# Building and sustaining the Homosaurus, an inclusive LGBTQ+ linked data vocabulary

**D St-Amour**

The Homosaurus Collective / Western University

**Keahi Adolpho**

The Homosaurus Collective / Virginia Commonwealth University

**Marika Cifor**

The Homosaurus Collective / University of Washington

**Michael Radmacher**

The Homosaurus Collective / University of Victoria

**B.M. Watson**

The Homosaurus Collective / Western University

**Crystal Yragui**

The Homosaurus Collective / University of Washington

**Travis Wagner**

The Homosaurus Collective / University of Illinois, Urbana-Champaign

**ABSTRACT**

This article describes the Homosaurus, a LGBTQ+ community-driven linked data vocabulary. Authored by members of the Homosaurus Collective, it details the project's origins, governance history, and transformation from an institutional thesaurus into an international linked data vocabulary.

## Introduction

The Homosaurus is an international and community-driven linked data vocabulary of LGBTQ+ terminology. It functions as a critical supplement and alternative to Library of Congress Subject Headings (LCSH), and emerged within a broader effort across galleries, libraries, archives, museums and special collections (GLAMS) to identify and address harmful or exclusionary descriptive language. Within GLAMS, the Homosaurus supports researchers and knowledge workers by providing community-

defined terms that can be integrated into cataloging, metadata creation, and discovery systems. Its structure is designed to be interoperable with widely used metadata frameworks.

By foregrounding community-authored and defined terminology, the Homosaurus reduces risks of erasure, misclassification, and descriptive bias that have historically shaped knowledge organization systems. The project reflects participatory ontology-building practices in which community expertise is treated as authoritative in determining terminology. This approach aligns with broader ethical frameworks that prioritize self-identification, transparency, and accountability in term development and descriptive practice. As a result, the Homosaurus operates not only as a technical metadata tool, but also as an intervention on the historical inequities embedded in subject analysis and other resource description systems.

## History

Like many LGBTQ+ entities, the Homosaurus emerges from the dedication and stewardship of its earlier generations. The project originated in 1982, when staff at Homodok - the University of Amsterdam Documentation Center for Gay and Lesbian Studies - began developing a list of Dutch-language keywords used to describe resources in their bibliographic database. In 1993, Homodok merged with Anna Blaman Huis (formerly Lesbian Archives Leeuwarden), forming the IHLIA (now IHLIA-LGBT Heritage[1]), which faced an urgent need to describe its newly combined and growing collections. Staff identified structural limitations within conventional subject heading thesauri, particularly in their ability to represent diverse queer identities and associated lived experiences. These limitations had direct implications for information retrieval, as researchers, students, and other users encountered difficulties locating materials on topics. In many cases, queer concepts were either absent from legacy vocabularies or described in ways that reduced (and even prevented) discoverability. Thus, in 1993, IHLIA developed the Dutch language *Queer Thesaurus* based largely on the combined subject lists used by Homodok and Anna Blaman Huis. In 1997, *Queer Thesaurus* was translated into English (Cifor and Rawson, 2023).

*Queer Thesaurus* reflected the content held in IHLIA, and was thus focused on gay and lesbian terms, with deeper coverage on white gay cisgender men. Between 2013 and 2015, Jack van der Wel and Ellen Greenblatt revised and substantially expanded the *Queer Thesaurus*, adding and linking hundreds of new terms in an effort towards both a more structured hierarchy and broader LGBTQ+ inclusion, forming the first iteration to use the name Homosaurus. This transformation reflected both shifts in community language and evolving expectations around digital metadata environments, while moving the vocabulary deliberately beyond its primarily gay and lesbian focus. During this period, the Homosaurus also began to expand beyond its role as a standalone, in-house reference tool. K.J. Rawson recognized its unique potential for describing materials in the then-nascent Digital Transgender Archive

---

[1] https://ihlia.nl/en/

(DTA)[2], an online repository for trans-related historical materials. Rawson's application of the Homosaurus within the DTA played a significant role in driving its transformation into a linked data vocabulary, enabling the Homosaurus to function within semantic web environments and supporting broader integration into digital collections and discovery systems (Cifor and Rawson, 2023).

In 2016, van der Wel and Rawson established the Homosaurus Editorial Board, creating a formal governance structure for the vocabulary. The board was comprised of eight to ten queer- and trans-identified information professionals and academics who met monthly to review term suggestions, make revisions, and discuss emerging language. This period also prompted an important reconceptualization of the vocabulary's scope: rather than attempting to comprehensively describe all resources that might be held in relevant collections, the board determined that the Homosaurus should function as an LGBTQ+-specific vocabulary designed to be used in conjunction with broader thesauri. This approach allowed for greater depth in representing LGBTQ+ terminology and made the Homosaurus usable by any GLAMS - including those operating within LCSH-dominated library catalogs - seeking to describe LGBTQ+ resources (Cifor and Rawson, 2023).

**Present**

Between 2015 and 2025, expansion and development continued as the Homosaurus saw a remarkable uptick in adoption and use. In 2016, the Homosaurus became an approved thesaurus source for MARC records (code: homoit) (Library of Congress, Network Development and MARC Standards Office, 2016), which vastly expanded the use of the vocabulary across the world. To further develop and diversify the vocabulary, Homosaurus used supplemental funding from Northeastern University, Rawson's home institution. This made it possible to support collaborations with community-based organizations such as AVEN[3], the Black Lesbian Archive[4], COLAGE[5], InterACT[6], the Leather Archives and Museum[7], and the Sex Worker Activist Histories Project[8].

Throughout this period of growth, the Editorial Board also received a number of requests for potential translation projects. As a result, in 2023, Marika Cifor and K.J. Rawson (in partnership with the San Francisco Public Library, UCLA Chicano Studies Research Center and Library, and Arizona Queer Archives) were awarded the Homosaurus' first external funding; a three-year research and development grant from U.S. National Endowment for the Humanities to create a Latin American Spanish-language version of the Homosaurus and to build a multi-language backend

---

[2] https://www.digitaltransgenderarchive.net/
[3] https://www.asexuality.org/
[4] https://blacklesbianarchives.wixsite.com/info
[5] https://colage.org/
[6] https://interactadvocates.org/
[7] https://leatherarchives.org/
[8] https://swahp.ca/

infrastructure for the Homosaurus. Despite the grant being rescinded by the Trump administration in April 2025, Rawson and Cifor along with postdoctoral fellow, Ernesto Cuba, student collaborators, Sofia Zamora, Sandy Alcantara, and Mar Munné, and translator Ana Portnoy Brimmer, were able to complete the project's primary goal of developing a Spanish language Homosaurus. Version 4 was released in early 2025 and reflected both direct translation of terms from the English language Homosaurus and concepts unique to Spanish-speaking LGBTQ+ communities. There are now versions available on the Homosaurus website in English and Latin American Spanish, with Bengali, Castilian Spanish, Dutch, German, Hindi, Japanese, and Swedish versions planned or in process.

As growth continued and and collaboration expanded, it became clear that the Editorial Board model was not a sustainable solution for continued governance. Furthermore, members observed that the Editorial Board model did not adequately reflect the Homosaurus' values of equity, justice and self-determination. Starting in 2024, the Editorial Board developed a new leadership structure, which was formalized as the Homosaurus Collective in late 2025.

## Governance & Decision Making

The Homosaurus Collective, which presently includes 40+ members based in 9 countries, is subdivided into three interdependent boards: Organizing, Term Development, and Implementation (see Figure 1 below). The Organizing Board provides coordination, visioning, and guidance across the project, and helps to develop and grow the Collective, seeking funding, hosting summits, and identifying partnerships. This board contains the Systems Committee, which focuses on software functionality and coordinates with developers. The Term Development Board supports cross-language work and provides coordination among Term Development Committees. This board oversees the Dutch, English, French, German, Hindi and Bengali, Japanese, Spanish, and Swedish Language-Specific Committees, which are sustained long-term, as well as Ad-Hoc Thematic Committees, such as the Pasifika/Oceania committee, which are temporary in nature. The Implementation Board works with outside organizations on the implementation of the Homosaurus, developing both internal and external training practices, creating implementation resources, and tracking use and scholarly application of and writing on the Homosaurus.

Because Homosaurus usage is primarily gathered through user self-identification via our community platforms, the precise number of institutions using the Homosaurus today is difficult to ascertain. According to the data we have, our user community includes over 150 institutions, ranging from large, public library systems and national libraries, to small, grassroots community-based queer archives in seven or more countries. To date, we have been able to track at least 255 mentions of the vocabulary's use across scholarly platforms representing at least seven different countries, and both discussion and usage of the Homosaurus continues to increase across a range of LIS and GLAMS settings. For example, in their recent study of
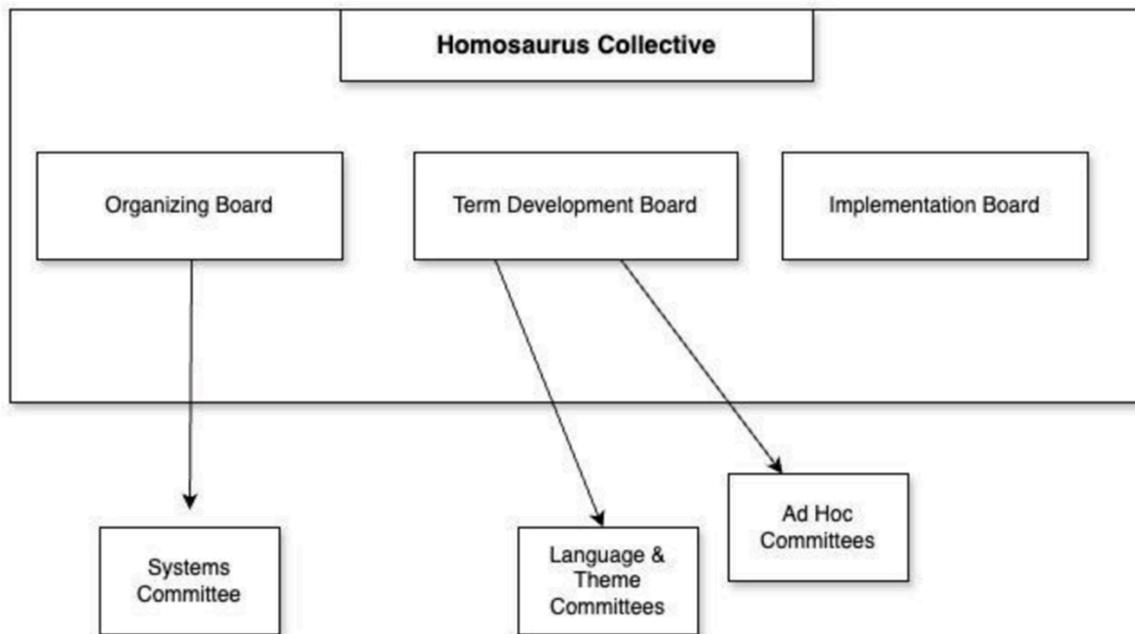
**Figure 1**: Homosaurus Collective structure. Diagram created by Chloe Noland, Homosaurus Term Development Board.

archivists working with LGBTQ+ collections and materials, Wagner et al. (2025) note that more than two thirds of the 29 archivists they interviewed utilized the Homosaurus as part of their archival description work. We also know that the Homosaurus has been incorporated into a number of grant funded projects, including a recent National Leadership Grant funded by the U.S. Institute of Museum and Library Services led by co-principal investigators Brian Dobreski, Karen Snow, and Heather Moulaison-Sandy. While findings are still forthcoming, this research will help expand the quantitative and qualitative uses of the Homosaurus (Gallant, 2024).

## Community Driven Term Curation

Community input has defined the Homosaurus vocabulary development process since its inception, with the vocabulary's ongoing growth and refinement shaped by the lived expertise and feedback of the communities it represents and serves. New terms, as well as revisions to existing terms, may be suggested by community users, community members, or by Homosaurus Collective members. Currently, the bulk of these requests come through the contact form on the Homosaurus site, via email, and on Homosaurus community social spaces. Once a new term or revision is proposed, if it is not duplicative, it is added to the language committee's term development spreadsheet for consideration.

Members of the English Term Development Committee discuss whether a term should be developed based on:

- prevalence (*Is the term currently or has it been previously widely used by LGBTQ+ communities it is relevant to?*),

- significance and relevance (*Is the term relevant in some way to LGBTQ+ concepts, issues, communities, and resources?*),
- coverage in other vocabularies (*Is the term present in mainstream vocabularies, and if so, is this coverage adequate? Does the coverage or preferred term differ from community usage?*).

Once a term is deemed to be in scope and/or appropriate for inclusion, a member of the English Term Development Committee drafts a term (either individually or in collaboration with others) including its preferred and alternative labels, scope note, relationships to other terms in the vocabulary, language of use, and any exact or close matches for the term in LCSH. Once the term is drafted, it is moved to the asynchronous approval queue if it is ready for voting or the terms for board discussion queue if the term may be controversial or need special attention. Presently, all terms need full-committee approval to be considered ready for publication. Some terms are approved as drafted, whereas others are collaboratively revised by the rest of the committee. Since the transition to the Homosaurus Collective structure, all new or revised terms drafted by the English Term Development Committee are shared with the rest of the language-specific term development committees for a final round of review before being added to the next semi-annual Homosaurus release. Currently, most other language-specific term development committees are focused on translating existing English Homosaurus terms rather than drafting original terms, though we expect more original terms from non-English committees moving forward and are prioritizing creating the infrastructure for supporting this work.

## Linked Data & Technical Architecture

Early multi-language efforts of Homosaurus were largely focused on front-end language translations that identified one-to-one comparable terms. These direct translation efforts were successful at expediting greater front-end language accessibility. However, these efforts still used English as foundational ground truth, which was incompatible with the cross-cultural and translinguistic goals of Homosaurus.

Under the Editorial Board structure, Homosaurus was primarily treated as an English-language vocabulary with translations in other languages. Under the new structure, the Term Development Board has been working towards a more international framework. Each language committee brings not only translation but genuine cultural and linguistic expertise, with the goal of building a diffused and collaborative model that resists mapping onto an English-language centre. In practice, this means that terms no longer must be included or first developed in the English vocabulary to exist in another language. Additionally, the Term Development Board is working towards developing translation guidelines in acknowledgement that there are concepts and terminology that may only be relevant in certain language versions or cultural and/or geographic contexts.

Homosaurus' pivot away from English as the basis for multi-language translation is supported by linked data's paralinguistic structure. Linked data enables Homosaurus to assign each concept in the vocabulary a unique alphanumeric uniform resource identifier (URI), which allows concepts to exist in a stable form detached from any particular linguistic tradition. If the concept resonates in a specific cultural or linguistic environment, it can then be expressed by attaching a label appropriate for the sociocultural context. This linked data structure thus shifts Homosaurus' efforts away from front-end language translation to back-end vocabulary building.

The Metadata Application Profile details the technical architecture and supporting data model that supports the vocabulary (Roles, et al., 2022). Concepts are modeled as SKOS Concepts, with alphanumeric URIs distinct from individual language labels and their corresponding HTTP URLs. Relationships between Homosaurus concepts (and mapping from Homosaurus to LCSH concepts) are expressed hierarchically in SKOS and available as structured data in N-Triples, JSON-LD, TTL, and CSV formats. Controlled language tags facilitate the vocabulary's multilingual back-end in compliance with IETF BCP 47[9] (Roles et al., 2022).

The Homosaurus website is a Ruby on Rails application available on GitHub (Anderson, 2024). The code is maintained by its author, Steven Carl Anderson. The website uses Blazegraph for querying, Solr for record display, and Postgres for data storage and preservation metadata (Watson, Noland, and Billey, 2021). The Homosaurus publishes new versions for major changes and releases for vocabulary changes, both of which are documented on the Releases page[10].

## Future Directions

The Homosaurus is a living vocabulary, designed to evolve with the people and communities it represents and serves. As LGBTQ+ language, identities, and cultures shift over time, so too does the vocabulary - with new terms added and existing terms revised on an ongoing basis. This commitment to responsiveness is not incidental but foundational: the Homosaurus exists to reflect community knowledges.

A central dimension of this growth has been the expansion of the Homosaurus into new languages, an undertaking which, under the previous structure, volunteers typically completed in isolation from other members. Working more closely on the Term Development Board has surfaced needs in translation workflows and implementation support that is both multilingual and not limited to MARC cataloging environments. Effectively collaborating across time zones (we currently have collaborators in Hawai'i, UTC-10:00, and Japan, UTC+9) and creating the framework to support and grow this cross-language work is an ongoing project for the Term Development Board.

---

[9] https://www.rfc-editor.org/info/bcp47
[10] https://homosaurus.org/releases

Alongside this linguistic expansion, the Homosaurus has worked to broaden participation in its governance and long-term sustainability. The transition to a Collective and distributed model of leadership has allowed stewardship of the vocabulary to be shared across a wider and more geographically diverse group of contributors. While ongoing technical infrastructure development has kept pace with the vocabulary's growing scope, new institutional homes and collaborations - particularly in Canada, the European Union, and beyond - will help to both safely anchor the project and to extend its reach.

Looking ahead, the Homosaurus Collective has an ambitious wish list for the future:

- dedicated staff and infrastructure to support day-to-day operations,
- expanded resources for community collaborations,
- continued growth of language boards,
- support for implementation in non-English and non-MARC cataloging environments, and
- contributions from software developers to strengthen the technical foundation of the project.

Realizing this vision will require sustained investment and partnership; however, the Homosaurus is driven by a clear purpose: to continue building a vocabulary that does not merely correct existing information systems, but actively works to make them more permeable, more equitable, and more genuinely reflective of the diverse LGBTQ+ communities it exists to serve.

## References

Anderson, S. (2024) *DigitalTransgenderArchive/homosaurus_site*. Available at: https://github.com/DigitalTransgenderArchive/homosaurus_site [Accessed 20 March 2026]

Cifor, M. and Rawson, K. J. (2023) 'Mediating queer and trans pasts: The Homosaurus as queer information activism', *Information, Communication & Society*, 26(11), pp. 2168-2185. Available at: https://doi.org/10.1080/1369118X.2022.2072753

Cushman, E. (2013) 'Wampum, Sequoyan, and story: decolonizing the digital archive', *College English*, 76(2), pp. 115–135. Available at: https://www.jstor.org/stable/24238145 [Accessed 20 March 2026]

Gallant, K. (2024) SIS assistant professor receives IMLS grant to help continue research on inclusive cataloging in libraires. Available at: https://cci.utk.edu/blog/2024/08/15/sis-assistant-professor-receives-imls-grant-to-help-continue-research-on-inclusive-cataloging-in-libraries/ [Accessed 20 March 2026]

Library of Congress, Network Development and MARC Standards Office (2016) *Techinical notice (September 15, 2016): Additions to Source Codes for Vocabularies, Rules, and Schemes*. Available at: https://www.loc.gov/marc/relators/tn160914src.html [Accessed 20 March 2026]

Roles, C. F., Rawson, K. J., Anderson, S., and Homosaurus Editorial Board (2022) *Homosaurus Metadata Application Profile*. Available at: https://docs.google.com/document/d/1ndif61THKvrKRG8m6Y9gBVAOzwDWa95O9TbJV8tFTTA [Accessed 20 March 2026]

Wagner, T. L., Allgood, E. M., and Caballero, M. (2025) '"The finding aid is the first thing that people see, we don't want to put anyone off viewing the collection": how practitioners navigate queerness in finding aids', *Archival Science*, 25, article 43. Available at: https://doi.org/10.1007/s10502-025-09513-w

Watson, B. M., Noland, C., and Billey, A. (2021) 'The Homosaurus', *Catalogue & Index*, 202(11), pp. 44-47. Available at: https://www.cilip.org.uk/resource/collection/CC45A47F-BE48-4BA1-B6D4-25BF10F1BC41/C&I202Homosaurus.pdf [Accessed 20 March 2026]