

Research Report

# Wikidata Use in Cultural Institutions

Wikimedia Deutschland e.V.

# Wikidata Use in Cultural Institutions – Intro

## Executive Summary

We talked to 16 users who worked at different cultural (“GLAM”<sup>1</sup>) institutions to find out about “How and why do people in cultural institutions use Wikidata?” and thus learn more about participants’ motivations, activities and problems. We did the research from June 2019 - September 2019.

## Insights

Our main insights about the motivations, activities and problems of our participants were:

- Participants want to share data since they perceive their institutions and their context using more and more digital systems. They also see direct advantages of sharing data, like greater use of the institution’s data.
- Participants are motivated by the plan to re-import improved data from Wikidata back into local databases. Participants called this a “roundtrip” (→ [People would like to “roundtrip” data](#)).
- It is hard for participants to come up with data structures that
  - represent the collections and
  - match the requirements of Wikidata(→ [Getting data to Wikidata the “right way”](#))
- What happens with imported data is hard to monitor. This makes it difficult for participants to spot new problems or improvements and to participate in the community (→ [Improvements or degradation of data](#)).
- Currently, there are few ways to metricize engagement on Wikidata.
- Many open data initiatives start with image imports to Wikimedia Commons and then transition (partly) to Wikidata (→ [Intertwinement with Commons](#)).

---

<sup>1</sup> The acronym GLAM stands for “Galleries, Libraries, Archives, Museums”

## What supports Wikidata Use?

- Data is often already available in GLAM institutions
- Participants are enthusiastic about using Wikidata
- Umbrella organizations and big institutions pushing for open data (→ [Kinds of institutions](#))
- Hopes for quality improvement via future roundtrip(→ [People would like to “roundtrip” data](#)).
- Support in Wikidata-appropriate modeling, reconciliation and import by service providers, Wikimedia Chapters or community members (→ [Kinds of institutions](#)).
- Participants are often “onboarded” to Wikidata via Commons (→ [Intertwinement with Commons](#)).
- In some areas, Wikidata has better usability or feature set than participant’s current software (→[Institution’s data](#)).

## What hinders Wikidata Use?

- Data modeling is hard and needs both GLAM-skills and Wikidata skills (→ [Practices of Modeling](#)).
- Difficulties in understanding the culture and practices on Wikidata
- Tech-resources (→ [Data Imports](#)) and consulting needed (→ [Data Imports](#), [Kinds of institutions](#))
- Impact of Wikidata use hard to measure, but metrics are often needed in organization (→ [Provide Metrics](#))
- Often no immediate and easy-to-demonstrate benefit by using Wikidata
- Problems of continuous engagement due to difficulties in monitoring data of interest. (→ [Monitoring Data](#))

# Content

## [Wikidata Use in Cultural Institutions – Intro](#)

### [Executive Summary](#)

#### [Insights](#)

#### [What supports Wikidata Use?](#)

#### [What hinders Wikidata Use?](#)

### [Content](#)

### [Research Methods](#)

### [Participants](#)

#### [How we contacted the people who participated in the research](#)

### [Research Interests](#)

#### [Data Quality](#)

#### [Motivations and activities in Wikidata work](#)

## [Findings](#)

### [Motivation to use Wikidata](#)

### [Participants' tasks in institutions](#)

#### [Position in the institution](#)

#### [Data imports](#)

#### [Workflows](#)

### [It is not obvious how to get your data into Wikidata “the right way”](#)

#### [Practices of Modeling](#)

#### [Open questions of modeling](#)

### [People would like to “roundtrip” data](#)

#### [Current ways to roundtrip data](#)

### [Improvements or degradation of data after import](#)

#### [Fears to make changes harder](#)

#### [Monitoring data](#)

##### [Detecting merges](#)

##### [Finding divergence between local data and data on Wikidata](#)

#### [Provide Metrics](#)

### [Intertwinement with Commons](#)

## [Summary](#)

## [Appendix A: Background information](#)

### [Types of data institutions use](#)

### [Modeling and reconciliation: Processes, terms and tools](#)

#### [Modeling](#)

#### [Reconciliation](#)

##### [Example: Reconciliation with Open Refine](#)

### [Identifiers](#)

[Kinds of Institutions](#)

[Appendix B: Research Methods](#)

[Research Questions](#)

[Iteration 1: Data quality and image of the institution](#)

[Iteration 2: Workflows and Goals](#)

[Authors and License](#)

## Research Methods

Our research was explorative and aimed to be able to describe and explain motivations, activities and problems of our participants when using Wikidata for their work. We started our exploration with a focus on data quality management practices and then broadened the scope.

Our data was collected in conversations with our participants. Since all participants were remote, we used tools like google meet or skype for your conversations.

The interviews were semi-structured. We used an interview guide which listed the topics we wanted to explore. The interviews were 30min to 1h:30min long. The length of the interview depended on the mutually available time frame and the questions explored. In almost all interviews, two researchers were present: one person leading the interview, the other person take taking notes. The notes were [shared with the participant<sup>2</sup>](#) after the interview via the collaborative text editing tool *google docs*. They could correct, comment and supplement them.

After the interview, the data was pseudonymized as an additional safeguard (The notes were already taken pseudonymously, but given that they were written live, it can happen that e.g. names are written down instead of placeholders ). The text was analyzed by iterative thematic coding in the CAQDAS application [Quirkos<sup>3</sup>](#): text segments were given one or multiple codes according to their content so that repeating topics could be identified across interviews. The depth and specificity of codes was iteratively improved.

Based on the themes constructed in the analysis, we created a slidedeck which we sent to the participants for a [member check<sup>4</sup>](#). Four participants responded and further improved our interpretations.

## Participants

How we contacted the people who participated in the research

Initially, participants were people we already knew and who were also known to contribute to cultural data or work with cultural institutions. These participants then referred us to people they worked with if they knew a person they considered particularly knowledgeable in the topic we asked about. In this case, they often offered to initiate the contact. One

---

<sup>2</sup> <https://fordes.de/posts/OpenFieldnotesInRemoteInterviews.html>

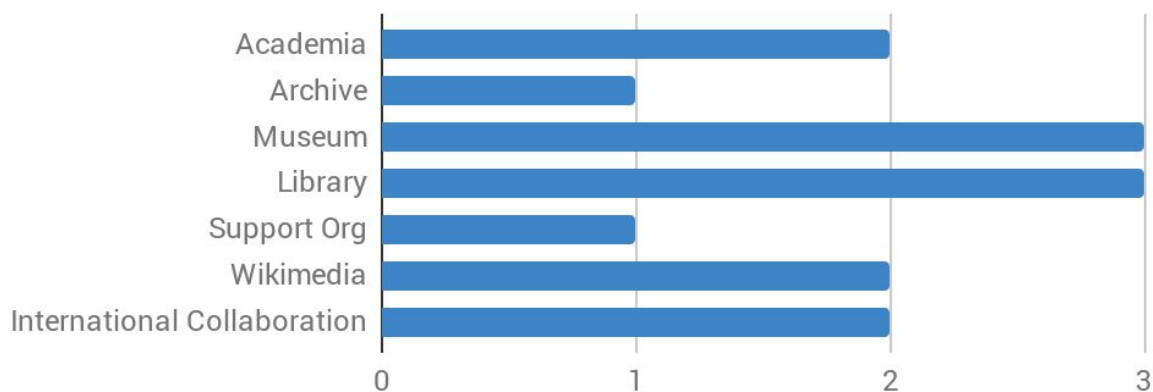
<sup>3</sup> <https://www.quirkos.com/learn-qualitative/features.html>

<sup>4</sup> [https://en.wikipedia.org/wiki/Member\\_check](https://en.wikipedia.org/wiki/Member_check)

participant got in the research via a post on Facebook to a Wikidata/GLAM group, all others were known to us or referrals.

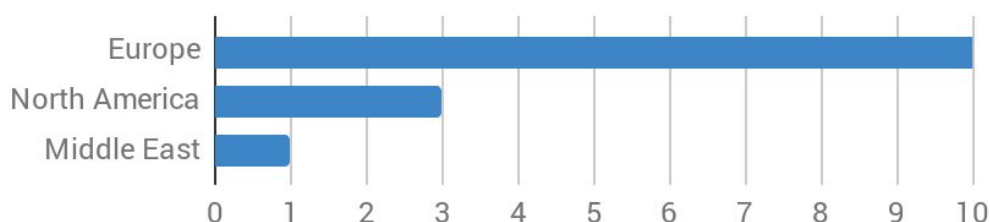
The requirements for participation was work in or with a cultural institution and use of Wikidata. Participants from Wikimedia organizations and one organization which supports GLAM institutions with digitalization were included to get an overview of practices and problems, since they work with several different institutions.

A large share of our participants came from museums, and libraries.



Participants often had formal roles at the intersection of data(management) and technology (like “database manager”). Even in cases they had no technology-focussed job title, most participants were interested in digital data and digitalization. Some of the participants were [“Wikimedians in residence”](#).

Due to the language-independence of Wikidata, participants came from different countries, most, however, came from Europe:



We talked to 17 people. Data from 16 of them was used in the current analysis, data from one participant was not used since they did not use Wikidata in their work. Two of the interviews were done with two participants together.

## Research Interests

In the research we focussed on two larger themes:

1. Participants perceptions of threats to data quality
2. general motivations, activities and problems when using Wikidata as part of their work.

The research focus shifted over time: We initially wanted to understand whether organizations had concerns regarding data quality and reputation when donating and using Wikidata's data (See: [Research questions iteration 1](#)). It turned out that participants had rather other concerns connected to their workflows and organizational needs and we wanted to learn more about these (See: [Research questions iteration 2](#)).

## Data Quality

We started the research based on a feature idea for “signed statements”. The idea of signed statements is that people in a (cultural) institution could cryptographically sign edits with their institution's cryptographic key. As long as the value which was signed stays the same, the cryptographic signature is “intact”, if the value has been edited by someone else. Thus, our way to describe our research interested to our participants and the first question we asked was this:

*“I heard that people working with/in institutions import data to Wikidata. They are sometimes concerned with other editors changing this data (possibly wrongly). I'd be interested if this impression is correct and what needs, activities and problems are connected to this.”*

## Motivations and activities in Wikidata work

We started to look into the related issues that participants brought up when they talked about data quality. These were concerning issues of

- workflows
- collaboration within the institution or with other institutions
- current or future advantages of working with and contributing to linked open data
- problems

Thus, our interest shifted to

*“How and why do people in cultural institutions use Wikidata?”*



# Findings

## Motivation to use Wikidata

When discussing reasons for use, a large part<sup>5</sup> of the participants mentioned that they want to provide data they have to more people. Participants also talked about telling others about the work they do e.g. via blog posts and sharing visualizations to increase visibility of their institution. Open data strategies of overarching organizations like library consortia, local, national or international laws and policies are also a positive influence on Wikidata use.

## Participants' tasks in institutions

### Position in the institution

Participants were often enthusiastic about free knowledge, Wikimedia projects or digitalization. This is not surprising since participants volunteered to participate and were recruited based on use of Wikidata.

Particularly in later conversations, we tried to find out about collaborations across institutions and work with peers within their institution. Participants often described peers as supportive. There were also mentions of peers being indifferent to open data or being critical of the participant's work and digitalization in general.

P4 about people who see no sense in importing to Wikidata:

[they would say] *“why would I do this when I already have it in my catalog?!”*

### Data imports

A major motivation for participants was the use of Wikidata for providing and sharing data on cultural artifacts<sup>6</sup>. Some also wanted to collaborate with the Wikidata community and hoped for peer-production-based improvement and [enrichment of their data](#)<sup>7</sup>.

Importing data to Wikidata needs resources for data cleaning, modeling and work with technology. So not all datasets that the institutions would like to donate are imported<sup>8</sup>.

---

<sup>5</sup> 9 Participants

<sup>6</sup> 9 participants mentioned sharing and providing data as motivation

<sup>7</sup> Also called “[Roundtrip](#)”, which was mentioned by 6 people

<sup>8</sup> One person explicitly talked about the lack of resources to import existing data, while another talked about the skills and time needed to successfully import data. This plausibly matches the described need for domain and technology expertise in the import workflow.

## Workflows

The workflow for importing existing collection data from the institutions internal collection management system into Wikidata usually followed these steps:

### 1. **Export**

From Collection Management System to XML or CSV

### 2. **Modeling**

GLAM expertise needed to answer: What data of the institution can and should be public? What makes sense, from a professional point of view, to be seen by others and be useful for them? What would be a good way to structure it in the professional domain?

Wikidata Expertise needed to answer: Are the possible imports relevant enough for Wikidata? Are there already similar things on Wikidata which can serve as template? What is the correct Property? Do I need new Properties?

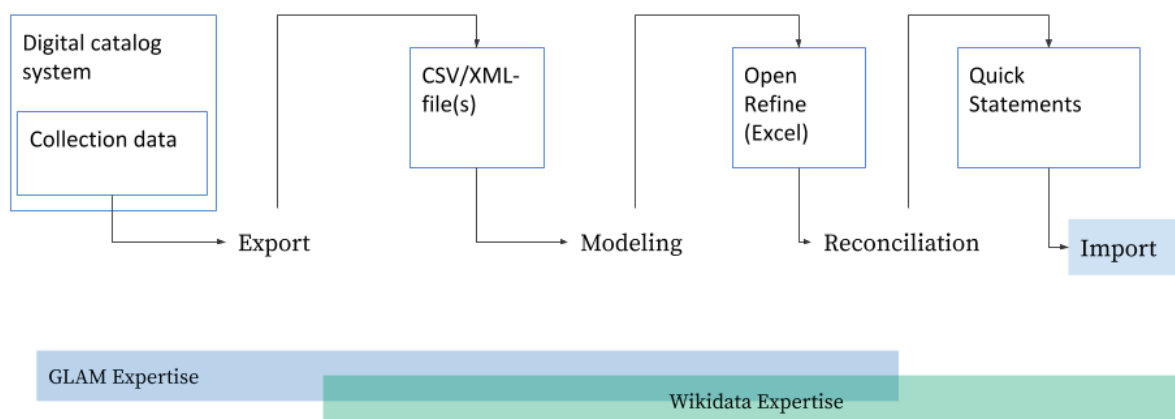
### 3. **Reconciliation**

Matching the exported data with Items, Properties and values on Wikidata, e.g. using Open Refine or, more rarely manually in Excel.

Wikidata and GLAM expertise is needed to ensure that data is matched to the concepts which are described in the original data.

### 4. **Import**

Upload to Wikidata e.g. using Quick Statements or Open Refine



The workflow often involves different roles: The GLAM person exports the data and a chapter, service provider, or community member does the reconciliation and import and sometimes also supports with modeling.

## It is not obvious how to get your data into Wikidata “the right way”

For getting one’s data into Wikidata, one needs to “model” and “reconcile” it.

[Modeling data](#) was frequently<sup>9</sup> described as difficult. This is plausible, given that Wikidata itself has few mechanisms to suggest which data to input. People need to find practices they can rely on to ensure a modelling which is both correct according to their professional discipline as well as implicit and explicit Wikidata usage conventions.

### Practices of Modeling

Here are some ways that people use to find out which Properties and values they should use to model a new Item on Wikidata.

- Use Examples: Look at existing Items, similar to the ones you want to create (e.g. P14)
- Use Wikiprojects<sup>10</sup> and consider their modeling suggestions (e.g. P8, P14)
- Have discussions with other community members about new Properties or if an existing Property fits the intended use (overloading, new creation, within the intended use?) (e.g. P06)
- Iterating: model a type of Item, make a small import of Items of that type and get feedback from other editors. If feedback is positive, import more Items using the same modelling. If the feedback is negative, consider the feedback, iterate the modeling and repeat the cycle. (e.g. P05)
- Utilize existing experiences: People who have experiences in imports and Wikidata can just do what they did the last time, if that worked well. (e.g. P13, see also [Import Workflow](#))
- Share tasks: Get help from chapters, volunteers or service providers and share tasks. The export is done by the GLAM, the import by the provider, modeling and reconciliation based on mutual feedback. (e.g. P8 gets support, P06 provides such support, see also [Import Workflow](#)))

---

<sup>9</sup> 8 people described modelling challenges, 3 of them described modelling as a crucial problem.

<sup>10</sup> Wikiprojects are informal associations that focus on a specific topic, e.g. the [Art-focussed “Sum of all Paintings”](#). They are represented by a page on Wikidata where project members keep relevant information about the project and on how to contribute to the project’s area of interest. Often, suggestions for modeling typical items in the project domain are given.

And to make it a good start painting item add the following statements:

Property	Data type	description	required
creator (P170)	Item	maker of this creative work or other object (where no more specific property exists)	yes
inception (P571)	Point in time	date or point in time when the subject came into existence as defined	yes
inventory number	String	identifier for a physical object or a	yes

(Cropped screenshot from the Wikiproject “Sum of All Paintings”-page on 2019-09-02, content under [CC-BY-SA 3.0](#))

There seems to be no standard process, participants described many different approaches.

Open questions of modeling

Participants expressed several concerns about modeling. These are some of the questions that were raised during the interviews:

- Can participants assume modeling of a type of Items to be like the models they knows from their profession or is there a “Wikidata-way” which is different?
- How can they find the Property for their modeling needs (if it exists)? (which concerns [reconciliation](#))
- How can they find other Items like the ones they want to model to find out how such Items are modelled on Wikidata?
- If they found an Item similar to the ones they want to import – is the existing modeling the right way to do it, or are there other (better) ones?
- Where can they ask people who have more Wikidata expertise?  
[A Property’s talkpage](#)<sup>11</sup>? [The Project Chat](#)? [A Wiki-Project](#)<sup>12</sup>? Or rather a facebook or Telegram group? Asking assumes knowledge that these institutions exists as well as the skills to communicate “appropriately” according to existing users and to interpret the content in a way considered “right”.

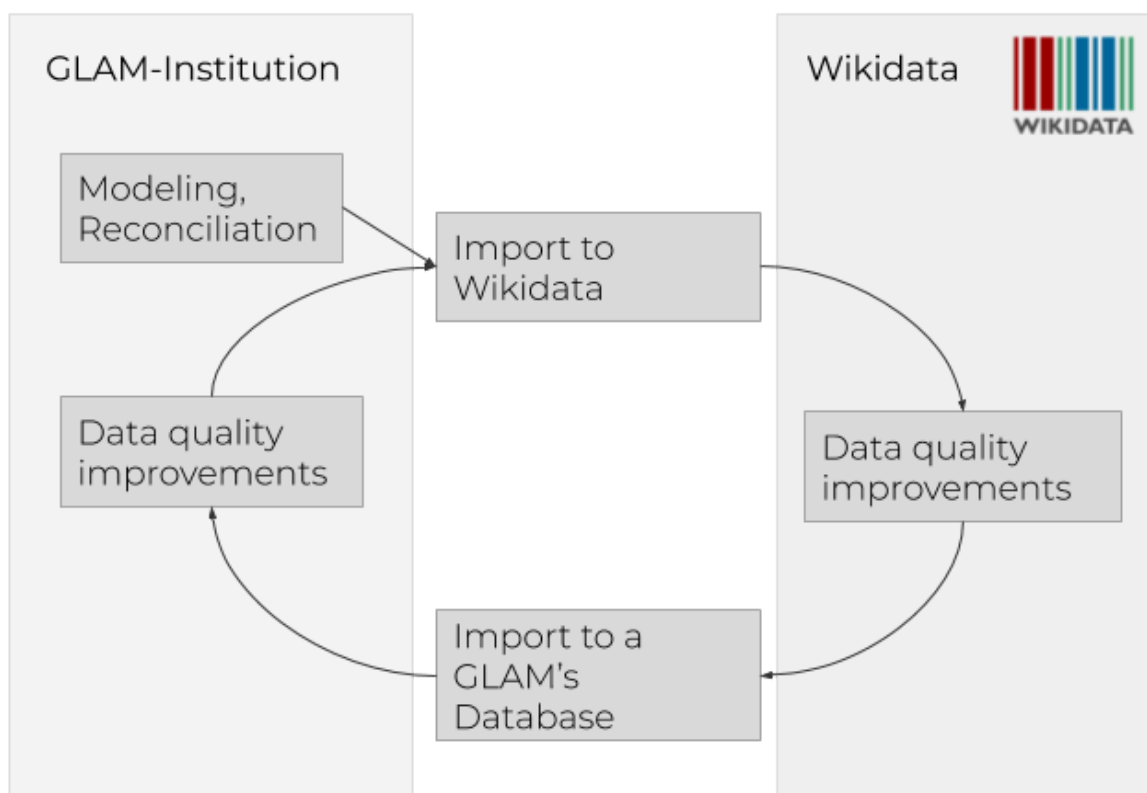
One participant remarked that “*You have to be psychic*” to know what to do [P01]

<sup>11</sup> e.g. [https://www.wikidata.org/wiki/Property\\_talk:P610](https://www.wikidata.org/wiki/Property_talk:P610)

<sup>12</sup> e.g. [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Books](https://www.wikidata.org/wiki/Wikidata:WikiProject_Books)

## People would like to “roundtrip” data

A roundtrip, ideally, improves data by exchanging updated and extended data between the GLAM’s local collection database and Wikidata<sup>13</sup>. A significant proportion of participants mentioned roundtrips as a goal<sup>14</sup>.



Participants hope that roundtrips lead to more and richer data in their local databases.

*If roundtrips work, that will sell everyone on it – no questions asked!” [P08]*

Some<sup>15</sup> participants explicitly pointed out that they considered some of their institution’s data not to be of great quality and that the data on Wikidata might be much better. Mentioned reasons for this were that the people who create the data in the institutions are necessarily generalists, not subject matter experts. A person also pointed out that authority data already can contain information taken from Wikipedia.

While the idea of roundtrips was seen positive in general, one participant described that

<sup>13</sup> In member-checking we learned that the term round trip is sometimes also used for an automatic import of new collection data into Wikidata, even if there is no re-import.

<sup>14</sup> 7 of 17 participants mentioned round trips as something they would like to be able to do.

<sup>15</sup> 3 people mentioned that their data might have quality problems directly.

re-importing data not created in their institution was seen as strange, “stupid” or dangerous by some other colleagues.

## Current ways to roundtrip data

It was a bit surprising that roundtrips were mentioned often, but that only a few institutions do them already. During the interviews, some participants referred to the [Swedish National Heritage Board](#)<sup>7</sup> and their [activities with it](#)<sup>7</sup> in the context of Commons metadata and the authority data used for that. One participant mentioned that their institution is “about to do the first roundtrip” [in this institution]. One participant mentioned to have started roundtrips recently. In the memberchecks, we also learned that the Metropolitan Museum of Art also does roundtrips.

Three participants mentioned that they manually correct data when they notice that people improved it (see the [section on Monitoring data](#) on how participants found changes in their data).

## Improvements or degradation of data after import

### Fears to make changes harder

Initially, we started with the assumption that degradation of once-imported data would be a big concern. We discussed a feature called “Signed Statements” with participants. The feature would allow to “sign” a statement. Any subsequent change to data would break the “signature”<sup>16</sup>.

However, among the people to whom we talked about this<sup>17</sup>, there were different understandings of what such a “signed statements” feature might do and participants did not see the proposal above as a solution to a relevant problem they had. One person explicitly mentioned their concerns of such a features making improvements to data less likely by deterring users with the signed data:

*“Wikidata is sold on these grounds ...[that it is] a good way to get your data corrected. If that’s the the use-case you’re going for, then you wouldn’t want to sign the statement” [P03]*

Since participants often hoped to re-import improved and corrected data from Wikidata via [roundtrips](#), deterring changes should be avoided.

---

<sup>16</sup> Making it behave less like a signature but more like a metaphorical sealing-wax.

<sup>17</sup> We talked about signed statements with 8 participants, also because we actively brought it up, since finding out about the possible usefulness of the feature was [one of our original research interests](#).

## Monitoring data

While participants had no interest in protecting data from being edited, they were interested in getting to know what happens with data they imported or at Items they contributed to. However, to monitor changes to data was described as difficult or that it needed to be done manually.

### *Detecting merges*

A frequent concern<sup>18</sup> were merges of several Items into one. From the perspective of the participants, these Items often referred to different concepts and should not be merged. One example was that an editor suggested to merge two items concerning two very similar looking art works. However, the two Items were deliberately different Items for two different works. The negative effects of such merges would have been minimal, but the participant suggested that in the future, their institution would build on Wikidata for services which would then deliver results not matching their collection.

The three participants who described the problems around merges in some detail went into discussions with other community members and could help them understand why they thought the Items should be separate.

Participants did not name any tools that would help to detect such merges. One participant mentioned that they got notified by another community member about the pending idea to merge Items.

### *Finding divergence between local data and data on Wikidata*

Changes of imported data could, from the perspective of the participants, indicate fixed mistakes or a degradation of the data. A degradation was particularly seen in merges of Items they considered to be deliberately separate entities in their local data. One participant said it would not be a “true representation of our data” anymore. Changes to ontology was a concern for one participant, another was concerned about changes which used properties in a way that was seen as unsuitable. Several participants voiced that they would like to keep track of changes better, but that it currently is a manual procedure for them.

## Provide Metrics

In many institutions, impact metrics are important. Metrics were mentioned by a majority of participants<sup>19</sup>. However, metrics are gathered mainly for imported media on Commons or the views of Wikipedia articles but not on Wikidata use. One participant saw the problem

---

<sup>18</sup> Mentioned by 5 participants

<sup>19</sup> 9 Participants mentioned metrics, for one user they were essential.

in the lack of tools, while another suggested that Wikidata and its CC0 model makes it hard to track data use since people could copy the data freely without mentioning the creator. Providing impact metrics seems to be an important part of getting endorsement and support from management. Thus, the difficulties of gathering impact metrics of Wikidata use might hinder its adaption.

## Intertwinement with Commons

For cultural institutions work on Wikimedia Commons seems to be *the* gateway to work on Wikidata<sup>20</sup>. If collection items were photographed or scanned, the images were uploaded to Commons, while Metadata gets stored on Wikidata.

Via [Structured Data on Wikimedia Commons](#) there is now also a direct connection between the two systems.

When participants talked about both systems, several remarked that it is easier for non experts or higher-ups to understand why one should contribute to Commons and is easier to prove impact on Wikimedia Commons since there are metrics-providing tools like GLAMerous or GLAMtools.

## Summary

The people we talked to were enthusiastic about using Wikidata and used it to share data of their institutions with more people. Another motivation is improving their institution's data by re-importing data from Wikidata.

It was hard for participants to model data according to both requirements of their professional discipline and Wikidata's (implicit) rules. Also, after the import, it is hard for people to see what happens to the Items and statements they added. This would be relevant to spot degradations and to participate in the community. A related challenge is creating metrics for impact of work in Wikidata.

Many of our participants and their institutions started their work with free knowledge on Wikimedia Commons. They then transitioned to Wikidata, since Wikidata could deal with the metadata of the depicted items.

---

<sup>20</sup> 9 participants mentioned Wikimedia Commons as how they got involved in Wikidata.



## Appendix A: Background information

To keep the main report short, we did not include explanations of terms and concepts that our participants used. These are named and explained in this section.

### Types of data institutions use

The GLAMs we talked to broadly seemed to deal with two types of data:

**Collection Data:** Data about the things the institutions has in its collection. For paintings it would be what materials were used, who the painter was and when the painting was created.

The collection data is stored in the collection management system<sup>21</sup>.

**Authority File:** List of unambiguous, standardized terms used to describe the items in the collection. E.g. is the famous painter called “Pablo Picasso”, “Picasso, Pablo” “Pablo Diego José Francisco Picasso”? And what would be a lesser known Artists, who shares the same first- and last name be named?

The authority files might be managed in a collection management system but might also be stored in lists or spreadsheets.

Institutions might create their own authority files or use and contribute to shared ones like the Library of Congress Name Authority File ([LCNAF](#)).

In the member check, a person pointed out that a third type of data are “**controlled vocabulary**” which define relationships between concepts. While the authority file would list how which artist is named, the controlled vocabulary would say what an “artist” is.

### Modeling and reconciliation: Processes, terms and tools

#### Modeling

...is structuring your data into property-value pairs. It needs to be specific enough to describe the item itself well, but general enough to apply to many items in your collection. There are different possible ways to structure data. E.g. movie awards could be modeled by having an Item like “Academy Award 2019” having a “won by” property which lists all the winners. But one could also give movies a property called “won award” linking to “Academy Award” and a qualifier with “Year: 2019”. Parallel structures with different modelings in

---

<sup>21</sup> Examples for collection management systems are [Adlibz](#) and [tmsz](#) (“the museum system”)

parallel are possible. However, this makes querying Wikidata hard (as described by P1 and P2)

Questions that need to be asked for a successful modeling are “Are these values relevant for Wikidata?”, “What information is essential about this type of Item?” and “How are Items of the same or similar type modeled already?”

A way to communicate modeling-suggestions are wikiprojects like “[Sum of all Paintings](#)” which suggest properties and their value types. However, people need to know that the project page and the suggestions exist and they need to use them manually – there is no templating mechanism or something similar.

And to make it a good start painting item add the following statements:

Property	Data type	description	required
<a href="#">creator</a> (P170)	Item	maker of this creative work or other object (where no more specific property exists)	yes
<a href="#">inception</a> (P571)	Point in time	date or point in time when the subject came into existence as defined	yes
<a href="#">inventory number</a>	String	identifier for a physical object or a	yes

*(Cropped screenshot from Sum of All Paintings page on 2019-09-02, content under [CC-BY-SA 3.0](#))*

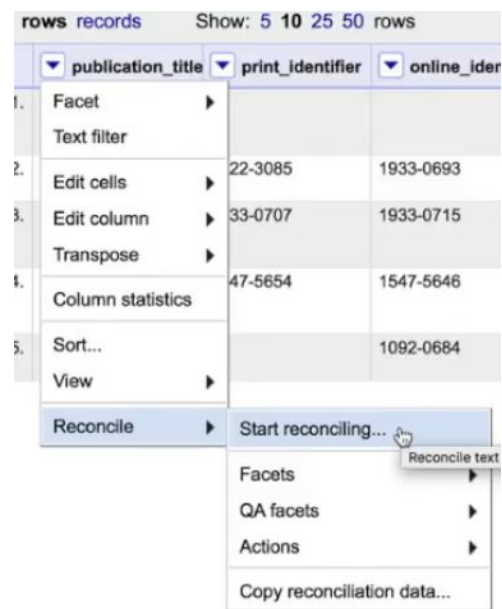
## Reconciliation

...is mapping the properties and values of your modeled data to existing Properties and values on Wikidata or plan to create new ones.

Questions that need to be asked are “To what Property on Wikidata does that property in my data match?” “Can I use this Property for my data or should I rather ask to create a new Property to avoid misusing the current Property?” or “What Items match these items in my database?”<sup>1</sup>

Reconciliation and modeling are intertwined and for doing them successfully, one needs two types of expertise: Expertise in the field that the data describes and expertise in how Wikidata and particular Properties should be used.

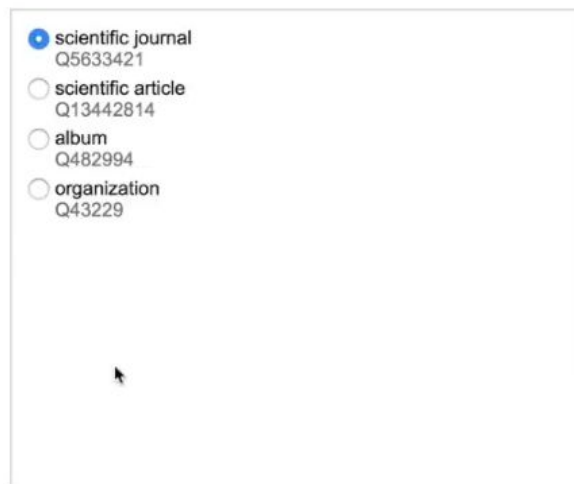
*Example: Reconciliation with Open Refine*



(Cropped frame from a video by Owen Stephens, [CC-BY 3.0, https://youtu.be/q8ffvdeyuNQ?t=209](https://youtu.be/q8ffvdeyuNQ?t=209))

The software Open Refine can semi-automatically reconcile: The content of a column can be matched with Items on Wikidata.

Reconcile each cell to an entity of one of these types:

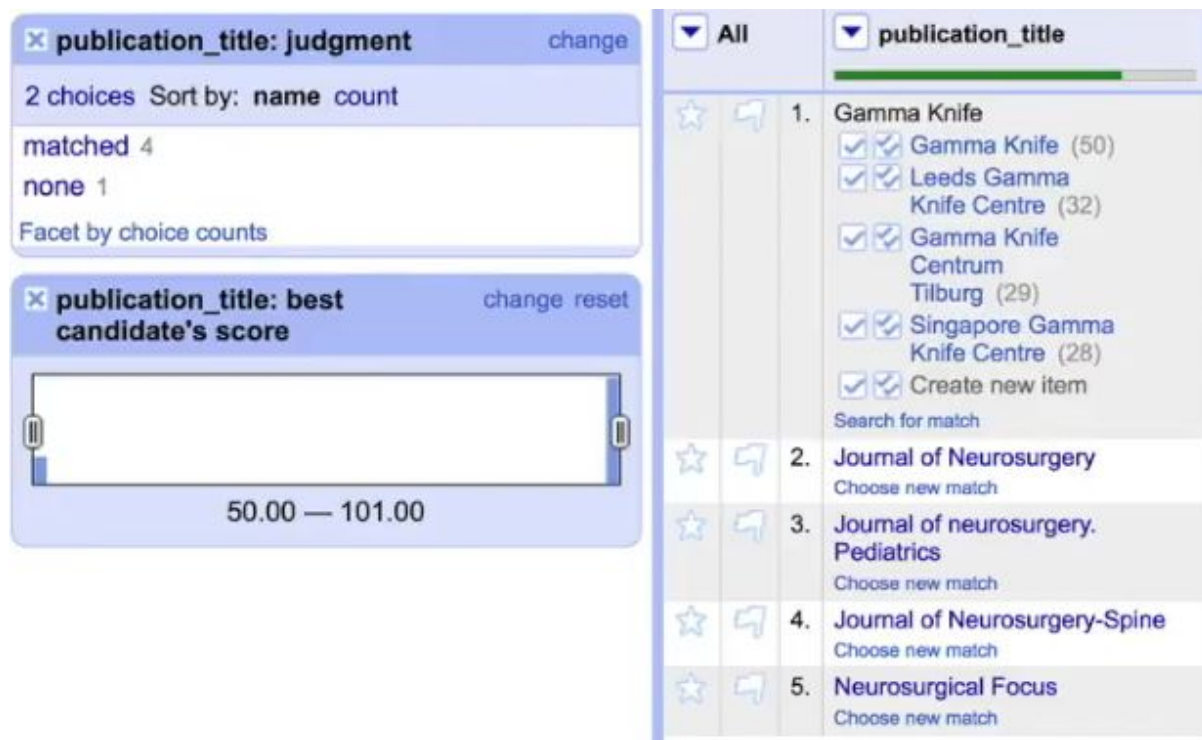


- Reconcile against type:
- Reconcile against no particular type
- Auto-match candidates with high confidence

(Cropped frame from a video by Owen Stephens, [CC-BY 3.0, https://youtu.be/q8ffvdeyuNQ](https://youtu.be/q8ffvdeyuNQ))

After choosing the type of the Item and/or doing other adjustment to improve the matching, Open Refine attempts to match the content in each cell with an Item on Wikidata.

While in the examples, most Items could be matched (“matched 4”), one could not (“Gamma Knife”) and needs to be matched with one of the possible options (or another Item entirely)



(Cropped frame from a video by Owen Stephens, [CC-BY 3.0, https://youtu.be/q8ffvdeyuNQ](https://youtu.be/q8ffvdeyuNQ))

## Identifiers

While not being a core topic in the conversations themselves, *identifiers* are an important part of the data management infrastructure. Examples for identifiers are

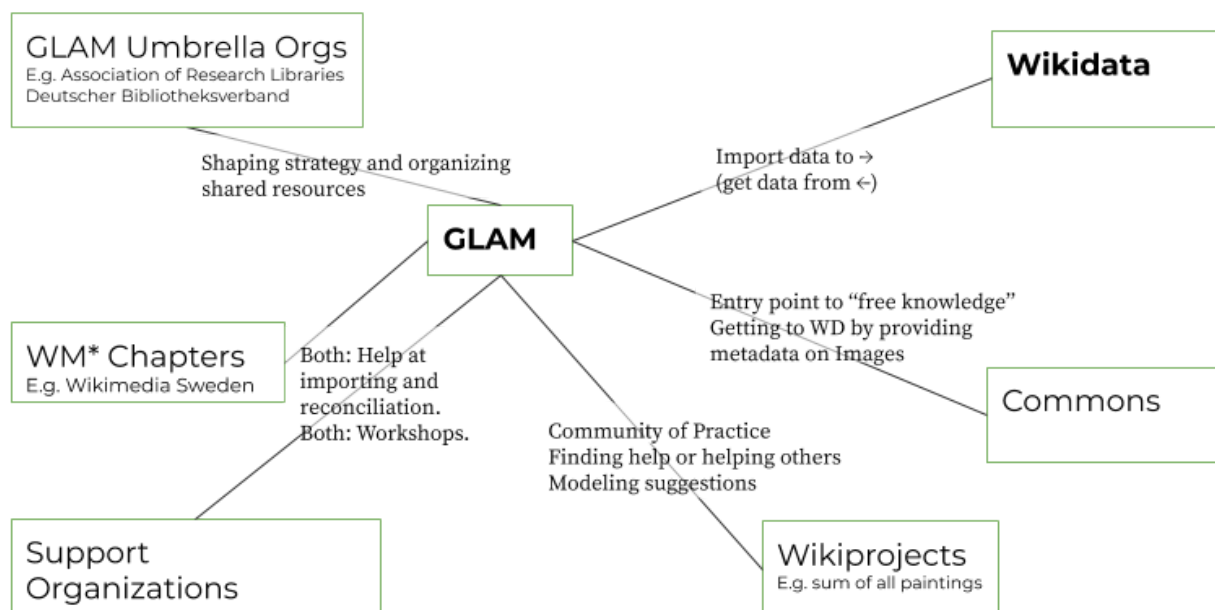
- Wikidata Item codes (like Q42)
- International Standard Book Number (ISBN) for books
- Library of Congress Control Number (LCCN) for books and authors

People add Wikidata identifiers to corresponding Items in their databases and/or add their institutions identifiers to Items in Wikidata. Using such identifiers, one can match the local items of the institutions database with corresponding Items on Wikidata automatically, which is important e.g. to do [data roundtrips](#). Some<sup>22</sup> people pointed out that now the library of congress has added Wikidata identifiers to their items. Given that the Library of Congress is a major institution and source of data, this is seen as an important endorsement of Wikidata and will further spread the use of Wikidata identifiers in other databases.

<sup>22</sup> 2 people mentioned this explicitly

## Kinds of Institutions

Based on the interviews, we created a map of typical actors that were mentioned. This is not true in all details for all organizations. The related actors can differ e.g. according to type of institution<sup>23</sup> and country<sup>24</sup> and but gives a good overview of actors one might want to consider.



## Appendix B: Research Methods

### Research Questions

Iteration 1: Data quality and image of the institution

We initially started with this short summary of our research interests:

*The Wikidata team wants to help institutions to maintain and raise the quality of data they imported. A feature related to this on our roadmap, for example, would be "Signed Statements".*

- *How and why is data imported*
- *What are the motivations of the institution?*

<sup>23</sup> As inferred from the interviews and as confirmed by a member check, libraries, e.g. seem to have stronger umbrella organizations in many cases

<sup>24</sup> E.g. Institutions in Germany are more likely to be financed by the state than in the US

- *What are the concerns of the institution around working on and with Wikidata*
- *How (if at all) do the institutions think about their name and reputation being connected to the data they imported?*

## Iteration 2: Workflows and Goals

We developed the second iteration according to the various motivations and activities we got to know about. In later interviews, we thus used this guide for semi-structured interviews:

- *What do you do in your work with open data?*
- *What are (de)motivations to use Wikidata?*
- *Workflows:*
  - *How and why is data imported?*
  - *Do you need to model data? If yes, how?*
  - *What happens once the data is imported?*
  - *Do you trip data back to supplement original data or the like?*
- *Institution:*
  - *What are concerns of the institution around working on and with Wikidata*
  - *What taints the quality of data? What would be a “bad” changes to data?*
    - *What would happen if you do not notice these changes?*
  - *How are roles distributed in looking at Quality of items? Do you need to be domain and tech expert? Or can an expert in WD work well with an expert in [content domain like libraries or genes]*
- *specific QA questions:*
  - *Do you know about signed statements?*
  - *What would it do for you, you think?*

# Authors and License

Research and Report was conducted and written by Jan Dittrich, Wikimedia Deutschland e.V., the UX Team at the Software Engineering department and the Department of Bildung-Wissenschaft-Kultur at Wikimedia Deutschland, e.V.

This research would not have been possible without the support of our participants at several GLAM institutions who shared their motivations, activities and problems of Wikidata use in their work.

November 2019



This work by Wikimedia Deutschland e.V. is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

Wikimedia Deutschland  
Gesellschaft zur Förderung Freien Wissens e.V.

Postfach 61 03 49 / 10925 Berlin  
Tempelhofer Ufer 23-24 / 10963 Berlin

Telefon: +49 (0)30 219 158 26-0  
Telefax: +49 (0)30 219 158 26-9

E-Mail: [info@wikimedia.de](mailto:info@wikimedia.de)  
Website: <http://wikimedia.de>  
Blog: <http://blog.wikimedia.de>  
Twitter: <http://twitter.com/WikimediaDE>  
Facebook: <http://facebook.com/WMDDeV>