific understanding of historical and cultural semantics.[7] [10]

The lack of interlingual equivalence in the terminology used to describe and analyze cultural heritage is not a new problem, but its consequences remain profound. Describing cultural phenomena across languages means engaging with terminological systems rooted in distinct intellectual traditions – including divergent philosophical, religious, and historical assumptions. This challenge is particularly acute in cultural heritage studies, where language not only conveys information but also reflects values, worldviews, and disciplinary norms. [11]

Several key dimensions underscore the difficulty of translating and aligning cultural heritage terminology across languages and knowledge systems. [12]

First, semantic divergence: terms in one language often carry culturally specific connotations that are not directly translatable. These meanings are embedded in local traditions, histories, and values, making literal translation insufficient and potentially misleading. [13]

Second, conceptual asymmetry: some terms refer to phenomena that are culturally unique and have no direct equivalents in other languages. This applies not only to non-European or indigenous knowledge systems – such as the spiritual dimensions of Australian Aboriginal rituals or the philosophical aesthetics of Chinese calligraphy – but also to regionally specific European contexts. A single term in one language might [14]

[5] NFDI4Culture is funded by the German Research Foundation (DFG) – 441958017.
[6] Cf. Lindenthal / Sandrock 2026.
[7] Cf. Mayr / Petras 2008.

cover a broad conceptual range, while its counterpart in another language may be narrower – or vice versa – resulting in either a loss of nuance or problematic generalization. A striking example is the ›Kanzelaltar‹ – a distinctive feature of Lutheran church architecture in which pulpit and altar are physically and symbolically integrated to express the centrality of the Word in liturgical practice. For terms such as ›Kanzelaltar‹, which are deeply rooted in specific regional or confessional traditions, the question arises: should the term be translated? And if so, how? (Fig. 1) Or is it more appropriate to retain the original and provide contextual explanation? These are not merely linguistic issues but touch on the broader question of whether certain terms in art-historical or heritage-related discourse are, in fact, ›untranslatable‹. The lack of conceptual equivalence can lead not only to misunderstanding and misinterpretation but also to the erosion of precision in the description and communication of cultural content.

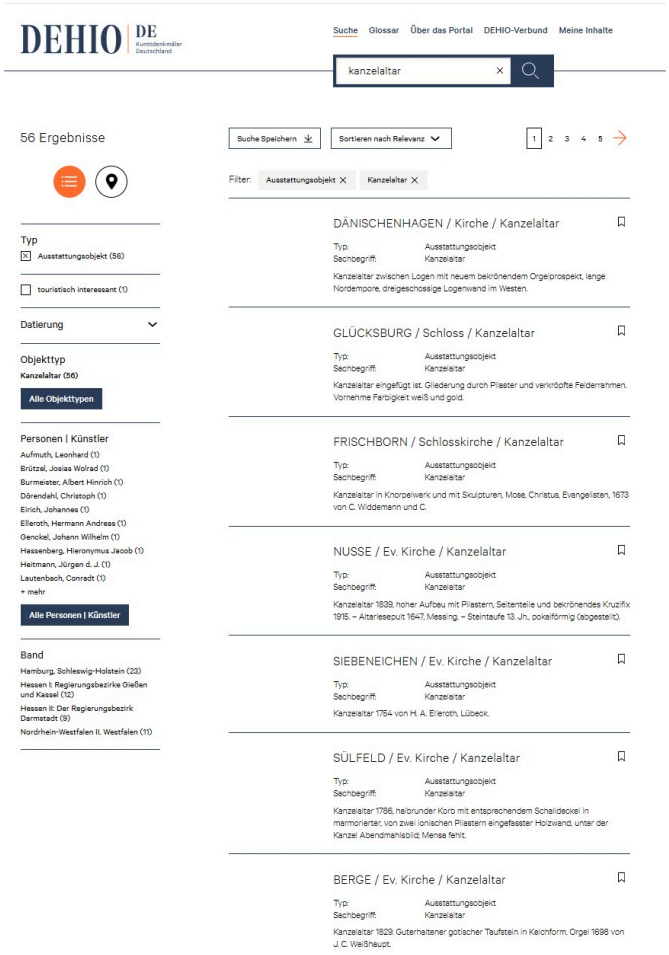


Figure 1: Search results for the term ›Kanzelaltar‹ in DEHIO Germany. [Screenshot: Ksenia Stanicka-Brzezicka 2025]

Third, technical and professional vocabularies – such as those used in heritage conservation, museum documentation, or art history – pose challenges of their own. These terminologies are often shaped by national scholarly traditions, institutional practices, and disciplinary norms. Even within a single language, such as German, significant terminological variation can exist between academic communities or institutions. International standardization efforts in digital infrastructures or collaborative documentation projects often struggle to accommodate these semantic and conceptual inconsistencies. [15]

A particularly thorny issue lies in the high degree of specificity that many local terms possess. This precision, while essential for scholarly accuracy, frequently hinders interoperability. The Polish term ›hamburka‹ offers a telling example: it refers to a specific type of upholstered armchair popular in Central Europe in the late 19th and early 20th centuries. Although etymologically derived from ›Hamburg‹, the term has no direct [16]

historical or material connection to the city. Instead, ›hamburka‹ is deeply rooted in regional craftsmanship, domestic typologies, and culturally specific notions of home life. No direct equivalent exists in international vocabularies like the AAT, making semantic mapping difficult without erasing cultural specificity.

Similar problems arise in the classification of regionally distinct architectural phenomena, such as Protestant churches in Silesia and Lusatia. Terms like ›border churches‹, ›Churches of Peace‹, ›refuge churches‹, and ›grace churches‹ (Fig. 2) refer not merely to building types but to historically specific responses to religious persecution, political agreements, or legal constraints. For instance, the Churches of Peace in Silesia (Fig. 3) were constructed under strict imperial conditions: outside city walls, within one year, without towers, and using only perishable materials. These constraints produced unique architectural solutions that defy standard typological categories and highlight the limits of generalized classification systems. [17]

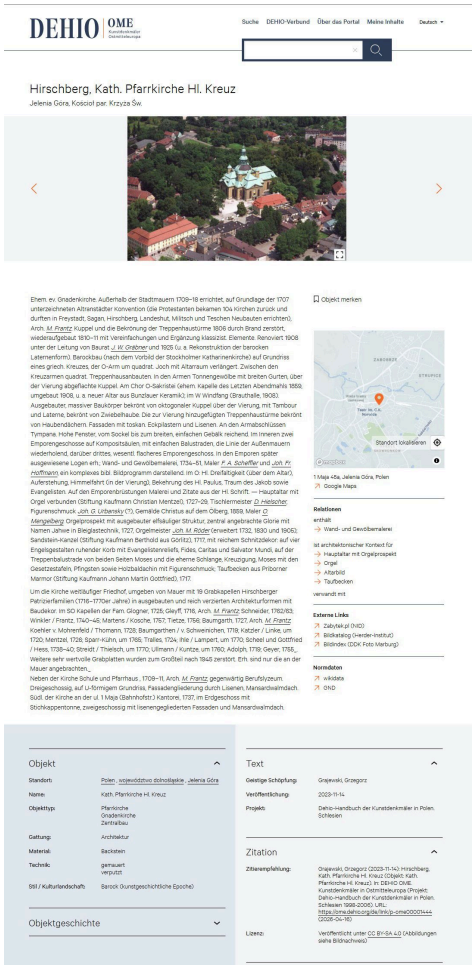
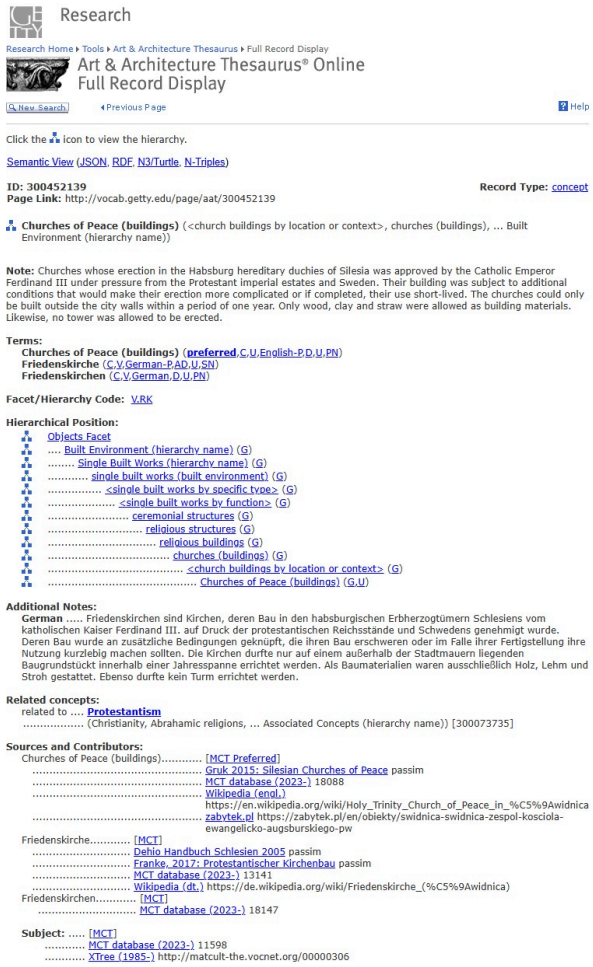


Figure 2: Grace Church in Jelenia Góra (Hirschberg), Poland, dataset from OME. [Screenshot: Ksenia Stanicka-Brzezicka 2025]


Research

Research Home ▸ Tools ▸ Art & Architecture Thesaurus ▸ Full Record Display

Art & Architecture Thesaurus® Online
Full Record Display

New Search | Previous Page | Help

Click the icon to view the hierarchy.

Semantic View (JSON, RDF, N3/Turtle, N-Triples)

ID: 300452139 **Record Type:** concept
Page Link: http://vocab.getty.edu/page/aat/300452139

Churches of Peace (buildings) (<church buildings by location or context>, churches (buildings), ... Built Environment (hierarchy name))

Note: Churches whose erection in the Habsburg hereditary duchies of Silesia was approved by the Catholic Emperor Ferdinand III under pressure from the Protestant imperial estates and Sweden. Their building was subject to additional conditions that would make their erection more complicated or if completed, their use short-lived. The churches could only be built outside the city walls within a period of one year. Only wood, clay and straw were allowed as building materials. Likewise, no tower was allowed to be erected.

Terms:
Churches of Peace (buildings) (**preferred**,C,U,English-P,D,U,PN)
Friedenskirche (C,V,German-P,AD,U,SN)
Friedenskirchen (C,V,German,D,U,PN)

Facet/Hierarchy Code: V.RK

Hierarchical Position:
Objects Facet
.... Built Environment (hierarchy name) (G)
..... Single Built Works (hierarchy name) (G)
..... single built works (built environment) (G)
..... <single built works by specific type> (G)
..... <single built works by function> (G)
..... ceremonial structures (G)
..... religious structures (G)
..... religious buildings (G)
..... churches (buildings) (G)
..... <church buildings by location or context> (G)
..... Churches of Peace (buildings) (G,U)

Additional Notes:
German ..... Friedenskirchen sind Kirchen, deren Bau in den habsburgischen Erbherzogtümern Schlesiens vom katholischen Kaiser Ferdinand III. auf Druck der protestantischen Reichsstände und Schwedens genehmigt wurde. Deren Bau wurde an zusätzliche Bedingungen geknüpft, die ihren Bau erschweren oder im Falle ihrer Fertigstellung ihre Nutzung kurzlebig machen sollten. Die Kirchen durfte nur auf einem außerhalb der Stadtmauern liegenden Baugrundstückt innerhalb einer Jahresspanne errichtet werden. Als Baumaterialien waren ausschließlich Holz, Lehm und Stroh gestattet. Ebenso durfte kein Turm errichtet werden.

Related concepts:
related to .... **Protestantism**
..... (Christianity, Abrahamic religions, ... Associated Concepts (hierarchy name)) [300073735]

Sources and Contributors:
Churches of Peace (buildings)..... [MCT Preferred]
..... Gruk 2015: Silesian Churches of Peace passim
..... MCT database (2023-) 18088
..... Wikipedia (engl.) https://en.wikipedia.org/wiki/Holy_Trinity_Church_of_Peace_in_%C5%9Awidnica
..... zabytek.pl https://zabytek.pl/en/obiekty/swidnica-swidnica-zespol-kosciola-ewangelicko-augsburskiego-pw
Friedenskirche..... [MCT]
..... Dehio Handbuch Schlesien 2005 passim
..... Franke, 2017: Protestantischer Kirchenbau passim
..... MCT database (2023-) 13141
..... Wikipedia (dt.) https://de.wikipedia.org/wiki/Friedenskirche_(%C5%9Awidnica)
Friedenskirchen..... [MCT]
..... MCT database (2023-) 18147

Subject: ..... [MCT]
..... MCT database (2023-) 11598
..... XTree (1985-) http://matcult-the.vocnet.org/00000306

Figure 3: The term ›Church of Peace‹ (Gnadenkirche), AAT. [Screenshot: Ksenia Stanicka-Brzezicka 2025]

Such examples underscore a key tension in cultural heritage data modeling: how to retain the semantic richness of localized, historically grounded knowledge while ensuring compatibility with structured, machine-readable systems. This tension becomes particularly salient in the context of digital transformation, where humanities research increasingly relies on computational tools and interoperable infrastructures. Given these dynamics, maintaining conceptual discipline in terminology is essential. Controlled vocabularies play a dual role: they serve not only to label, classify, and retrieve information, but also to mediate between different knowledge systems. However, for them to fulfil this role effectively, they must accommodate cultural and linguistic diversity while adhering to the logic of structured data. This calls for flexible, multilingual approaches to vocabulary development and mapping – approaches that respect disciplinary depth without sacrificing interoperability. [18]

2.2 Typologies of Domain-Specific Vocabularies: National, Project-Based, Thematic

Domain-specific vocabularies in the humanities arise from diverse institutional, methodological, and linguistic settings and may, in broad terms, be described as national, project-based, or thematic in nature. National vocabularies are often the result of coordinated efforts by cultural heritage institutions or academic bodies to provide a unified terminology for a country's collections and documentation systems. A prominent example is the Tezaurus Dziedzictwa Kulturowego (Thesaurus of Cultural Heritage), developed by the University of Wrocław, which offers a standardized, Polish-language controlled vocabulary tailored to the needs of museums, libraries, and archives across Poland. In contrast, project-based vocabularies often arise from specific research initiatives with defined goals and temporal scopes. These vocabularies reflect the conceptual [19]

focus and disciplinary requirements of a given project. The MCT, rooted in art history and material culture studies, builds on the long-standing documentary practice of the DDK and was further developed within the DEHIO-OME project to address the need for a nuanced vocabulary describing architectural and object-related phenomena in East Central Europe. A third category, thematic vocabularies, is organized around a specific subject area that may cut across institutional or national boundaries. The British Museum Materials Thesaurus, for example, focuses on substances and material properties relevant to museum collections, conservation science, and cultural heritage documentation.

2.3 Extending Instead of Building Anew: Making Better Use of Reference Vocabularies

[20] Collections and research projects rely on controlled vocabularies to structure and describe their data. While there is a general openness to using established reference vocabularies, these often lack the necessary granularity to represent domain-specific concepts in sufficient detail. This frequently leads to the idea of developing a custom vocabulary, especially since tools for building and managing vocabularies have become increasingly accessible. However, practice and experience have shown that it is often more effective to extend and refine existing vocabularies rather than starting from scratch. The ongoing development of the MCT for its application in projects such as DEHIO provides a particularly illustrative example of this approach.

[21] Connecting to and building upon reference vocabularies offers clear advantages. It reinforces all of the aforementioned aspects – semantic interoperability, standardization, and the leveraging of existing community efforts – thereby situating individual datasets within a broader disciplinary and infrastructural framework. As a result, compatibility with other systems and datasets is ensured – a key prerequisite for collaborative research, sustainable data integration, and the development of interoperable infrastructures.

3. Conceptual and Technical Foundations of Thesaurus Mapping

3.1 Conceptual Grounding and Mapping Challenges

[22] Mapping controlled vocabularies is not a purely technical task – conceptual decisions fundamentally shape it. Each vocabulary embodies specific aims, disciplinary assumptions, and structuring principles. Vocabularies may differ significantly depending on context: whether they are intended for cataloguing or research, whether they operate on a concrete or abstract level, how they organize knowledge – by form, function, material, or historical period – and whether they are embedded in particular cultural or linguistic traditions. These differences influence not only the content of a vocabulary but also the logic by which it structures and relates terms. As a result, mapping between vocabularies requires more than algorithmic matching; it demands interpretive judgment and a deep understanding of both the source and target domains.

[23] Because of structural and semantic divergences, mapping vocabularies requires more than matching similar labels. Terms must be analyzed in context, including their position within hierarchies, their conceptual scope, and their relations to other terms. A common challenge is that vocabularies may categorize differently – for example, by function versus material – or vary in granularity. Such mismatches demand critical examination during the mapping process. Concept-based mapping helps ensure that alignments are semantically accurate and not misleading.

3.2 Semantic Relationship Types, SKOS Mapping Properties, and Technical Standards

[24] To effectively model and implement mappings, a typology of semantic relationships is essential. The Simple Knowledge Organization System (SKOS) provides a well-established standard framework for expressing these relationships in RDF format. SKOS is not only a technical serialization format, but also a conceptual model that defines core types of semantic relationships relevant for mapping. This dual role enables the publication and integration of vocabularies as *Linked Open Data*. SKOS allows vocabularies to be machine-readable, interoperable, and linkable across domains. Linked Open Data is a paradigm for publishing structured data using *persistent URIs*, with vocabularies like the Getty AAT or collaboratively curated, multilingual Wikidata serving as central reference points for mapping. By linking local terms to such authorities via SKOS mapping properties, institutions can enhance semantic interoperability, enable cross-dataset discovery, and situate their vocabularies within a broader knowledge graph.

[25] The most important categories within SKOS are hierarchical relations. These describe broader and narrower concept structures and allow for representing differences in levels of abstraction. For example, *skos:broadMatch* identifies a concept that is broader than the local concept. An example is the AAT term ›blankets (coverings)‹ which is broader than the German term ›Bettdecke‹, which refers specifically to bed coverings. Conversely, *skos:narrowMatch* indicates a more specific, but not fully equivalent, concept – such as mapping ›Postgebäude‹ (a general category for postal buildings) to ›post offices‹, which refers narrowly to service-oriented postal facilities. Such mappings are particularly sensitive to differences in granularity or underlying classification systems. Careful analysis is required to avoid incorrect generalizations or omissions.

[26] SKOS also includes equivalence relations that indicate semantic identity or high similarity. The relation *skos:exactMatch* is used when two concepts are functionally and semantically identical – for example, ›metropolitan areas‹ and ›Metropolregion(en)‹, provided that definition and scope fully coincide in both vocabularies. In practice, however, this relation should be applied cautiously. More commonly, *skos:closeMatch* is used when concepts are very similar but not fully interchangeable. For instance, ›Aquarell‹ closely corresponds to the English ›watercolor‹, although differences in usage, definition, or linguistic nuance may exist.

[27] Furthermore, SKOS recognizes associative relations, which denote thematic but non-hierarchical links between concepts. *skos:relatedMatch* indicates a contextual or semantic connection without implying equivalence or hierarchy. An example would be the relationship between ›Protestantism‹ and ›Churches of Peace‹, which reflects a historical and thematic link without a structural connection. These relations can facilitate content discovery and navigation, but should be applied selectively to avoid diluting conceptual precision. (Fig. 4)


Research

Research Home › Tools › Art & Architecture Thesaurus › Full Record Display

Art & Architecture Thesaurus® Online
Full Record Display

New Search | Previous Page | Help

Click the icon to view the hierarchy.

Semantic View (JSON, RDF, N3/Turtle, N-Triples)

ID: 300073735 **Record Type:** concept
Page Link: http://vocab.getty.edu/page/aat/300073735

Protestantism (Christianity, Abrahamic religions, ... Associated Concepts (hierarchy name))

Note: The general term for types of Christian faith originating from the Reformation. Although the early forms of Protestantism were those who followed Luther, Calvin, and Zwingli, the term now includes most non-Roman Catholic or non-Orthodox denominations. Protestants want to be closer to the style of faith of the early Church which they feel has been obscured in Catholic practices. The term derives from the word 'protestari' which means not only to protest but to avow or confess. Common characteristics of Protestantism include the justification by faith alone, the authority of scripture, and the priesthood of all believers, in which not only the clergy are able to hear the confession of sin.

Terms:
Protestantism (**preferred**,C,U,LC,English-P,D,U,N)
Protestant (C,U,English,AD,U,N)
新教 (C,U,Chinese (traditional)-P,D,U,U)
抗羅宗 (C,U,Chinese (traditional),UF,U,U)
xin jiao (C,U,Chinese (transliterated Hanyu Pinyin)-P,UF,U,U)
xin jiao (C,U,Chinese (transliterated Pinyin without tones)-P,UF,U,U)
hsin chiao (C,U,Chinese (transliterated Wade-Giles)-P,UF,U,U)
protestantisme (C,U,Dutch-P,D,U,U)
protestant (C,U,Dutch-AD,U,U)
protestantismo (C,U,Spanish-P,D,U,U)

Facet/Hierarchy Code: B,BM

Hierarchical Position:
Associated Concepts Facet
.... Associated Concepts (hierarchy name) (G)
..... <religions and religious concepts> (G)
..... religions (belief systems, cultures) (G)
..... Abrahamic religions (G)
..... Christianity (G)
..... Protestantism (G)

Additional Notes:
Chinese (traditional) ..... 泛稱源自宗教改革的各式基督信仰。雖然早期的新教形式是指遵循路德(Luther)、喀爾文(Calvin)、以及茲溫利(Zwingli)的派別，但現在新教通常涵蓋包含了大部分非羅馬天主教或非正教的派別。新教徒希望能更接近早期教會信仰，認為天主教將之窒得等同難解。本詞衍生自"protestari"，不僅意指抗議，更要坦白或懺悔。新教的常見特色包括因純信而稱義、經典的權威性、信徒皆牧師、所以不是只有牧師才能聽取赦罪懺悔。
Dutch ..... De algemene term voor typen christelijk geloof die zijn ontstaan tijdens de Reformatie. Hoewel de vroegste vormen van protestantisme werden beoefend door de volgelingen van Luther, Calvijn en Zwingli, wordt de term tegenwoordig gebruikt voor de meeste overtuigingen die niet rooms-katholiek of orthodox zijn. Protestanten willen dichter bij de geloofsstijl van de vroege kerk staan, die volgens hen in de katholieke praktijk overschaduwd is geraakt. De term is afkomstig van het woord 'protestari', wat niet alleen 'protesteren' betekent, maar ook 'belijden'. Enkele algemene kennerken van het protestantisme zijn: rechtvaardiging door het geloof alleen, de berusting van alle gezag bij de Bijbel en de leer dat alle gelovigen zelf hun zonden aan God kunnen belijden, waarbij het aanhoren van de biecht niet alleen aan priesters is voorbehouden.
Spanish ..... El término general para tipos de fe cristiana originadas a partir de la Reforma. Aunque las formas tempranas del protestantismo fueron aquellas que obedecían a Lutero, Calvino y Zwinglio, el término ahora incluye la mayoría de las denominaciones no Romanas o no Ortodoxas. El caracter común de los protestantes incluye la justificación de la fe por la fe, la autoridad de la sagrada escritura y el sacerdocio de todos los fieles, por el que cada creyente puede oir la confesión de los pecados, no siendo exclusivo del clero. protestantes quieren estar cerca al estilo de fe de la primera Iglesia el cual sienten ha sido opacado en las prácticas católicas. El término deriva desde la palabra "protestari" que significa no sólo protestar sino que reconocer o confesar. Característica común.

Related concepts:
related to .... **border churches (buildings)**
..... (<church buildings by location or context>, churches (buildings), ... Built Environment (hierarchy name)) [300452140]
related to .... **Churches of Peace (buildings)**
..... (<church buildings by location or context>, churches (buildings), ... Built Environment (hierarchy name)) [300452139]
related to .... **Grace Churches (buildings)**
..... (<church buildings by location or context>, churches (buildings), ... Built Environment (hierarchy name)) [300452132]
related to .... **Refuge Churches (buildings)**
..... (<church buildings by location or context>, churches (buildings), ... Built Environment (hierarchy name)) [300452138]

Figure 4: The term ›Protestantism‹ with relations, AAT. [Screenshot: Ksenia Stanicka-Brzezicka 2025]

The further technical foundation relies on the international standard ISO 25964, which defines best practices for thesaurus construction and mapping. It offers guidelines for vocabulary structure, display, and maintenance, and emphasizes the conditions necessary for interoperability with other vocabularies. It also outlines requirements for mapping, such as the inclusion of concept definitions, hierarchical depth, and relation types. [28]

3.3 Transparency, Consistency, Documentation

Standards and best practices are essential for guiding the development of vocabularies. Aligning with them ensures consistent and coherent data representation, which in turn enhances clarity, reduces ambiguity, and improves the understanding and reuse of data. [29]

By adopting existing reference vocabularies, one can leverage the collective expertise of the communities that have developed them. This allows users to work with extensive, curated, and regularly updated terminologies without having to build them from scratch. Reference vocabularies evolve continuously, reflecting advances in knowledge and shifts in terminology conventions. Integrating them into one's data practices helps ensure accuracy, relevance, and semantic richness. They also provide a broader conceptual context by linking related terms and concepts, enabling more nuanced queries and cross-domain analysis. Contributing to and expanding these vocabularies strengthens the underlying semantic network and enhances its functionality, benefitting both individual projects and the wider community. [30]

But developing and maintaining a comprehensive vocabulary requires not only significant time, effort, and expertise but also well-defined workflows, guiding principles, and standardized documentation of decisions. These elements are crucial to ensure transparency, consistency, and the long-term sustainability of the vocabulary. [31]

4. Case Study: Mapping the Material Culture Thesaurus (MCT) to the AAT

Recognizing these problems within NFDI4Culture has led to a measure ›Cross-linking of subject-specific vocabularies with international reference vocabularies‹, which ran from 2021 to 2023 at the Herder Institute for Historical Research on East Central Europe – Institute of the Leibniz Association. [32]

Within the framework of the measure, a mapping of selected concepts from the MCT to the AAT was tested. New concepts were also indexed, submitted to the AAT, and integrated into the thesaurus – all following guidelines of the Getty Vocabulary Program. Contrary to what one might expect, this is not a straightforward task. It requires adapting one's software to support both the export and import of data and workflows that enable simultaneous editing in both vocabularies. [33]

4.1 Background of the MCT: Development, Structure, and Aims

The MCT is the result of an initiative to catalogue the photographic holdings of the DDK, to digitize the Dehio handbooks for Germany (DEHIO-digital) and to prepare new Dehio volumes within the framework of Dehio East Central Europe (DEHIO OME)[8], a project carried out at the HI. Based on their long-standing experience in the documentation of art history and cultural heritage, the MCT is intended to be a multilingual application thesaurus to support the indexing and publication of information on cultural heritage objects, maintained following the ISO 25964 standard. The development is carried out jointly by the two institutions, as tasks of content-related work priorities, extension and quality assurance, in the interest of both partners. Governance, usage contexts, and administrative rights have been defined in an agreement. [34]

As part of the defined project measure, 24 terms were selected from the German-language volumes of the Dehio-Handbuch der Kunstdenkmäler and elaborated specifically for contribution to the Getty Vocabulary Program. These concepts primarily refer to art objects and phenomena found particularly in Central and Eastern Europe. They were indexed using German and English terms, bilingual definitions, and bibliographic references. [35]

An equally important goal of the MCT is to add translations in other national languages, such as Polish, Lithuanian, Latvian, and Estonian, especially as Dehio volumes for these countries are being prepared. This multilingual approach will enhance the accessibility and usability of the thesaurus across different linguistic and cultural contexts. At the same time, it presents an additional challenge for mapping, as conceptual alignment must account for linguistic nuances and culturally specific terminology. [36]

[8] Cf. Brzezicki / Nürnberger 2022.

4.2 Technical Implementation and Workflow

[37] The technical implementation of a thesaurus mapping project requires a structured workflow adapted to the technical architecture of the vocabulary and the organizational or disciplinary context in which it is used. While certain principles apply universally to vocabulary work – such as the requirement that terms be unambiguous – many aspects of cross-linking depend on the concrete use case. Foundational principles must therefore be translated into individual workflows that are closely aligned with the project's objectives, methods, and data models.

[38] A central consideration when designing a workflow is the distinction between editorial processes – such as vocabulary development, maintenance, and mapping – and the parallel application of the vocabulary during indexing. This separation is crucial for determining how the mapped vocabulary is integrated into systems and used effectively in practice. If the vocabulary is to be implemented and used immediately, this has direct consequences for editorial processes – for instance, how term statuses are managed or how new terms are introduced without compromising semantic clarity. Elements such as hierarchical context, scope notes, and homonym qualifiers can be employed to maintain semantic clarity and prevent ambiguity.

[39] Based on our case study experience, several recommendations can be formulated for structuring such processes. First and foremost, a formal mapping plan should be developed. This document outlines the project's goals, scope, timeline, and costs. Furthermore, a thorough understanding of the data model is essential; the structural and semantic design of the thesaurus must be transparent and comprehensible for everyone involved.

[40] In projects of this nature, rights management and access control play a crucial role. It is necessary to define the scope of the collaborators' permissions – whether they are allowed to add, map, or authorize terms – and to implement appropriate technical infrastructure to support these processes. Equally important is the composition of the working group. Ideally, it brings together subject specialists, translators familiar with both source and target languages, and experts who combine linguistic and domain-specific knowledge. This interdisciplinary setup is essential for ensuring the semantic precision of mappings.

[41] Technical requirements for software systems should also be defined early on, including support for *persistent identifiers*, collaborative editing environments, and version control. As in other digital humanities projects, administrative measures are required: setting up the team, clarifying responsibilities, securing licenses, and establishing clear communication channels all contribute to a sustainable project structure. Adherence to general standards, such as the FAIR and CARE principles, as well as formal cooperation agreements and shared platforms, are essential prerequisites.[9]

[42] The practical execution of mapping involves three core processes: validation, matching, and the designation of preferred terms. Term validation refers to the initial identification and curation of relevant concepts from literature, domain expertise, or existing vocabularies. This process includes defining scope notes, clarifying semantic relationships, and organizing terms hierarchically. Matching entails identifying equivalent or related terms across vocabularies. While automated tools and alignment algorithms can provide initial suggestions – especially when leveraging natural language processing techniques – semantic matching ultimately remains a human-driven interpretive task. It requires careful consideration of the meaning, context, and use of each term.[10]

[9] Cf. Lindenthal / Sandrock 2026.
[10] Cf. Stiller et al. 2014; Suominen et al. 2022.

[43] Once equivalent or related terms are identified, mapping establishes the formal semantic relationships between them. These may be equivalence relations (e.g. skos:exactMatch, skos:closeMatch), hierarchical links (e.g. skos:broadMatch, skos:narrowMatch), or associative connections (skos:relatedMatch), depending on the conceptual structure of the vocabularies involved. Determining the preferred term involves selecting the most appropriate term for each concept within the controlled vocabulary, guided by clarity, conciseness, neutrality, and usage frequency. Community acceptance, language conventions, and cultural sensitivities also play a role – especially in multilingual or cross-cultural contexts.

[44] Overall, our experience confirms that thesaurus mapping is not a purely technical process but one that demands conceptual clarity, editorial coordination, and semantic literacy. As a key component of interoperable, well-structured research data infrastructures, it should be approached with the same methodological care as other scholarly activities.

4.3 The Mapping Exercise: Challenges of Semantic, Linguistic, and Editorial Interoperability

[45] The mapping exercise revealed several core challenges. A key complexity arose from the need to align two independently developed hierarchical thesauri. Both the MCT and the AAT feature structured semantic relationships – such as broader and narrower terms – but are based on differing classificatory principles. Reconciling these models required more than lexical matching; it involved the systematic comparison of hierarchical logics, including polyhierarchies and facet-based categorizations. In many instances, one thesaurus exhibited a higher degree of granularity or captured conceptual nuances not represented in the other. These asymmetries often resulted in mappings to broader proxy terms – or, in some cases, in the absence of any suitable match. Such decisions raised important questions about how to preserve semantic richness while enabling interoperability.

[46] Some concepts from the MCT, particularly those embedded in specific historical or cultural contexts, lacked direct equivalents in the AAT. These ›non-match‹ cases exposed the limitations of existing reference thesauri and pointed to the need for their further development or contextual extension. Multilingual scope notes added another layer of complexity. Regional differences in conceptual understanding and language usage often surfaced, requiring both linguistic and disciplinary expertise – especially for culturally embedded terms such as the Polish ›hamburka‹. Editorial workflows had to address these issues through iterative revisions, dual-control mechanisms, and rigorous quality assurance protocols.

[47] A particularly delicate aspect was the translation of scope notes. Here, the central editorial decision concerns the type of translation: literal, free, or pragmatic. Literal translation, while appropriate in technical or scientific contexts, can lead to awkward or misleading results in cultural vocabularies due to structural and idiomatic differences between languages. Free translation, which focuses on conveying meaning rather than form, is generally recommended for scope notes, as it ensures accessibility and readability in the target language. Pragmatic translation allows for additional explanations or culturally adapted formulations, making it particularly useful when the target-language audience operates within a different conceptual tradition. Institutions such as Getty explicitly allow and categorize these translation types to ensure consistency across multilingual environments. Whichever approach is chosen, it must be documented transparently and remain faithful to the syntactic and semantic logic of the target vocabulary – especially in hierarchical placement.

[48] One of the most prominent conceptual challenges is polysemy: the coexistence of multiple meanings within a single term. A typical example is ›church‹, which can denote both a building and an institution. In the AAT, these are distinguished via parenthetical qualifiers (e.g., ›church (building)‹ vs. ›church (organization)‹). Accurate mapping depends on this kind of explicit disambiguation.

Closely related is the issue of semantic granularity. Different domains and languages often operate at varying levels of specificity. While it is technically possible to map a narrower concept to a broader one, this must be documented using SKOS properties (e.g., skos:broadMatch, skos:narrowMatch) to preserve semantic transparency and avoid oversimplification. [49]

Even within expert-controlled vocabularies, idiomatic terms can present difficulties. Expressions like ›flying buttress‹ refer to culturally and historically specific architectural elements that may lack direct equivalents in other languages or be prone to misinterpretation without proper context. Such cases highlight the necessity of domain-specific interpretation throughout the mapping process. [50]

A key challenge in mapping lies in the linguistic standardization of terms across different languages. Variations in grammar, orthography, and stylistic conventions complicate efforts to consistently align vocabularies. Mapping projects must therefore establish clear editorial guidelines that conform to AAT conventions (if mapping on AAT) and are appropriate for the target language. In fusional languages such as German, for example, grammatical gender and inflection introduce additional complexities not anticipated by the ISO standard's assumption of language-neutral preferred labels. [51]

Compound terms, common in German, add another layer of complexity to the standardization process. Forms such as ›Abfertigungshalle‹, ›Brunnenhalle‹, or ›Fahrzeughalle‹ consist of a head (›Halle‹) combined with a modifying element that specifies function or context. Crucially, these compounds are not classified according to the properties of the modifier (Fahrzeug, Brunnen), but according to the head, which determines their categorical and ontological status. The modifier merely restricts or refines the interpretation without altering the fundamental category. Any mapping strategy must account for this head-driven structure in order to avoid systematic misclassification. [52]

Additional technical and editorial challenges arise from diacritics and special characters, which are essential in many languages for grammatical correctness but may interfere with data entry, indexing, and software compatibility. Their use must be carefully standardized to support interoperability while preserving linguistic accuracy. [53]

Finally, the normalization of spelling, punctuation, and capitalization is vital. Inconsistencies – such as ›plein-air‹ versus ›plein air‹ or ›plein air painting‹ – can significantly impact data retrieval and integration. Without robust editorial policies for normalization, even semantically rich vocabularies risk fragmentation and reduced discoverability. [54]

These editorial and technical standards form the groundwork upon which all subsequent mapping and translation decisions must be based. Without a stable and normalized linguistic foundation, even the most carefully considered semantic alignments risk being inconsistent or unintelligible. Once this foundation is in place, the next crucial step is to verify the meaning and usage of each term in both source and target languages to ensure conceptual accuracy and contextual appropriateness. [55]

Before any translation or mapping decision is made, the meaning of a term must be verified in both the source and the target language. This should begin with general dictionaries (for German e.g. Duden, Langenscheidt Großwörterbuch) and continue with domain-specific glossaries and scholarly lexicons. Even within a single language, conceptual overlaps and differences may occur. A good example is the term ›watercolor‹ vs. ›Aquarellfarbe‹: while all Aquarellfarben are watercolors, not all watercolors are understood as Aquarellfarben in German. In English, ›watercolor‹ can refer both to the material (watercolor paint) and to the resulting artwork (watercolor painting). Controlled vocabularies such as the AAT make such distinctions [56]

explicit, listing ›watercolor (paint)‹ under materials and ›watercolors (paintings)‹ as a product or object class. This highlights the importance of investigating each term's semantic range and functional role within its linguistic and disciplinary context.

Digital resources offer valuable support in this process. Wikipedia and Wikidata, for instance, allow for multilingual access to concept definitions. While Wikipedia provides contextual and narrative information, Wikidata offers structured, machine-actionable data. As a language-independent, collaborative knowledge base, Wikidata is increasingly relevant for multilingual terminology work[11]. It allows for the modeling and alignment of concepts across languages and domains. Its openly licensed, exportable content can be linked to other open datasets in the Semantic Web, making it an important hub for the development of multilingual authority data. [57]

Other tools, such as Linguee, also offer contextual insights by showing how terms are used in actual textual environments. While not authoritative, such resources can be helpful for identifying usage patterns and translation conventions in specialized contexts. [58]

Ultimately, the transfer of terms in the context of cultural heritage is less a matter of literal translatability than of hermeneutic mediation. Where semantic or conceptual equivalence cannot be achieved, the aim should not be substitution but explanation – making difference visible and comprehensible without flattening cultural specificity. In this sense, terminological work in the cultural domain is not merely linguistic but epistemological in nature: it reveals how we structure, share, and interpret knowledge across languages and cultures. [59]

5. Strategies and Solutions: Lessons Learned and Mapping Guidelines

The mapping process between complex, multilingual vocabularies reveals not only technical and semantic challenges but also underscores the importance of robust strategic approaches. This section outlines key insights and practical guidelines developed through the mapping exercise. These strategies address core dimensions such as quality assurance, editorial workflows, semantic transparency, and documentation standards – each critical to ensuring that vocabulary alignment efforts remain reliable, transparent, and sustainable over time. [60]

5.1 Quality

In the context of vocabulary mapping and terminology management, quality refers to the degree to which data is accurate, consistent, semantically coherent, and fit for purpose. It also includes the clarity and traceability of editorial decisions. Quality is not merely an outcome but an ongoing process that requires both methodological rigor and continuous critical reflection – even when working with terminology in one's own language. [61]

Key criteria for quality include accuracy, completeness, consistency, and relevance. These must also extend to the documentation of mappings: metadata, source references, and explanatory notes should accompany each mapping decision to ensure transparency and reproducibility. It is important to distinguish between quality aspects that can be verified manually – such as the semantic adequacy of a match – and those that allow for automated validation, such as label format or character encoding. [62]

[11] Cf. Heath / Bizer (eds.) 2011; Vrandečić / Krötzsch 2014.

In practice, quality control can be embedded in editorial workflows through software-supported configurations: for example, defining rules for the use of preferred terms, managing homonyms, and ensuring that broader generic concepts are consistently applied. Regular audits, peer review, and version control systems further contribute to long-term quality assurance. [63]

5.2 Practical Recommendations

Effective vocabulary mapping not only depends on conceptual and editorial strategies but also on the practical tools that support these processes. A key aspect is the critical evaluation of available software: tools must be assessed concerning their performance, usability, and their fit for specific workflows. Essential functionalities include data export mechanisms and clustering tools that assist in organizing related concepts or uncovering semantic patterns – facilitating both the refinement of vocabularies and the transfer of terminological structures to other contexts. [64]

In order to ensure long-term usability and adaptability, vocabulary work must be embedded in flexible workflow strategies. These workflows should account for the fact that both the vocabulary and the databases in which it is implemented may evolve over time. Since workflow development is itself an iterative process, the steps taken and decisions made should be thoroughly documented. It is particularly helpful if the software environment is intuitive and provides inbuilt instructions, thereby reducing the need to consult external guidelines during routine tasks. [65]

Editorial rules must be comprehensive and well-structured. They should provide clear procedures for adding and validating new terms, including rules for assigning homonyms and synonyms, and for determining hierarchical relationships within the thesaurus. Approval processes should be defined with minimal requirements that ensure consistency and transparency. Mapping practices require equally clear rules, particularly regarding the use of SKOS mapping properties, the classification of terms within hierarchies, and the citation of authoritative sources. In addition, the intended scope and granularity of mappings – i.e., the coverage of equivalence – should be explicitly defined. [66]

The vocabulary mapping relies heavily on tools and software functionalities that support the organization, searchability, and integration of data. Ideally, these functionalities are embedded in the system and directly accessible to users. Among the most valuable features are autocomplete and suggestion mechanisms for term entry, the ability to compile and manage lists of terms and objects requiring translation or alignment, and access to reference structures that offer alternative terms, synonyms, and associative relationships. Supplementary resources such as knowledge bases, FAQs, and user forums can further support collaborative knowledge-building and the resolution of terminological issues. In addition, customizable user interfaces – featuring configurable dashboards, reusable templates, and intuitive navigation – can significantly streamline the workflow. A well-designed access rights management system is also essential, allowing for the differentiation of user roles for term proposal, review, translation, and publication, and supporting accountability through change tracking and version control. [67]

Ultimately, it is necessary to define the objectives and processes that each functionality should support. In many cases, different technical means can serve the same purpose – for example, lists may be generated via a dedicated module or dynamically created by applying specific search filters, such as the status or type of a given term. The choice of method should be aligned with the overall workflow strategy and the specific needs of the user community. [68]

When mapping to a widely used external vocabulary such as the AAT, additional considerations apply. The Getty Vocabulary Program recommends a conservative approach whereby only substantive changes are made to the master AAT. This policy aims to preserve the stability and reliability of the AAT as a standard reference, ensuring that descriptors, record identifiers, and hierarchical structures remain consistent. Avoiding modifications to these core elements helps maintain interoperability and compatibility with other systems and datasets that rely on the AAT. Therefore, vocabulary mapping efforts must carefully balance the need for local adaptation with respect for the authoritative nature of established reference vocabularies, integrating changes primarily through mapping relations rather than direct alteration. [69]

5.3 Scenarios for Mappings

Typical scenarios for vocabulary mapping arise in multilingual projects and heritage documentation contexts. All mapping strategies typically result in asymmetric thesauri – term counts and scope vary across languages and institutions. A fundamental requirement for any mapping effort is the use of persistent identifiers (URIs) to ensure sustainable, system-independent, and temporally stable referencing of concepts. Only through such identifiers can concepts be reliably linked, machine-actionable, and integrated into semantic networks. [70]

One scenario involves the direct mapping of a local vocabulary to an established international reference thesaurus such as the AAT. This approach is particularly feasible in thematically focused research projects, where domain-specific expertise is available and where mappings are often limited to a single facet or conceptual area. Scholars working in these contexts are usually able to supply detailed scope notes and ensure semantic accuracy. However, in broader institutional documentation practices, this strategy proves more challenging. The thematic diversity of documentation, combined with limited technical and editorial resources, often precludes the consistent implementation of direct mappings. Translation projects, such as ›AAT-Deutsch‹, help mitigate this issue by providing accessible reference points and ready-to-use terminologies that can serve as anchors for mappings, even when full integration is not possible. It is important to note that only actual concepts – i.e., terms with unique identifiers – can be mapped to the AAT. Structural elements such as guide terms or hierarchy labels are used for navigation within the AAT but do not themselves constitute concepts and therefore cannot serve as targets for external referencing. Moreover, the polyhierarchical structure of the AAT requires careful attention to the semantic context of any given term, as meaning can shift depending on hierarchical embedding. [71]

In many documentation workflows, terms are simultaneously mapped to more than one reference vocabulary, for instance, to both the AAT and the German Integrated Authority File (Gemeinsame Normdatei, GND). Such parallel mappings allow for multilingual access while aligning local documentation with national and international standards. The GND has gained importance in recent years, particularly in the context of library and archival documentation, and increasingly in museums. However, this approach introduces a different set of challenges. Differences in the structure, scope, and editorial histories of vocabularies can complicate mapping efforts. Missing scope notes, inconsistencies in hierarchies, and variations in terminological granularity often require manual interpretation and validation. At the same time, double mappings offer productive opportunities for quality control and enrichment: mappings can serve to identify inconsistencies, reveal semantic gaps, or validate existing structures through comparison. Still, they also increase the risk of ambiguity and misalignment. [72]

A different scenario can be observed in collaborative projects where local vocabularies are developed or even jointly developed by multiple institutions and simultaneously mapped to an external reference system. This approach is especially appropriate for institutions with shared documentation needs and complementary collections, provided that a combination of technical infrastructure and curatorial coordination is in place. An illustrative example is the mapping of the MCT to the AAT. The MCT is set to adopt elements of the AAT's hierarchical structure, with planned measures aiming to align its organization accordingly while [73]

integrating additional terms that reflect the specific collection contexts of the participating institutions. Editorial responsibilities are clearly defined, and mapping candidates for the AAT is submitted following Getty's editorial guidelines.

[74] Finally, an important strategy lies in the systematic translation of international vocabularies such as the AAT into other languages. These translation projects – such as ›AAT-Chinese‹[12], ›AAT-French‹[13], or AAT-Deutsch[14] – respond to the needs of national documentation systems, which depend on localized access to terminologies for practical and legal reasons. In contrast to research publications, which often rely on English-language terms, institutional collection management systems require multilingual vocabularies. Translation projects operate largely autonomously, but following Getty's editorial and technical standards. While Getty provides tools, training, and review, final editorial oversight remains with Getty to ensure consistency with its standards. The translations may also introduce terms that are specific to national or regional contexts, as long as they can be accommodated within the AAT's conceptual structure. A full translation of the AAT is unlikely, given the dynamic nature and ongoing expansion of the thesaurus. However, the structural compatibility of local vocabularies with the AAT enables the integration of context-specific terms and supports the long-term enrichment of the global thesaurus through locally driven use cases.

5.4 Data Contributing to Getty Vocabulary Program and Other Reference Vocabularies

[75] As has already been observed, mapping local vocabularies to established reference systems often brings semantic gaps to light – whether due to the broader conceptual scope of the reference vocabulary or the presence of locally specific terms not covered by the target system. In the latter case, contributions of new data offer a valuable strategy for enriching shared terminological resources. Major institutions responsible for reference vocabularies, such as the Getty Research Institute, generally support data contributions under clearly defined conditions. While comprehensive national translation initiatives continue to play a significant role in the expansion of reference vocabularies, smaller-scale contributions by individual institutions or consortia are increasingly common and explicitly encouraged. These contributions help close semantic and linguistic gaps, provide regionally specific concepts, and reinforce the interoperability of data infrastructures.

[76] The ›Getty Vocabulary Program‹ exemplifies this open and collaborative approach to vocabulary development.[15] It maintains a set of global, multilingual, and semantically structured vocabularies – namely the Art & Architecture Thesaurus (AAT), the Union List of Artist Names (ULAN), the Thesaurus of Geographic Names (TGN), the Cultural Objects Name Authority (CONA), and the Iconography Authority (IA) – all of which are open to contributions. The integration of externally contributed terms enhances the global applicability of these vocabularies while supporting the needs of specialized or regionally grounded documentation practices. To ensure quality and consistency, contributions must comply with a set of criteria, including relevance to the scope of the vocabulary, provision of minimum required metadata, and adherence to the prescribed formats. Accepted formats include standardized spreadsheets, XML schemas, and web forms designed for smaller data volumes. In order to meet the metadata requirements of these exchange formats, local thesaurus management tools must provide at least partial alignment with the metadata structure expected by the Getty system.

[12] By the Taiwan e-Learning and Digital Archives Program, TELDAP. Cf. Harpring 2010a.
[13] By the Canadian Heritage Information Network, CHIN.
[14] ›AAT-Deutsch‹, initiated by the Institut für Museumsforschung in Berlin between 2012 and 2014, continues this work and had processed approximately 12,700 terms in 2023, of which around 8,500 are publicly available.
[15] Cf. Harpring 2018.

[77] Each contribution undergoes an internal review process before publication. This review assesses the semantic accuracy, grammatical consistency, and authoritative sourcing of the submitted terms. For inclusion in the AAT, in particular, newly proposed concepts must be documented in English-language sources to ensure compliance with the thesaurus's editorial guidelines. This requirement, however, poses a significant challenge – if not a paradox – when attempting to include terms for local or culturally specific phenomena, which are often poorly represented or entirely absent in English-language sources. After successful review, the contributed terms are assigned permanent URIs, integrated into the hierarchical structure, and published in various semantic formats including JSON-LD, RDF / XML, and Turtle. The process is governed by a formal Data Contribution and License Agreement, and contributors must ensure that mappings in their local vocabularies are updated accordingly.

[78] Based on our experience with this process, automatic data export is unlikely to be feasible for most vocabularies unless they are specifically optimized for this purpose from the outset. In cases involving smaller datasets, manual preparation and submission remain a viable and pragmatic solution. For long-term projects, however, it is essential to align local data structures with the criteria of the reference vocabulary to facilitate more efficient future contributions.

[79] Although editorial procedures vary between systems, all require potential contributors to address key parameters such as expected metadata content, required data formats, and mapping of internal fields to external schemas. Semantic compatibility and technical interoperability must be carefully established before submission. In practice, the approval and publication of new terms in such vocabularies can take several months, depending on the scale of the contribution and the priorities of the receiving institution. Nevertheless, these contribution pathways offer significant opportunities for small-scale or regional projects to participate in the development of global knowledge infrastructures, helping to ensure that cultural and linguistic diversity is reflected in shared digital vocabularies.

6. Potentials and Limitations of AI in Vocabulary Work

6.1 Preconditions for Effective AI Use: Curated, Context-Rich Data and Humanities

[80] Artificial Intelligence is evolving rapidly, reshaping how we communicate, make decisions, and produce knowledge. At the heart of this transformation lies data – massive, complex, and ever-growing. For AI to be meaningful, trustworthy, and effective in communicative contexts, it must be grounded in curated, context-rich data, especially in structured vocabulary data that captures the semantic, historical, and cultural dimensions of the concepts involved. Vocabulary work – organizing, enriching, and contextualizing terms – is therefore not an auxiliary task, but a core precondition for AI systems that aim to interpret, translate, or generate human language.

[81] In digital art history and the digital humanities more broadly, vocabularies and authority data are not only tools for organizing knowledge – they are essential methodological instruments. They enable more transparent, interpretable, and controllable applications of AI by providing structured, semantically rich reference points. When published as Linked Open Data, they contribute to the creation of interconnected knowledge spaces that reflect the contextual complexity of historical and cultural entities.

To identify and link the same entities across different datasets, various tools and technologies can be applied – ranging from simpler formats like *BEACON* [16] to more advanced software for data matching, such as R [17], SAS [18], SPSS [19], or OpenRefine [20]. These tools support different algorithms and methods for entity resolution. Historical data in particular are often characterized by linguistic variation, inconsistent naming conventions, and duplicate or ambiguous records, all of which make entity resolution especially complex and error-prone.[21] [82]

From a research perspective, having shared methodological frameworks and interoperable data models emerges as a key criterion that makes digital analysis methods and their results more comparable, transparent, and trustworthy. This is especially important in digital research and when applying AI methods, as it helps make results more reliable – or even enables their generation in the first place. At the same time, research practice today is marked by a high heterogeneity of perspectives and approaches – a hallmark of humanities scholarship. However, this diversity can also limit the comparability of research results and sometimes cause failures in digital research projects when methods are insufficiently standardized or widely known.[22] [83]

But data alone is not enough. Without frameworks for understanding meaning, ambiguity, and cultural context, even the most sophisticated AI risks replicating biases, flattening complexity, or producing outputs that miss the point. [84]

This is where the humanities offer essential expertise. Disciplines such as philosophy, linguistics, cultural studies, and history bring deep insight into how meaning is constructed, how language functions, and how context shapes interpretation. These strengths are not merely complementary to AI – they are foundational. The humanities help us ask better questions about the assumptions built into data, the categories we use to classify the world, and the narratives we construct with the help of machines.[23] [85]

Linguistics, for example, contributes not only to structural language analysis but also to an understanding of semantics and pragmatics – key to disambiguating meaning. Hermeneutics teaches that texts, laws, and policies are never neutral and require context-sensitive interpretation. Narratology reveals how stories function and how narrative coherence influences comprehension, crucial for conversational agents and content generation tools. Disciplines such as cultural studies and art history demonstrate how concepts like ›identity‹, ›culture‹, or ›art‹ are historically and socially constructed rather than fixed or universal. This understanding is crucial for fields such as museum documentation and knowledge organization, where the classification and interpretation of cultural objects rely on these often-contested terms. This perspective is essential when designing AI systems intended to operate across diverse cultural and disciplinary domains. At the same time, discourse analysis reveals how language actively shapes perceptions of meaning and value, offering critical tools for prompt engineering, human-AI interaction design, and the semantic evaluation of AI-generated content in heritage and visual knowledge contexts. [86]

[16] BEACON is a simple text-based format developed in the library sector, especially in the German-speaking context, for publishing and linking authority data (e.g. persons) via persistent identifiers such as GND or VIAF. It enables lightweight semantic connections across web resources. Cf. Wikimedia 2025.
[17] R is a free software environment for statistical computing and graphics. It is widely used for data analysis, including entity matching.
[18] SAS (Statistical Analysis System) is a commercial analytics platform used in professional data environments. It offers advanced data management and matching capabilities, especially for large-scale or enterprise-level datasets.
[19] SPSS (Statistical Package for the Social Sciences) is a software package used for statistical analysis in social science. It includes tools for data transformation, matching, and deduplication.
[20] OpenRefine is an open-source desktop application for data cleaning and transformation. It supports operations like clustering, reconciliation, and linking to external databases (e.g., Wikidata), making it useful for entity resolution and standardization.
[21] Cf. Moeller / Purschwitz 2025, p. 7. See also: Stiller et al. 2014; Suominen et al. 2022.
[22] Cf. Lemaire et al. 2025, p. 5.
[23] Cf. McShane / Nirenburg 2021.

In all these ways, the humanities help to enrich the data that fuels AI. They offer frameworks for identifying relevance, nuance, and meaning – qualities that raw data alone cannot provide. As we move toward more socially embedded, ethically grounded AI systems, the humanities open up new opportunities: to design technologies that are not only functional but also reflective, inclusive, and culturally aware. By integrating humanistic knowledge into the foundations of AI, we unlock the potential to build systems that truly understand the worlds they are meant to engage with. [87]

This foundational role of structured vocabularies becomes particularly tangible in the cultural heritage sector. Resources such as the Getty Vocabularies, ICONCLASS, and various domain-specific repositories illustrate how disciplinary knowledge can be encoded into machine-actionable formats – capturing semantic relationships, cultural knowledge, and scholarly conventions. [88]

By doing so, these vocabularies improve the accuracy, interpretability, and relevance of AI systems – especially in information retrieval and cross-lingual / cultural interoperability. They form the backbone of responsible, transparent, and context-aware AI development. [89]

The EuropeanaTech community – a network of heritage technologists, researchers, and developers within the Europeana Network Association – actively promotes these ideas through R&D task forces, publications, and strategic initiatives focused on enhancing data quality and AI readiness in cultural heritage contexts. In parallel, initiatives like NFDI4Culture recognize and support this infrastructure by advocating for hybrid workflows: AI drives scalability and efficiency, while expert-driven editorial processes ensure semantic integrity and cultural sensitivity. [90]

In sum, while AI holds significant promise for enhancing and scaling vocabulary work, its full potential will only be realized through continued investment in the foundational development of curated, expert-driven vocabularies. Rather than replacing human knowledge, AI in this field should be understood as amplifying it – and as requiring it at every stage. [91]

6.2 Integrating NLP and ML into Vocabulary Enrichment: Potentials, Limits, and Human Oversight

Building on this foundation, recent developments in Artificial Intelligence – particularly in *Natural Language Processing (NLP)* and *Machine Learning (ML)* – offer promising but still limited tools for advancing vocabulary work in the cultural heritage sector. These technologies are increasingly explored for tasks such as term mapping, multilingual translation, semantic pattern detection, and lexical expansion. Their usefulness lies not in replacing human interpretation, but in supporting and accelerating editorial workflows – provided they are trained on curated, semantically rich, and historically contextualized data.[24] [92]

In this setting, NLP and ML can assist by identifying synonym clusters, suggesting cross-lingual correspondences, or detecting latent semantic structures across large datasets. These functionalities have the potential to enhance consistency and efficiency, especially in multilingual environments. However, their performance is highly dependent on data quality and expert oversight: only under these conditions do automated suggestions become reliable and meaningful. [93]

[24] Cf. Stiller et al. 2014; Suominen et al. 2022; Mayr / Petras 2008.

Moreover, many vocabulary-related tasks in the humanities involve semantic ambiguity, shifting historical meanings, and disciplinary nuance – challenges that current AI models are not equipped to navigate autonomously. Linguistically similar terms may diverge in meaning across fields or periods; taxonomic categories may overlap or reflect contested knowledge regimes. In such cases, interpretive judgment remains essential to ensure that alignments, translations, or enrichments do not distort or oversimplify complex knowledge structures. [94]

Given the evolving nature of AI and NLP technologies, a human-in-the-loop approach currently represents the most viable model for integrating automated processes into vocabulary development. In this paradigm, machine-generated suggestions are systematically filtered, validated, and contextualized by domain experts. Such a hybrid model ensures both scalability and the preservation of semantic accuracy. Under these conditions, NLP and ML may function as catalysts for methodological innovation – augmenting rather than displacing the expert knowledge foundational to vocabulary work. In this context, AI technologies and *(Named) Entity Recognition* techniques demonstrate considerable potential, although further refinement and domain-specific adaptation remain necessary. [95]

6.3 The ›pre-AI‹ Nature of the Case Study as Groundwork for Future Applications

The case study presented in this article – mapping the MCT to the AAT – illustrates a stage of vocabulary work that might be termed ›pre-AI‹: a phase focused on establishing the semantic, hierarchical, and conceptual coherence of vocabularies before they are suitable for machine-assisted processing. This involves aligning terms across languages and traditions, reconciling different classification logics, and resolving ambiguous or overlapping categories – challenges that require domain-specific knowledge and interdisciplinary negotiation. While some examples of such complexities have already been discussed, including those drawn from the classification of Protestant churches in Silesia, the range and density of problematic cases warrant further illustration. [96]

Particularly in specialist thesauri, terminology tends to be highly nuanced and context-dependent. For example, in German, terms such as ›Wandbild‹, ›Wandgemälde‹, ›Gewölbebild‹, ›Deckenmalerei‹, ›Wandmalerei‹, and ›Deckenbild‹ refer to subtly different types of mural or ceiling paintings, each embedded in distinct art-historical and linguistic conventions. Attempts to clarify these distinctions using popular AI-based language models – for example, by prompting them for definitions, translations, or contextual explanations – have not yielded satisfactory results.[25] These models often fail to recognize domain-specific usages, collapse semantically distinct concepts into generic categories, or provide inconsistent responses. Moreover, the richness and granularity of German vocabulary in this field frequently lacks direct equivalents in English, further complicating cross-lingual alignment. Additional examples include terms such as ›Blockbau‹, ›Ständerbau‹, ›Bohlenständerbau‹, ›Dreiständerbau‹, and ›Vierständerbau‹, which denote both construction techniques and resulting architectural forms. These require careful semantic modeling to reflect their dual function within thesaurus hierarchies. Such cases underscore not only the methodological precision required in controlled vocabulary development, but also the indispensable role of human expertise in preparing these resources for meaningful and reliable AI-supported applications. [97]

[25] Prompt-based queries were tested using large language models such as ChatGPT (GPT-4, OpenAI) and DeepL Write between April and June 2025. While results occasionally yielded general definitions, they lacked the terminological precision, conceptual differentiation, and hierarchical embedding needed for thesaurus-level interoperability. The limitations highlight the need for curated, expert-validated vocabulary structures as a prerequisite for domain-aware AI processing.

7. Conclusion

[98] Vocabulary alignment constitutes a crucial step toward realizing FAIR data principles within the digital research infrastructures of the humanities. The integration of subject-specific vocabularies into established international reference frameworks – such as the AAT – enhances the accessibility, discoverability, and reusability of cultural heritage data by bridging specialized disciplinary terminologies with broader semantic standards.

[99] The examples discussed in this article demonstrate that thesaurus mapping transcends purely technical or computational challenges; it is a complex intellectual endeavor demanding deep terminological expertise, linguistic sensitivity, and methodological rigor. These intricacies highlight the indispensable role of human expertise in ensuring semantic coherence, particularly given the specificity and nuance inherent to disciplinary vocabularies, even as technological tools and AI-driven methods continue to advance.

[100] Looking ahead, investment in curated, domain-specific vocabularies will form the basis for interoperable, AI-ready cultural data ecosystems. By fostering hybrid workflows that synergize expert curation with machine-assisted processes, future projects can capitalize on the complementary strengths of human and artificial intelligence. This approach promises to accelerate innovation in vocabulary work, supporting more robust and scalable infrastructures.

[101] Ultimately, the findings presented here contribute valuable insights to the broader field of digital humanities infrastructure and offer practical guidance for initiatives aiming to align and integrate specialized disciplinary vocabularies with international reference standards – essential steps toward unlocking the full potential of cultural data in the digital age.

Bibliography

Sławomir Brzezicki / Ulrike Nürnberger: Dehio-Handbuch trifft digiCULT. digiCULT.web als Instrument für die Dokumentation von Bau und Kunstdenkmälern. Das Wissensportal Dehio Ostmitteleuropa (Dehio OME). 2022. PDF. [online]

Stephanie Russo Carroll / Ibrahim Garba / Oscar L. Figueroa-Rodríguez / Jarita Holbrook / Raymond Lovett / Simeon Materechera / Mark Parsons / Kay Raseroka / Desi Rodriguez-Lonebear / Robyn Rowe / Rodrigo Sara / Jennifer D. Walker / Jane Anderson / Maui Hudson: The CARE Principles for Indigenous Data Governance. In: Data Science Journal 19 (2020). DOI: 10.5334/dsj-2020-043

Martin Doerr: Semantic Problems of Thesaurus Mapping. In: Journal of Digital Information 1 (2001), no. 8. PDF. [online]

Getty Vocabulary Contribution Portal. Last update: 11.07.2023. HTML. [online]

Patricia Harpring (2010a): Development of the Getty Vocabularies: AAT, TGN, ULAN, and CONA Art Documentation. In: Journal of the Art Libraries Society of North America 29 (2010), no. 1, pp. 67–72. [online]

Patricia Harpring (2010b): Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works. Los Angeles 2010. PDF. [online]

Patricia Harpring: Linking the Getty Vocabularies: The Content Perspective, Including an Update on CONA. In: 2018 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC). Conference Proceedings (San Francisco, 27.–30.10.2018). New York, US 2018, pp. 34–41. DOI: 10.23919/PNC.2018.8579460

Tom Heath / Christian Bizer (eds.): Linked Data. Evolving the Web into a Global Data Space (= Synthesis Lectures on Data, Semantics, and Knowledge, 1). San Rafael, US-CA 2011. HTML. DOI: 10.2200/S00334ED1V01Y201102WBE001

Antoine Isaac / Ed Summers: SKOS Primer. Simple Knowledge Organization System Primer. In: Semantic Web Deployment Group (ed.): W3C Working Group Note 18.08.2009. Last update: 30.10.2023. HTML. [online]

ISO 25964-1. Information and Documentation – Thesauri and Interoperability with Other Vocabularies. Part 1: Thesauri for Information Retrieval. Last update: 07.01.2026. HTML. [online]

ISO 25964-2. Information and Documentation – Thesauri and Interoperability with Other Vocabularies. Part 2: Interoperability with Other Vocabularies. Last update: 05.09.2023. HTML. [online]

Angela Kailus: Handreichung für ein FAIRes Management kulturwissenschaftlicher Forschungsdaten. NFDI4Culture. 01.03.2023. Version 2.0.1: 04.09.2025. HTML. [online]

Marina Lemaire / Anne Voigt / Ursula Lehmkuhl: Whitepaper: Datenkompetenzen für die historisch arbeitenden Disziplinen (= 4memory Working Paper Series, 3). Zenodo. 23.05.2025. DOI: 10.5281/zenodo.15479671

Robert Lew: Dictionaries and Lexicography in the AI Era. In: Humanities and Social Sciences Communications 11 (2024). DOI: 10.1057/s41599-024-02889-7

Jutta Lindenthal / Jessica Sandrock: digiCult Thesaurus-Handbuch. Last update: 28.03.2026. HTML. [online]

Phillip Mayr / Vivien Petras: Cross-Concordances: Terminology Mapping and its Effectiveness for Information Retrieval. arXiv. 23.06.2008. DOI: 10.48550/arXiv.0806.3765

Marjorie McShane / Sergei Nirenburg: Linguistics for the Age of AI. Cambridge, US-MA etc. 2021. DOI: 10.7551/mitpress/13618.003.0016

Alistair Miles / Sean Bechhofer: SKOS Reference. Simple Knowledge Organization System Reference (= Semantic Web Deployment Group). W3C Recommendation. 18.08.2009. HTML. [online]

Katrin Moeller / Anne Purschwitz: Kontrollierte Vokabulare und Normdaten der historisch arbeitenden Disziplinen (= 4Memory Working Paper, 5). Zenodo. 02.07.2025. Version 2.0: 02.07.2025. PDF. DOI: 10.5281/zenodo.15745568

Juliane Stiller / Vivien Petras / Maria Gäde / Antoine Isaac: Automatic Enrichments with Controlled Vocabularies in Europeana: Challenges and Consequences. In: Marinos Ioannides / Nadia Magnenat-Thalmann / Eleanor Fink / Roko Žarnić / Alex-Yianing Yen / Ewald Quak (eds.): Digital Heritage. Progress in Cultural Heritage: Documentation, Preservation, and Protection. 5th International Conference, EuroMed 2014. Conference Proceedings (Limassol, Cyprus, 03.–08.11.2014). Cham, CH 2014, pp. 238–247. DOI: 10.1007/978-3-319-13695-0_23

Osma Suominen / Mona Lehtinen / Juho Inkinen: Annif and Finto AI: Developing and Implementing Automated Subject Indexing. In: JLIS.It 13 (2022), no. 1, pp. 265–282. DOI: 10.4403/jlis.it-12740

Denny Vrandečić / Markus Krötzsch: Wikidata: A Free Collaborative Knowledgebase. In: Communications of the ACM 57 (2014), no. 10, pp. 78–85. DOI: 10.1145/2629489

Wikimedia: Dynamic Links to External Resources. Last update: 05.01.2025. HTML. [online]

List of Figures

Figure 1: Search results for the term ›Kanzelaltar‹ in DEHIO Germany. [Screenshot: Ksenia Stanicka-Brzezicka 2025]

Figure 2: Grace Church in Jelenia Góra (Hirschberg), Poland, dataset from OME. [Screenshot: Ksenia Stanicka-Brzezicka 2025]

Figure 3: The term ›Church of Peace‹ (Gnadenkirche), AAT. [Screenshot: Ksenia Stanicka-Brzezicka 2025]

Figure 4: The term ›Protestantism‹ with relations, AAT. [Screenshot: Ksenia Stanicka-Brzezicka 2025]