




Slide 1


An Introduction to the Wikidata Thesis Toolkit




Helen Williams







Ruth Elder



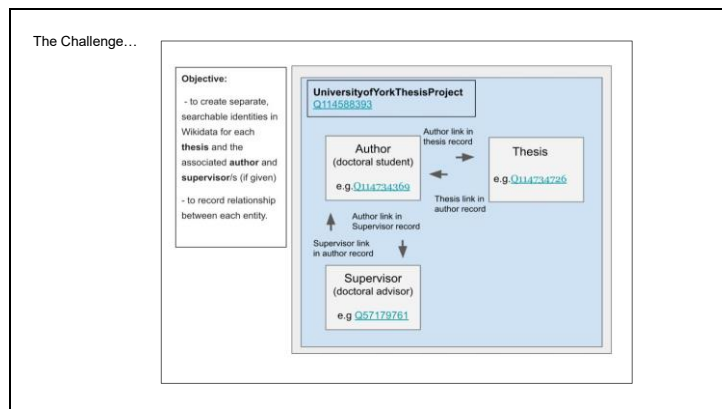
Ruth

Good morning from myself and Helen, and welcome to this session on introducing the Wikidata toolkit.

As a very short preamble, the toolkit is built on the initial work of Helen, (starting back in lockdown in 2019), to develop a process to upload theses metadata into Wikidata - work which was prompted by her interest in the growth of Wikidata as a topic of conversation within the metadata community.

Subsequent work by myself to develop a similar process flow for York helped to refine the process further, and has since led to a fruitful and engaging collaboration between the two libraries resulting in the publication of the toolkit earlier this year, which we are very happy to be sharing with you today.

Slide 2



Wikidata is described as a:

- Structured database operating as the central data store for all Wikimedia projects.

And a

- ‘Free and open knowledge base containing multilingual data that can be read, edited and re-used by humans and machines, supporting global access to information.

So what exactly was the task in hand that Helen and I had challenged ourselves with in order to take advantage of these attributes of Wikidata?

Well, that was to develop a sustainable process to input pre-existing theses metadata already held in digital institutional repositories or datasets, into Wikidata in order to promote original research to the widest possible audience through signposting back to the repository.

And to make that achievable through creating separate, searchable identities in Wikidata for each doctoral thesis and its associated author and supervisor (if given), and to record the relationship between each entity... as represented graphically on the slide.

“Improving access and discovery of our collections online, enabling local and global audiences to discover, use and enjoy them, using an ‘open as possible’ approach”

University of York Library, Archives and Learning Services (01753 7526)

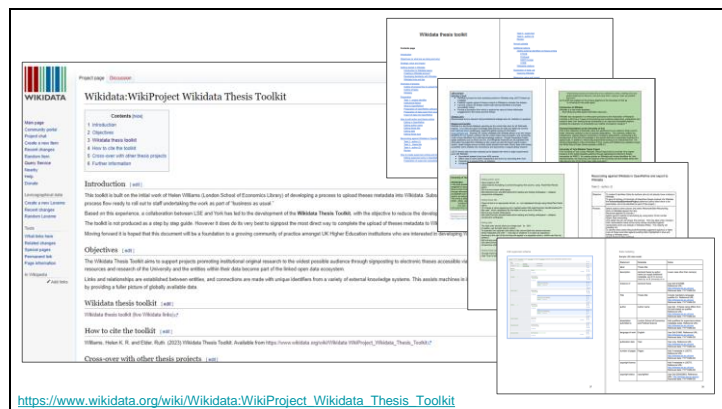


So if that was the challenge ... in a world of competing demands and limited resources - why should any of us say yes to it, and invest staff effort and resource in using Wikidata, and specifically at adding institutional doctoral theses metadata into Wikidata?

And the short answer is really - because it supports priorities which are likely to feature in all of our strategic plans...

- It makes scholarly content more widely accessible and visible,
- It promotes the role of the institution as a provider of open knowledge to local and global audiences,
- It enables wider public access to research outputs
- Through the creation of unique identifiers for research outputs, institutional content and the entities within it become part of the Linked Open Data ecosystem.
- Google Knowledge Graphs, digital assistants, and Wikipedia infoboxes are all populated, in part, with information harvested from Wikidata, so this work impacts on search engine results by providing a fuller picture of globally available data.
- And at a more local level, many libraries will be looking to expand the work of their metadata teams beyond traditional cataloguing and develop staff skill sets to future proof roles – Wikidata is a good access point to introducing linked data work and expand the range of staff digital skills, confidence and experience.

Slide 4



And so through a series of discussions sharing our experiences as self-taught Wikidata beginners, and all the trials, errors, false starts and journey’s down cul-de-sacs we both experienced, Helen and I linked up in December 2022, and have since co-created the Wikidata Thesis Toolkit, (and met face to face for the first time this morning!)

The aim of creating this toolkit is to reduce the development burden for other institutions looking to establish similar projects. It is designed as a guide, rather than a step-by-step handbook, and I am sure there are other different, potentially better ways of doing things, but it reflects our knowledge and experience at the time.

The toolkit is now available through a Wikidata project page, and has recently been shared with the wider academic library and Wikipedia communities.

And our hope is that the document will be a starting point to a growing community of practice amongst UK HE and GLAM institutions who are interested in developing Wikidata work and sharing experiences with one another.

And at this point, I will hand over to Helen to continue...

Helen

In support of the open research agenda we want to help you promote doctoral dissertations to the widest possible audience by increasing their visibility and accessibility - because as more doctoral research metadata is included in Wikidata there’s the potential for more unexpected connections and relationships to be surfaced, meaning the data can be explored in new ways, beyond institutional silos, to make sense of combined cultural heritage.

So how is the toolkit going to be of practical use to you?

We’ve started by signposting to resources which will introduce you to the basics of Wikidata, and provide foundational knowledge - that will take 30 minutes to an hour, so it’s a really low barrier as a starting point. There are links to some further reading if you have time, but if you want to jump straight in there’s guidance on setting up an account and practical tasks to develop basic familiarity.

There’s a process overview, followed by the actual steps to add metadata to Wikidata, including data preparation, editing in OpenRefine, reconciling names with Wikidata, and

creating Qids for individuals and thesis titles. And some bonus guidance on linking theses to external identifiers, creating links to your theses on relevant Wikipedia pages, and using SPARQL to visualise your data in ways that aren't usually possible through your institutional repository.

Institutional rankings <https://w.wiki/jwZ>

The screenshot shows the Wikidata Query Service interface. At the top, there are navigation buttons for 'Examples', 'Query Builder', and 'Help'. Below these is a SPARQL query editor with the following code:

```

1 SELECT (COUNT(?thesis) As ?count) ?institutionLabel WHERE {
2   ?thesis wdt:P31 wd:Q187685;
3   wdt:P4101 ?institution
4   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_L
5 } GROUP BY ?institutionLabel ORDER BY DESC(?count)

```

Below the query editor is a small image with the word 'TOGETHER' and several colorful figures holding hands. To the right of the query editor is a table of results:

count	InstitutionLabel
9882	Delft University of Technology
7464	University of Auckland
6612	University of Washington
6220	University of Leeds
5675	University of Oxford
4505	University of Canterbury
4422	Vanderbilt University
4147	London School of Economics and Political Science
3627	Massey University
3026	National University of La Plata
3381	University of Otago

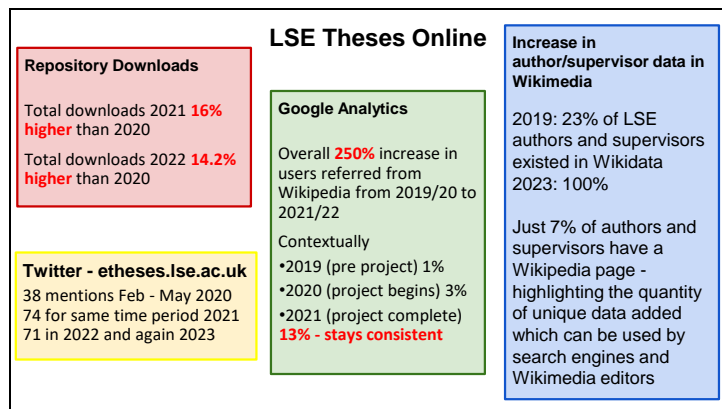
The toolkit also provides some guidance on how you can measure the impact of your work. There are a couple of challenges with this. Firstly getting your metadata into Wikidata is about enabling other sources to make use of it, so it's not always easy to measure impact. Secondly, terminology - in the university environment the word impact tends to refer to an effect on the wider world in terms of demonstrable impact in life changes or opportunities - so at LSE for measuring what I would usually refer to as the impact of our work we talk about reach and engagement.

There's some obvious immediate value in that you and colleagues are developing digital skills, working with and contributing to the broader network of Linked Open Data and promoting your unique collections beyond the usual library silos - and you can shout about those things to institutional research and PhD teams, and alumni, to raise the profile of the Library and demonstrate the value of metadata in ensuring that your unique content can be understood in the semantic web environment.

But senior managers are more likely to want evidence about why time and resources should be invested - and that requires some analysis, which isn't something I've seen reported for many Wikidata projects. So I'm going to whip you through how we've approached this, and if you've got further ideas then we'd be really keen to hear about them.

Depending on where you work you might find that you're asked to look at what comparator institutions are doing, or where your work sits with what others are doing, so institutional rankings for doctoral theses in Wikidata might be useful... though they'll obviously change over time, and might be less valuable if you're a smaller institution.

Slide 6



Repository downloads can be a useful measurement.

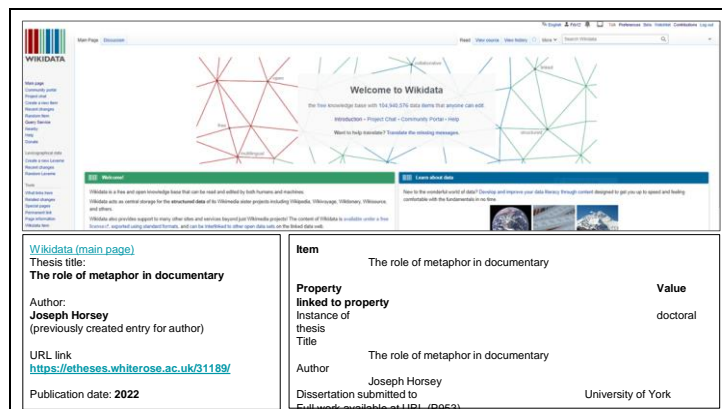
We've seen some encouraging figures from Google Analytics

Our Twitter mentions pretty much doubled over the time period we looked at - you could use paid resources to analyse this more extensively.

We've looked at the extension of institutional names in Wikimedia which highlights the quantity of unique data we've added which can now be used by search engines and Wikimedia editors. And finally we've just started looking at Overton and Scopus to see where LSE theses have been cited in policy documents and at Scopus to see where they're cited in other research - really early days so no reportable figures on that yet.

We'll come back to visualisations, but first time for a live demo!

Slide 7



Ruth

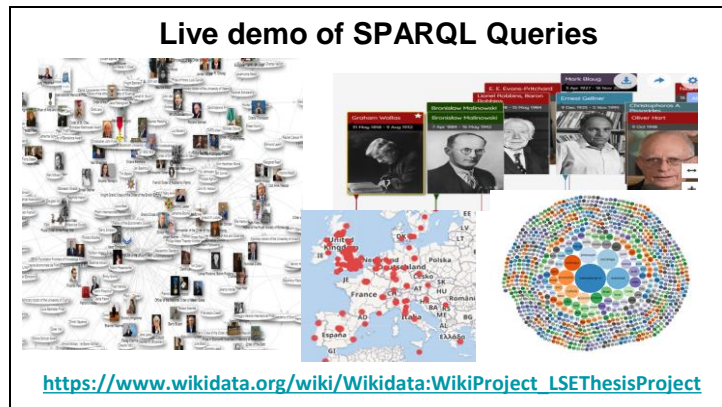
So without the aid of a safety net I am now going to show you how to add a single thesis entry into Wikidata – so fingers crossed the technology behaves at this point...

Starting your Wikidata journey by adding individual manual entries for theses and their authors is a brilliant way of finding your way around Wikidata, and to understand how the elements work together before moving onto bulk edits and uploads.

And making mistakes and learning how to correct them is all part of the process, so I suggest you embrace them as all part of the learning journey.

- Once you have set up your account and logged into Wikidata you will face this main Wikidata page.
- We are going to add a basic thesis record to Wikidata ...
- And our first step is by checking if the thesis is already entered into Wikidata, by entering the title into the search box at the top right. **The role of metaphor in the documentary**
- As it comes up with no entry being listed, we have the option to create a new item....
- And a very basic record framework comes up here – which by default identifies that the entry is in English,
- Enter the **title of the thesis: The role of metaphor in the documentary**
- And a brief description – which here we are using **doctoral thesis** for.
- So this has created the briefest of records – but we do want to add some additional information to this
- And we do this in the form of adding statements.
- Statements link a property to the item (in this case the thesis entry) and a value to the property.
- And multiple statements are entered within the WD entry for a specific item –
- So to add additional information:
- So next an example of adding the **property of author** into the record, and associating it with the **value** of “**Joseph Horsey**” and publish
- The **property of title** – and the **value** of “**The role of metaphor in documentary**”
- We will add the property of **Publication Date** and add the value of **2022**.
- The **property of dissertation submitted to** and the **value** of **University of York**
- And the **property of Full work available at URL**, and the
- **value** of **https://etheses.whiterose.ac.uk/31189/**

- And if we now do a search for the thesis title of **The role of metaphor in documentary**, we have the beginnings of a decent record which we can add to further, or anyone else is able to edit and make additions to.



Helen

Because the data on Wikidata is all described in the same structured format it can be queried using the built in SPARQL query tool.

If you're a beginner you don't even need to know SPARQL, you can borrow and edit queries already written by other people.

So, if you want to query thesis data for your institution, you can use some of the existing queries on LSE's project page - and just substitute the Qid for your own institution. And the beauty is that you're linking your data with all the other data in Wikidata to make connections you might not previously have been able to - for example

https://www.wikidata.org/wiki/Wikidata:WikiProject_LSEThesisProject

We can see what awards have been won by our thesis authors and supervisors - and if Ruth wants to see that data for York we just substitute the Qid for her project (Q114588393)

Or we want to see a map of where people have gone on to be employed... and we can swap that to show the data for York.

Or we can show that as a table instead of map by changing the default view.

Obviously once you've completed your thesis project the thesis data for your institution is complete when you run queries - but data you link with, like awards or employers, isn't necessarily complete - but it can still be useful - you just need to make that clear.

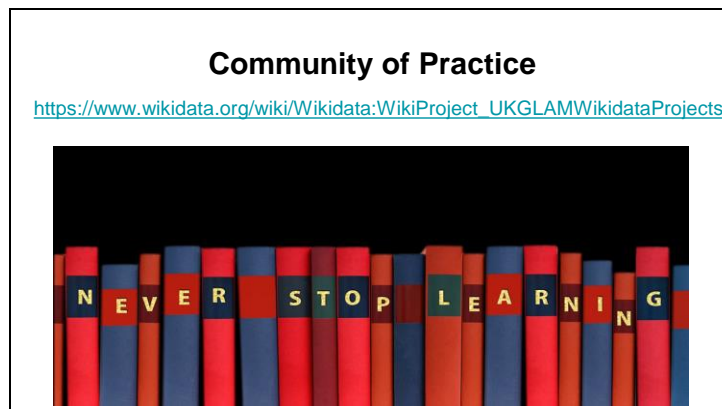
Reflections and call to action!

- Jump in and have a go
- Enjoy it and have fun
- Don't be afraid to make mistakes and learn from them
- Be confident you can make an impact
- Innovate and share ideas
- Get involved in the Wikimedia community



We have both so enjoyed working with Wikidata - we both find it one of the most satisfying parts of our jobs.

We would really encourage you to jump in and have a go... you don't have to know everything to make an impact - just start where you are, play around, learn as you go along - with the help of the toolkit and the wider community - and as your skills develop you'll doubtless come up with new innovative ideas, which we'd really encourage you to share with the community.



There are plenty of sources of help out there...

The discussion pages on Wikidata itself, WikiEdu, the OpenRefine community, the LD4 Wikidata Affinity Group and Wikimedia experts - and we'd also really like to encourage a UK community of practice so that we can share what we're working on, reduce the development burdens for each other in specific areas, and see opportunities for collaborative projects and exploring links between our data.

We'd really like that to be a community-led initiative so we're going to use Wikidata itself as our forum, and from there anyone can link to their projects and initiate discussion.

I've set up a page, so please jump in and let's get that going as an active community space.
https://www.wikidata.org/wiki/Wikidata:WikiProject_UKGLAMWikidataProjects

Slide 11

Further Information & Links	
Wikidata Thesis Toolkit Project Page	Wikidata Thesis Toolkit (PDF)
Wikidata:WikiProject:UniversityofYorkthesisproject	Wikidata:WikiProject LSEThesisProject
New Zealand Thesis Project (Wikipedia) Wikidata:WikiProject NZThesisProject	Wikidata:WikiProject Source MetaData/Theses by institution
<p>Clark, Jason A., Williams, Helen K.R., and Rossmann, Doralyn. 'Wikidata and Knowledge Graphs in Practice: Using Semantic SEO to Create Discoverable, Accessible, Machine-readable Definitions of the People, Places, and Services in Libraries and Archives'. 1 Jan. 2022 : 1 – 14. https://content.iopress.com/articles/information-services-and-use/isu220171, accessed December 16 2022</p>	
<p>Williams, Helen K. R., Wikidata: what? why? how? Catalogue and Index (203). pp. 28-35. ISSN 2399-9667 (2020) http://eprints.lse.ac.uk/110987/1/Williams_wikidata_what_why_how_published.pdf, accessed November 30 2022</p>	
<p>Williams, Helen K. R., LSE's adventures in Wikidata-land: tears and triumphs down the rabbit hole. Catalogue and Index, 206. pp. 2-6. ISSN 2399-9667 (2022) http://eprints.lse.ac.uk/114976/1/, accessed November 30 2022</p>	

Slide 12

Slide	Attribution
1	
3	https://www.wallpaperflare.com/book-book-open-book-pages-wind-read-sky-blue-education-wallpaper-gbsha
5	https://pixabay.com/illustrations/together-earth-people-board-school-2450090/
6	https://w.wiki/w/Z
8	Wikidata SPARQL query service
9	https://pixabay.com/photos/play-stone-network-networked-1237457/
10	https://pixabay.com/photos/adult-education-a-book-books-2706977/

Image attribution