

Open Tool Registries! Resolving the Directory Paradox with Wikidata

Till Grallert <till_dot_grallert_at_hu-berlin_dot_de>, Humboldt-Universität zu Berlin  <https://orcid.org/0000-0002-5739-8094>

Sophie Eckenstaler <sophie_dot_eckenstaler_at_uni-marburg_dot_de>, Deutsches Dokumentationszentrum für

Kunstgeschichte – Bildarchiv Foto Marburg  <https://orcid.org/0000-0001-8633-0120>

Claus-Michael Schlesinger <post_at_esthet1cs_dot_net>,  <https://orcid.org/0000-0001-6718-5773>

Nicole Dresselhaus <nicole_dot_dresselhaus_at_hu-berlin_dot_de>, Humboldt-Universität zu Berlin  <https://orcid.org/0009-0008-8850-3679>

Isabell Trilling <isabell_dot_trilling_dot_1_at_hu-berlin_dot_de>, Humboldt-Universität zu Berlin  <https://orcid.org/0009-0005-7177-0293>

Sophie Stark <sophie_dot_stark_at_student_dot_hu-berlin_dot_de>, Humboldt Universität zu Berlin  <https://orcid.org/0009-0005-6876-3404>

DOI: <https://doi.org/10.63744/c6fmeh495g4w>

Abstract

This paper introduces the conceptual framework for open and community-curated tool registries, posing that such registries provide fundamental value to any field of research by acting as curated knowledge bases about a community's past and current methodological practices as well as authority files for individual tools. The modular framework of a basic data model, SPARQL queries, bash scripts, and a prototypical web interface builds upon the well-established and open infrastructures of Wikimedia, GitLab, and Zenodo for creating, maintaining, sharing, curating, and archiving linked open data. We demonstrate the feasibility of this framework by introducing our concrete implementation of a tool registry for digital humanities, initially repurposing data from existing silos, such as TAPoR and the SSH Open Marketplace, and retaining the established TaDiRAH classification scheme while being open to communal editing in every aspect.

Introduction

The digital humanities have a strong culture of discussing their tooling. However, apart from projects documenting their workflows with fine-grained metadata, the field has not yet developed widely adopted practices for unambiguously citing and referencing such tools or keeping a record of their use and utility for particular research projects [Schindler et al. 2021] [Alireza et al. 2022] [Henny-Krahmer and Jettka 2022] [Ruth, Niekler, and Burghardt 2022] [Irrera et al. 2023]. If at all, authors mostly deploy URLs to websites or code repositories^[1], which — like all URLs — are prone to link rot and do not further contextualise a given software. Tool registries thus address an obvious and concrete need of research communities and individual practitioners for transparency and guidance. They collect basic information on software (e.g. names of developers, licenses, version numbers, dependencies, links to code repositories) and add some classification as to their purpose within the research process. In fields with rapidly evolving and complex technology stacks, such as the digital humanities, registries, done right, could serve as one-stop shops for providing a comprehensive and up-to-date overview of (computational) methods and their implementation in specific software, and thus provide a central service within the knowledge ecology. Ideally, they would form part of the semantic web that might guide researchers to projects, publications, and tutorials, and allow them to evaluate the utility of a particular tool for their own needs.

Tool Registries in the Digital Humanities

Registries of computational tools have grown into a well-established genre within the digital humanities (see [Grant et al. 2020] for a history of tool registries in DH): from DiRT (*Directory of Research Tools*) to Bamboo and the Canadian *Text Analysis Portal for Research* (TAPoR 3), from large EU projects like the *Social Sciences and Humanities Open Marketplace* (SSHOM) to DARIAH's (*Digital Research Infrastructure for the Arts and Humanities*) now defunct *Tools E-Registry for E-Social science, Arts and Humanities* (TERESAH), or, in Germany, multiple independent registries developed as part of the *National Research Data Infrastructure* (NFDI) and the Specialized Information Services (*Fachinformationsdienste* (FID) [NFDI4Objects 2025] [Registry - NFDI4Culture n.d.] [Home | nfdi.software n.d.]). In practical terms, these tool registries are of a dual nature. On the one hand, they embody a field's methodological knowledge, the archive of scholarship as *techné*,^[2] connecting information about tools to information about their use and purpose. On the other hand, registries are part of the socio-technical infrastructures of the information age to maintain and access the record of that knowledge. It is evident that, at least in the practical reality of limited resources, their aims must remain aspirational in both their content and their infrastructural implementation.

2

Most tool registries approach knowledge about *techné* and the infrastructure needed to maintain this knowledge through an archival framework, with its fundamental hierarchy and power relations between knowledge production and knowledge consumption housed and controlled by institutions largely unaccountable to the *démós* and enforced by the *archon* (chief magistrate) (c.f. [Derrida 1996] [Vismann 2009] [Ebeling 2009]). This is not to say that the archive is governed by ill will. To the contrary. But academic knowledge ecologies are both inherently conservative and hierarchical. Within such a framework, registries seem a sufficiently familiar task to which we can readily apply our well-established models and practices: pro-bono expert peer-review, grant-funded project cycles, curatorial hierarchies, and tightly controlled vocabularies and technical infrastructures. Unfortunately, these commonly result in proprietary data silos and slowly dying interfaces.

3

Quinn Dombrowski (2021) has coined the term *directory paradox* to describe the inevitable failure of such approaches to knowledge curation to answer the scholarly need for always up-to-date information in quickly developing fields. Against the backdrop of their own involvement with DiRT and the TaDiRAH classification scheme, Dombrowski argues that despite the considerable resources poured into existing tool registries, they have all failed as infrastructures and part of larger scholarly ecosystems, as long as their content and siloed technical infrastructure depend on unpaid volunteer and grant funding [Dombrowski 2021] (c.f. [Bernardou et al. 2018]). However, we do not agree with a view that “such ‘registries’ only have value if the data that underpins them is constantly updated” [Hughes, Constantopoulos, and Dallas 2015, p.156]. Instead, they can, and probably should, be considered snapshots or documentations of a field's historical practices at given points in time and space (c.f. [Grant et al. 2020]).

4

We argue that the resulting tension between researchers and library patrons seeking up-to-date information to solve an issue at hand and curators, who are painfully aware of any given catalogue entry's quality as a historical artefact bound up in the contingencies of its making, cannot be resolved. We, therefore, propose to shift our perception and start treating catalogues and registries as the imperfect but living documents they are — digital commons that need communal care and upkeep.

5

Gathering and curating information about a large and ever-growing number of tools requires substantial resources and skilled domain experts in a broad range of fields. The common solution to this problem is to build on existing bodies of knowledge or, in other words, to continuously repackage already existing data in new interfaces. Our libraries' cataloguing ledgers have been superseded by index cards, which were then scanned into a (Card) Image Public Access Catalogue (CIPAC or IPAC), transcribed into digital texts, and ingested into database systems that power OPACs (Open Public Access Catalogue) (c.f. [Oberhauser 2003]). Today, these data are feeding large commercial aggregators, and they are harvested for training large language models and retrieval-augmented generation systems.

6

SSHOM is an example of such a genealogy of tool knowledge. As part of the *European Open Science Cloud* (EOSC) under the European Commission's *Horizon 2020* programme, SSHOM was built by DARIAH, the *Common Language Resources and Technology Infrastructure* (CLARIN), and the *Consortium of European Social Science Data Archives* (CESSDA) between 2019–23 and is considered a flagship service for DARIAH. In May 2023, SSHOM hosted information on almost 1,700 items. More than 1,200 of these were ingested from TAPoR over a few days in late 2022.^[3]

7

TAPoR, in turn, has a strong focus on textual research. Its current iteration integrated yet another popular tool registry, DiRT, in 2017/18, which itself had originated from earlier tool registries such as Bamboo [Rockwell 2006] [Dombrowski 2014] [Dombrowski 2021] [Grant et al. 2020]. Our Wikidata-based tool registry, outlined on the following pages, follows this example and integrates SSHOM and TAPoR into a linked-open-data environment.

Yet another tool registry?

The conceptual and technical approach to building a Wikidata-based, open tool registry is rooted in the multidisciplinary background and diversity of job roles of those involved. As a team, we needed both a means to curate tool boxes for specific research questions and disciplinary assemblages of methods as well as an authority file for unambiguously identifying software in research outputs and lab notes. Neither requirement was satisfied by existing tool registries.

The first use case is one we frequently encounter in consulting and service contexts in the library [Grallert et al. 2024]. Digital scholarship librarians or, more broadly, digital scholarship experts offering research support services, bring their own expertise regarding specific fields of research, scientific methods, and digital tools. Library teams — including teams of one — integrate consulting services with other formats like workshops. They might even offer basic infrastructure as digital scholarship support and act as technical partners for research projects. Institutionalized teams or larger institutional contexts often work with team-specific toolsets and technological stacks, which need to be documented for further reference, e.g. when documenting use cases or research output. Researchers, in turn, seek information regarding the possibilities for implementing computational workflows adapted to specific research questions and research data along the full data life-cycle from aggregation to publication. In such contexts, curated lists of tools reflect team expertise and recommendations for tools optimized for local target groups. It is essential that contributing to such lists is fairly simple. It ought to be governed by expertise and experiences with particular tools and not tied up in institutional hierarchies.

Secondly, we were looking for a publicly-accessible authority file of tools, a single source-of-truth about software, which would allow us to unambiguously identify and reference tools employed in the course of a research project (c.f. [Irrera et al. 2023, p. 26:4]). Such an authority file would also greatly facilitate the ability to quickly and transparently curate the tool lists mentioned in our first use case without a need for manually maintaining and updating redundant information: a catalogue to follow up on methods and tools mentioned in other scholars' research output.

The approach proposed here is designed to address both use cases by providing the basic infrastructure to put together a well-curated data set for local communities and make it possible to store aggregated and enhanced data sets ready for reuse.

Design goals

The design goals for our tool registry are tailored towards librarians, researchers, and institutions. All of these groups possess domain-specific knowledge about tools and methods. At the same time, they could make good use of descriptions and documentation provided by peers. Therefore, all components of the system — data, software, documentation — shall be open, re-usable, adaptable, and sustainable without continuous access to labour and funding for specific teams of curators. Data produced through the application of modules or components shall satisfy the FAIR criteria [FAIR Principles 2020]. Users shall also be able to browse and compare data through workflow descriptions or basic modules, e.g. a web frontend with a search interface and browsing functions.

Our approach is rooted in *minimal computing* practices [Gil and Ortega 2016] [Risam and Gil 2022] and the *Endings Principles for Digital Longevity* [Endings Project Team 2023]. We acknowledge the resources at hand and the constraints under which we operate. As a small team of humanities scholars on fixed-term contracts in various grant-funded research and infrastructure projects, without access to institutional commitments beyond our own labour time, we learned “how to produce, disseminate, and preserve digital scholarship ourselves, without the help we can't get, even as we fight to build the infrastructures we need at the intersection of, with, and beyond institutional libraries and schools” [Gil and Ortega 2016, p. 29]. We optimised for sustainability and adaptive reuse by consequently separating data and interfaces into independent components, minimising complexity, and opting for openness on every level.

Relying on existing infrastructures with proven track records whenever possible reduces overhead for hardware and software to the bare minimum.

Core components

Our tool registry is built around Linked Open Data and Wikidata, which provides the socio-technical infrastructure that sustains our effort. Ultimately, our main contribution is conceptual, designing data models, queries, and workflows that instantiate a catalogue of cultural artefacts on top of Wikidata. In order to demonstrate the feasibility of the conceptual framework, we also contributed major data sets from existing tool registries to Wikidata.

14

Wikidata

Wikidata is the largest public knowledge graph with hundreds of millions of items published under a public domain license.^[4] Such a graph structures information about real-world entities following the Linked Open Data paradigm in statements, which construct relations between items, such as “This item is called ‘Gephi’”, “Gephi is a piece of software”, “Gephi can be used for social network analysis”, etc. Established in 2012 by the Wikimedia Foundation, Wikidata is part of the Wikiverse and the structured sister to Wikipedia. Wikibase, the software it runs on, is maintained as free and open-source software (FOSS) (c.f. [Vrandečić, Pintscher, and Krötzsch 2023]). Most importantly, Wikidata is also an international community of active and passionate volunteers contributing their time and knowledge to a shared knowledge ecology.

15

Wikidata addresses many of the weaknesses of existing tool registries, as outlined above. Firstly, Wikidata inherently adheres to the FAIR principles by making all data findable, accessible, interoperable, and reusable [FAIR Principles 2020]. Even better, Wikidata provides five-star Linked Open Data (LOD) [Berners-Lee 2009]:

16

1. Data are available on the web with a public domain licence (CC0);
2. Data are machine-readable structured data;
3. Data are serialised in non-proprietary formats;
4. Data adhere to open standards; and
5. Data link to other people’s data to provide context.

Data can be accessed and linked to through multiple APIs (application programming interfaces), including SPARQL endpoints, with canonical identifiers for all items and properties and permanent, versioned URIs for each edit.^[5]

17

Secondly, Wikidata is a well-established, open infrastructure with large communities of users, contributors and maintainers, which far exceeds anything we as scholars could ever achieve on an infrastructural level. As such, it is increasingly adopted as a platform to share authority data by GLAM institutions and research projects [Allison-Cassin and Scott 2018] [Odell, Lemus-Rojas, and Brys 2022] [Sardo and Bianchini 2022] [Zhao 2022] [Fagervig 2023] [Grallert 2025]. Thirdly, the underlying Wikibase software is user-centered, despite Wikidata’s seeming emphasis on data. It provides easy tools and protocols for collaboratively editing and contributing data. The threshold for contributing to a tool registry on Wikidata is thus rather low, compared with solutions that employ technical and organisational gatekeeping mechanisms like login or expert verification.

18

Relying on Wikidata as a data store positions tool data as part of the commons. All data on tools can thus become objects of distributed and collaborative efforts to aggregate, sustain, and develop knowledge on tool usage in digital humanities. Most importantly, we are not limited to this particular field. Linked Open Data and Wikidata allow us to engage and join efforts with other communities working with structured descriptions of software, e.g. due to the growing necessity for process metadata or software preservation purposes [Christophersen et al. 2023].

19

Such openness comes with its own challenges. Libraries and large infrastructure projects, as well as researchers curating information are commonly wary to cease authority over “their” data sets to volunteer-driven, radically open projects. While active misinformation campaigns do not constitute a significant problem on Wikidata, misidentification of items and erroneous edits occur on a daily basis. We strongly believe that the benefits of using Wikidata as a platform

20

and open knowledge graph outweigh the potential imperfection of individual data points by far. We argue that imperfect information is already a major “feature” of tool registries. If at all, Wikidata makes it easier to correct data points and keep information up to date. We also address the fluidity of data by saving weekly, versioned screenshots of the data that constitutes “our” tool registry to publicly-funded research data repositories.

It is important to point out that Wikidata does not remove the need for data stewardship. Somebody has to put in the necessary labour. But unlike siloed custom infrastructures, sharing, distributing, and transferring such stewardship for the commons is a central pillar of the Wikiverse.

21

Data models

We provide a basic object-oriented data model for tools as objects as a reference model. We consider this reference model as a minimal basic set that can and should be extended with domain specific and research specific elements. This means that while everyone is able to use their own models, all information added to Wikidata will immediately become part of the shared tool registry as long as these models respect the basic reference model.

22

Wikidata follows no strict ontology but instead applies an open, bottom-up approach, which is famously referred to as a *folksonomy* (a term coined by [VanderWal 2004]), with a very basic and deliberately generic data model of *items* (identifiers starting with “Q”) and *properties* (identifiers starting with “P”) connecting items to other items or *values* of various data types (strings, dates, integers ...) through *statements*. All statements can carry properties as qualifiers and references to the source for this information. A large part of the statements in Wikidata are properties linking to other platforms and authority files through external identifiers. All properties might come with a set of formal constraints, which, if violated, will result in various warning flags in the Wikidata web interface. All items should carry two special string-value properties for multilingual human-readable labels and descriptions. The only mandatory property is instance of (P31) linking to one or more items, without which an item would not be connected to the larger knowledge graph. But even this constraint is not strictly enforced on a technical level (c.f. [Hosseini Beghaeiraveri et al. 2023]).

23

Such a generic and unconstrained data model makes it hard to query Wikidata with a predictable result, but it provides a flexible system for any conceivable, community-driven data modelling: Data models exist largely through communities of practice adhering to them and providing the necessary SPARQL queries for building sub graphs. To this end, we set up a WikiProject documenting our approach and data model as a port of entry to build a community of people and institutions wanting to maintain a tool registry [Wikidata:WikiProject DH Tool Registry 2024]. A community-oriented documentation and discussion process through the WikiProject provides a channel for working on the common reference model and evolving the data model regarding mandatory and optional qualities as well as additional modules developed for specific purposes. As an added bonus, WikiProjects rank highly on search engines.

24

Our basic data model

We differentiate between a number of concepts to model the relation between a tool and its purpose:

25

1. Research **tools** comprise both **methods** and concrete **software**.
2. **Methods** are informed by **theories** and have a *purpose*.
3. **Methods** are implemented through (multiple) layers of **software**, which, in turn, require *hardware* and infrastructural resources such as electricity, internet connectivity or licences and which interact with data *formats* and *serialisations* (reading and writing).
4. **Software** is written in programming languages and can be interacted with through *interfaces*. Command line interfaces and application programming interfaces require knowledge of programming languages to interact with them.
5. **Methods**, **languages**, and **formats** rely on and implement abstract *concepts*.

The tool registry is concerned with only a subset of this larger ontology: *software* and *methods*. The basic data model for *software* requires no mandatory information in addition to the Wikidata base model of *label*, *description*, and instance of (P31) beyond the has use (P366) property that associates *software* with one or more *methods*.

26

The basic data model for *methods* is even more rudimentary. Here the only mandatory property is a TaDiRAH ID (P9309) that proclaims equivalence to a concept within “The Taxonomy of Digital Research Activities in the Humanities” (TaDiRAH, see below). Figure 1 shows a schematic of this data model conceptualising “Gephi” as an instance of “Software” that can be used for “network analysis.” Additionally, the source for this claim is identified through a reference to the URI of an entry in the SSHOM.

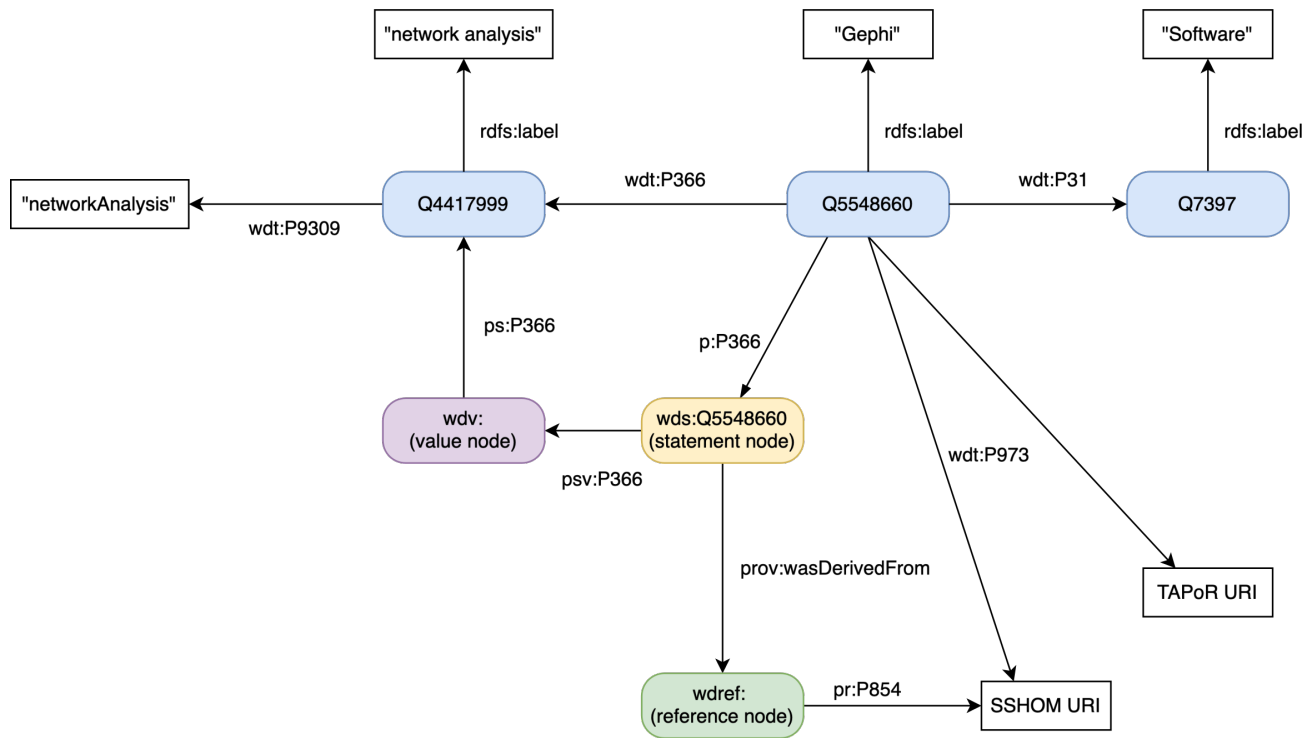


Figure 1. Schematic of our basic data model, using Gephi (Q5548660) as an example.

TaDiRAH mapping

Like all archives, tool registries depend on classification schemes and taxonomies in order to file, retrieve, and produce knowledge about an item in their collection. Over the course of the last decade “The Taxonomy of Digital Research Activities in the Humanities” (TaDiRAH) has become the most widely adopted classification scheme in the digital humanities and is used in its current version (>2.0) for author-assigned classifications of conference submissions to the *Alliance of Digital Humanities Organizations* (ADHO) and *Digital Humanities im deutschsprachigen Raum* (DHD) conferences as well as SSHOM (TAPoR uses the older, incompatible version of TaDiRAH).

TaDiRAH was developed from the mid-2010s onwards in the tradition of John Unsworth’s “scholarly primitives” ([Hughes, Constantopoulos, and Dallas 2015, pp. 155–157]; For the genealogy of TaDiRAH see [Borek et al. 2016]). Members of DARIAH-DE and BambooDiRT developed TaDiRAH on the basis of the *ICT Methodology*, which itself had evolved from the AHDS Taxonomy of Computational Methods (2003–) under the auspices of the Oxford University Digital Humanities Programme. The first version of TaDiRAH (v0.5) was released in 2016 [Borek et al. 2015] [Borek et al. 2016] as a four-level hierarchy of *goals*, *methods*, *techniques*, and *objects*. The current iteration (>v2.0) was released in 2021 [Borek et al. 2021] and has been redesigned from the ground up as a SKOS vocabulary (Simple Knowledge Organization System). Unlike earlier iterations, it only covers methods and techniques as a single type of entities, which can nest in three hierarchical layers. Therefore, v2.0 is not backwards compatible with earlier versions.

The major strengths of TaDiRAH are its wide adoption across the field of digital humanities and its integration into wider Linked Open Data (LOD) infrastructures as part of the DARIAH Vocab services, which is maintained by the *Austrian Centre for Digital Humanities and Cultural Heritage* (ACDH-CH) of the *Austrian Academy of Sciences*. In addition, a partial mapping between TaDiRAH and Wikidata has been implemented on both sides. The Wikidata property TaDiRAH ID (P9309) was created in 2015 by Adam Schiff, Principal Cataloger at the University of Washington Libraries in Seattle,

to state equivalence between a Wikidata item and a TaDiRAH class.

However, TaDiRAH has major flaws rather similar to those of tool registries:

31

1. Development depends on voluntary labour and thus the vocabulary has been dormant for a number of years.
2. The current version and its documentation have been hard to find. Until September 2025, when the team finally updated the GitHub repository to v2.0, search engines usually returned the incompatible v0.5 (see also [Zhao 2022]).
3. The SPARQL endpoint and API, which would serve the classification scheme as RDF, are frequently down.

Despite of these weaknesses, potential competitors and successors have not been successful. The “NeDiMAH Methods Ontology for the Digital Humanities” (NeMO) [Hughes, Constantopoulos, and Dallas 2015, pp. 165–167] [Bernardou et al. 2018, pp. 4–5], developed with substantial funding from the European Science Foundation by the *Network for Digital Methods in the Arts and Humanities* (NeDiMAH) from 2011 to 2015 [Hughes, Constantopoulos, and Dallas 2015] is a CIDOC CRM-compliant ontology and provided a mapping from TaDiRAH but has seen no traceable adoption.^[6] This might be due to information on NeMO — quite fittingly, given its name — being incredibly hard to come by. The official website is outdated and has not been updated with the project results, and access to the documentation, although officially licensed under “CC BY-NC-SA 4.0”, requires a login on <http://nemo.dcu.gr/resources/>.

32

Given its wide adoption, including our seeding data set of tools from SSHOM and TAPoR, we have therefore opted to retain TaDiRAH classification as the main organising principle for our tool registry. The system is flexible enough to adopt other classification systems should they emerge, as long as their identifiers are mapped to a Wikidata property.

33

The TaDiRAH classification of items is implemented through the has use (P366) property, linking to other Wikidata items that carry a TaDiRAH ID (P9309). To this end, we completed the mapping between Wikidata and TaDiRAH so that all TaDiRAH classes can now be found on Wikidata (see below).

34

Optional parts of the basic data model and curating collections of tools

The basic data model leverages the weaknesses of an open-world knowledge graph without a formal ontology to our advantage as it can be easily extended to accommodate additional information on individual items, such as licenses, version numbers, URIs of a source code repository, etc. For this purpose, our basic data model proposes a number of optional statements.

35

A core requirement for our tool registry is the ability for diverse communities to curate their own collections. This can be done through the property *collection* (P195) pointing to one or more Wikidata items describing the collection and potentially naming curators and contributors (Figure 2). We have implemented two such collections, *SSHOM* (Q131847864) and *TAPoR* (Q3979414), to document existing tool registries that formed the basis of our data set (see below). Note that such collections are not necessarily limited to tools classified with TaDiRAH.

36

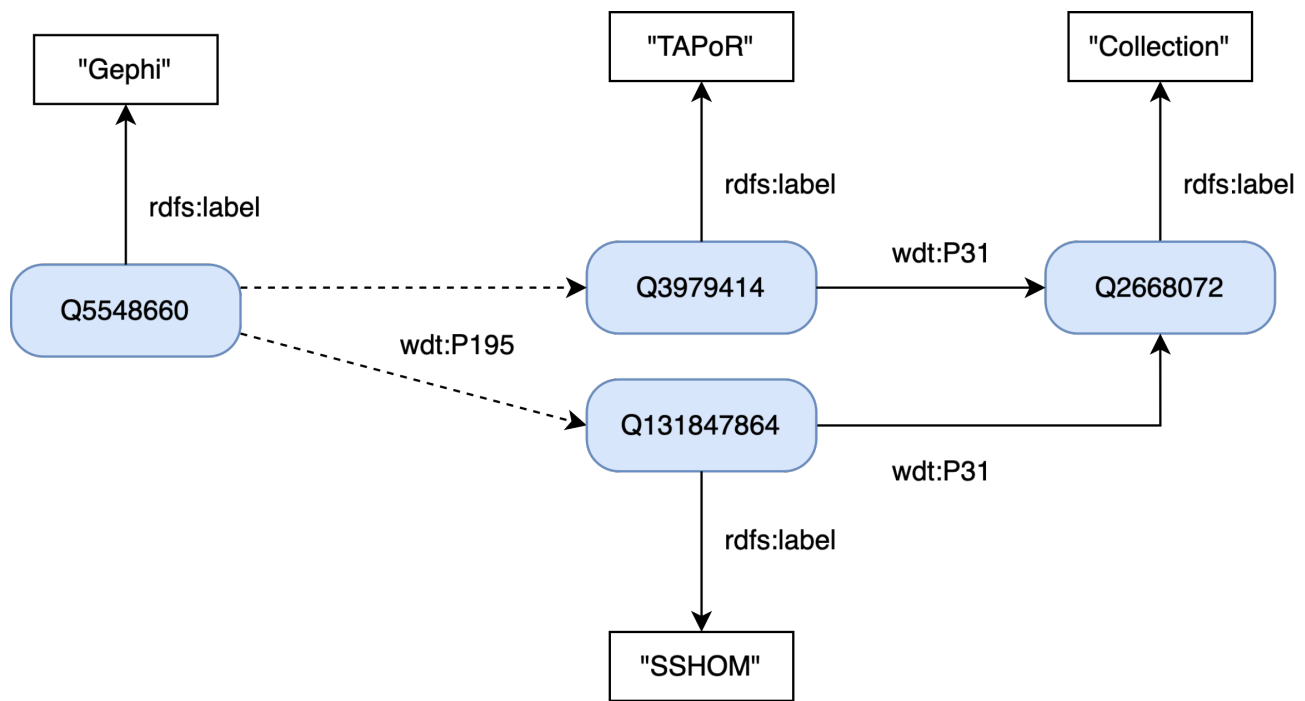


Figure 2. Schematic of the data model for modelling registries as collections, using Gephi (Q5548660) as an example

Domain-specific extensions

This approach also allows us to accommodate the needs of other communities that might want to model additional, domain-specific relations. One such extension has been implemented by some of the authors as part of a survey of the fields of digital humanities and digital history. There, we model the relation between methods (classified with TaDiRAH), research output in the form of publications and conference papers, and their authors. This extension then allows to ask for exemplary applications of a method and the tools (potentially) used by specific scholars [Grallert, Trilling, and Skibba 2025], which, in turn, allows us to track the popularity of tools over time or to build training curricula based on the relevance of particular tools for specific fields (see for instance `method_authors.rq` from [Grallert 2024]).

37

Frontends

Wikidata provides several interfaces: a web frontend, APIs, and a SPARQL endpoint. All of these interfaces come with constraints or a non-trivial amount of complexity that necessitate prior knowledge about data structures, protocols, query languages, or hardware-specific extensions to such standards.

38

Wikidata's default interface is the *Linked Data Interface* (Figure 3). It emphasises the platform's use as an authority file by focusing on individual items. The *Linked Data Interface* functions both as a view and as an input mask, enabling rather efficient data processing for a graphic user interface. This is greatly facilitated by the simple presentation of a list of statements currently available for any specific item, including all qualifiers and a summary of existing references. However, this reduced presentation has its own disadvantages. For example, the rather technical design of the interface, which is based on the principles of Linked Open Data (LOD), can be perceived as cumbersome and less intuitive, especially for new users and those unfamiliar with knowledge graphs.

39

WIKIDATA

Search Wikidata

Search

Gephi (Q5548660)

Item

Discussion

Read

View history

network analysis and visualization software package

edit

In more languages

Configure

Language	Label	Description	Also known as
default for all languages	No label defined		
English	Gephi	network analysis and visualization software package	
German	Gephi	Softwarepaket zur Netzwerkanalyse und Visualisierung	
French	Gephi	logiciel libre d'analyse et de visualisation de réseaux	
Bavarian	Gephi	No description defined	

All entered languages

Statements

instance of

free software

0 references

edit

software

0 references

edit

desktop application

0 references

edit

+ add value

inception

2008

1 reference

edit

+ add value

has use

digital capturing

1 reference

edit

+ add value

Figure 3. Figure 3: The item Gephi (Q5548660) in Wikidata's Linked Data Interface.

But most importantly, by focusing on individual items, this interface does not provide direct access to the underlying knowledge graph. There is no obvious or simple means for accessing and visualising specific subsets of the knowledge graph — e.g. all tools that can be used for social network analysis. Without such a means to return curated tool lists, Wikidata could solely serve as an authority file.^[7]

Query tools with SPARQL

As stated above, the data model is fundamentally instantiated through SPARQL queries linking tools to classifications or collections. We therefore provide a set of modular queries for use in various applications and to ease access to SPARQL for colleagues less comfortable writing their own queries from scratch [Grallert 2024]. Each query first asks for all items with a TaDiRAH ID, that is items which our model considers to be *methods*. It then queries for all items that point to at least one of those methods through the has use (P366) property and which are instances of “Software” or its subclasses:

```

SELECT DISTINCT
  ?tool ?toolLabel # only get Software-ID and Software-Name
WHERE {
  ?method wdt:P9309 ?tadirahID. # Variable method is a tadirah-
method
  ?tool wdt:P366 ?method; # Variable tool 'has method'
method

```

40

41

```

(wdt:P31/(wdt:P279*)) wd:Q7397. # and tool is child of "Software"
SERVICE wikibase:label {
  # set wikibase-service to auto-language with fallback english
  bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en".
  # get the tool-label (=name) of our tool.
  ?tool rdfs:label ?toolLabel.
}
}

```

This query returns the IDs of items and their labels as a potential sanity check for human readers. The ID can then be used in further SPARQL queries, API calls, or the plethora of tools for interacting with Wikidata hosted on Toolforge, ranging from *Scholia*, a long-running project for querying and visualising scientometrics [Nielsen, Mietchen, and Willighagen 2017], to *Reasonator*, which displays Wikidata items in a view optimised for their particular item-type and enhanced with some basic reasoning.

Another option is to directly query for tools in curated collections, such as SSHOM (Q131847864) and TAPoR (Q3979414), as mentioned above.

42

```

#title:Tools in the SSHOM
#defaultView:Table
PREFIX collection:<http://www.wikidata.org/entity/Q131847864> # a specific
collection
SELECT
  ?tool ?toolLabel
WHERE {
  ?tool wdt:P195 collection: ;      # items in the collection
    wdt:P31/wdt:P279* wd:Q7397. # limit tools to software in the
broadest sense
  SERVICE wikibase:label
    bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en".
    ?tool rdfs:label ?toolLabel.
}
}
LIMIT 3000

```

However, writing and adapting such queries require profound knowledge of the SPARQL query language, which makes it difficult to use for many, if not most potential users. A number of research teams are testing the application of LLMs or LLM-based systems for interacting with knowledge graphs and SPARQL endpoints in natural language with promising results [Rony et al. 2022] [Taffa and Usbeck 2023] [Liu et al. 2024] [Rangel et al. 2024]. All current LLMs are pretty good at generating syntactically correct SPARQL but cannot be relied upon to correctly identify entities and properties of any given knowledge graph. Based on our anecdotal testing, the prototypical SPINACH chat bot, which was specifically designed for querying Wikidata with SPARQL, outperforms any general purpose LLM and is a great entry point for those with only cursory knowledge of SPARQL [Liu et al. 2024]. As always, one needs to check the suggestions for plausibility and the more specific the natural language instructions, the better will be the results.

Tool Registry Frontend Module

In order to overcome the limitations of the Linked Data Interface and SPARQL, we developed *ToolFindr*, a prototypical web frontend for the tool registry to be used in the context of the Kompetenzwerkstatt Digital Humanities (KDH) at the university library of Humboldt-Universität zu Berlin. The idea of using Wikidata as a database and authority file for web applications has successfully been implemented by multiple other projects. Good examples are the *Archivführer*

43

Kolonialzeit and *Scholia*. The core argument for this architecture is that one can build custom interfaces for specific communities based on standardised and well-documented APIs while retaining all the advantages of Wikidata communities of contributors and its centralised and persistent data storage without being locked into its user interface.

ToolFindr offers specialised search and some facettted browsing options for digital humanities tools, as well as a more reader-friendly individual view of the tools compared to Wikidata’s native linked data interface. It also enables the integration of corporate design elements and the provision of additional background information relevant to stakeholders in the library. We took particular care to strictly maintain the separation between data and presentation levels in order to ensure data integrity and reusability.

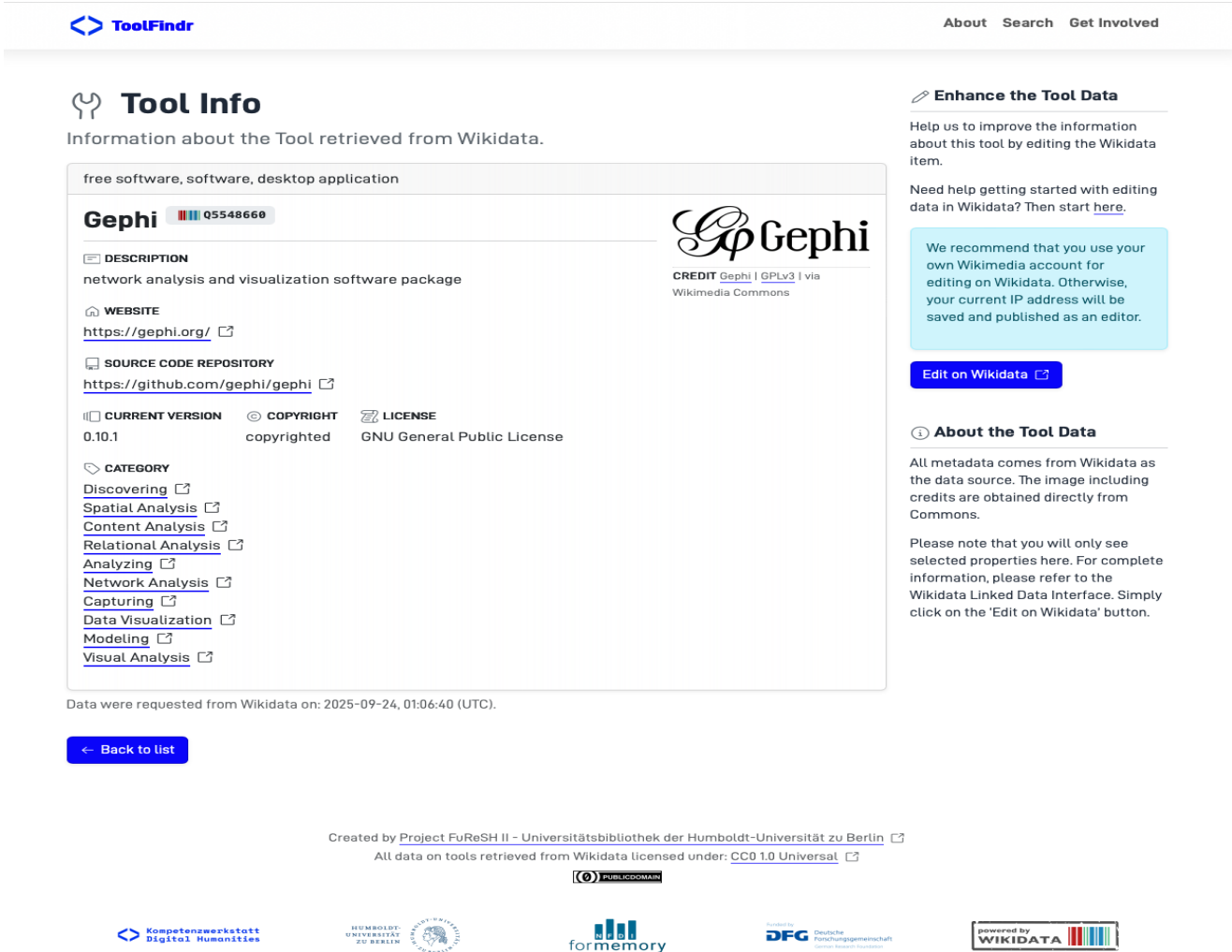


Figure 4. Figure 4: The front-end application developed in the KDH uses Wikidata as a database and authority file.

In order to facilitate installation and operation — and to minimise the hurdles for subsequent use by other projects — the frontend is implemented as a static website and published via GitHub pages [Eckenstaler and Schlesinger 2025]. This decision minimises the maintenance effort and makes it easy to deploy the frontend.

Data and Workflows

Data sources

As outlined above, tool registries tend to build upon existing data sets. We followed this well-established practice and populated our tool registry with the existing knowledge about tools within the DH community by integrating and linking the TAPoR and SSHOM data sets and their classification of tools according to TaDiRAH into Wikidata. This approach

was facilitated by all three platforms providing machine-actionable data through APIs and serialised as JSON.

Unfortunately, TAPoR's APIs are entirely undocumented and we only discovered them through other projects interested in the usage of tools within the DH community, namely the ToolXtractor [Fischer, yoannspace, and laureD19 2022] (see also [Barbot et al. 2019a] [Barbot et al. 2019b] [Fischer and Moranville 2020]. The TAPoR API provides at least two endpoints. https://tapor.ca/api/tools/by_analysis returns a full list of all tools with minimal information for each entry. Importantly, it links tools to categories of research activities, e.g. *OpenRefine* to "Enrichment." <https://tapor.ca/api/tools/{ID}> then returns detailed information for individual tools, based on the tool ID obtained through the first endpoint.^[8]

47

SSHOM's API, on the other hand, is well documented. Again, one has to query for all tools (through <https://marketplace-api.sshopencloud.eu/api/tools-services?approved=true>) in order to retrieve the necessary IDs for individual tools, which then allow to use more specific API endpoints such as <https://marketplace-api.sshopencloud.eu/api/tools-services/{ID}>. In either case, the API returns plenty of data, including links to equivalent concepts on Wikidata and classifications such as TaDiRAH or the *Austrian Fields of Science and Technology Classification 2012* on the DARIAH Vocab services. The APIs, however, do not allow to query for tools by concepts from these classification schemes. It seems that this is only possible through the web interface and URL query parameters, (i.e. <https://marketplace.sshopencloud.eu/search?f.activity={activity}>) where the camel case in TaDiRAH concepts is replaced with "+" and all component terms are capitalised, i.e. opticalCharacterRecognition needs to be translated to Optical+Character+Recognition.

48

TaDiRAH is a SKOS vocabulary currently hosted through the DARIAH Vocab services. Even though this platform is envisioned as part of the Semantic Web and should provide Linked Open Data, it has a patchy track record of doing so. At the time of writing in 2024, neither the service's API and SPARQL endpoint nor their documentation can be reached. Fortunately, we were able to download all data as RDF in late 2022.

49

Data import

There are generally two main methods for adding and editing Wikidata items in bulk: QuickStatements and OpenRefine. We chose the latter as we had to reconcile our initial data sets with existing items to avoid creating duplicates. In addition, OpenRefine provides means for extensive data manipulation and a graphical user interface, which some of our team members were already familiar with. Edits on Wikidata are generally non destructive. Wikibase, the software running Wikidata, provides a mature system of user management and version control, which allows to revert to earlier states.^[9]

50

Reconciliation required a lot of semi-automated data cleaning and manual decision making. The necessary schemas for mapping our input data sets to the data model were created iteratively through the OpenRefine GUI and in tandem with the data model. In total, we have mapped 85 missing TaDiRAH classes (mostly through linking TaDiRAH IDs to existing Wikidata items), added almost 700, and edited more than 1200 tools from the SSHOM and TAPoR data sets.^[10]

51

Adding and editing data

Wikidata allows anyone with an internet connection to edit and contribute without further ado. Registration is not required, but the platform will log the editor's public IP address in such cases. With our approach, existing queries will pick up new information as soon as they are added to Wikidata. Tracking changes to "our" data set, we see constant improvements of data through a combination of individual edits through the Wikidata user interface, bulk edits, and bots. The latter, for instance, periodically query linked GitHub repositories for new releases and update items accordingly.

52

Export and publication of stable data sets

One of the primary challenges in working with Wikidata in academic environments lies in its open-editing nature; anyone can modify data entries, which might raise concerns about reliability and accuracy. This characteristic necessitates additional measures to ensure the stability and trustworthiness of the information it contains, especially for academic use where data integrity is paramount for reproducibility.

53

To address this issue, we regularly pull copies of our data set from Wikidata in formats such as RDF (serialised as Turtle, .ttl) or JSON-LD using a combination of SPARQL queries and API calls. We publish these versioned releases on platforms like GitLab or GitHub and archive them on public repositories such as Zenodo [Dresselhaus and Grallert 2024].

54

This approach allows the data sets to be cited with a DOI, ensuring that specific versions can be referenced reliably. Wikidata itself provides persistent identifiers for citing particular versions of data objects, but through the Linked Data Interface the statements of such persistent versions of items still point to the canonical ID of other items. The archived data set therefore includes full copies of first-level objects (tools) and their first-degree neighbours. Adding higher-degree neighbours would exponentially increase the size of the data set. This, we argue, represents a sensible balance between the scholarly urge for comprehensive data sets and the requirement to minimise our footprint on infrastructure maintained by others. Full access to historical perspectives on the full data set would require to archive the entire subgraph with all dependencies, which would yield a great amount of Wikidata itself with billions of triples (c.f. [Hosseini Beghaeiraveri et al. 2023]).

55

Discussion and outlook

In this paper, we have presented a conceptual framework and concrete implementation of what we believe to be a sustainable approach to the directory paradox as formulated by [Dombrowski 2021]. Our conceptual framework and implementation of a Wikidata-based tool registry for digital humanities attends to communities who require a flexible and open authority file of cultural artefacts as well as curated lists of subsets. The Linked Open Data paradigm and the Wikiverse allow anyone to contribute their data, curate their own subsets, or to query the graph for connections beyond the initial scope of a tool registry. We are fully aware that the fundamental flexibility and openness of Wikidata and the absence of hierarchical access control and formal ontologies, such as CIDOC-CRM, might impede adoption among libraries and infrastructure providers. However, the QIDs of tools can simply function as a persistent identifier (PID) to connect otherwise disparate and siloed registries and knowledge graphs.

56

While TaDiRAH is currently the only classification scheme for research in the digital humanities employed across multiple venues, it can be superseded or amended by future classification schemes for the digital humanities when they emerge. Equally, other fields could use their own classification schemes without breaking the data model and registry introduced in this paper by adapting and modifying the relevant SPARQL queries.

57

We acknowledge that the reliance on SPARQL is a major obstacle to wider adoption. We thus perceive of the modular frontend and a graphical user interface for building queries as the most important venues of future development. We can imagine a frontend with editing functionality based on HTML forms, JSON schemas, and the Wikidata REST API currently under development [Starting fresh: The Wikibase REST API 2023] in order to technically enforce data models.

58

Acknowledgments

Funding

This work was funded by the German Research Foundation (DFG) through a collaboration between the NFDI consortium 4Memory (www.4memory.de, DFG project no. 501609550) and *Future e-Research Support in the Humanities II* (FuReSH II, DFG project no. 466522693).

59

Author contributions (CRedit)

- Conceptualization: Till Grallert, Sophie Eckenstaler, Claus-Michael Schlesinger
- Data Curation: Till Grallert, Isabell Trilling
- Investigation: Till Grallert, Sophie Eckenstaler, Isabell Trilling, Sophie Stark
- Software: Sophie Eckenstaler (Frontend, SPARQL), Nicole Dresselhaus (Archiving), Till Grallert (SPARQL, R)
- Writing, original draft: Till Grallert, Claus-Michael Schlesinger, Sophie Eckenstaler, Nicole Dresselhaus

- Writing, review & editing: Till Grallert, Claus-Michael Schlesinger

Data availability

All data and code is available on Zenodo:

60

- SPARQL queries: [Grallert 2024].
- Data model and JSON schemas for use with OpenRefine: https://scm.cms.hu-berlin.de/methodenlabor/p_publish2wikidata.
- Front end: [Eckenstaler and Schlesinger 2025].
- Weekly screenshots of the data set as exported from Wikidata: (Grallert and Dresselhaus, 2024–).

Notes

[1] We ourselves are guilty of this referencing practice, e.g. [Schlesinger, Gäckle-Heckelen, and Burkard 2023] (grallert, 2022).

[2] This is inspired by Derrida's (1996) call for an *archivology* and his genealogy of the archive as grounded in the Latin *archivum* from Greek *arché* as *commencement* and *commandment* [Derrida 1996], Foucault's *Archaeology of Knowledge* [Foucault 2002, esp. 126–131], and Heidegger's writing on technology [Heidegger 2000] (c.f. [Ihde 2010, esp. 32–35, 62]).

[3] This information is transparently provided on the item level through the API (see below). The website is much quieter about this fundamental heritage and only lists TAPoR 3 as one of 15 “trusted sources” [About the data population 2023].

[4] See <https://www.wikidata.org/wiki/Wikidata:Statistics> for up-to-date statistics.

[5] In fact Wikidata recommends to always reference the versioned permanent URL instead of the canonical ID; see <https://www.wikidata.org/w/index.php?title=Special:CiteThisPage&page=Q25212027&id=2313077072&wpFormIdentifier=titleform> for citations of <https://www.wikidata.org/w/index.php?title=Q25212027&oldid=2313077072>.

[6] The academic knowledge graph OpenAlex finds only 9 references.

[7] Hosseini Beghaeiraveri et al. (2023) show that subsetting massive knowledge graphs such as Wikidata has become more relevant in recent years [Hosseini Beghaeiraveri et al. 2023]. However, the selection of approaches and tools evaluated in their study also demonstrate that there is no standardised procedure to be easily applied to the tool registry.

[8] Information includes data on tools such as URLs, source code repositories, email addresses of creators, image URLs, metadata on when the TAPoR entry was last updated and by whom, as well as rating on the platform.

[9] There are some tools for editing and reverting batch edits, but at the time of writing in 2024 they had become dysfunctional and were looking for new maintainers.

[10] The JSON schemas for uploads from OpenRefine to Wikidata can be found at https://scm.cms.hu-berlin.de/methodenlabor/p_publish2wikidata.

Works Cited

About the data population 2023 About the data population (2023). *Social Sciences & Humanities Open Marketplace*. Available at: <https://marketplace.sshopencloud.eu/about/data-population> (Accessed: June 4, 2024).

Alireza et al. 2022 Alireza, Z., Seung-Bin, Y., Fischer, F., Đurčo, M., and Wieder P. (2022) “Measuring the use of tools and software in the digital humanities: A machine-learning approach for extracting software mentions from scholarly articles”, DH2023, 28 July. Available at: <https://dh-abstracts.library.virginia.edu/works/12004>

Allison-Cassin and Scott 2018 Allison-Cassin, S. and Scott, D. (2018) “Wikidata: A platform for your library's Linked Open Data”, *The Code4Lib Journal* [Preprint], (40). Available at: <https://journal.code4lib.org/articles/13424> (Accessed: April 26, 2024).

Barbot et al. 2019a Barbot, L., Fischer, F., Moranville, Y., and Pozdniakov, I. (2019a) *Tools mentioned in the proceedings of the annual ADHO conferences (2015–2019)*. Available at: <https://lehkost.github.io/tools-dh-proceedings/index.html> (Accessed: July 26, 2022).

Barbot et al. 2019b Barbot, L., Fischer, F., Moranville, Y., and Pozdniakov, I. (2019b) *Which DH tools are actually used in research? weltliteratur.net: A Black Market for the Digital Humanities*. Available at: <https://weltliteratur.net/dh-tools-used-in-research/> (Accessed: July 26, 2022).

Bernardou et al. 2018 Bernardou, A., Champion, A., Dallas, C., and Hughes, L.M. (2018) “Introduction: A critique of digital practices and research infrastructures”, in A. Bernardou et al. (eds.) *Cultural Heritage Infrastructures in Digital Humanities*. London: Routledge (Digital Research in the Arts and Humanities).

- Berners-Lee 2009** Berners-Lee, T. (2009) "Linked Data - Design Issues", *W3: World Wide Web Consortium* Available at: <https://www.w3.org/DesignIssues/LinkedData#fivestar> (Accessed: May 23, 2024).
- Borek et al. 2015** Borek, L., Hastik, C., Khramova, V., and Geiger, J. (2015) "TaDiRAH". Digital Humanities Taxonomy Group. Available at: <https://github.com/dhtaxonomy/TaDiRAH> (Accessed: May 29, 2024).
- Borek et al. 2016** Borek, L., Dombrowski, Q., Perkins, J., and Schöch, C. al. (2016) "TaDiRAH: A case study in pragmatic classification", *Digital Humanities Quarterly*, 10(1). Available at: <https://dhq.digitalhumanities.org/vol/10/1/000235/000235.html>.
- Borek et al. 2021** Borek, L., Hastik, C., Khramova, V., Illmayer, K., and Geiger, J.D. (2021) "Information organization and access in digital humanities: TaDiRAH revised, Formalized and FAIR", in *Information between Data and Knowledge*. Glückstadt: Werner Hülsbusch (Schriften zur Informationswissenschaft, 74), pp. 321–332. Available at: https://epub.uni-regensburg.de/44951/1/isi_borek_et_al.pdf.
- Christophersen et al. 2023** Christophersen, A., Colón-Marrero, E., Dietrich, D., Falcao, P., Fox, C., Hanson, K., Kwan, A., and McEniry, M.. (2023) "Software metadata recommended format guide", *softwarepreservationnetwork.orgZenodo*. Available at: <https://doi.org/10.5281/zenodo.10001787>.
- Derrida 1996** Derrida, J. (1996) *Archive fever: A Freudian impression*. Translated by E. Prenowitz. Chicago: University of Chicago Press.
- Dombrowski 2014** Dombrowski, Q. (2014) "What ever happened to Project Bamboo?", *Literary and Linguistic Computing*, 29(3), pp. 326–339. Available at: <https://doi.org/10.1093/lc/fqu026>.
- Dombrowski 2021** Dombrowski, Q. (2021) "The directory paradox", in A.B. McGrail, A.D. Nieves, and S. Senier (eds.) *People, Practice, Power: Digital Humanities outside the Center*. Minneapolis: University of Minnesota Press (Debates in the Digital Humanities). Available at: <https://dhdebates.gc.cuny.edu/read/people-practice-power/section/ca87ec4c-23a0-452d-8595-7cfd7e8d6f0c> (Accessed: February 24, 2022).
- Dresselhaus and Grallert 2024** Dresselhaus, N. and Grallert, T. (2024) "P_wikidata2gitlab". Berlin: NFDI4Memory, Humboldt-Universität zu Berlin. Available at: https://scm.cms.hu-berlin.de/methodenlabor/p_wikidata2gitlab (Accessed: November 12, 2024).
- Ebeling 2009** Ebeling, K. (2009) "Das Gesetz des Archivs", in M.K. Ebeling, S. Günzel, and A. Assmann (eds.) *Archivologie: Theorien des Archivs in Wissenschaft, Medien und Künsten*. Berlin: Kulturverlag Kadmos, pp. 61–88.
- Eckenstaler and Schlesinger 2025** Eckenstaler, S. and Schlesinger, C.-M. (2025) "ToolFindr". Kompetenzwerkstatt Digital Humanities (KDH). Available at: <https://github.com/FuReSH/toolfindr> (Accessed: September 24, 2025).
- Endings Project Team 2023** Endings Project Team (2023) "Endings principles for digital longevity". Available at: <https://endings.uvic.ca/principles.html> (Accessed: May 14, 2024).
- FAIR Principles 2020** FAIR Principles (2020). *GO FAIR*. Available at: <https://www.go-fair.org/fair-principles/> (Accessed: August 5, 2020).
- Fagervig 2023** Fagervig, A. (2023) "Wikidata for authority control: Sharing museum knowledge with the world", *Digital Humanities in the Nordic and Baltic Countries Publications*, 5(1), pp. 222–239. Available at: <https://doi.org/10/gs56rd>.
- Fischer and Moranville 2020** Fischer, F. and Moranville, Y. (2020) "Tools mentioned in DH2020 abstracts." *weltliteratur.net: A Black Market for the Digital Humanities*. Available at: <https://weltliteratur.net/tools-mentioned-in-dh2020-abstracts/>.
- Fischer, yoannspace, and laureD19 2022** Fischer, F., yoannspace and laureD19 (2022) "ToolXtractor". Available at: <https://github.com/lehkost/ToolXtractor> (Accessed: May 28, 2024).
- Foucault 2002** Foucault, M. (2002) *The archaeology of knowledge*. Translated by A.M.S. Smith. London: Routledge.
- Gil and Ortega 2016** Gil, A. and Ortega, É. (2016) "Global outlooks in digital humanities: Multilingual practices and minimal computing", in C. Crompton, R.J. Lane, and R. Siemens (eds.) *Doing digital humanities: Practice, training, research*. Abingdon: Routledge, pp. 22–34.
- Grallert 2022** Grallert, T. (2022) "Open Arabic periodical editions: A framework for bootstrapped scholarly editions outside the global north", *Digital Humanities Quarterly*, 16(2). Available at: <https://dhq.digitalhumanities.org/vol/16/2/000593/000593.html>.
- Grallert et al. 2024** Grallert, T., Eckenstaler, S., Tirtohusodo, S., and Schlesinger, C-M.. (2024) "Ob Werkzeugkoffer, Werkstatt oder Baumarkt: Offene, community-kuratierte Tool Registries mit Wikidata". DHd2024, Zenodo, 29 February.

Available at: <https://doi.org/10.5281/zenodo.10698252>.

Grallert 2024 Grallert, T. (2024) "Tool registry: SPARQL queries". Available at: https://scm.cms.hu-berlin.de/methodenlabor/tr_sparql (Accessed: January 21, 2025).

Grallert 2025 Grallert, T. (2025) "Adding every Arabic periodical published before 1930 to Wikidata: Moving the scholarly crowd-sourcing project Jarā'id to the digital commons", *Transformations: A DARIAH Journal*, 1: Workflows, pp. 1–39. Available at: <https://doi.org/10.46298/transformations.14749>.

Grallert and Dresselhaus 2024– Grallert, T. and Dresselhaus, N. (2024–) "Tool registry for digital humanities". Zenodo. Available at: <https://doi.org/10.5281/zenodo.14259806>.

Grallert, Trilling, and Skibba 2025 Grallert, T., Trilling, I. and Skibba, A. (2025) *Wikidata:WikiProject Field Survey Digital Humanities / Digital History*. Wikidata. Available at: https://www.wikidata.org/w/index.php?title=Wikidata:WikiProject_Field_Survey_Digital_Humanities/_Digital_History&oldid=2296127677 (Accessed: November 5, 2024).

Grant et al. 2020 Grant, K. et al. (2020) "Absorbing DiRT: Tool directories in the digital age", *Digital Studies / Le champ numérique*, 10(1). Available at: <https://doi.org/10.16995/dscn.325>.

Heidegger 2000 Heidegger, M. (2000) "Die Frage nach der Technik", in F.-W. von Herrmann (ed.) *Vorträge und Aufsätze*. Frankfurt: Klostermann (Gesamtausgabe), pp. 5–36.

Henny-Krahmer and Jettka 2022 Henny-Krahmer, U. and Jettka, D. (2022) "Softwarezitation als Technik der Wissenschaftskultur - Vom Umgang mit Forschungssoftware in den Digital Humanities". DHd 2022 Kulturen des digitalen Gedächtnisses. 8. Tagung des Verbands "Digital Humanities im deutschsprachigen Raum" (DHd 2022), Zenodo, 7 March. Available at: <https://doi.org/10.5281/zenodo.6328047>.

Home | nfdi.software n.d. Home | nfdi.software (n.d). Available at: <https://nfdi.software/> (Accessed: October 17, 2025).

Hosseini Beghaeiraveri et al. 2023 Hosseini Beghaeiraveri, S.A., Labra-Gayo, J.E., Waagmeester, A., Ammar, A., Gonzalez, C., Slenter, D., Ul-Hasan, S., Willighagen, E., McNeill, F., and Gray, A.J.G. (2023) "Wikidata subsetting: Approaches, tools, and evaluation", *Semantic Web*, Preprint, pp. 1–27. Available at: <https://doi.org/10.3233/SW-233491>.

Hughes, Constantopoulos, and Dallas 2015 Hughes, L., Constantopoulos, P. and Dallas, C. (2015) "Digital methods in the humanities: Understanding and describing their use across the disciplines", in S. Schreibman, R. Siemens, and J. Unsworth (eds.) *A New Companion to Digital Humanities*. Chichester: Wiley, pp. 150–170. Available at: <https://doi.org/10.1002/9781118680605.ch11>.

Ihde 2010 Ihde, D. (2010) *Heidegger's technologies: Postphenomenological perspectives*. New York: Fordham University Press (Perspectives in Continental Philosophy).

Irrera et al. 2023 Irrera, O., Mannocci, A., Manghi, P., and Silvello, G. (2023) "A novel curated scholarly graph connecting textual and data publications", *Journal of Data and Information Quality*, 15(3), pp. 26:1–26:24. Available at: <https://doi.org/10.1145/3597310>.

Liu et al. 2024 Liu, S., Semnani, S.J., Triedman, T., Xu, J., Zhao, I.D., and Lam, M.S. (2024) *SPINACH: SPARQL-Based Information Navigation for Challenging Real-World Questions*. Available at: <https://doi.org/10.48550/arXiv.2407.11417>.

NFDI4Objects 2025 NFDI4Objects (2025) *Verzeichnis und Evaluation vorhandener Tools, Standards und Daten*. NFDI4Objects. Available at: https://nfdi4objects.net/portal/trails/directory_and_evaluation_of_existing_tools_standards_and_data/ (Accessed: October 17, 2025).

Nielsen, Mietchen, and Willighagen 2017 Nielsen, F.Å., Mietchen, D. and Willighagen, E. (2017) "Scholia, scientometrics and Wikidata", in E. Blomqvist et al. (eds.) *The Semantic Web: ESWC 2017 Satellite Events*. Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 237–259. Available at: https://doi.org/10.1007/978-3-319-70407-4_36.

Oberhauser 2003 Oberhauser, O.C. (2003) "Card-image public access catalogues (CIPACs): An international survey", *Program: Electronic Library and Information Systems*, 37(2), pp. 73–84. Available at: <https://doi.org/10.1108/00330330310472867>.

Odell, Lemus-Rojas, and Brys 2022 Odell, J., Lemus-Rojas, M. and Brys, L. (2022) *Wikidata for Scholarly Communication Librarianship*. PUI University Library. Available at: <https://doi.org/10.7912/9Z4E-9M13>.

Rangel et al. 2024 Rangel, J.C., de Farias, T.M., Sima, A.C., and Kobayashi, N. (2024) *SPARQL Generation: An analysis on fine-tuning OpenLLaMA for Question Answering over a Life Science Knowledge Graph*. Available at:

<http://arxiv.org/abs/2402.04627> (Accessed: June 11, 2024).

Registry - NFDI4Culture n.d. Registry - NFDI4Culture (no date). Available at: <https://nfdi4culture.de/de/ressourcen/registry.html> (Accessed: October 17, 2025).

Risam and Gil 2022 Risam, R. and Gil, A. (2022) "Introduction: The questions of minimal computing", *Digital Humanities Quarterly*, 16(2). Available at: <https://dhq.digitalhumanities.org/vol/16/2/000646/000646.html>.

Rockwell 2006 Rockwell, G. (2006) "TAPoR: Building a portal for text analysis", in R.G. Siemens and D. Moorman (eds.) *Mind technologies: Humanities computing and the Canadian academic community*. Calgary: University of Calgary Press, pp. 285–289. Available at: <https://www.deslibris.ca/ID/415538> (Accessed: June 4, 2024).

Rony et al. 2022 Rony, M.R.A.H., Kumar, U., Teucher, R., Kovriguina, L., and Lehmann, J., (2022) "SGPT: A generative approach for SPARQL query generation from natural language questions", *IEEE Access*, 10, pp. 70712–70723. Available at: <https://doi.org/10.1109/ACCESS.2022.3188714>.

Ruth, Niekler, and Burghardt 2022 Ruth, N., Niekler, A. and Burghardt, M. (2022) "Peeking inside the DH toolbox - Detection and classification of software tools in DH publications", in *Workshop on Computational Humanities Research*. Available at: <https://www.semanticscholar.org/paper/Peeking-Inside-the-DH-Toolbox-Detection-and-of-in-Ruth-Niekler/a9017fcb60338fb23bea1c6519da5f40cabbe839> (Accessed: March 21, 2024).

Sardo and Bianchini 2022 Sardo, L. and Bianchini, C. (2022) "Wikidata: A new perspective towards universal bibliographic control", *JLIS : Italian Journal of Library, Archives and Information Science = Rivista italiana di biblioteconomia, archivistica e scienza dell'informazione*, pp. 291–311. Available at: <https://doi.org/10/gn7n2m>.

Schindler et al. 2021 Schindler, D., Bensmann, F., Dietze, S., and Krüger, F. (2021) "SoMeSci- A 5 star open data gold standard knowledge graph of software mentions in scientific articles", in. New York, NY, USA: Association for Computing Machinery (CIKM '21), pp. 4574–4583. Available at: <https://doi.org/10.1145/3459637.3482017>.

Schlesinger, Gäckle-Heckelen, and Burkard 2023 Schlesinger, C.-M., Gäckle-Heckelen, M. and Burkard, F. (2023) "Onboarding digital humanities students with a shared working environment for introductory courses: Concept, implementation, and lessons learned", *IDEAH*, 3(4). Available at: <https://doi.org/10.21428/f1f23564.4c0566bc>.

Starting fresh: The Wikibase REST API 2023 Starting fresh: The Wikibase REST API (2023). *Wikimedia Tech News*. Available at: <https://tech-news.wikimedia.de/2023/09/07/starting-fresh-the-wikibase-rest-api/> (Accessed: March 20, 2025).

Taffa and Usbeck 2023 Taffa, T.A. and Usbeck, R. (2023) *Leveraging LLMs in scholarly knowledge graph Question Answering*. Available at: <http://arxiv.org/abs/2311.09841> (Accessed: July 18, 2024).

VanderWal 2004 VanderWal, T. (2004) *Feed On This*. Available at: <https://www.vanderwal.net/random/entrysel.php?blog=1562> (Accessed: February 6, 2025).

Vismann 2009 Vismann, C. (2009) "Arché, Archiv, Gesetzesherrschaft", in M.K. Ebeling, S. Günzel, and A. Assmann (eds.) *Archivologie: Theorien des Archivs in Wissenschaft, Medien und Künsten*. Berlin: Kulturverlag Kadmos, pp. 89–103.

Vrandečić, Pintscher, and Krötzsch 2023 Vrandečić, D., Pintscher, L. and Krötzsch, M. (2023) "Wikidata: The making of", in *Companion Proceedings of the ACM Web Conference 2023*. New York, NY, USA: Association for Computing Machinery (WWW '23 Companion), pp. 615–624. Available at: <https://doi.org/10.1145/3543873.3585579>.

Wikidata:WikiProject DH Tool Registry 2024 Wikidata:WikiProject DH Tool Registry (2024). *Wikidata*. Available at: https://www.wikidata.org/w/index.php?title=Wikidata:WikiProject_DH_Tool_Registry&oldid=2277445090 (Accessed: November 5, 2024).

Zhao 2022 Zhao, F. (2022) "A systematic review of Wikidata in digital humanities projects", *Digital Scholarship in the Humanities*, pp. 1–22. Available at: <https://doi.org/10.1093/llc/fqac083>.

Recommendations

DHQ is testing out three new article recommendation methods! Please explore the links below to find articles that are related in different ways to the one you just read. We are interested in how these methods work for readers—if you would like to share feedback with us, please complete our short evaluation survey. You can also visit our documentation for these recommendation methods to learn more.

SPECTER Recommendations

Below are article recommendations generated by the SPECTER model:

1. TaDiRAH: a Case Study in Pragmatic Classification, 2016, Luise Borek, Technical University of Darmstadt; Quinn Dombrowski, Miami University; Jody Perkins, Miami University; Christof Schöch, University of Würzburg
2. Open Data in Cultural Heritage Institutions: Can We Be Better Than Data Brokers?, 2020, S.L. Ziegler, Louisiana State University Libraries
3. Beyond Gutenberg: Transcending the Document Paradigm in Digital Humanities, 2014, David Schloen, University of Chicago; Sandra Schloen, University of Chicago
4. Towards a National Data Architecture for Cultural Collections: Designing the Australian Cultural Data Engine, 2024, Rachel Fensham, University of Melbourne; Australian Cultural Data Engine; Tyne Daile Sumner, Australian National University; Australian Cultural Data Engine; Nat Cutter, University of Melbourne; Australian Cultural Data Engine; George Buchanan, RMIT University; Rui Liu, University of Melbourne; Justin Munoz, Independent Scholar; James Smithies, Australian National University; Ivy Zheng, University of Newcastle; David Carlin, RMIT University; Erik Champion, University of South Australia; Hugh Craig, University of Newcastle; Scott East, University of New South Wales; Chris Hay, Flinders University; Lisa M. Given, RMIT University; John Macarthur, University of Queensland; David McMeekin, Curtin University; Joanna Mendelssohn, University of Melbourne; Deborah van der Plaats, University of Queensland
5. Working on and with Categories for Text Analysis: Challenges and Findings from and for Digital Humanities Practices, 2023, Dominik Gerstorfer, Technical University of Darmstadt; Evelyn Gius, Technical University of Darmstadt; Janina Jacke, Kiel University

DHQ Keyword Recommendations

Below are article recommendations generated by DHQ Keywords:

1. Introduction, 2017, Emily Christina Murphy, Queen's University; Shannon R. Smith, Bader International Study Centre, Herstmonceux Castle, Queen's University
2. Linked data from TEI (LIFT): A Teaching Tool for TEI to Linked Data Transformation, 2022, Francesca Giovannetti, University of Bologna; Francesca Tomasi, University of Bologna
3. Minimizing Computing Maximizes Labor, 2022, Quinn Dombrowski, Stanford University
4. The Sound of Many Hands Clapping: Teaching the Digital Humanities through Virtual Research Environment (VREs), 2012, Craig Bellamy, VerSI, University of Melbourne, Australia
5. Dendrography and Art History: a computer-assisted analysis of Cézanne's Bathers., 2019, Melinda Weinstein, Lawrence Technological University; Edward Voss, University of Iowa; David Soll, University of Iowa

TF-IDF Recommendations

Below are article recommendations generated by the TF-IDF Model:

1. Towards a National Data Architecture for Cultural Collections: Designing the Australian Cultural Data Engine, 2024, Rachel Fensham, University of Melbourne; Australian Cultural Data Engine; Tyne Daile Sumner, Australian National University; Australian Cultural Data Engine; Nat Cutter, University of Melbourne; Australian Cultural Data Engine; George Buchanan, RMIT University; Rui Liu, University of Melbourne; Justin Munoz, Independent Scholar; James Smithies, Australian National University; Ivy Zheng, University of Newcastle; David Carlin, RMIT University; Erik Champion, University of South Australia; Hugh Craig, University of Newcastle; Scott East, University of New South Wales; Chris Hay, Flinders University; Lisa M. Given, RMIT University; John Macarthur, University of Queensland; David McMeekin, Curtin University; Joanna Mendelssohn, University of Melbourne; Deborah van der Plaats, University of Queensland
2. TaDiRAH: a Case Study in Pragmatic Classification, 2016, Luise Borek, Technical University of Darmstadt; Quinn Dombrowski, Miami University; Jody Perkins, Miami University; Christof Schöch, University of Würzburg
3. From the Rock Wall to Wikipedia: Sites of Digital Connection and Repair in the Community-Engaged Classroom, 2026, Courtney Rivard, UNC-Chapel Hill; Alexandra Odom, UNC-Chapel Hill; Carly Schnitzler, Johns Hopkins University; Katherine Stein, UNC-Chapel Hill; Rolando Rodriguez, UNC-Chapel Hill; Kathryn Wall, The Marian Cheek Jackson Center for Saving and Making History; Anna Spencer, The Marian Cheek Jackson Center for Saving and Making History; Kyle Cunningham, UNC-Chapel Hill; Sage Clausen, UNC-Chapel Hill; Sarah Davey, UNC-Chapel Hill
4. Old Content and Modern Tools – Searching Named Entities in a Finnish OCR'd Historical Newspaper Collection 1771–1910, 2017, Kimmo Kettunen, National Library of Finland, Mikkeli, Finland; Eetu Mäkelä, University of Helsinki, Helsinki Centre for Digital Humanities; Teemu Ruokolainen, National Library of Finland, Mikkeli, Finland; Juha Kuokkala, University of Helsinki, Department of Modern Languages, Helsinki, Finland; Laura Löfberg, Department of Linguistics and English Language, Lancaster University, UK
5. A Critical Collection History of Nineteenth-century Women's Letters: Overcoming the Occluded Archive with Data-Driven Methods, 2026, Ilona Pikkanen, The Finnish Literature Society; Matti La Mela, Uppsala University; Hanna-Leena Paloposki, Independent Scholar; Jouni Tuominen, University of Helsinki and Aalto University



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

