

Linking Women Editors of Periodicals to the Wikidata Knowledge Graph

Katherine Thornton^{a,*}, Kenneth Seals-Nutt^b Marianne Van Remoortel^c Julie M. Birkholz^d and Pieterjan De Potter^e

^a *Stories Services Collaborative, USA*

E-mail: katherine.thornton@yale.edu

^b *Stories Services Collaborative, USA*

E-mail: kenneth@seals-nutt.com

^c *Department of Literary Studies, Ghent University, Belgium*

E-mail: marianne.vanremoortel@ugent.be

^d *Ghent Centre for Digital Humanities, Ghent University and KBR- Royal Library of Belgium's Digital Research Lab, Belgium*

E-mail: julie.birkholz@ugent.be

^e *Ghent Centre for Digital Humanities, Ghent University, Belgium*

E-mail: pieterjan.depotter@ugent.be

Abstract. Stories are important tools for recounting and sharing the past. To tell a story one has to put together diverse information about people, places, time periods, and things. We detail here how a machine, through the power of Semantic Web, can compile scattered and diverse materials and information to construct stories. Through the example of the WeChangEd research project on women editors of periodicals in Europe from 1710 - 1920 we detail how to move from archive, to a structured data model and relational database, to a Linked Open Data model and make this information available on Wikidata, to the use of the Stories Services API to generate multimedia stories related to people, organizations and periodicals.

As more humanists, social scientists and other researchers choose to contribute their data to Wikidata we will all benefit. As researchers add data, the breadth and complexity of the questions we can ask about the data we have contributed will increase. Building applications that syndicate data from Wikidata allows us to leverage a general purpose knowledge graph with a growing number of references back to scholarly literature. Using frameworks developed by the Wikidata community allows us to rapidly provision interactive sites that will help us engage new audiences. This process that we detail here may be of interest to other researchers and cultural heritage institutions seeking web-based presentation options for telling stories from their data.

Keywords: Wikidata, Linked Open Data, Literary Studies

1. Introduction

The use of linked data models in the humanities and in cultural heritage institutions to structure, store, share and link knowledge on our historical past has seen a marked increase of interest and implementation [1]. Evidence of this is the growing size of Wikimedia Foundation's collaborative multilingual knowledge graph of Wikidata[2]. Sharing information in this way,

as Linked Open Data (LOD), provides opportunities for increasing its accessibility and find-ability as well as technologies for efficiently integrating and implementing previously unstructured, siloed data, at lightning speed. Despite these affordances, there remains a gap in access between those familiar with Semantic Web principles, who can implement SPARQL queries to explore data, and those new to LOD. In working to filling this gap, we questioned how can we generate multimedia stories from LOD.

*Corresponding author. E-mail: katherine.thornton@yale.edu.

To tell a story one has to put together diverse information about people, places, time periods, and things. We detail here how a machine, through the power of Semantic Web, can compile scattered and diverse materials and information to construct stories. Through the case of the ERC Starting Grant project "Agents of Change: Women Editors and Socio-Cultural Transformation in Europe, 1710-1920" (acronym WeChangEd), we detail how to move from archive, to a structured data model and relational database, to Linked Open Data on Wikidata, to a Stories Services API powered application to tell machine-readable stories of women editors in Europe. We show that the LOD-powered WeChangEd Stories can be an important tool for re-counting and sharing the past.

2. WeChangEd: Women Editors in Europe

The WeChangEd research project¹ at Ghent University, Belgium from 2015 - 2021 investigates the impact of women editors on public debate and processes of socio-cultural change in Europe in the eighteenth to early twentieth centuries. Women's access to public life and to political power and decision-making was limited in this period. Women were often excluded from formal education, and lacked many fundamental legal, political, and financial rights, most notably the right to vote. Results from the WeChangEd project have revealed that the print medium of the periodical offered women an alternative means to make their voices heard far beyond their immediate sphere of influence. As editors of their own periodicals, women were able to establish transnational networks of intellectual exchange across Europe, engage in cultural transfer, and position themselves in contemporary debate as makers of culture, arbiters of social values, and proponents of women's rights.

This project is carried out with a team of seven researchers and seven student interns with complementary language skills and methodological expertise in literary studies, the digital humanities and the social sciences. This has resulted in two main outputs: 1) a comprehensive database of women editors and their periodicals; 2) a series of thematic sub-projects (in the form of four doctoral dissertations², a forthcoming

special issue, journal articles³, and blog posts⁴) examining a number of key areas in which European women editors of the early eighteenth to early twentieth centuries affected change, including deliberative democracy, salon culture, fashion, and women's rights. By exploring how these processes unfolded in the press through practices of textual transfer both among women and in the larger publishing landscape, we aim not only to initiate a shift in our thinking about the participation of women in society and print culture but also to pave the way for further transnational research on the periodical press.

3. WeChangEd Data Model

To accurately take stock of women editors in Europe over the period 1710-1920, the WeChangEd team developed a LOD model [4]. Using the collaborative object-oriented relational database-based research environment nodegoat [5] researchers started to collect, organize, and structure this information. This included a unique identifier for each person, organization, and periodical, allowing us to classify information not only on women editors themselves, but also the people in their lives (e.g. partners, colleagues, co-editors, family, and so forth), as well as information on the periodicals they edited, and the organizations that these periodicals may have emerged from, were supported by, or served as official organ for. This resulted in a dataset which includes around 1700 persons, 1600 periodicals, 200 organizations and numerous links between these entities. This data is available as CSV files upon request at [6], but it is also publicly available on Wikidata as we detail below. Inherent to the WeChangEd project is facilitating future research on these women editors and periodicals and also increasing the knowledge on women in this period in particular, that are often absent from mainstream records on these periodi-

Ghent University, 2020.; Bezari, Christina, "'Restless Agents of Progress': Female Editorship, Salon Sociability and Modernisation in Spain, Italy, Portugal, and Greece (1860-1920)." Unpublished doctoral dissertation, Ghent University, 2020.; D'Eer, Charlotte. "Women Editors in the German-Language Periodical Press (1740-1920): Transnational Emotional Networks." Unpublished doctoral dissertation, Ghent University, 2020.; Forestier, Eloise. "Women Editors Conducting Deliberative Democracy: A Transnational Study of Liberty, Equality, and Justice in Nineteenth-Century Periodicals." Unpublished doctoral dissertation, Ghent University, 2020.

³e.g.[3]

⁴<https://www.wechanged.ugent.be/blog/>

¹<https://www.wechanged.ugent.be/>

²Mariia Alesina. "Femininity at the Crossroads: Negotiating National and Gender Peripherality in the Russian Fashion Journal *Modnyi Magazin* (1862-1883)." Unpublished doctoral dissertation,

cals given their lacking of fundamental rights that allowed them to hold formal positions as editors. Thus the project team wanted to ensure that the information on these editors would be more easily identifiable and accessible beyond a private database or in a non-digitized document in a library or archive. Thus, there was a need to ensure both the findability of the information on the web as well as the user friendliness in accessing and exploring this information. Consequently the WeChangEd team elected to make this data available on Wikidata and develop stories from the LOD using the Stories Services API, which resulted in what we detail here further in the WeChangEd Stories App.

4. Semantic Web

The vision of the semantic web is an accumulation of interconnected data from heterogeneous sources connected to points of reference for which we define meaning. This entity- and link-based architecture allows for navigation of data from many databases or collections via known points of representation. An organization well-known for cultivating the technologies necessary for the Semantic Web is the World Wide Web Consortium or W3C. In 2009, members of a W3C group stated that: "Semantic Web is the idea of having data on the Web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration, and reuse of data across various applications." [7]. The Wikidata knowledge base fulfills the requirements outlined by the W3C in that each resource has a unique identifier and is linked to other resources by properties, and that all of the data is machine actionable as well as editable by both humans and machines.

5. The Wikidata Data Model

Wikidata is the knowledge base of structured data that anyone can edit [8]. This community-edited knowledge base contains multilingual structured data from many domains, from computing to biodiversity to cultural heritage [9]. A sister project to the Wikipedias, Wikidata is a project of the Wikimedia Foundation. The data in Wikidata is available under the Creative Commons Zero (CC0) license, meaning it is free for anyone to reuse for any purpose. Data is available in multiple formats (CSV, JSON, RDF, etc.) and can

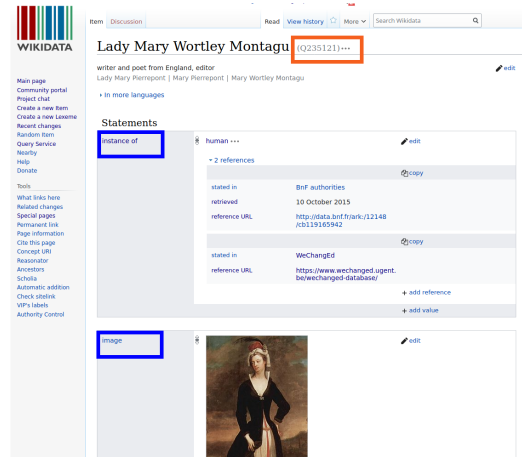


Fig. 1. Properties, statements, references on the item for Lady Mary Wortley Montagu

be accessed via an application programming interface (API) or from a SPARQL endpoint.

The data model of Wikidata consists of Items, Properties and unique identifiers. In Figure 1 we see a screenshot of the Wikidata item for Lady Mary Wortley Montagu, a person that is also in the WeChangEd database. The Wikidata identifier for this item is Q235121. Two properties "instance of" and "image" are used as the predicates of statements. Three properties "stated in", "retrieved", and "reference URL" are used as qualifiers within the reference that provides sourcing information for the claims.

There are more than 8,000 properties in use in Wikidata at the time of this writing. These properties are used to express statements of fact about items. Aligning data with Wikidata requires selecting properties to express the types of statements you'd like to make about your dataset.

6. Getting WeChangEd Data into Wikidata

The WeChangEd team partnered with two Wikidata experts in order to align their data with Wikidata, write the data to Wikidata, and create a Wikidata-powered application for visualization of the data. We outline the steps we took to accomplish this data contribution. The first step was to access the data available from the WeChangEd database. It is possible to export data from nodegoat, the platform used to store the WeChangEd database, in a number of different formats: CSV, ODT, and JSON. Data export in the JSON-LD format is possible as well, but this is undoc-

umented and cumbersome. The data was exported as CSV files and a Python script was used to detect inconsistencies in relation properties (relations in node-goat have to be added in both directions and can have different properties). After a few iterations of updating the source data, exporting, and checking, the exported CSV files were reshaped to prepare the data for wikification.

We then used the OpenRefine software tool to reconcile the people, organization, and periodical entities with Wikidata⁵. OpenRefine includes a reconciliation extension tailored to Wikidata. We used OpenRefine in order to determine which of the entities described in the WeChangEd dataset already exist in Wikidata, and which entities we would need to create new items for in Wikidata. The reconciliation process proposed matches between entities described in the CSV files with entities described in Wikidata. These proposed matches were created on the basis of string matching. We then manually reviewed the matches proposed by OpenRefine to verify that the entities were correct matches. This is an important quality control step that helps reduce the risk of creating duplicate items in Wikidata.

We proposed a new property in Wikidata for WeChangEd to help us identify the items in the dataset and write SPARQL queries related to the data. We proposed the property as an external identifier using Wikidata's community property proposal process⁶. Each item for each resource from the WeChangEd database has a statement with this property that indicates the identifier in the original database. This property is very helpful in writing SPARQL queries about this set of items important to the WeChangEd project.

After reconciling the WeChangEd data with Wikidata we then needed to write statements to Wikidata to contribute the data. We used a tool called WikidataIntegrator (WDI) to write the WeChangEd data to Wikidata. WDI is a python library for interacting with data from Wikidata [10]. WDI was created by the Su Lab of Scripps Research Institute and shared under an open-source software license via GitHub⁷. Using WDI as a framework, we wrote eight scripts for our bot to use to transfer data from the WeChangEd database to Wikidata.

⁵<https://openrefine.org/>

⁶https://www.wikidata.org/wiki/Wikidata:Property_proposal/WeChangEd_ID

⁷<https://github.com/SuLab/WikidataIntegrator>

We used the WDI library to prepare scripts for the bot to write statements with our selected properties and values from the WeChangEd data. We proposed the bot plan to the community on March 6, 2020 and it was approved on March 21, 2020⁸. The bot review process helps to ensure that the work a bot does will not cause problems for other editors and that the proposed data model that the bot will follow will fit with other data models in Wikidata.

We used an object-oriented-programming approach with the django-wikidata-api⁹ to generate SPARQL queries that model the WeChangEd dataset into python classes based on the entities we contributed to Wikidata. The combination of django-wikidata-api and WDI allowed us to mitigate the risk of duplicating items upon re-execution of the bots. This approach also allowed us to use both the Mediawiki API (via WDI) for contributing new statements as well as the SPARQL endpoint (via django-wikidata-api) for data retrieval and reporting after the bots were finished.

We used the EditGroups tool¹⁰ to tag our bot runs. By integrating a code snippet into our WDI bots, each batch of edits was assigned a unique identifier and a page on the EditGroups tool. We could then programmatically treat that set of edits as a unit. This is useful if a set of edits needs to be reverted. In the process of reviewing one group of edits from one of our bots we noticed an error in the structure of a qualifier. We were able to use EditGroups to revert the entire set with a single click. Use of the EditGroups tool allowed us to save us considerable time in avoiding manual reverting, which is tedious and error-prone.

We used the wbstack platform to create an instance of Wikibase for testing¹¹. The wbstack service provides a hosted version of Wikibase that users can load with their own data. Wikibase is the software used to support Wikidata itself. We were able to create an empty knowledge base that works as Wikidata does where we could run the bots as a testing step. This allowed us to ensure that our bots were operating as we expected them to operate. Wikidata also provides a testing area where people can experiment before taking actions in Wikidata itself¹².

⁸https://www.wikidata.org/wiki/Wikidata:Requests_for_permissions/Bot/WeChangEdBot

⁹<https://github.com/kennethsn/django-wikidata-api>

¹⁰https://www.wikidata.org/wiki/Wikidata:Edit_groups

¹¹<https://www.wbstack.com/>

¹²<https://test.wikidata.org/>

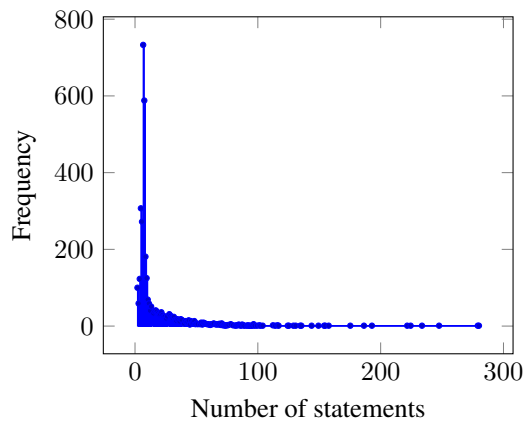


Fig. 2. Frequency of the number of statements per item in Wikidata for WeChangEd dataset

Aligning the WeChangEd data with Wikidata required finding patterns to express relationships in the original data and mapping them to items and properties from the Wikidata knowledge base. We created a data model using relevant Wikidata properties to represent the data created by the WeChangEd project team. In the WeChangEd data information was collected about the start and end dates of an editor's involvement with various periodicals. In Wikidata start and end dates are represented as qualifiers to other statements. In this case we created statements on the items for the periodicals using the property 'editor' and then applied the 'start date' and 'end date' qualifiers to those statements.

Once we aligned the data with Wikidata, we were able to create an average of seven statements per item. The chart in Figure 2 represents the frequency of the number of statements across items with a WeChangEd identifier in July, 2020. Many of these items now have more than seven statements. Some have dozens and a few items already have hundreds of statements. After contributing their dataset to Wikidata, the WeChangEd team now has additional information about many of the resources to consult.

6.1. How this Data Enhanced Wikidata

Donating this data to Wikidata contributed to the diversity and breadth of data in the knowledge base. At the time of this writing, of the set of items in Wikidata that describe people, eighty percent of those items represent males. When WeChangEd completed this donation it resulted in 1044 WeChangEd Persons and related bibliographical data, 1687 WeChangEd Periodi-

cals, and 219 WeChangEd Organizations to Wikidata. The majority of these were women. This is a contribution that works to counteract the gender gap in Wikidata [11, 12].

The WeChangEd team assembled extensive data about changes in the editors' names over time. These changes were due to marriage, noble titles, use of pseudonyms, or being referred to by their husband's names. We added each of these names as aliases in their Wikidata items. This means that end-users of Wikidata who search for these people will now be more likely to find the correct Wikidata item regardless of the version of the name they are using to search.

Of the 1,687 periodical titles in the WeChangEd dataset, roughly 1,550 of them were new for Wikidata. In August, 2020 there were more than 53,000 periodical titles in Wikidata. The WeChangEd data donation increased the coverage of periodicals from the 1700s, 1800s, and early 1900s. Researchers who consult Wikidata now benefit from readily available information about where these publications were published, who edited them and dates of their runs.

Of the 219 organizations in the WeChangEd dataset, roughly 150 were new for Wikidata. Many of these organizations had a mission related to women's rights. The WeChangEd data donation helps complete gaps in coverage of organizations with specific social functions.

This project enhanced Wikidata by serving as a model for using software tools to help automate data creation that is relatable for humanists. WikidataIntegrator has been used by many groups in the biological and life sciences to import datasets [10, 13–15]. There are also several projects that center topics relevant to the humanities in relation to Wikidata [16–18]. Our use of WikidataIntegrator in the field of literary studies demonstrates workflows that may be useful for other humanists who are not yet familiar with Wikidata.

6.2. How Alignment with Wikidata Enhanced WeChangEd

Once we contributed the WeChangEd data to Wikidata we quickly experienced the benefits of a collaborative knowledge base in the form of error correction and enrichment. There were a small number of cases where the dates of birth and the dates of death for an individual were reversed in the nodegoat database. These were quickly identified by a Wikidata editor who regularly runs a query to catch items where the death date is chronologically before the birth date. This user mes-

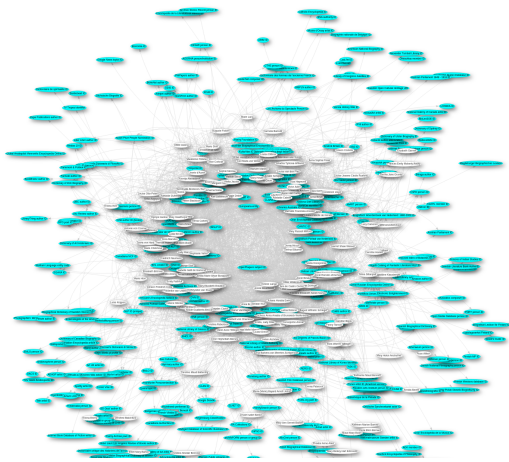


Fig. 3. Graph of external identifiers for people in the WeChangEd dataset available in Wikidata in July, 2020

saged us to let us know of the error and also corrected the values in Wikidata.

The WeChangEd team recorded Virtual International Authority File (VIAF) and International Standard Name Identifier (ISNI) identifiers for many of the resources in their dataset. These external identifiers helped us confirm matches with existing Wikidata items. Wikidata is recognized as a hub of external identifiers [19]. External identifiers are a type of property in Wikidata, there are thousands of these properties in use. After aligning the WeChangEd data with Wikidata, we not only knew the VIAF and ISNI identifiers for resources, we also had access to all of the identifiers that were already in Wikidata to describe them. In Figure 3 we see a graph visualization of the external identifiers for people available from Wikidata. The process of aligning the WeChangEd data with Wikidata expanded our ability to quickly find pathways to additional sources of information about these editors.

For resources in the WeChangEd dataset for which we had references to scholarly works, we were able to connect the statements to their supporting publications as seen in 4. Each of these publications have their own Wikidata item, thus metadata about the publications is available for consultation and reuse. Metadata about these publications is used in the WeChangEd stories application¹³.

After contributing the WeChangEd data to Wikidata we were able to write SPARQL queries to ask questions of this data in combination with other data in the knowledge base. We wrote a query to return items that

¹³<https://stories.wechanged.ugent.be/>



Fig. 4. Usage of the 'described by source' property

have WeChangEd Id and image¹⁴. This query is a request for all items in Wikidata from the WeChangEd dataset that have a link to one or more images in Wikimedia Commons, the media repository of the Wikimedia Foundation. This query allows the WeChangEd team to quickly find images they can use to illustrate their projects and publications.

We also wrote a query for items from the WeChangEd dataset that have something named after them¹⁵. These include streets, buildings, and prizes, among other things. Some of these connections were made by other editors and some were made through our process of aligning the data with Wikidata. This is a nice example of a query that we expect to return additional results over time, as the data in Wikidata grows. The results of this query could be useful in creating map visualizations of editors' lives or as a launching point for additional research into a person, organization or periodical.

In August, 2020 there were more than 550 works that have a resource from the WeChangEd dataset listed as a main subject¹⁶. We anticipate that the result set returned by this query will grow over time as more editors create connections between works and their main subjects. These works are biographies, obituaries or other types of reference materials that serve as stepping stones for further research about the resources in WeChangEd. It is possible that facts drawn from these reference works will be added to Wikidata as statements, which will increase the number of machine-readable statements we all will have access to about these topics.

Contributing the WeChangEd data to Wikidata allows us to use tools in the Wikimedia ecosystem to interact with this data. We can now use the MediWiki

¹⁴<https://w.wiki/xZb> A query for all items with a WeChangEd id and an image.

¹⁵<https://w.wiki/xZc> Query for all items in Wikidata from the WeChangEd dataset that have one or more other items named after them.

¹⁶<https://w.wiki/ZJG> Query for works with a resource from WeChangEd as main subject.

API, the Wikidata Query Service, and these tools allow us to get the data out in a range of formats.

7. Digital Storytelling

Thanks to the many options for getting data out of Wikidata, we were able to create an application for the display of the WeChangEd dataset. This application fetches data from Wikidata and presents it in the format of an interactive website. The web application allows users to explore the WeChangEd dataset¹⁷. This application was inspired by Science Stories [20]. The WeChangEd Stories application is built within the same framework that powers Science Stories, and the data drawn from Wikidata is combined with data from Wikipedia and Wikimedia Commons.

The WeChangEd Stories application centers the lives of the editors in the WeChangEd dataset. For each editor, the application presents a series of moments highlighting the organizations, places, and other people with which the person is associated. There is a moment that displays all publications that the person edited, authored, or of which the person is a main subject. The moments can include video or images of the person. These are the stories we can compile from this LOD.

We used Python as the primary programming language for the backend of the application. We used the Django framework for the Stories Service layer. For storing configurations about the presentation metadata of each story we used a Postgresql database. For added performance in our API, we use Redis to cache Wikidata SPARQL query results used throughout the application. To offload long-running processes we have a Celery server for queuing complex SPARQL queries, and use API polling in the frontend once the tasks have been executed.

The Stories Services team maintains a package for working with Wikidata data in a Django application¹⁸. The Django-wikidata-api library is designed loosely around the core Django ORM, but with support for interacting with wikibases via SPARQL queries instead of a relational database. The WikidataItemBase python class within the package has an interface for determining which statements are needed to represent a dataset, generating OpenAPI documentation, and constructing serializers compatible with the Django Rest

Framework¹⁹ to provide JSON responses of the data. We use ReDoc as our API documentation framework and Swagger for an API playground. The API uses both Token Authentication, for admin users to manage their story configurations, and API Key Authorization to render stories hosted from another application.

There are two primary functions of our frontend: 1) rendering a story and 2) managing a the presentation of a story. For Story rendering, we developed react-stories-api²⁰, a React.js open source component library of all visual story elements seen on the WeChangEd application. This allows us to decouple the infrastructure from the presentation so that the WeChangEd team can host the stories on any domain while the data is powered through the Stories Services API. The goal for our development of the react-stories-api library was for there to be little maintenance as possible for people to host their own stories application as a statically-served single-page application (SPA). The library comes bundled with all the necessary rendering components as well as the API client itself inside a single StoriesAPIStory component. The react-stories-api also provides a component for presenting all Stories within a Collection along with search and pagination support. The WeChangEd Stories website leverages both components to create a full user experience including routing and navigation with no backend server or data store needed. We use Material Design²¹ as our design framework and most of the components in the library are built with the Material-UI core library²².

For managing the collection and story presentation, we developed a Publisher Workspace to serve as the visual frontend of API operations. While the Stories Services API layer powers the data in the WeChangEd website, the Publisher Workspace is where admin users can rearrange the ordering of moments, modify the story metadata itself, and most importantly, enhance the stories with curated content such as images, videos, and links found outside of Wikidata. The Publisher Workspace is built using React.js and has react-stories-api as a core dependency to provide publishers real-time previewing of their story selections using the same presentation library as their hosted site.

¹⁹<https://www.django-rest-framework.org/>

²⁰<https://github.com/kennethsn/react-stories-api>

²¹<https://material.io/>

²²<https://material-ui.com/>

¹⁷<https://stories.wechanged.ugent.be/>

¹⁸<https://github.com/kennethsn/django-wikidata-api>

8. The Case for Linked Data in Research Projects

Creating LOD as part of our research projects allows us to connect data to webs of other data [1]. All items with a WeChangEd ID are now connected by properties to other items in the Wikidata knowledge base. The SPARQL query language allows us to ask questions that make use of any of these connections. Rather than maintaining a database silo that only the original research team can consult, we have contributed the WeChangEd data to an international, public database that anyone with access to the internet may consult. More people can now share in the knowledge produced by this research team.

Researchers and technologists have created many visualization options for LOD. Once data is published as LOD, we can reuse frameworks and packages for presenting visualizations of this data for a variety of purposes. For example, the Wikidata Query Service provides support for creating graphs, charts, bubble diagrams, network graphs, and image grids from a menu in the user interface for the SPARQL endpoint. Researchers do not have to create visualizations using other software tools, they can select the options they would like directly from Wikidata Query Service [21]. This allows researchers who are new to data visualization to access additional formats when communicating their results.

Publishing LOD allows us to communicate with a wider audience. Anyone searching Wikidata itself, or any project reusing Wikidata data, will find references to the WeChangEd project on any of the statements the project team contributed. This audience is wide and still growing, as the data will persist for future searches for years to come.

From the perspective of digital preservation, donating the WeChangEd dataset to Wikidata will result in greater longevity of the data. A project-team based preservation strategy is more costly and time-intensive to maintain because the team would have to provide server space for the database and train people to maintain the data. Wikidata will be maintained for the duration of the Wikimedia Foundation itself.

This project is an example of an interactive application built to showcase a specific view of a Wikidata subgraph. The subgraph highlighted is the set of resources with a WeChangEd identifier.

9. Conclusion

Archival work in retracing people of the past is often a laborious and time intensive task, where the discovering of related items is not always immediate. Leveraging Semantic Web technologies in the WeChangEd project allowed us to have a wider audience for our work and to develop an interactive, user-friendly presentation formats for our research to tell stories from Linked Open Data (LOD). The model described here from how to move from archive, to a structured data model and relational database, to a LOD model and make this information available on Wikidata, to the use of the Stories Services API to generate multimedia stories can be an important tool for recounting and sharing the past.

Publishing this data as LOD, ensures that this information can be integrated and accessible in the growing knowledge base of the Semantic Web and more specifically Wikidata; compared to what is often traditionally done in the Humanities of locally archiving databases and or making CSV files available on websites or upon request which creates silos of knowledge. Anyone who consults Wikidata now has access to the dataset created by the WeChangEd team about women editors in Europe and the organizations and periodicals with which they were affiliated. This dataset is woven into the Wikidata knowledge base via property relationships with other types of resources. Thanks to Wikidata we now know the geo-coordinates of all of the locations in the WeChangEd dataset. These affordances for both Wikidata and the research project were outlined in detail in Section 6/

As more humanists, social scientists and other researchers choose to contribute their data to Wikidata we will all benefit. As researchers add data, the breadth and complexity of the questions we can ask about the data we have contributed will increase. As more data is curated in Wikidata additional relationships between resources are created through the use of properties. Relationships between editors in the dataset and other people described in Wikidata will continue to grow over time. The sets of information connected to the WeChangEd dataset will be enriched and expanded over time, making this an evolving dataset.

A research team partnering with Wikidata experts was an effective collaboration. This partnership allowed us all to learn from one another. The Wikidata experts relied on the WeChangEd team for domain knowledge and feedback on data modeling decisions. The WeChangEd team was able to continue their nor-

mal research activities without having to learn every detail of how to work with Wikidata.

In addition, the WeChangEd Stories App, developed using the Stories Services API, to merge this distributed data, afforded an unique, interactive and user friendly way to view LOD data on women editors. It also provided a new entrance point into the data and discovering the diverse and distributed material on the Web of women editors specifically. It showcases a new way of telling stories from machine readable information, in a visual appeasing and more accessible manner than the technical SPARQL query. Building applications that syndicate data from Wikidata allows us to leverage a general purpose knowledge graph with a growing number of references back to scholarly literature. Using frameworks developed by the Wikidata community allows us to rapidly provision interactive sites that will help us engage new audiences.

10. Acknowledgements

We would like to thank a number of people and organizations for helping us realize this work; for both cheering us on to generously supporting us with their expertise. First, we would like to thank the Su Lab for creating the WikidataIntegrator, the Python library we used to contribute WeChangEd data to Wikidata. We thank the Wikidata community for sharing thoughts, feedback, data curation and data modeling work. We thank Wikimedia Deutschland²³ for supporting the Wikidata project. We are very thankful for all the fruitful discussions between attendees that emerged during Ghent University's Wikidata/ Wikibase Workshop in 2019²⁴. In addition, we would specifically like to thank Andra Waagmeester- you are awesome!, the entire WeChangEd team and those that contributed their knowledge on women editors to the WeChangEd Database, and Geert Van Pamel from Wikimedia Belgium. The work of the third and four author was supported by the H2020 European Research Council under the ERC Starting Grant agreement no. 639668.

References

- [1] S. Van Hooland and R. Verborgh, *Linked Data for Libraries, Archives and Museums: How to clean, link and publish your metadata*, Facet publishing, 2014.
- [2] S. Allison-Cassin, A. Armstrong, P. Ayers, T. Cramer, M. Custer, M. Lemus-Rojas, S. McCallum, M. Proffitt, M. Puente, J. Ruttenberg et al., ARL White Paper on Wikidata: Opportunities and Recommendations (2019).
- [3] M. Van Remoortel, Women Editors and the Rise of the Illustrated Fashion Press in the Nineteenth Century, *Nineteenth-Century Contexts* **39**(4) (2017), 269–295.
- [4] J. Schelstraete and M. Van Remoortel, Towards a sustainable and collaborative data model for periodical studies, *Media History* **25**(3) (2019), 336–354.
- [5] P. van Bree and G. Kessels, nodegoat: a web-based data management, network analysis & visualisation environment, 2013. <http://nodegoat.net>.
- [6] M. Van Remoortel, J.M. Birkholz, J. Schelstrate, M. Alesina, C. Bezari, C. D'Eer and E. Forestier, WeChangeD Database, 2019. <https://www.wechanged.ugent.be/wechanged-database/>.
- [7] W3C, Semantic Web Activity, 2009. <https://www.w3.org/2001/sw/>.
- [8] D. Vrandečić and M. Krötzsch, Wikidata: a free collaborative knowledgebase, *Communications of the ACM* **57**(10) (2014), 78–85.
- [9] F. Erxleben, M. Günther, M. Krötzsch, J. Mendez and D. Vrandečić, Introducing Wikidata to the linked data web, in: *International semantic web conference*, Springer, 2014, pp. 50–65.
- [10] A. Waagmeester, G. Stupp, S. Burgstaller-Muehlbacher, B.M. Good, M. Griffith, O.L. Griffith, K. Hanspers, H. Hermjakob, T.S. Hudson, K. Hybiske et al., Science Forum: Wikidata as a knowledge graph for the life sciences, *ELife* **9** (2020), e52614.
- [11] M. Klein, H. Gupta, V. Rai, P. Konieczny and H. Zhu, Monitoring the gender gap with wikidata human gender indicators, in: *Proceedings of the 12th International Symposium on Open Collaboration*, 2016, pp. 1–9.
- [12] P. Konieczny and M. Klein, Gender gap through time and space: A journey through Wikipedia biographies via the Wikidata Human Gender Indicator, *New Media & Society* **20**(12) (2018), 4608–4633.
- [13] T.E. Putman, S. Lelong, S. Burgstaller-Muehlbacher, A. Waagmeester, C. Diesh, N. Dunn, M. Munoz-Torres, G.S. Stupp, C. Wu, A.I. Su et al., WikiGenomes: an open web application for community consumption and curation of gene annotation data in Wikidata, *Database* **2017** (2017).
- [14] F.Å. Nielsen, D. Mietchen and E. Willighagen, Scholia, scientometrics and wikidata, in: *European Semantic Web Conference*, Springer, 2017, pp. 237–259.
- [15] A. Waagmeester, E.L. Willighagen, A.I. Su, M. Kutmon, J.E.L. Goyo, D. Fernández-Álvarez, Q. Groom, P.J. Schaap, L.M. Verhagen and J.J. Koehorst, A protocol for adding knowledge to Wikidata, a case report, *BioRxiv* (2020).
- [16] A. Heftberger, J. Höper, C. Müller-Birn and N.-O. Walkowski, Opening up Research Data in Film Studies by Using the Structured Knowledge Base Wikidata, in: *Digital Cultural Heritage*, Springer, 2020, pp. 401–410.
- [17] E. Hyvönen, P. Leskinen, M. Tamper, H. Rantala, E. Ikkala, J. Tuominen and K. Keravuori, BiographySampo—publishing and enriching biographies on the semantic web for digital humanities research, in: *European Semantic Web Conference*, Springer, 2019, pp. 574–589.
- [18] F.Å. Nielsen, Literature, Geolocation and Wikidata., in: *Wiki@*

²³<https://wikimedia.de/wiki/Hauptseite>

²⁴<https://www.wikidata.org/wiki/Wikidata:Events/Belgium/UniversiteitGent/Wikidata2019/Wikibaseworkshop/2019>

- [19] J. Neubert, Wikidata as a Linking Hub for Knowledge Organization Systems? Integrating an Authority Mapping into Wikidata and Learning Lessons for KOS Mappings., in: *NKOS@TPDL*, 2017, pp. 14–25.
- [20] K. Thornton and K. Seals-Nutt, Science Stories: Using IIF and Wikidata to Create a Linked-Data Application., in: *International Semantic Web Conference (P&D/Industry/BlueSky)*, 2018.
- [21] F.Å. Nielsen, Ordia: A Web Application for Wikidata Lexemes, in: *European Semantic Web Conference*, Springer, 2019, pp. 141–146.
- [22] M. Van Remoortel, *Women, Work and the Victorian Periodical: Living by the Press*, Springer, 2015.
- [23] M. Van Remoortel, Who do you think they were? What genealogy databases can do for Victorian periodical studies, in: *Re-searching the nineteenth-century periodical press: case studies*, Routledge, 2018, pp. 131–144.
- [24] J. Schelstraete and M. Van Remoortel, Towards a sustainable and collaborative data model for periodical studies, *Media History* **25**(3) (2019), 336–354.
- [25] J.M. Birkholz and A. Meroño Peñuela, Decomplexifying networks: a tool for RDF/Wikidata to network analysis, 2019.
- [26] D. Abián, G. Candela, J. Birkholz, M. Dolores Saez, P. Escobar, S. Chambers, I. Martinez-Sempere and J. Vicente Berna-Martinez, Wikidata/Wikibase Workshops: Lessons Learned, in: *WikidataCon 2019*, 2019.
- [27] T. Putman, K. Hybiske, D. Jow, C. Afrasiabi, S. Lelong, M.A. Cano, G.S. Stupp, A. Waagmeester, B.M. Good, C. Wu et al., ChlamBase: a curated model organism database for the Chlamydia research community, *Database* **2019** (2019).