Encontros Bibli

WIKIDATA FOR STRUCTURING ACADEMIC INSTITUTIONAL PROFILES: A NEUROMAT SCHOLARLY PUBLICATIONS CASE STUDY

Wikidata para estruturação de perfis institucionais acadêmicos: um estudo de caso de publicações acadêmicas do NeuroMat

Erika Guetti Suca
Universidade de São Paulo
Departamento Ciências da Computação
São Paulo, SP, Brasil
eguetti@usp.br
https://orcid.org/0000-0003-2934-9778

Éder Porto Ferreira Alves
Wiki Movimento Brasil
São Paulo, SP, Brasil
eder.porto@wmnobrasil.org
https://orcid.org/0000-0001-8723-1431

Paul R. Burley
Northwestern University, Northwestern University Libraries, Acquisitions and Metadata Services,
Evanston, IL, EUA
p-burley@northwestern.edu
https://orcid.org/0000-0001-7365-7482

João Alexandre Peschanski
Wiki Movimento Brasil
São Paulo, SP, Brasil
joalpe@wmnobrasil.org
https://orcid.org/0000-0002-2352-1787

A lista completa com informações dos autores está no final do artigo

ABSTRACT

Objective: The objective of this work is to identify competencies in the construction and analysis of individual and academic institutional profiles using Wikidata.
Methods: We present a case study utilizing structured semantic data related to the scholarly output of researchers at NeuroMat, a research institution in Brazil. The case study illustrates how Wikidata was used to construct and analyze individual and academic institutional profiles.
Results: We highlight: (1) enriching bibliographic data related to the published scientific research of members at Research, Innovation and Dissemination Center for Neuromathematics (NeuroMat); (2) harvesting publication and author metadata from external databases into structured, linked data for Wikidata; (3) augmenting, editing, and querying NeuroMat author and publication data in Wikidata; (4) utilizing NeuroMat data in Wikidata to create visualizations for institutional and individual analyses of scholarly production. These visualizations will demonstrate publication impact and connections between authors and institutional activities at NeuroMat.
Conclusions: We illustrate that Wikidata, as a free and open linked data platform, allows Global South institutions to decenter themselves from the hegemony of closed, proprietary Global North scholarly publishing and metadata.
KEYWORDS: Wikidata. Global South. Semantic Web. Academic profiles. Scholarly publishing.

RESUMO

Objetivo: O objetivo deste trabalho é identificar competências na construção e análise de perfis individuais e institucionais acadêmicos usando o Wikidata.
Métodos: Apresentamos um estudo de caso utilizando dados semânticos estruturados relacionados à produção acadêmica de pesquisadores do NeuroMat, uma instituição de pesquisa no Brasil. O estudo de caso ilustra como o Wikidata foi usado para construir e analisar perfis individuais e institucionais acadêmicos.
Resultados: Destacamos: (1) o enriquecimento de dados bibliográficos relacionados à pesquisa científica publicada pelos membros do Centro de Pesquisa, Inovação e Difusão Científica em Neuromatemática (NeuroMat); (2) a coleta de metadados de publicações e autores de bases de dados externas em dados estruturados e vinculados para o Wikidata; (3) o aumento, edição e consulta de dados de autores e publicações do NeuroMat no Wikidata; (4) a utilização dos dados do NeuroMat no Wikidata para criar visualizações para análises institucionais e individuais da produção acadêmica. Essas visualizações demonstrarão o impacto das publicações e as conexões entre autores e atividades institucionais no NeuroMat.

Conclusões: Ilustramos que o Wikidata, como uma plataforma de dados vinculados livre e aberta, permite que instituições do Sul Global se desvinculem da hegemonia das publicações acadêmicas e metadados fechados e proprietários do Norte Global.
PALAVRAS-CHAVE: Wikidata. Sul Global. Web Semântica. Perfis acadêmicos. Publicação acadêmica.

1 INTRODUCTION

The Research, Innovation and Dissemination Center for Neuromathematics (NeuroMat) is an interdisciplinary research center created in 2013 and located at the University of São Paulo. It is dedicated to integrating mathematical modeling and theoretical neuroscience. Among the core missions of NeuroMat is the creation of a new mathematical system to understand neural data and the development of neuroscientific open-source computational tools. NeuroMat also promotes the dissemination of knowledge and open science (Santos, 2019). In order to achieve their objectives, we suggest some initiatives that could be implemented by the institution, namely, the creation of mechanisms to enable visualization, access, enrichment, and documentation of data and metadata obtained from scholarly data of the NeuroMat scientific production. This data includes traditional periodical indexing elements, such as author(s), title(s), name of publication, dates, enumeration, and subject terms. Wikidata can be used as a subject schema; it includes additional elements such as academic journal rating, geographic location of work, abstract, copyright information, funding sources, author networks, and teaching affiliations.

Vught and Huisman (2014) define an institutional profile to set out “what the institution does, how good it is at it and how it can be compared to other institutions.” They divide the institutional profile into activity profiles, which are descriptive; and performance, which are evaluative. We define the creation of an academic institutional profile as the process of structuring information about scholarly works via metadata related to institutional intellectual output. We aim to create an institutional profile as an academic narrative of the institution's scholarly work, which may allow us to present previously selected and organized scholarly data by following related events or experiences according to our interest in scholarly dissemination. In the present work, we will create NeuroMat's institutional academic profile by using the mechanisms offered by the Wikidata platform.

We explore Wikidata as a platform that allows open and free metadata publication, in contrast to the closed, proprietary systems of current academic publishing. We look to the Wikidata platform to analyze NeuroMat scholarly publication output for data analysis and visualization for institutional data not available on academic publishing platforms. This work

highlights the capabilities of both Wikidata and linked data in general to stimulate insights into NeuroMat research and documentation of the research progress of the institution. We note, tangentially, that Wikidata operates outside of Global North publishing and established metadata platforms; it also provides a platform that is context-specific and accessible to Global South institutions.

2 MOTIVATION

The goal of this work is to provide mechanisms to structure semantic data in Wikidata to analyze the relevance of research of scientific production at NeuroMat via publishing. We focus on the institutional production carried out by its researchers rather than on their individual activities. We also aim to present, in a clear and objective way, the evolution of scientific works of the institution. This makes it possible to understand the development and connections among publications. It is essential to create queries based on specific bibliographic and author data. In view of these aims, we identified several limitations in existing scholarly publication data management services routinely used to analyze NeuroMat's scientific production, e.g., Google Scholar, etc.

The user interface of Wikidata is simplistic, has a low barrier to access, and presents a low learning curve for the entry of data elements into its database. In contrast, there is little documentation available about the Google Scholar API and its functionality is likewise limited (Mingers; O'Hanley; Okunola, 2017). Similarly, we identified many of the same limitations in other institutional analysis tools. Semantic Scholar, Microsoft Academic, and Scopus are closed data platforms; their rigidity of input and access do not offer the possibility for detailed assessments of pre-established queries as standards in academic data management tools.

In this paper, we introduce the Wikidata platform for the analysis of institutional academic profiles. Wikidata is presented in the literature as a flexible tool for graphing, data modeling, and analysis with great potential for academic data analysis. Wikidata can positively impact the dissemination of NeuroMat research because it is the central data repository for all Wikimedia projects. It can enable the connection of NeuroMat data with a large number of libraries, structured vocabularies, and other types of bibliographic identifiers. Wikipedia, and increasingly Wikidata, ranks very high in Google searches. In the following section, we will explain the basic concepts behind Wikidata technology.

3 BACKGROUND

Wikidata is a platform inspired by the vision of the Semantic Web, which was designed to allow computers and people to interact over a network. The Semantic Web presents knowledge in a way that it can be processed by machine, enables data to be linked from a source to any other source, and to be understood by computers. The Semantic Web provides the opportunity to create inference rules and to manage automated reasoning both between people and between machines, so they could perform increasingly sophisticated tasks (Berners-Lee; Hendler; Lassila, 2001).

From a technical standpoint, the Semantic Web architecture is based on standard formats for exchanging data on the Internet. We highlight the model RDF (Resource Description Framework) and SPARQL (SPARQL Protocol and RDF Query Language) developed and standardized by the World Wide Web Consortium (W3C). RDF is a standard model for data interchange on the Web. All Semantic Web information is stored and represented in RDF. SPARQL is the query language of the Semantic Web. This language was specifically designed to query data across various systems, which may include multiple RDF graphs, and thus provide interoperable ways to collect data.

RDF is a generic metadata architecture that allows describing entities in the Web through standards aimed at machine processing. A basic statement in RDF is a triple of elements with the following structure: *subject*, *predicate,* and *object*. An RDF triple models the statements as a graph. A subject node, an object node, and a predicate arc are structured from the subject node to the object node. An RDF assertion expresses a relationship between two resources. A resource can be either a concrete or abstract entity– an object, event, situation, or concept. A scientific article, a book's author, and a university are considered resources. Feelings and colors are also resources. The Semantic Web provides the identification of these resources by assigning a unique identifier called a Uniform Resource Identifier (URI).

When RDF databases (or RDF triplestore) are created together with formal semantics (i.e., a set of facts), some kind of classification, relationships, instances, and rules with the purpose of solving a problem, they are called Knowledge Bases. In other words, Knowledge Bases add a semantic model to the data, which makes it possible to interact and respond to user queries while allowing the inference of new facts, helping to narrow solutions.

A Knowledge Graph is built via a Knowledge Base and represents a collection of interlinked concept descriptions that define possible entity and property type. Knowledge Graphs represent and transmit knowledge of the real world, whose nodes represent interest entities and whose edges represent potentially different relations between these entities. They are often associated with linked open data projects and are utilized across multiple domains. They provide a consistent view that interlinks the knowledge provided by different systems and collaborators (Hogan *et al.*, 2021).

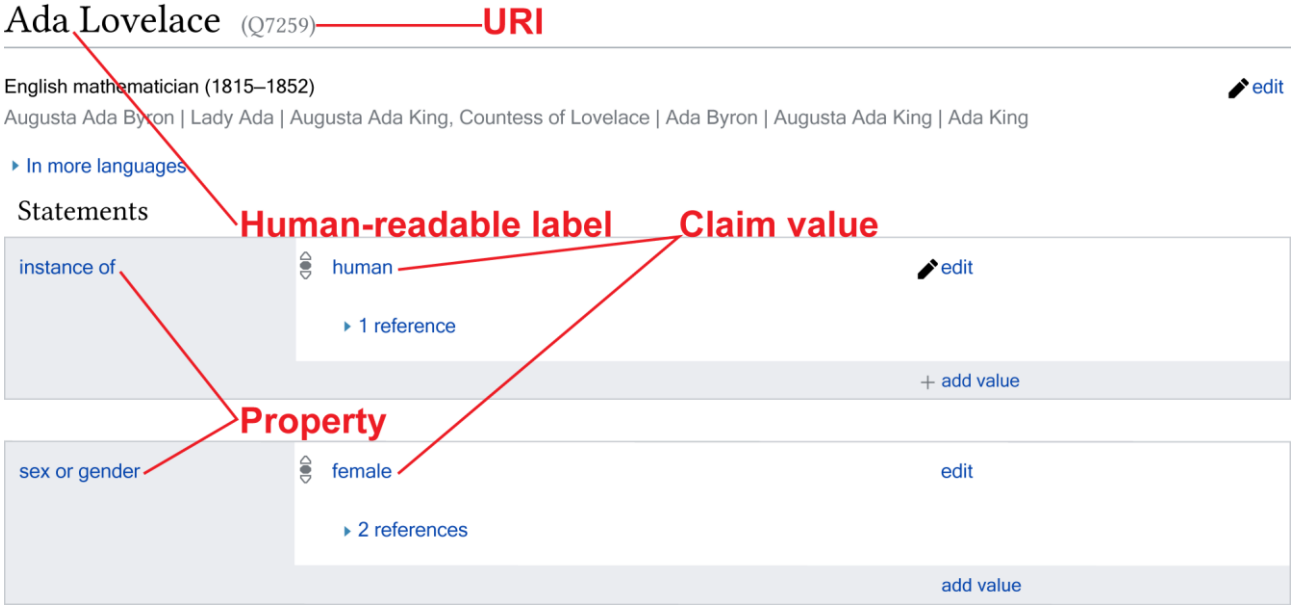
Wikidata is a collaborative Knowledge Graph hosted by the Wikimedia Foundation. It operates as the central data store for Wikimedia projects. It is a free and open-source software (Erxleben *et al.*, 2014). All data is CC-0 license. Wikidata is multilingual and represents many domains. It offers appropriate APIs (Application Programming Interface) for linking its data. Wikidata provides perhaps the simplest interface and a powerful query service for anyone to edit the Semantic Web, enabling users to visualize and use data in infinite ways (Vrandečić, 2013; Vrandečić; Krötzsch, 2014).

The Wikidata community has grown greatly, and its users have improved the platform via their experience and technical knowledge. Information in Wikidata is created, curated, and maintained by a community of about 23,000 active contributors allowing access to more than 2,000 Knowledge Bases (Veen, 2019).

A Wikidata entry consists of concepts belonging to a specific item. The item additionally has history and discussion pages to track edits and deliberate on issues related to an item, or related items if appropriate. The history page allows users to see when an edit was made (specific date and time), the editor username, and size (in bytes) of the contribution. The editing history in Wikidata thus allows the reversion, merge, or deletion of an item when needed.

Figure 1 shows the Wikidata representation of the scientist Ada Lovelace. Ada Lovelace's URI is Q7259. A Wikidata statement is a property-value pair that connects an item (entity or resource) or a literal. Statements can be enriched with qualifiers, references, or ranks. Properties are used to state any relationship between entities. Items and properties are identified by alphanumeric codes, i.e., URIs, in which a letter Q is for items, and P for properties followed by a number. For example, the URI Q5 specifies a Human entity. P31 determines the property *instance of*. Editors can add human-readable labels and descriptions to items and properties in any of the 358 languages used in Wikidata. Facts about items and properties are asserted by means of statements.

Figure 1 - A Wikidata Item with its main elements highlighted. In the picture, references are collapsed; they can be shown by clicking on the related link



Source: Piscopo and Simperl (2018).

4 RELATED WORKS

We discuss the related works following the main approaches found in the literature from the perspective of an academic database, an academic data search engine, and, finally, from its role as a bibliometric analysis tool.

4.1 Wikidata as an academic database

The proposal by Lemus-Rojas and Odell (2018) used Scholia to provide visualizations of faculty profiles and bibliographic data. Scholia is a service that creates visual scholarly profiles for topics, people, organizations, publications, and other preconfigured queries in Wikidata. Similarly, Lemus-Rojas and Pintscher (2017) claimed that the connection between libraries and their collections in Wikidata facilitated their works being discovered. They explored integration with other systems to perform broader queries, which would not be possible using only local systems. The work of Lemus-Rojas *et al.* (2022) focused on improving knowledge about women researchers at the university. They emphasized the connection between academic institutions and scholars' publications, enabling their entries to continue to be enhanced by other contributors linking them to an ever-growing network of related entities.

The approach of Alves, Burley and Peschanski (2021) introduced a process for large-scale contributions of articles from scholarly publications into Wikidata. Their work described tools and processes used in Wikidata, Zotero, and Google Sheets, focusing on cultural heritage data. This work provided a massive Wikidata-feeding methodology accessible to Global South institutions. Likewise, the work of Peschanski (2021) presented GLAM[1]-Wikis as initiatives for the collaborative digital dissemination of collections and knowledge of cultural institutions of Brazil through Wikimedia projects. They worked mainly with data from the Museum of Veterinary Anatomy, the Paulista Museum, both part of the University of São Paulo, and the National Archive of Brazil. Their work transferred almost 35,000 images to Wikimedia Commons media and hundreds of thousands of characters of text were added to the Portuguese-language Wikipedia. It included the development of tools that relied on Wikidata for scientific dissemination.

The proposal presented by Allison-Casin and Scott (2018) aimed to improve the visibility of their data through the organization of Wikipedia edit-a-thons. They recognized that Wikidata offered a low-barrier and notable results in making that data not only visible, but reusable as linked open data.

These works emphasize how a partnership between knowledge organizations and Wikidata can improve the visibility, accessibility, and reuse of metadata collection. Once metadata related to their collections are uploaded to Wikidata, networked contributors can augment, edit, and otherwise improve Wikidata items in an iterative manner. Once the data is available, however, search and retrieval may still be not so intuitive without some prior knowledge of both Wikimedia and external technology and tools.

4.2 Wikidata as an academic data search engine

The research of Tharani (2021) presents a systematic review of the application of Wikidata in libraries. The author describes several advantages of Wikidata such as data integration, interoperability, data enrichment, the improvement of catalog interfaces, and the visualization of multilingual metadata. The research and proposal elaborated by Samuel (2021) presents Wikidata as a robust, intuitive, and standards-based query engine to retrieve and visualize structured data, as well as improving the online presence and discoverability of scholarly data. The author explained how Wikidata manages to amplify the

[1] GLAM is an acronym for “galleries, libraries, archives, and museums.”

reach of existing collections through Wikipedia and utilize digital assistants to use linked structured data to answer questions. These are balanced against the willingness of library administrations to utilize Wikidata-based metadata for library operations and workflows.

Allison-Cassin et al. (2019) presents recommendations for GLAMs on the use of Wikidata and highlights opportunities for research on library involvement, community-based collections, community-owned infrastructure, and collective collections. The study considers engagement with Wikidata as a competitive way to invest institutional resources.

Clark, Williams and Rossmann (2022) studied how metadata can represent an organization equitably and how this improves the reach of global information communities. Their analysis of knowledge graphs implementation demonstrated that the discovery and interpretation of academic data were improved by bots and search engines that index data. Metadata can increase the visibility of libraries and extend their reach to establish equitable access to research and data. In a wider sense, it points to improved accessibility and scope of cooperative global information communities.

Peschanski (2021) introduced Wikidata-based GLAM content that had a reach of almost 50 million views on average per month in 2019. It was inspired by the application of structured narratives with information extracted from Wikidata. Structured narratives are based on automated texts created from the fusion of predetermined templates with pre-established Wikidata query data (Azzellini; Peschanski; Paixão, 2019). They also augment museum content production, and provide information such as location, collection metrics, number of annual visitors, current curator, architectural style, awards received, heritage situation, and other data points.

The above studies demonstrate the potential of semantic search via Wikidata. SPARQL queries provide an interconnection between heterogeneous vocabularies. This supports the elimination of ambiguity, and, thus, improves the quality of search results. Against this backdrop, we propose the analysis of scholarly output in Wikidata for numerous purposes, chiefly building individual and institutional academic profiles.

The work of Nielsen, Mietchen and Willighagen (2017) explains how the Wikidata Query Service (WDQS) and Scholia can provide many different scientometrics analyses of the data available in Wikidata. The work presents numerous properties on Wikidata for deep linking to external resources, e.g., for DOI, PMID, PMCID, arXiv, ORCID, Google Scholar, VIAF, Crossref funder ID, ZooBank, and Twitter. It shows that Wikidata can act as a hub for scientometrics studies in resources.

The bibliographic data in Wikidata are still quite limited, but it will likely continue to grow considerably in the future. In this regard, eventually, it may be a necessity to use personalized bibliometric indicators for the analysis of specific interests.

5 METHODOLOGY

This work is exploratory research, with a main objective to enable the construction of an institutional academic profile of NeuroMat. We present our methodology in three phases. The first phase outlines structuring data and metadata of academic publications. This phase presents the steps of our method of semantic structuring of academic information from NeuroMat in Wikidata, some of which we were able to automate. The second phase is a process of linking the metadata that needed to be built or refined before to be associated with scholarly publications via Wikidata. The third and final phase presents an analysis with proposed bibliometric indicators based on the potential of functionality and analytical outcomes of Wikidata.

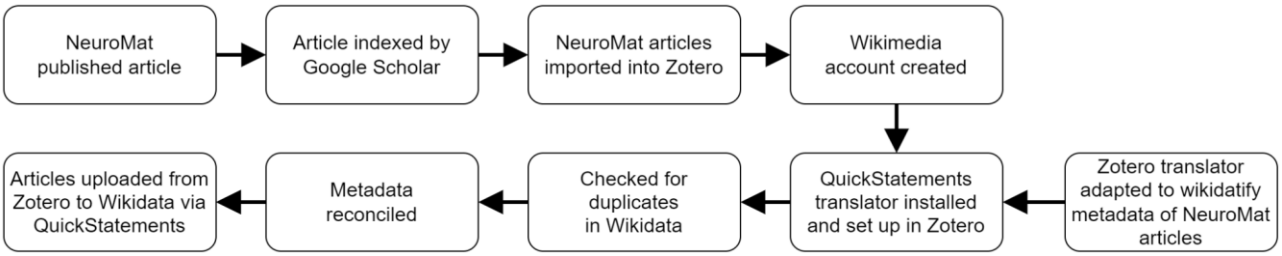
The reconciliation of metadata is a process to create an accurate correspondence between metadata in one database (i.e., Zotero) and another (i.e., Wikidata). The objective of the reconciliation process is to guarantee the semantic integrity of the correspondence in the Zotero translator. For example, an article whose author is homonymous to another one requires us to decide which author is the correct one. That is done by querying the Wikidata database in search of possible duplicates of items, investigating their relations, and deciding if it is the same entity we are trying to find or if it is another one.

To enrich the bibliographic data related to NeuroMat's published scientific research, we adapted the process of ingestion of academic articles metadata into Wikidata proposed by Alves, Burley and Peschanski (2021), as illustrated by Figure 2. The adaptation involved modifying the Zotero (2006) translator to reconcile author names, journals of publication, and subject terms of each article with their corresponding Wikidata identifier (QID). This allowed us to map important textual metadata of the NeuroMat bibliographic production in Zotero and better structure the Wikidata item in the translation from Zotero into QuickStatements format. QuickStatements (Manske, 2013) is a tool that can edit and create Wikidata items from a set of text commands, written in a specific syntax that relates the item being edited, the property being added, and its value to produce a semantic relation in Wikidata. The new article added to Wikidata is created with information that has been

correlated between Zotero and Quickstaments. However, the technical linkage between these tools is still quite new, and only some essential metadata is linked between the two platforms.

We added a key Wikidata property, "*part of* (P361)", to identify the relation between NeuroMat articles and the institution. The Wikidata property "*published in* (P1433)" has also been linked to all articles. This attribute declares the journal or conference proceedings where the article was published. Additionally, Wikidata content is available, among other means, through an API for Google Sheets. We created online spreadsheets to manage the progress of the articles created. Using this API allows us to collaborate with other users in real time. More information about this tool is outlined in the work of Steiner (2016). To reconcile articles cited, whose identifier is "*cites work* (P2860)", in a NeuroMat article, we used a mixture of manual and automatic methods using Python scripts. We used the articles' unique identifiers, such as DOI, ISBN, PMIDs, and arXiv IDs to ensure that no duplicate item was created on Wikidata. There are more than ten thousand articles cited in the NeuroMat academic production and this reconciliation is still in progress.

Figure 2 - Ingesting NeuroMat scholarly articles into Wikidata.



Source: Produced by authors.

Wikidata has tools that can assess the completeness of an individual article or set of articles. The completeness of metadata of a Wikidata item is directly related to the accuracy and efficiency of its discovery, retrieval, and reuse. Figure 3 shows the completeness of the main properties of NeuroMat articles. For instance, the *Count* column shows all the 463 NeuroMat articles registered on Wikidata. All the articles have the properties "*part of*", "*instance of*" and "*title*" filled. Figure 4 shows the completeness of NeuroMat properties for each article; blue cells have the value filled, and red cells represent that the value is missing or is not applicable.

Metadata is essential to the organization of and access to information in a digital environment. Structured metadata related to scientific production aims to minimize the risk of duplicate information and maximize the likelihood of data persistence and accessibility.

More information about the reconciliation of NeuroMat publications can be found at the project GitHub repository (Suca, 2021).

Figure 3 - Completeness of the main properties of NeuroMat scientific production in Wikidata

Name	Count	title (P1476)	author (P50)	author name string (P2093)	published in (P1433)	main subject (P921)	cites work (P2860)	DOI (P356)
NeuroMat (Q18477654)	463	100.0% (463)	96.54% (447)	32.4% (150)	99.14% (459)	35.21% (163)	69.33% (321)	77.54% (359)
Totals (all items)	463	100.0% (463)	96.54% (447)	32.4% (150)	99.14% (459)	35.21% (163)	69.33% (321)	77.54% (359)

Source: Wikidata:GLAM/CEPID NeuroMat/Completeness of properties (2021a).

Figure 4 - Visualization of the completeness of the main properties in NeuroMat articles

label	title	author	author name string	published in	main subject	? P2860	DOI	instance of	part of	page(s)	language of work or name
2D Quantitative Imaging of Magnetic Nanoparticles by an AC Biosusceptometry Based Scanning Approach and Inverse Problem	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
A Bayesian approach for convex combination of two Gumbel-Barnett copulas	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✗
A Naturalistic Assessment of the Organization of Children's Memories Predicts Cognitive Functioning and Reading Ability	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✗

Source: Wikidata:GLAM/CEPID NeuroMat/Completeness of properties per publication (2021b).

In the last phase we began the analysis using basic counting indicators such as the number of articles per author, the number of authors per article, the number of citations per article, and other metrics. Next, we researched in particular the perspective of evaluating scientific production in the Brazilian context (Machado et al., 2021). We observed the importance of the Qualis CAPES system in academia and the lack of this data element in other platforms. We decided to add this property to the articles created. Finally, we carried out several tests, taking advantage of Wikidata's graphical potential, and presenting the data

in a way that would maximize the value of the information and help to create insights as much as possible.

6 RESULTS

We can improve the indexing and online discovery of an article by the linkage of bibliographic metadata. For example, the article *Infinite Systems of Interacting Chains with Memory of Variable Length—A Stochastic Model for Biological Neural Nets* (Galves; Löcherbach, 2013), the cornerstone of NeuroMat's research, has a unique identifier (URI) of https://www.wikidata.org/wiki/Q56592766 and includes all elements of full-level bibliographic description. The Wikidata property "instance of" indicates that the item is a scholarly article; the property “published in” that it is a bibliographical component part of the publication Journal of Statistical Physics; and the property "part of" that the work is associated with NeuroMat research production. The property "cites work" has values for the Wikidata URIs of bibliographical citations in the article, in the specific order that they appear in the article.

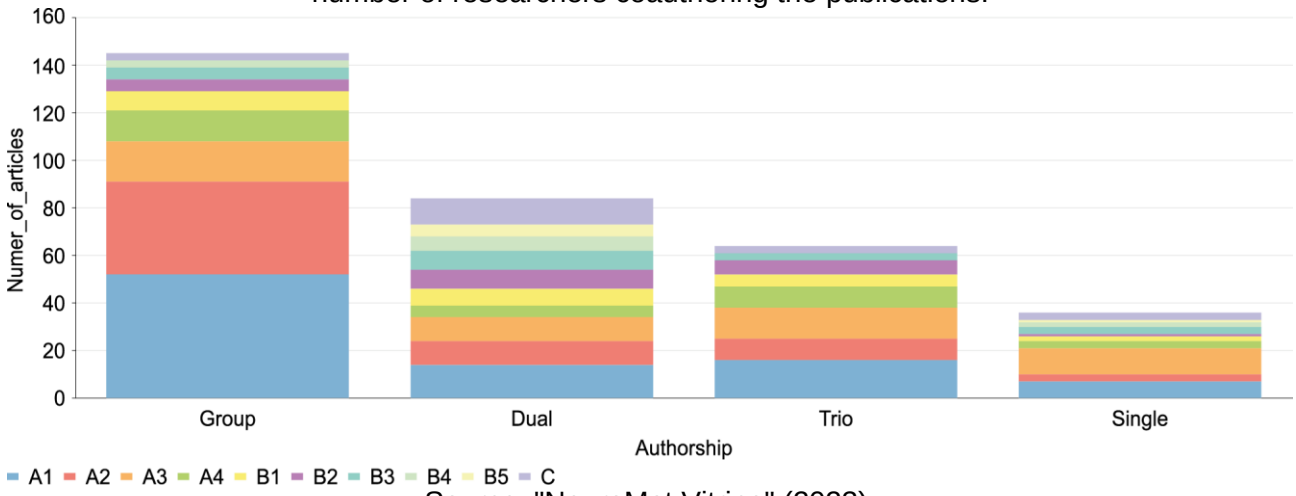
The Wikidata item, importantly, utilizes the "*main subject"* property to provide subject analysis of the article. It is defined in Wikidata as the *primary topic of a work*. Values for the *"main subject"* property are drawn from Wikidata itself; these subject terms are, therefore, drawn from an unstructured set of vocabulary terms. The subject terms, by extension, reflect the existing and emerging lines of research in NeuroMat's works over time. We note that the "*main subject*" property is the most relevant for analysis of the evolution of NeuroMat's research, despite the time-consuming nature of its production in Wikidata.

Figure 5 shows the number of articles with different Qualis Rank scores per type of authorship, and identifies the number of authors per article. Qualis Rank is a Brazilian system created and maintained by the Coordination for the Improvement of Higher Education Personnel (CAPES) to evaluate local and international academic journals. The Qualis Rank is a reference for the evaluation of postgraduate programs called *stricto sensu* modality in Brazil. These postgraduate programs are usually focused on research development and teaching; the average duration is two years for a master's, and four years for a doctorate. Qualis uses three bibliometric indicators to rank an academic journal: the Cite Score, the Impact Factor, and the h5 index. The Qualis classification model is not absolute and is constantly evolving. The last current rank is from 2019 (CAPES, 2020) and

ranges from the highest ratings of A1, A2, A3, and A4 to the lowest ratings of B1 to B4 and C, according to the Qualis 2019 classification system.

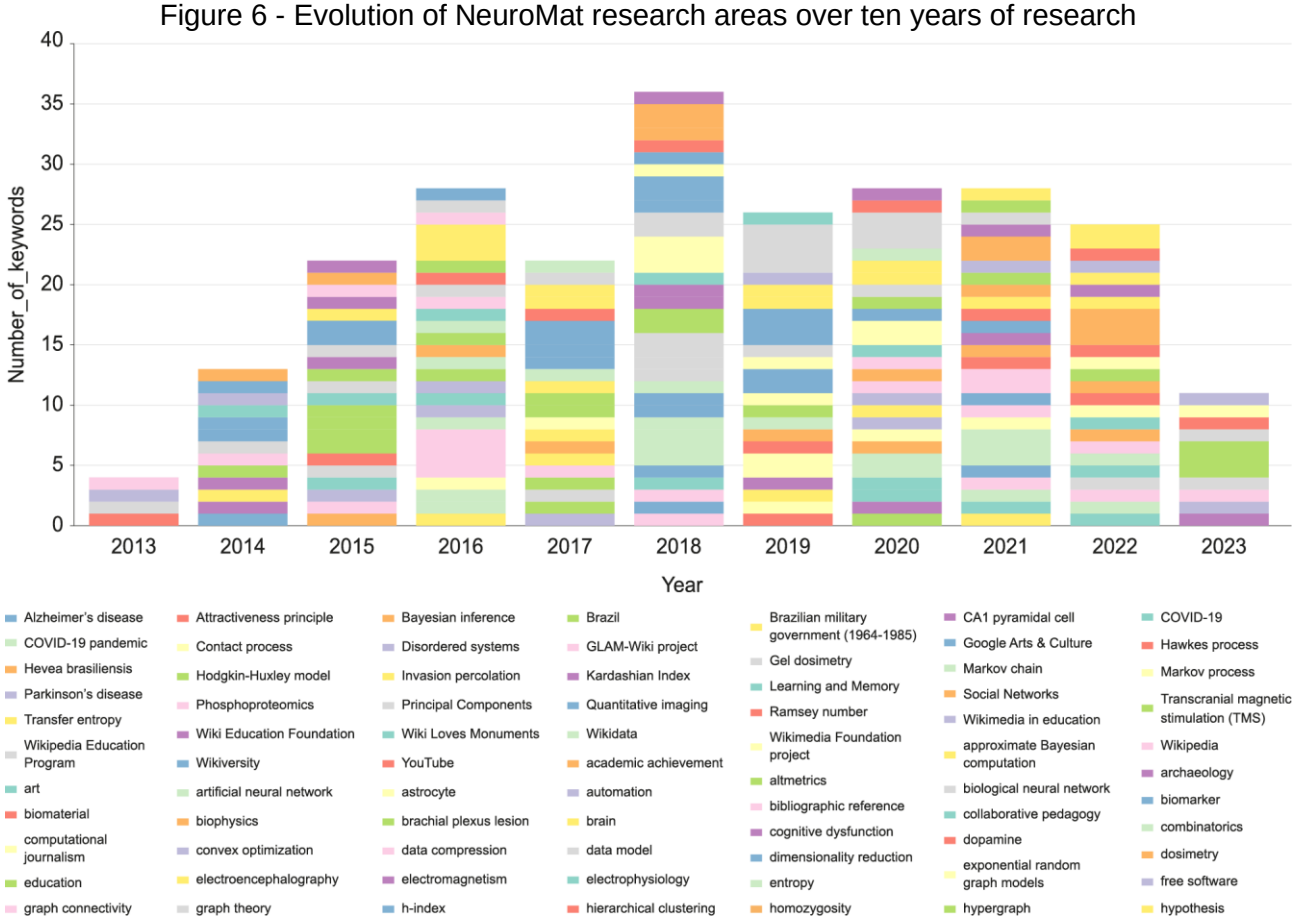
We proposed the property "Qualis Rank" on Wikidata and assigned it to all journals that have published NeuroMat articles. In addition, we registered at least 160 researchers: 7 associate, 23 master, 114 doctoral, and 42 post-doctoral. We correlated information such as the start and end date as a researcher linked to NeuroMat, the place of study (in Brazil or abroad), and the gender of the author, among other professional and biographical information. We added the *Curriculum Lattes* property to every NeuroMat author in an effort to complement the items and connect the item with other information structured databases. The Curriculum Lattes platform integrates various curricular databases, research groups, and science and technology institutions into a single information system that operates in Brazil. Complementary properties of items related to NeuroMat publications, such as the ones mentioned above, helps us accurately reconcile and retrieve information.

Figure 5 - Types of authorship in articles from Qualis A1, A2, A3 and A4 journals, defined by number of researchers coauthoring the publications.



Source: "NeuroMat Vitrine" (2023).

Figure 6 - Evolution of NeuroMat research areas over ten years of research

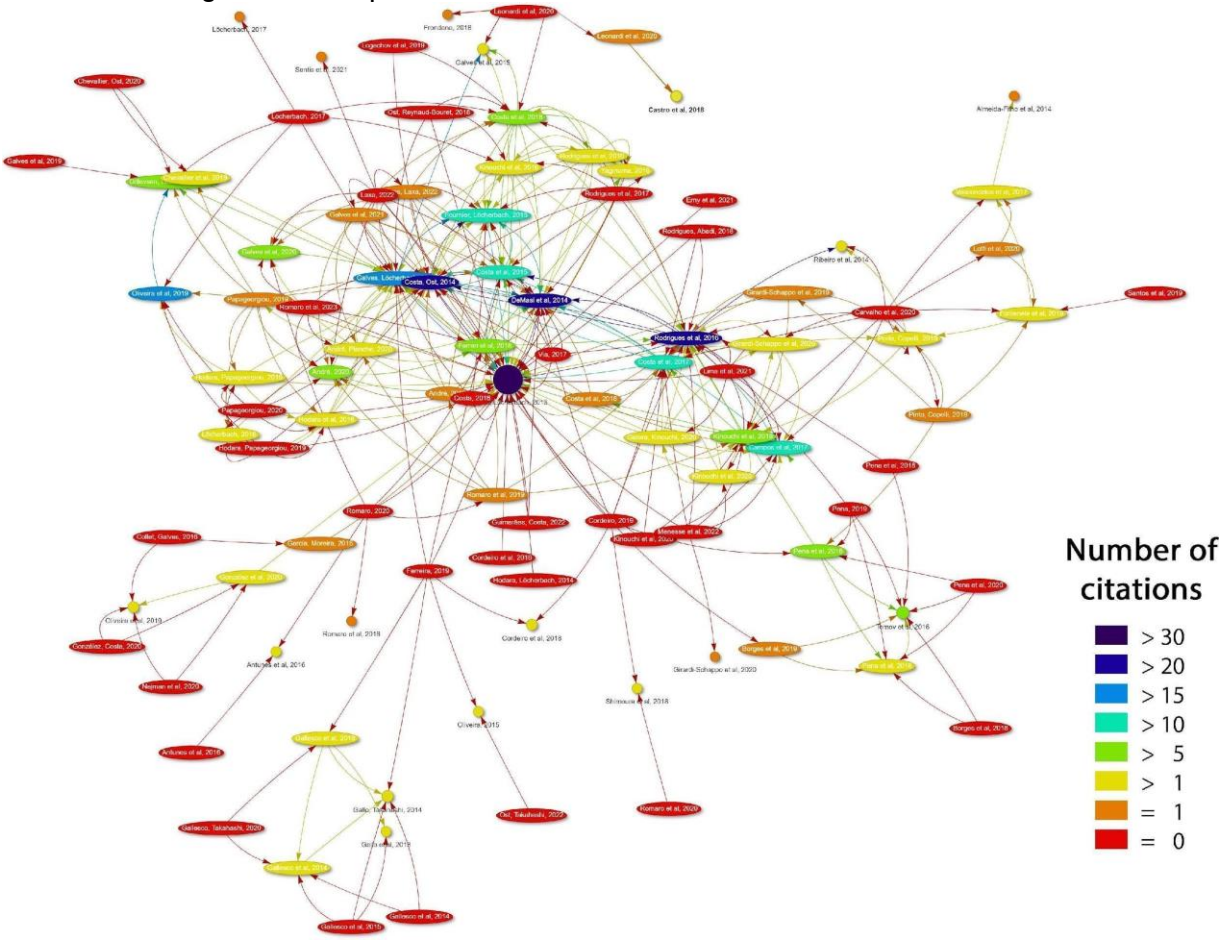


Source: "NeuroMat Vitrine" (2023).

We have proposed bibliometric indicators and complementary properties that help us visualize, explore and analyze the institution's scholarly metadata in order to retrieve information about NeuroMat research and better understand innovations in its research areas, influence of authors, and relevance of articles. Figure 6 illustrates the evolution of research areas and research topics developed by NeuroMat since 2013. Among the most frequent topics in NeuroMat publications are spiking neural networks, random graphs, Markov process, COVID-19, and other related fields.

Figure 7 shows the citations graph of NeuroMat publications that cite the seminal work of *Infinite Systems of Interacting Chains with Memory of Variable Length—A Stochastic Model for Biological Neural Nets*. The color-coded nodes of the graph represent the number of internal citations received by each article: red indicates no internal citation; orange exactly one citation; and increasing shades of yellow, green and blue indicate progressively higher number of publications. In Figure 8, we can see a graph illustrating co-authorship, and thus collaborative networks within NeuroMat.

Figure 7 - The publication network of NeuroMat foundational article



Source: "NeuroMat Vitrine" (2023).

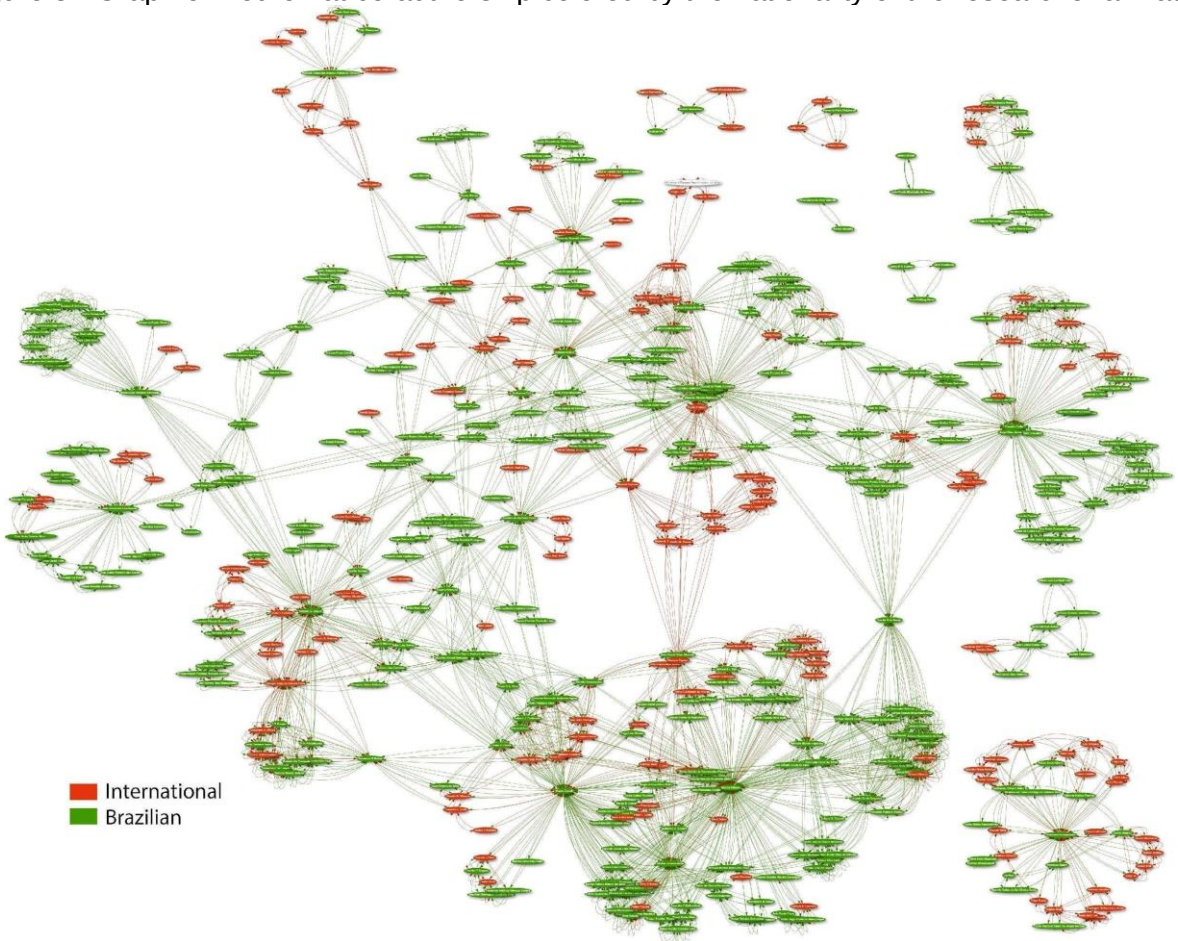
The graph presents groups of authors with common scientific interests and demonstrates that collaborative research takes place within a multidisciplinary network of research areas. The color-coding in the nodes of the graph represents the type of institutional affiliation of the author: green when they have at least one Brazilian connection, and orange if they only have international affiliations. Overall, these figures offer a valuable glimpse into the collaborative networks of authors and citations, which provides useful analysis on researchers, institutions and stakeholders that participate in NeuroMat work.

Wikidata has an internal query service to search, retrieve, and visualize data, as seen in the charts and figures above[2]. Wikidata also has integration with external tools, for example, Scholia, a tool that creates visual scholarly profiles for authors, institutions, and other topics using bibliographic metadata. We used Scholia to see the organizational profile

[2] A separation of Wikidata's scholarly data from the main query service is scheduled for February 2025. This will address challenges related to managing large volumes of data in the current system, as scholarly articles represent nearly one third of the items on Wikidata. Links on the Vitrine will be updated as this separation unfolds.

for NeuroMat ([2023]), which includes several predefined and applicable queries. It is updated automatically when new information is added to Wikidata, and can be embedded into other web-based applications with the use of iframes or plain text.

Figure 8 - Graph of NeuroMat co-authorship colored by the nationality of the researcher affiliation



Source: "NeuroMat Vitrine" (2023).

Based on this feature and the work of (Azzellini; Peschanski; Paixão, 2019), we created a webpage that uses structured narrative templates populated with information extracted from Wikidata, called "NeuroMat Vitrine" (2023). Its purpose is to summarize and display the institution's history; demonstrate the impact and relevance of its research, national and international collaboration; and present statistical information and interpretation. The page is updated automatically as soon as new information becomes available on Wikidata. At present, NeuroMat Vitrine is only available in Portuguese.

7 CONCLUSION

The NeuroMat Wikidata case study highlighted bibliometric analysis institutional academic profiles in a linked data environment. We viewed Wikidata from various perspectives: as an academic database, as an academic data search engine, and as a bibliometric analysis tool.

Wikidata, as an academic database, is a powerful platform for sharing and discovering institutional bibliographical data. It presents great potential for advanced linked data integration. Wikidata enables relatively easy and straightforward data enrichment, which is a process that may be unaffordable, too technically difficult, or inaccessible in the traditional library environment. Wikidata, most importantly, holds the potential for shared, dynamic data amongst the full international library community. This is not possible in current library platforms, and is unlikely a goal of any library system vendors at present.

Wikidata, as an academic data search engine, is an intelligent digital content infrastructure. It can enrich relationships among academic data, metadata, and other bibliographic identifiers. Wikidata provides a low barrier to entry for libraries to adopt linked open data. It offers a query service that uses standard computational language of the Semantic Web (SPARQL) to perform queries that allow implementing personalized searches. Its limitations are noted in this area, but the maturation of data models, growth of tools, and sustained international cooperation hold great promise.

Wikidata, as a bibliometric analysis tool, enables the construction of indicators such as the network of co-authors, citation network, types of authorship, and other elements illustrated above. Wikidata additionally, in the Brazilian context, empowers the implementation of a region-specific bibliometric indicator (Qualis Rank). This analysis would be very difficult to obtain from other platforms due to the rigidity of the analysis possibilities with predetermined indicators.

8 FUTURE WORKS

We note, first and foremost, the need for research, development, and documentation of Wikidata bibliographic models across all publication types, both within the platform and in alignment with established international bibliographic standards. Wikidata has existed for only eleven years, in contrast to more than a century of established indexing practices. The

development of bibliographic models in Wikidata will, by necessity, occur with consensus among both Global North and Global South GLAM institutions, researchers, and existing national and international practices. We recognize that this has caused conflict within the Wikidata community, but look to the internationalization of GLAM metadata environments and new levels of cooperation.

We note the technical and resource limits of populating all elements for NeuroMat publication data into Wikidata in the case study. NeuroMat publications have approximately 10,000 references, some without readily available metadata. The case study illustrates the full complexity of metadata harvesting and reconciliation. Wikidata has attracted ambitious projects among international GLAMs. The NeuroMat case study, in contrast, sharply illustrates that Wikidata projects require significant resources–namely, staff time, knowledge of the platform, and the expectation of work-arounds for tools not yet developed on the platform.

We note the possibility of analyzing the generation of academic genealogy: the study of intellectual heritage that is perpetuated through academic mentoring relationships. We look at the possibility to identify wider academic trends, research leadership nodes, communities of authors, and degrees of participation of different types of researchers.

The NeuroMat case study echoes Machado, Rodrigues and Barros's (2021) call for approaches to evaluate research from the perspective of social impact, notably by creating indicators for measuring the level of openness of science production at the institution.

We believe further exploration of analysis of data and visualization in Wikidata can reveal levels of diversity in gender, ethnicity and language, topics historically ignored in conventional Global North contexts.

We look to a broader replication of the NeuroMat case study in other language and subject domain environments. Wikidata, as a free and open platform, has emerging yet strong communities outside Global North contexts. We especially note Wikidata's unique multilingual component, currently seen in few other GLAM metadata platforms; and more importantly, a platform not reliant on anglophone communities of metadata production and practices.

ACKNOWLEDGMENTS

This research was supported by FAPESP (grant 2013/07699-0). We appreciate the critical reviews of earlier versions of our work provided by participants at WikiCon Brasil, 2022 and the Primer Congreso Internacional Wikimedia, Educación y Culturas Digitales, 2023. We also thank the anonymous reviewers for their insightful suggestions.

REFERENCES

ALLISON-CASSIN, S. *et al*. **ARL white paper on Wikidata: opportunities and recommendations**. Washington, DC: Association of Research Libraries, 18 Apr. 2019. Available at: https://apo.org.au/node/254221. Accessed: 4 June 2024.

ALLISON-CASSIN, S.; SCOTT, D. Wikidata: a platform for your library's linked open data. **The Code4Lib Journal**, [*S. l.*], n. 40, 4 May 2018. Available at: https://journal.code4lib.org/articles/13424. Accessed: 4 June 2024.

ALVES, É. P. F.; BURLEY, P. R.; PESCHANSKI, J. A. Structuring bibliographic references: Taking the journal Anais do Museu Paulista to Wikidata. *In*: BRIDGES, L. M.; PUN, R.; ARTEAGA, R. A. (Orgs.). **Wikipedia and Academic Libraries**. Michigan: Michigan Publishing, 2021. Available at: https://doi.org/10.3998/mpub.11778416.ch17.en. Accessed: 4 June 2024.

AZZELLINI, É. C.; PESCHANSKI, J. A.; PAIXÃO, F. J. As potencialidades de narrativas estruturadas para o Jornalismo Computacional: competências jornalísticas na elaboração de textos gerados com bancos de dados. **Texto Livre: Linguagem e Tecnologia**, Belo Horizonte, v. 12, n. 1, p. 138–152, 2019. Available at: https://doi.org/10.17851/1983-3652.12.1.138-152. Accessed: 4 June 2024.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. **Scientific American**, [*S. l.*] v. 284, n. 5, p. 34–43, May 2001. Available at: https://doi.org/10.1038/scientificamerican0501-34. Accessed: 4 June 2024.

CAPES - COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR. **Orientações sobre o processo avaliativo CAPES Ciclo 2017-2020**. Brasília: [*s. n.*], 2020. Available at: https://www.gov.br/capes/pt-br/centrais-de-conteudo/documentos/avaliacao/ORIENTAES_PROCESSO_AVALIATIVO_INFORMATIVO_1.pdf. Accessed: 4 June 2024.

CLARK, J. A.; WILLIAMS, H. K. R.; ROSSMANN, D. Wikidata and knowledge graphs in practice: Using semantic SEO to create discoverable, accessible, machine-readable definitions of the people, places, and services in Libraries and Archives. **Information Services & Use**, [*S. l.*], v. 42, n. 3–4, p. 377–390, 16 dez. 2022. Available at: https://doi.org/10.3233/ISU-220171. Accessed: 4 June 2024.

ERXLEBEN, F. *et al*. Introducing Wikidata to the Linked Data Web. *In*: MIKA, P. *et al*. (Eds.). The Semantic Web – ISWC 2014. **Proceedings**... Cham: Springer International Publishing, 2014. Available at: https://doi.org/10.1007/978-3-319-11964-9_4. Accessed: 4 June 2024.

GALVES, A.; LÖCHERBACH, E. Infinite Systems of Interacting Chains with Memory of Variable Length—A Stochastic Model for Biological Neural Nets. **Journal of Statistical Physics**, [*S. l.*], v. 151, n. 5, p. 896–921, 21 Mar. 2013. Available at: https://doi.org/10.1007/s10955-013-0733-9. Accessed: 4 June 2024.

HOGAN, A. *et al*. Knowledge Graphs. **ACM Computing Surveys**, [*S. l.*], v. 54, n. 4, p. 1–37, 2 July 2021. Available at: https://dl.acm.org/doi/10.1145/3447772. Accessed: 4 June 2024.

LEMUS-ROJAS, M.; ODELL, J. Creating Structured Linked Data to Generate Scholarly Profiles: A Pilot Project using Wikidata and Scholia. **Journal of Librarianship and Scholarly Communication**, [*S. l.*], v. 6, n. 1, 10 dez. 2018. Available at: https://doi.org/10.7710/2162-3309.2272. Accessed: 4 June 2024.

LEMUS-ROJAS, M. *et al*. Leveraging Wikidata to Build Scholarly Profiles as Service. **KULA: Knowledge Creation, Dissemination, and Preservation Studies**, [*S. l.*], v. 6, n. 3, p. 1–14, 27 July 2022. Available at: https://doi.org/10.18357/kula.171. Accessed: 4 June 2024.

LEMUS-ROJAS, M.; PINTSCHER, L.. Wikidata and Libraries: Facilitating Open Knowledge. *In*: PROFFITT, Merrilee (Ed.). **Leveraging Wikipedia: Connecting Communities of Knowledge**. Chicago: American Library Association, 2017. Available at: https://hdl.handle.net/1805/16690. Acesso em: 4 June 2024.

MACHADO, R. N.; RODRIGUES, K. O.; BARROS, S. S. **Diálogos sobre bibliometria e cientometria**. Salvador, BA: Edufba, 2021.

MANSKE, M. **QuickStatements**. Toolforge, 2014. Available at: https://quickstatements.toolforge.org. Accessed: 4 June 2024.

MINGERS, J.; O'HANLEY, J. R.; OKUNOLA, M. Using Google Scholar institutional level data to evaluate the quality of university research. **Scientometrics**, [*S. l.*], v. 113, n. 3, p. 1627–1643, dez. 2017. Available at: https://doi.org/10.1007/s11192-017-2532-6. Accessed: 4 June 2024.

NEUROMAT Vitrine, 2023. Available at: https://vitrine.numec.prp.usp.br. Accessed: 4 June 2024.

NEUROMAT, [2023]. *In*: SCHOLIA. Available at: https://scholia.toolforge.org/organization/Q18477654. Accessed: 4 June 2024.

NIELSEN, F. Å.; MIETCHEN, D.; WILLIGHAGEN, E. Scholia, Scientometrics and Wikidata. *In*: BLOMQVIST, E. *et al*. (Eds.). **The Semantic Web: ESWC 2017 Satellite Events**. Lecture Notes in Computer Science. Cham: Springer International Publishing,

2017, v. 10577, p. 237–259. Available at: https://doi.org/10.1007/978-3-319-70407-4_36. Accessed: 4 June 2024.

PESCHANSKI, J. A. Variedades de processos de difusão digital colaborativa: descrição e análise de iniciativas GLAM-Wiki no Brasil. **Resgate: Revista Interdisciplinar de Cultura**, Campinas, v. 29, p. e021006, 10 June 2021. Available at: https://doi.org/10.20396/resgate.v29i1.8659966. Accessed: 4 June 2024.

PISCOPO, A.; SIMPERL, E. Who Models the World?: Collaborative Ontology Creation and User Roles in Wikidata. **Proceedings of the ACM on Human-Computer Interaction**, [*S. l.*], v. 2, n. CSCW, p. 1–18, Nov. 2018. Available at: https://doi.org/10.1145/3274410. Accessed: 4 June 2024.

SAMUEL, J. **Making Scholarly Articles Findable: Towards Ensuring F of FAIR Data Principles**. Presentation in WikidataCon 2021, 31 Oct. 2021. Available at: https://pretalx.com/wdcon21/talk/CUQXWE. Accessed: 4 June 2024.

SANTOS, J. C. F. **A ciência aberta e suas (re)configurações políticas, infraestruturas e prática científica**. Tese (Doutorado em Política Científica e Tecnológica) – Universidade Estadual de Campinas, Campinas, 2019. Available at: https://doi.org/10.47749/T/UNICAMP.2019.1083089. Accessed: 4 June 2024.

STEINER, T. **Wikipedia Tools for Google Spreadsheets**. In: Arxiv. 2016. Available at: DOI https://doi.org/10.48550/arXiv.1602.02506. Accessed: 4 June 2024.

SUCA, E. G. **VitrineNeuroMaT**. GitHub, 2021. Available at: https://github.com/eguetti/VitrineNeuroMat. Accessed: 4 June 2024.

THARANI, K. Much more than a mere technology: A systematic review of Wikidata in libraries. **The Journal of Academic Librarianship**, [*S. l.*], v. 47, n. 2, p. 102326, Mar. 2021. Available at: https://doi.org/10.1016/j.acalib.2021.102326. Accessed: 4 June 2024.

VEEN, T. van. Wikidata: From "an" Identifier to "the" Identifier. In: **Information Technology and Libraries**, [*S. l.*], v. 38, n. 2, p. 72–81, 17 June 2019. Available at: https://doi.org/10.6017/ital.v38i2.10886. Accessed: 4 June 2024.

VRANDEČIĆ, D. The Rise of Wikidata. **IEEE Intelligent Systems**, [*S. l.*], v. 28, n. 4, p. 90–95, July 2023. Available at: https://doi.org/10.1109/MIS.2013.119. Accessed: 4 June 2024.

VRANDEČIĆ, D.; KRÖTZSCH, M. Wikidata: a free collaborative knowledgebase. **Communications of the ACM**, [*S. l.*], v. 57, n. 10, p. 78–85, 23 Sep. 2014. Available at: https://doi.org/10.1145/2629489. Accessed: 4 June 2024.

VUGHT, F. van; HUISMAN, J. Institutional Profiles: Some Strategic Tools. **Tuning Journal for Higher Education**, [*S. l.*], v. 1, n. 1, p. 21, 7 Apr. 2014. Available at: https://doi.org/10.18543/tjhe-1(1)-2013pp21-36. Accessed: 4 June 2024.

WIKICONECTA; Editando o Wikidata. *In*: **WikiConecta, a Wikidiversity Course**. Available at:

https://pt.m.wikiversity.org/wiki/WikiConecta/O_Wikidata/Editando_o_Wikidata. Accessed: 22 Set. 2024.

EDITING WIKIDATA. In: **Wikidata, a free and open knowledge base. The University of Edinburgh**. Available at: https://www.ed.ac.uk/information-services/help-consultancy/is-skills/wikimedia/wikidata/editing-wikidata. Accessed: 22 Set. 2024.

WIKIDATA:GLAM/CEPID NeuroMat/Completeness of properties. *In*: WIKIDATA. [San Francisco, CA: Wikimedia Foundation, 26 May 2021a]. Available at: https://www.wikidata.org/wiki/Wikidata:GLAM/CEPID_NeuroMat/Completeness_of_properties. Accessed: 4 June 2024.

WIKIDATA:GLAM/CEPID NeuroMat/Completeness of properties per publication. In: WIKIDATA. [San Francisco, CA: Wikimedia Foundation, 6 June 2021b]. Available at: https://www.wikidata.org/wiki/Wikidata:GLAM/CEPID_NeuroMat/Completeness_of_properties_per_publication. Accessed: 4 June 2024.

ZOTERO [2006]. Available at: https://www.zotero.org. Accessed: 4 June 2024.

NOTAS

AGRADECIMENTOS
Os autores agradecem a Carybé G. Silva e Mike Peel o apoio na solução de problemas técnicos na interoperabilidade de plataformas e a Camilla Tsuji Oviedo Lopes pela atuação preliminar no projeto da Vitrine NeuroMat.

CONTRIBUIÇÃO DE AUTORIA
Concepção e elaboração do manuscrito: E. G. Suca, J. A. Peschanski
Coleta de dados: E. G. Suca, É. P. F. Alves
Análise de dados: E. G. Suca, É. P. F. Alves
Discussão dos resultados: E. G. Suca, J. A. Peschanski
Revisão e aprovação: J. A. Peschanski, P. R. Burley

CONJUNTO DE DADOS DE PESQUISA

FINANCIAMENTO
Este trabalho integra o projeto FAPESP 2013/07699-0.

CONSENTIMENTO DE USO DE IMAGEM
Não se aplica

APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA
Não se aplica

CONFLITO DE INTERESSES
Não se aplica

PUBLISHER
Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no Portal de Periódicos UFSC. As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES
Edgar Bisset Alvarez, Ana Clara Cândido, Patrícia Neubert, Genilson Geraldo, Jônatas Edison da Silva, Mayara Madeira Trevisol, Edna Karina da Silva Lira e Luan Soares Silva.

HISTÓRICO
Recebido em: 14-07-2024 – Aprovado em: 28-11-2024 – Publicado em: 29-12-2024