


# NZBSI Auckland Workshop script

 NZBSI Auckland Wiki Workshop Slides Currently attached to slides

## Slide 1 - Introduction

Kia ora koutou I'm Siobhan Leachman, if you don't know already I'm currently the Wikimedian in Residence for the Bioeconomy Science Institute. Here you can see my contact details. If you have any queries about Wiki, reach out anytime. My contract expires at the end of June but I'm very keen to support anyone who wants to engage further with Wiki at BSI so please feel free to reach out after that date if you have questions or need support.

Also these slides are filled with links that can help you find more information on Wiki and assist you to edit. I've included a bit.ly link to the slides and a QR code to help you get to these slides. I've also included a burly link to the spreadsheet we'll be using for editing activities but more on that later!

Please feel free to ask questions at any point in this presentation and also during our editing activities.

## Slide 2 - Wiki links

I'm including here three links - the first is to the Wikimedian in Residence page and the second is the Wiki page for this workshop and the last is to all the workflows and documentation I've created during this Wikimedian in Residence. All these pages are filled with links and resources that you can explore to support your understanding and editing of Wikidata.

## Slide 3 - Plan for today

This is the plan for today. For the first hour I'm going to take you through a presentation giving you first a basic overview of Wikidata. Then I'm going to attempt to explain why you should care about having you and your institution's data in Wikidata and some of the methods you can use to ensure this happens. I want to give you a few examples of data reuse once that data. I'll give you links to workflow documentation to assist you in editing both manually and in bulk via OpenRefine. I also want to give you some information on querying Wikidata and give examples of some basic Wikidata queries. Finally I've got some editing tasks for you to undertake so you get confident about editing Wikidata. Please feel free to ask me questions at any time.

## Slide 4 - Wikidata is a sister project

What is Wikidata? Wikidata is one of a suite of Wikimedia sister projects. Other projects include the various language Wikipedias and the image repository Wikimedia Commons. Wikidata is a knowledge base that has around 42,000 active editors. It has over 120 million items all of which are interconnected. Just under 4% of those items are taxa items, around 9% relate to people and over 30% relate to scholarly articles - so all very relevant to your work.

## Slide 5 - Wikidata

Wikidata acts as the data backbone for Wikipedia, and has a much lower notability criteria than Wikipedia. While Wikipedia provides text, Wikidata stores structured data that are human- and machine-readable. To summarise:

- Wikidata is a free, open **knowledge base** anyone can edit.
- It is **multilingual** by design, you can use it in many different languages.
- It can be read and edited by humans and applications, so it's easy to improve a single item or thousands.
- The data can be **queried** and because they are openly licensed (Creative Commons 0, CC0), the data can be **reused** without restriction.
- Wikidata doesn't just store linked data it also acts as a **hub** for external **identifiers**! It also acts as a hub to interconnect all the various WikiProjects - so the Wikidata item for a particular concept should also link to all the language Wikipedia articles for that concept and the Wikimedia Commons category for all the images of that concept.

Wikidata's structure, openness, and accessibility makes it a powerful way to answer really complex questions.

## Slide 6 - Structured and linked

So as I've said Wikidata is made up of structured data, that is, data that is linked to other data. Thing 1 is linked to Thing 2. That linked structure is often called a triple, because it's made up of three parts: subject, predicate, object. For example Ynes Mexia has the occupation botanist. So Mexia's wikidata item is linked to the Wikidata property occupation which is then linked to the Wikidata item for "botanist". It is this linking that makes Wikidata different from a database and turns it into a knowledge base.

Importantly all structured data statements made on Wikidata items should be referenced citing where the data has come from.

## Slide 7 - Multilingual data

Wikidata is multilingual. Each item expresses a concept represented by the Wikidata Qid number. This concept can be expressed in multiple languages.

## Slide 8 - Multilingual editing

You can also read and add data to Wikidata in many languages. This means the data in Wikidata can be contributed to and enriched by non-English speakers. Multiple non English speaking editors enrich data related to New Zealand biodiversity every day.

## Slide 9 - Anyone can edit

Anyone can edit Wikidata, whether or not you have an account. But if you do have an account, other editors are more likely to assume you are editing in good faith. If you don't have an account your IP address will be recorded in the history of those edits. The community tends to be more suspicious of these types of edits because of vandalism.

## Slide 10 - CC0 Data

All the data added to Wikidata are licensed for reuse under a CC0 license. This means the data is dedicated to the public domain and is able to be reused by anyone for anything.

## Slide 11 - Identifier hub

One of the most important elements of Wikidata is that it acts like an identifier hub. In this slide you can see all the identifiers added to the Wikidata item for the botanist Ynes Mexia. The text is small because there are SO MANY identifiers listed on her item. If you look closely you'll see library identifiers, archive identifiers, online encyclopaedia identifiers, botany database identifiers, natural history institution identifiers, and genealogical database identifiers. So Wikidata acts like a crosswalk or a bridge between one institution's data and another institution's data.

## Slide 12 - Weave a data ecosystem

When you are editing Wikidata you are helping weave a data ecosystem. When adding or enriching items on Wikidata, you are increasing the shared knowledge available to everyone - data that can be reused across multiple platforms. It can be used by humans searching for information, but also by machines, thus increasing the pace at which data can be interlinked.

## Slide 13 - 2023 Wikidata intensity map

This is a map of any Wikidata item which has geographical data linked to it. It gives an indication of Wikidata coverage of particular areas of the planet. I particularly want to draw your attention to the fact that South America and Africa are not as well covered as locations in the global north. This is definitely something to have in your mind when editing or querying Wikidata. Wikidata is biased in its coverage. But the great thing is that anyone can edit and fill gaps.

## Slide 14 - Why YOU should care

Why you should care about Wikidata. One of the core purposes of your institution is to enrich and share knowledge with multiple communities. I would argue that by adding BSI data to Wikidata you will be fulfilling one of the core purposes of your institution.

## Slide 15 - Wikidata can help

Wikidata can help make BSI data more findable and accessible, as well as help with making the data interoperable and reusable.

Wikidata can help make BSI data **Findable** via its Persistent Identifiers: Every Wikidata item, property, and reference has a unique persistent identifier. Items are described with rich metadata, including labels, descriptions, and aliases in multiple languages, making them easily discoverable. Wikidata is also a searchable repository which is indexed by web search engines.

Wikidata can help BSI data become more **Accessible** as its data is accessible via free, open protocols, including HTTP and REST APIs, allowing for retrieval without specialized software. The Wikidata Query Service (SPARQL) enables direct machine access and retrieval.

Wikidata can help BSI data become **Interoperable** as Wikidata uses the Resource Description Framework (RDF) format, which is a standardized format for knowledge representation. Wikidata also uses controlled vocabularies and links data to other authority files (e.g., VIAF, ORCID) to ensure interoperability.

And of course Wikidata can help BSI data become more **Reusable**. Wikidata content is released under the Creative Commons Public Domain Dedication (CC0), allowing maximum reusability. Wikidata also provides detailed provenance - it includes references supporting statements made. The data is also structured in a way that adheres to community-driven models.

## Slide 16 - Internet search engines AI

I want to emphasise the findability of Wikidata - multiple internet search engines use Wikidata and also Wikipedia to inform their search results. And nowadays of course Wikidata serves as a massive, community-curated knowledge graph for AI, reducing hallucinations by providing structured, real-time data for retrieval-augmented generation.

## Slide 17 - Multitudes of data

And Wikidata is huge and is still growing. It contains multitudes of data and that data is continuously being enriched.

## Slide 18 - Biodiversity Knowledge Graph

I'm going to now discuss some examples of data in Wikidata by looking at certain elements of the Biodiversity Knowledge Graph. The Biodiversity Knowledge Graph is made up of the interconnected network of taxa, taxonomic names, publications, people, locations, sequences, images, collections and so much more. Wikidata can play a role in ensuring that all these disparate parts are connected and making it easier to help others to build upon these connections and reuse this data to support research.

## Slide 19 - People

So let's look first at People in Wikidata and here I'm including specimen collectors, determiners as well as scientific authors and even folk in this room.

## Slide 20 - Wikidata item

This is what a Wikidata person item looks like. It has a label which can be in multiple languages with the wikidata item identifier called the Qid next to it. Items should have a description which, as I've explained previously, can be in multiple languages. It also has an alias section called the "also known as" section. When I edit I add aliases as it reduces the number of duplicate items added to Wikidata and helps people find the data. The Wikidata item contains data in the form of statements. These statements should be referenced. There are two main ways to reference a statement. Either through a "stated in" statement which references a source described in a Wikidata item or a "reference url" statement which links to a website address where the source of the data can be found.

There is a separate section at the bottom of the item for identifiers. This section is automatically created when an identifier statement is added to Wikidata. The Identifiers section is how Wikidata links data from across organisations.

## Slide 21 - Botanist item properties

Here is a figure showing just some of the properties that can be used to add data to a botanist person item. You can see the amount of data that can be added about a person can be extensive.

## Slide 22 - Wikidata policy: Living people

For living people Wikidata has a policy that requires editors to strive to add only information we have a high confidence that is verifiable and which doesn't violate a person's reasonable expectations of privacy. For deceased people there is no longer any expectation of privacy and so long as the data is from a reputable source it is able to be combined into a Wikidata item.

## Slide 23 - Closer look

Now I just want to take a closer look at types of data on peoples items that I think is relevant to BSI work. The three types I've picked are : data on the institutions that hold a person's natural history collections, significant people for that person and also employer information for that person.

## Slide 24 - Examples of Wikidata statements

Here you can see examples of all these types of data as laid out in a Wikidata item for a person. You can see "Collection items at" statements on the Wikidata item for the entomologist George Hudson. This shows several of the collections that hold Hudson's specimens - two of which are BSI collections.

You can also see a "significant person" statement on a Wikidata item linking it to the Wikidata item for Darren. You can see this statement has the qualifier "coauthor" but the qualifier can also indicate the person is a work colleague, a co-collector, a correspondent and so much more.

And lastly you can see the employer statement from a Wikidata item. Again giving information on where a person is employed.

## Slide 25 - Collection items at

Much of the “collection items at” statements added to Wikidata for natural history folk is as a result of Bionomia. Bionomia is website and tool that takes observation data from GBIF, extracts the name strings attached to the observation data and serves this up for volunteers like me to use to disambiguate and to attribute specimens that person via either the person’s Wikidata item if they are deceased or their ORCID if they are alive.

Bionomia then produces a Bionomia profile for the person based on this attribution data. This Bionomia profile uses not just Bionomia & GBIF data but also data from Wikidata and links to both Wikidata and Wikipedia. After attributions have taken place, Bionomia will list all the institutions where those attributed specimens are held. You can see here that Hudson has specimens in multiple institutions including Auckland Museum, the Natural History Museum in London AND the New Zealand arthropod collection.

## Slide 26 - “Collection items at” data for institutions

Bionomia also provides a profile for each dataset giving specimen occurrence data in GBIF. Here is the profile for the NZAC symbiota dataset. You can see that this dataset has 73% of its specimens linked to its collector or determiner. This is quite high as datasets go.

You can see all the people linked to the collection as collectors and identifiers. And there is also a list of frictionless data packages that can provide information on such things as the attributions made as well as problem determiner and collector dates. These data packages can be downloaded and reused and can also be used to improve the collection metadata. And all this is built with the help of Wikidata.

## Slide 27 - Adding Bionomia sourced “collection items at”

I’ve produced documentation on how to use those Bionomia frictionless data packages to ensure that “collection items at” data is added to collector Wikidata items. Wikidata can then be queried to get a list of all collectors who have contributed to an institution.

By adding this data Wikidata ensures the collection is linked to its collectors, and increases the likelihood of interested third parties coming across the collection and empowering them to reuse the data either via Wikidata or contacting the collection itself. It improves collection findability.

## Slide 28 - Significant person

Where a specimen has been attributed to two or more of its collectors, Bionomia will also generate a list of co-collectors. This information can be added to Wikidata via the “significant person” property to build up a relationship network giving information on who is collecting with whom.

The use of this property can help provide data on the networks of the person. The image here is of Alice Eastwood’s network, she’s a famous American botanist. I went through her finding aid for her archives and added all the significant people who already had a Wikidata item to her item as a significant person. Hopefully you can see the potential for digital humanities and history of science research based on just this one type of statement.

But it also can be used to add context to specimen collecting events.

## Slide 29 - Employer

And lastly we have the employer statement. Linking employees and those affiliated to a particular institution empowers Wikidata tools to show the impact of the institution. So, using the example of Manaaki Whenua, you can see this Scholia tool has produced a profile which then lists the employees and affiliated people, as well as topics these people have published on, recent publications and recent citations.

Scholia is a tool that is made up of multiple queries of Wikidata. Scholia can be used to help indicate the impact of an institution by linking to and summarising the published works of employees and affiliates. And the impact of those published works.

Just as a warning, currently the Scholia tool is under redevelopment. However this tool is vital to much work in Wikidata so developers are committed to ensuring its continuation.

## Slide 30 - Identifiers for people

As I’ve explained previously - one of the most important elements of Wikidata is that it acts like an identifier hub. So Wikidata acts like a crosswalk or a bridge between one institution’s data and another institution’s data. The identifiers highlighted in this list are examples of institutions that not only have their identifiers linked to in Wikidata they also pull data from Wikidata.

You can see Bionomia, which we’ve already discussed, the Biodiversity Heritage library which both pushes data into Wikidata but also pulls data out of Wikidata for catalogue enrichment. There is also Yale’s digital platform LUX which again pushes data out to Wikidata and also pulls data from Wikidata into their digital platform. And finally there is Te Papa which again links to Wikidata in their collection management system.

## Slide 31- Te Papa CMS

I just wanted to discuss Te Papa a bit more in detail as this is the direction I'd love the collections from BSI to move towards. Te Papa has been very proactive in adding their collector data into Wikidata. But Te Papa also adds the Wikidata item link to its collection management service person records. And if appropriate the CMS will ensure that this link is served up onto its front-facing website, roundtripping this data and making both Wikidata and Te Papa data more visible to multiple audiences.

## Slide 32 - Taxa related data

Moving on to another part of the Biodiversity Knowledge Graph - taxa and taxonomic names. As I've explained, around 4% of Wikidata is made up of items for taxonomic names. Here's a wikidata query giving all the properties related to taxa, this can be used to help improve taxa items.

## Slide 33 - Taxon item properties

Here's a visualisation of just some of those properties. These can be used to add information such as taxon authors, the year of publication, synonyms, type locality, taxon range, hosts, habitat, invasive or endemic nature, trait information, conservation status and so much more.

## Slide 34 - Example taxon Wikidata item

Here are just some of the examples of statements on a Wikidata species taxon item. You can see in the example that the taxon name statement links to the Wikidata item for its author, the year in which it was described and whether or not the taxonomic name is a recombination. References supporting the taxonomic name statement may include links to the Wikidata item for the publication that did the naming or recombination, as well as to databases such as Biota of New Zealand. Other statements highlighted here relate to the endemic nature of the species, the hosts of the species, as well as the type locality.

## Slide 35 - Researchers using Wikidata for trait data

As an example of researchers using taxon Wikidata items, here is a research paper that documents the use of Wikidata by botanists to add statements about trait data for flowering plants with blue flowers and then using Wikidata to assist in analysing that data.

## Slide 36 - Taxa item external identifiers

I also want to emphasise that Wikidata taxa items also have an identifier section. And of course this is where it is really important for BSI database identifiers to be listed.

The reason I want to link both the Biota of New Zealand and New Zealand Organism Register identifiers to Wikidata taxa items is to ensure that these data are findable. When it's easier for people to find, the easier it is for people to reuse that data. And in turn the easier it will be to build upon that data.

Also by adding the NZOR identifiers to taxa items empowers that identifier to be displayed on the Wikipedia article taxonbar. Readers and reusers of that Wikipedia article can then just click on the identifier link to be taken directly to the NZOR databases to find out more on the species.

## Slide 37 - Biota NZ id

I regard this linking of identifiers as SO important that I've spent a lot of time during this WiR linking the BiotaNZ identifier to taxa Wikidata items. I have also produced documentation on how to use OpenRefine to undertake this work. Not only does this improve the reach of the Biota of NZ website this also brings to light data errors both in Wikidata and in the Biota of NZ database itself & allows for error correction.

And in case you are wondering what OpenRefine is, it is a magic data cleaning and curation tool that is my tool of choice to bulk edit Wikidata.

## Slide 38 - Taxon identifiers on Wikipedia page

So this is what taxon identifiers look like on a Wikipedia article. This screenshot is of the taxon bar on the Wikipedia article for a moth species endemic to New Zealand. This taxon bar does not display all the identifiers available on Wikidata so it is missing the Biota of NZ identifier. However you can see that the NZOR database has identifiers on the current name as well as the synonyms for this taxon.

I just want to take this opportunity to get on one of my soap boxes and explain how the work of integrating Biota of NZ identifiers into Wikidata has yet again highlighted to me omissions in GBIF and the Catalogue of Life. GBIF's taxonomic backbone is informed by the Catalogue of Life initiative. The catalogue of Life aims to compile a single, integrated list of all known species worldwide. It does this by aggregating multiple datasets added to Checklist bank.

## Slide 39 - GBIF

Currently neither the Biota of NZ nor the NZOR is in Checklist Bank. To give you an example of why this matters, this is the profile of a species of moth in GBIF. GBIF currently aggregates 67 occurrences of this species under the genus name **Declana**. Because GBIF is taxonomically out of date and iNaturalist is taxonomically up to date, all 382 occurrences in iNaturalist are not yet aggregated in GBIF.

I want the Biota of New Zealand dataset shared with the Catalogue of Life by adding it to Checklist Bank. Doing this would go some way to help rectify the situation shown in this slide. I want to emphasise that it is not unusual to come across this issue for New Zealand endemic species.

I am asking BSI to share the openly licensed Biota of NZ dataset with Checklist Bank. Doing so will help keep Wikipedia articles up to date taxonomically, it will ensure the citizen science contributions to be aggregated, it will help institutions both in New Zealand and world wide to keep taxonomically up to date and will aid them in aggregating their collections to GBIF which in turn helps inform biodiversity research into New Zealand species.

## Slide 40 - Wikidata items for type specimens

Stepping off my soap box, I just want to let you know of some of the other work I've been doing related to taxa data. I've been collaborating with Auckland Museum to create a data model for Wikidata type specimen items.

This is needed as multiple institutions have been creating Wikidata items for type specimens but there was no agreed model everyone was following. The holotype specimens held by the NZAC now have Wikidata items created for them. I'm assisting Leanne with uploading the type specimen images into Wikimedia commons which can then be linked to those items.

## Slide 41 - Example of a type specimen item & image

Here is an example of a type specimen held at the NZAC as well as its associated image. Again I want to remind you of the reason for doing this work. Not only is this data used in Wikidata and Wikipedia, internet search engines like Google reuse this data to help inform search results. AI reuses this data. Creating these items shows who holds the type specimen and who to contact if researchers need more information or access to that specimen either digitally or physically.

Also while we are here - see the “collection date” statement on this type specimen image. You can see one date has been deprecated. This is because there was an error in the NZAC system as a result of the incorrect transcription of the handwritten date. While checking the data of this specimen I looked at both the digitised label and the original description of the species and realized there was an error in the date transcription. This has been fixed in the NZAC database but I wanted to show you that it is possible to deprecate a statement in Wikidata.

## Slide 42 - WiR documentation for type specimen data

This collaboration produced not just the data model but also crosswalk documentation from Wikidata to Darwin Core and an example upload. All the documentation has been placed into the repository Zenodo, so it can be reused by anyone but we are particularly hopeful other natural history institutions will use it to help guide their creation of type specimen Wikidata items. Thus making types easier to locate and also to link to the original descriptions of the taxonomic names the specimens relate to.

## Slide 43 - WiR documentation for specimen images

I've also produced documentation for the upload of specimen images into Wikimedia Commons. As a reminder Wikimedia Commons is the image repository for the Wikiverse. These documents and workflows include guidance for employees of institutions that wish to upload images into Wikimedia Commons, documentation on how to bulk upload specimen images into Wikimedia Commons via OpenRefine and a data model for the structured data that may be attached to specimen image files in Wikimedia Commons.

## Slide 44 - Publication related data

Again looking at another part of the biodiversity knowledge graph, I want to move on to publication data. Wikidata has a LOT of data on books, scientific articles, grey literature etc. A large proportion of Wikidata is scientific article metadata.

## Slide 45 - Scientific article example

Take for example this scientific article by Darren. If you click on the link in the slide it will take you to the full item. Not only does the item contain information that can be used to generate a citation, it also links to the URL for the full work, the Wikidata items for the authors, who funded the research, the collections it used, the publications it cites, the people acknowledged. The item also links to the multitudes of identifiers that relate to the article such as its DOI, ISBN number, zoobank publication id and so much more.

## Slide 46 - Findable data

I used this publication as an example in documentation showing how to go from unstructured data in a scientific document to findable, reusable and linked data by using Wikidata.

## Slide 47 - “Wikifying” an article

I wrote documentation on how to “Wikify” an article. Not only adding the publication citation data to Wikidata but also how to reuse appropriately licensed data, images and other information contained in the publication in Wikidata, Wikipedia and Wikimedia Commons.

## Slide 48 - Linking publications to collections via Wikidata

I have also created documentation on how to link publications to the natural history collections that provide data and specimens used in the research. This includes adding “cites work” statements which link the publication item to items for the natural history collection datasets and also publications cited by that article. “Uses” statements can be added where the publication says it used natural history collection material. And “acknowledged” statements can be added to the item where the publication thanks people and institutions.

## Slide 49 - Example scientific article linked to collections

So here for example you can see screenshots of an item for a publication that specifically states the collections it has used. The publication also points out it used data sourced from datasets published by those collections in GBIF. Once items are created for those GBIF datasets, the publication item can have statements added saying it cites those datasets.

By wikifying the article and linking that article item to the collections it uses or cites, this raises awareness of and the possibility for reuse of the publication and its data. It also helps inform others about the collections used and can assist with making visible and quantifying the impact of natural history collections.

## Slide 50 - Query Wikidata: Impact

Of course once the data is in Wikidata it can be queried. There are currently 31 scientific articles in Wikidata that cite the NZAC symbiota database.

## Slide 51 - Scholia profile for article

The scholia tool also works on articles and can help show the impact of an article via visualising various queries of Wikidata.

## Slide 52 - Scholia profile for author

And of course Scholia can visualise the impact of an author. Again this tool is based on the data in Wikidata. So if an author's articles are not in Wikidata or they are listed as only a name string and their Wikidata item is not linked to, this visualisation of impact won't be optimised.

## Slide 53 - Scholia co-author graph

But to give you an example of the types of visualisations in a Scholia profile for an author - here's Darrens co author graph. You can see the groups of people he likes to collaborate with to produce scholarly articles.

## Slide 54 - Other biodiversity data

So I've gone into quite a bit of detail about people related data, taxa related data, and publication related data in Wikidata. Of course this doesn't cover everything. There's also location data, institution data, conservation status data, research expedition data - all interlinking to other data in Wikidata.

## Slide 55 - Wikidata WikiProject Research Expeditions

I particularly want to mention research expedition data as this can also be used by BSI collections to give more context to the collecting events for their collections. It's also a great example of how a collaboration on Wikidata evolves into impact on Biodiversity data standards and inspires the creation of tools using correctly modelled data.

This slide shows our Wikidata WikiProject. A WikiProject is where editors come together to collaborate on improving knowledge about a particular subject - in this case research expeditions.

## Slide 56 - International collaboration

This group of Wikidata editors and museum professionals working in natural history institutions recognised the importance of research expeditions for collections and the wider natural history community. So we formed a WikiProject to help coordinate efforts to improve coverage of research expeditions in Wikidata and to create Wikidata Q identifiers for research expeditions. We quickly realised that a data model was needed to ensure data was added consistently.

As a result, a working group was proposed that was later approved as an official TDWG Biodiversity Data Standards **Task Group** extending the work we were doing in Wikidata and working towards an internationally agreed terminology and data model for research expeditions. This effort is still ongoing.

## Slide 57 - Reuse of Expedition Wikidata item QID

Once a Wikidata item exists, the data in that item and the item identifier or QID are then able to be reused by third parties. Examples of this reuse include by the Biodiversity Heritage Library, who ingests such Wikidata QIDs into its catalogue. Another example are Natural History Institutions who can use Wikidata QIDs in their Collection Management Systems. The Wikidata item QID for research expeditions can also be used as the Darwin Core “parent event identifier” and can be propagated by natural history institutions into the Global Biodiversity Information Facility (GBIF).

## Slide 58 - Expeditia

And of course once the data is in Wikidata a third party tool was created that reuses that data - this is Expeditia. This tool uses the data and images placed in Wikidata about expeditions to visualise that expedition on a map. Where possible it also pulls data from GBIF about specimens obtained during that expedition and links them to the expedition profile.

## Slide 59 - Reuses Wikidata, Bionomia and GBIF data

Expeditia can also show specimens collected during the research expedition, and place those specimens on the expedition journey map. The tool does this by taking collector and specimen information from GBIF. Much of this data has been linked to GBIF through the work of Bionomia volunteers linking the collector, as represented by their Wikidata identifier, to the specimens.

By combining both the structured data in Wikidata and information from GBIF, Expeditia can provide more context and information about research expeditions and their collections. It is a great example of reuse of data in Wikidata.

## Slide 60 - Wikidata is queriable

I now want to go into a bit more detail about Wikidata being queryable. If you know the SPARQL query language you can use the Wikidata Query service to query Wikidata. There is also the Wikidata query builder which allows you to visualise the data, for example here you can see a map of all the herbaria in New Zealand.

## Slide 61 - Querying Wikidata

As well as the Query builder and the query service you can also query Wikidata via Wikidata APIs. It's outside the scope of this workshop but I've put links to each of these pathways in the slides.

I've also put a link to a basic Wikidata query. You can see there is an examples tab where you can find other examples of Wikidata queries that can assist you in building your own.

In this query, I'm asking Wikidata to show me images of all the taxa listed in Wikidata with the type locality of the Wellington Botanic Garden. And I've asked Wikidata to show me the pictures of those taxa which have been placed on those items. The reason there are only moths in this result is because I've been working on New Zealand moths and I've added this data.

## Slide 62 - Taxa with type locality Wellington Botanic Garden

So that query returns these images of all the taxa listed in Wikidata with the type locality of the Wellington Botanic Garden.

## Slide 63 - Taxa hosted by plant species

Here is another simple query made using the Query builder asking for all the taxa hosted by a particular plant species. Again, in case you are wondering why only moths turn up, that's because I've been adding this data into Wikidata when creating or enriching Wikipedia articles and Wikidata items for New Zealand moths.

But if you have a data source this type data is SO easy to add to Wikidata via OpenRefine. And of course when data and images are added to Wikidata, folk start querying for it and reusing it.

## Slide 64 - Moth GIF

Over the years I've also added openly licensed images sourced from the NZAC to Wikimedia Commons and then reused these images in Wikipedia and Wikidata. And once again when you share, you see reuse. Here's a Gif of New Zealand moth images and as you can see most are sourced from the NZAC.

## Slide 65 - Reuse of Wikidata

In the final part of this presentation I want to highlight some examples of reuse of Wikidata.

## Slide 66 - Te Papa Collection Management System

I've already discussed how Te Papa reuses Wikidata QIDs for collectors and research expeditions in their collection management system which then gets populated into their collections online website.

## Slide 67 - iNaturalist

iNaturalist reuses the Wikipedia article, the Wikicommons images in that article and also the Wikidata identifiers displayed in the taxonbar. You can see here how the NZOR link is ingested.

## Slide 68 - Atlas of Living Australia

There is also the Atlas of Living Australian. Similar to iNaturalist the Atlas of Living Australia reuses the Wikipedia article, the Wikicommons images in that article and also the Wikidata identifiers displayed in the taxonbar. You can see here how the NZOR link is ingested.

## Slide 69 - Biodiversity Heritage Library catalogue

And then there is the Biodiversity heritage library. BHL not only ingests and displays the Wikidata QID for people and publications in its catalogue, it also ingests other identifiers from that Wikidata item such as the library identifiers from VIAF or the Library of Congress.

## Slide 70 - Other reuse examples

And of course there are multiple other reuses of Wikidata. Websites like the previously mentioned Bionomia. There is also JACQ, a pilot by the Consortium of European Taxonomic facilities.

## Slide 71 - Science stories

There is also the website Science Stories. This website uses data added to Wikidata and images added to Wikimedia Commons to create visualisations and help tell the stories of women scientists. There are hundreds of tools and websites that link to or reuse Wikidata items. So the impact of adding data to this platform is SO much larger than you might think.

## Slide 72 - Helpful general resources

Here's a list of general resources you can explore to learn more about editing Wikidata.

## Slide 73 - Other helpful links

And of course Youtube is your friend here so I've included some links to videos and useful "how to's".

I'd highly recommend you start by manual editing to get your head around the knowledge graph aspect of Wikidata. For those used to editing databases it takes a while to get used to the fact that it is the linking that is the priority and that not all data needs to be on a single item.

BUT once you're familiar with how Wikidata works and the schema used for different types of items I encourage you to start exploring bulk editing. As someone who once spent two weeks sorting out a large batch upload, I highly recommend starting with small batch uploads to test this out prior to undertaking larger uploads!

## Slide 74 - Wikidata in Brief

I've added this "Wikidata in Brief" handout so you can grab it from the slides - it provides context about Wikidata editing and provides links to helpful resources.

## Slide 75 - WiR documentation

And I've spent a lot of time creating multiple how to's and documentation. So I feel obligated to provide a link!

## Slide 76 - OpenRefine tutorial on Youtube

If you are familiar with OpenRefine you might be interested in these two OpenRefine tutorials based on collection data.

## Slide 77 - Questions

I recognise I've covered a LOT very quickly. Does anyone have any questions?

## Slide 78 - Break

5 min to stand up and stretch! It's been A LOT of information in a very short amount of time. After the stretch we'll move on to conflict of interest statements, adding gadgets and user scripts to your account to prepare you to more easily edit Wikidata.

# Preparing for Editing

## Slide 79 - Conflict of Interest

One subject that may very well apply to editors in this room is Conflict of Interest. Conflict of interest (COI) editing involves contributing to Wiki about yourself, family, friends, clients, employers, or your financial and other relationships.

However, it does get complicated with some editors and subjects, like for example Heidi Meudt.

If she's a botanist, and researches and publishes on forget-me-nots, can she write about forget-me-nots and cite herself? Can she edit items about species she's described or written about?

The answer is Yes she can, but on her userpage she's transparent about this conflict of interest, and she makes sure she takes a neutral point of view in articles and items, and cites ALL articles about the subject not just the ones she's written.

## Slide 80 - Do you have an account?

So now we want to get ready to do some editing. First off, hopefully everyone has a Wiki account. Please log into your account.

## Slide 81 - User page

Once logged in, the top of your Wikidata page should look something like this. You can see my Username is coloured blue and if I click on it, it will take me to my user page where I have introduced myself and added more information about what I am contributing to Wiki including conflict of interest statements. You can see I've actually given my real name. You DO NOT have to do this!

What I suggest you do now is click on your user name - especially if this link is currently red - which will take you to your user page. Then add a sentence about yourself introducing yourself to the Wiki community. You can be anonymous if you wish.

I suggest writing something like "Hi, I'm a beginner editor aiming to improve Wikidata" or something similar. Once you've finished that don't forget to press the blue "publish changes" at the bottom left of the page.

## Slide 82 - Gadgets

Once you are logged in, please go to your account dropdown at the top right and select Preferences and then select the Gadgets tab. Make sure the following gadgets are ticked. These will definitely help with your edits. You can add additional Gadgets and change other Preferences later as you continue your editing journey.

## Slide 83 - User Scripts

Prior to actually starting editing I also want to point out that User scripts exist. These are programmes produced by the community that can make your editing life useful. They are installed on your [common.js](#) page. You have to create that page yourself. I've put a link in the slides that will send you to your page. There is also a link to a list of user scripts that enhance the Wikidata user interface. You won't need or want to install all of them but the one I'd highly recommend is one called "script installer". This makes it easy to install any of these scripts with just a click of a button. So I'm going to pause now while you click on the link in the slides and go to your [common.js](#) page. Copy and paste the script installer and then press save. This will set you up to easily install any scripts you wish at a later date.

## Slide 84 - Other information

I would highly recommend you create and add content to your user page. This introduces you to the Wikidata editing community and will help ensure your edits are viewed as more trustworthy. You'll also notice it is good practice and highly recommended to have a conflict of interest statement on your userpage since you all work for the Bioeconomy Science Institute. Again this will ensure a higher level of trust and also will give other editors a contact point should they believe your edits infringe any of the Wikidata editing community norms.

Also be warned if you have only recently signed up for an account you may be restricted in tools you can use to edit. You may have to buddy up for some of the tasks.

## Slide 85 - Creating publication Wikidata items

The first thing we will be doing is creating a new Wikidata item for a publication. The links on this slide provide an example item to base your new item on, a link to the google sheet with the publications that need an item created for them and a link to two tools that can help generate new Wikidata items for articles via the article DOI.

If we could make sure we create the highlighted publication items in the list first that would be fabulous as we'll need to be able to link these later to new items we will create for taxa.

Although we are going to be manually editing Wikidata and creating new items I've included links to two tools as they are very useful for quick creation of scholarly article Wikidata items. The first is the BHL2Wiki tool. Although this is aimed at BHL generated DOIs this tool can be used for any article that has metadata in CrossRef. I prefer to use this tool as it has the advantage of linking to other articles the original article cites as well as authors based on their ORCID. I'll be demo-ing this tool and also giving you the chance to use it after you have got familiar with adding scholarly articles manually.

The old SourceMD tool can also be used for creating scholarly articles. For both these tools you may be asked to sign in to allow the tool to edit on your behalf. This is completely normal and the tool will only edit what you instruct it to. Also, as I've explained, you may not be able to use these tools if your account is under 4 days old and you have not yet reached the required minimum of 50 edits. But you can buddy up and see how it's done.

But first lets go to the spreadsheet and start manually creating a wikidata item for a scholarly article.

## Slide 86 - Live Demo

I'm going to do a live demonstration to show you how you add an article manually to Wikidata. This is mainly aimed at folk who have not yet edited Wikidata. For those of you who are experienced please go ahead and start creating a scholarly item from the list.

## Slide 87 - Your turn

## Slide 88 - BHL2Wiki tool

Now for those of you whose account is over 4 days old and you've already edited Wikidata I'd like you to pair up with those in the room who are not yet auto confirmed. We're now going to use the BHL2Wiki tool to create a scholarly article item. First put your name down beside the DOI you want to create a Wikidata item for in the spreadsheet.

## Slide 89 - BHL2Wiki markup for Quickstatements

The BHL2Wiki tool extracts data from CrossRef and creates markup that can be added to the Wikidata Quickstatements tool to automatically create a new item to Wikidata. Quickstatements is a powerful tool that can help you batch edit Wikidata.

## Slide 90 - Quickstatements Wikidata tool

However - the quickstatements tool can only be used if your account has autoconfirmed status. Your account needs to be over 4 days old and you must have done over 50 edits. If you are new to editing Wikidata you will be well on your way to the 50 edit mark by the end of this workshop so I'm keen to ensure you still see how this tool works. If you have a newish account and can't use Quickstatements please pair up with someone who can.

## Slide 91 - Log into Quickstatements

Once you've logged into Quickstatements and you have asked the BHL2Wiki tool to add the scholarly article data to the Quickstatement tool, it should look something like this.

## Slide 92 - Run Quickstatements

Press RUN at the bottom left hand side of the page. If you have a large batch you can press "run in background" BUT you are only creating one item so just let it run where you can see it.

## Slide 93 - Success!

Once the Quickstatements tool has created the item it should look something like this!

## Slide 94 - Possible Demo?

I could demo this if folk like? The link to the bhl2wiki tool is in the spreadsheet and if you are not logged into Quickstatements it will ask you to do this when you try and press the "open to quickstatements" button.

## Slide 95 - Your turn!

## Slide 96 - Creating a taxon Wikidata item

Here we will be creating new taxon items. I've included a link to an example item, the google sheet with the list of taxa that need to be created, and a link to documentation on Wikimedia Commons on how to manually create a new taxon item.

## Slide 97 - Live Demo

So now I'm going to do a live demonstration on how you add a taxon to Wikidata. Then it will be your turn!

## Slide 98 - Your turn!

## Slide 99 - Adding original description reference

Now I'd like you to go to an existing item as set out in the spreadsheet and add an original description reference to that taxon name statement on that item. I also want you to add qualifiers to the taxon name statement including the authors of that taxon name and the date of original description. Then referencing that statement with the paper that gives that original description. The items that need this have been listed in the spreadsheet.

On this slide I've also included a Wikidata query link that gives all those endemic to New Zealand taxa items that need the author qualifier adding to the taxon name statement.

## Slide 100 - Live Demo

## Slide 101 - Create people items

Finally we are now going to create an item for an entomologist or researcher. Again the list of people that need Wikidata items created for them is in the google sheets under the tab title people.

## Slide 102 - Live Demo

## Slide 103 - Your turn