

# Wikidata: Engagement with Australian GLAMs

Rebecca Hawcroft

July 2020

For Wikimedia Australia



## TABLE OF CONTENTS:

SECTION 1 INTRODUCTION .....	4
<b>Background</b> .....	<b>4</b>
<b>Outcomes</b> .....	<b>4</b>
SECTION 2 CURRENT CONTEXT .....	4
1. Map of Australian GLAMs and Wikidata July 2020.....	5
2. Wikidata and GLAM partnership pathways.....	6
3. Support documentation .....	7
4. Visual version of our partnership pathway .....	8
SECTION 3 INTERVIEWS WITH WIKIDATA USERS.....	9
1. Introduction.....	9
2. Questions:.....	9
3. Overarching principles.....	9
<b>Key Issues</b> .....	<b>9</b>
1. Develop and refine reasons why GLAMs should be involved in Wikidata. ....	10
2. Clarify Wikidata’s role as a data repository and the benefits/freedoms compared to other platforms. 10	
3. Provide examples of successful collaborations. ....	10
4. Provide clear channels of support and liaison between the GLAM and the open knowledge movement. 11	
5. Identify tools required for use of Wikidata and direction to clear and simple training .....	11
6. Establish learning communities for Wikimedians in libraries, cultural heritage, and research institutions 11	
7. Develop the value proposition. ....	11
SECTION 4 DEVELOPMENT OF CONTENT .....	12
<b>Target GLAMS</b> .....	<b>12</b>
<b>Draft Content</b> .....	<b>12</b>
1. Pre-upload- Preparing your data for Wikidata .....	13
2. Wikidata Value Proposition .....	17
SECTION 5 NEXT STEPS .....	19
<b>Content Testing</b> .....	<b>19</b>
<b>Final Outcomes</b> .....	<b>19</b>



## SECTION 1 INTRODUCTION

The following report has been prepared by independent consultant Rebecca Hawcroft for Wikimedia Australia. The aim of the project is to contribute knowledge of the GLAM sector, review existing Wikidata materials and conduct needs analysis to inform GLAM institutions about Wikidata, and to encourage GLAMs to incorporate Wikidata into their work. This report represents the first stage of that project.

### BACKGROUND

Wikidata is a collaboratively edited knowledge base hosted by the Wikimedia Foundation, the non-profit organization that also hosts Wikipedia and several other wiki-based knowledge projects. Wikidata serves as a central knowledge base for Wikimedia projects as well as being a freely available open database of linked open data for other projects.

Wikimedia Australia (WMAU) has a long history of working with the GLAM (Galleries, Libraries, Archives, and Museums) communities. Primarily these projects have focused on Wikipedia although there have been some examples of Wikidata projects. WMAU wants to build on those existing relationships to drive the uptake of the Wikidata platform. Australia currently has 1019 GLAM organisations with items on Wikidata; 552 Museums, 395 libraries and 19 archives.<sup>1</sup> This is well below the number of institutions participating from the United States and many European countries, indicating considerable potential for expansion of GLAM involvement in Wikidata.

Wikimedia Australia has allocated funding to investigate ways to improve the representation of Australian cultural heritage institutions (GLAMs) and their collections on the Wikidata platform.

The first stage of this project has involved surveying a range of people involved in Wikidata to understand the issues that might impede and/or encourage GLAMs to participate and how successful partnerships have worked. The findings of these investigations are summarised below.

The second stage of the project will involve the development of content storyboards for 'Why and how of Wikidata' that will be tested with a selection of GLAM institutions. Target GLAMs to test this material on are identified here. The development of final content will then follow, incorporating the feedback and responses received during the testing period.

### OUTCOMES

The desired outcome is the establishment of relationships with new GLAM partners and to greatly enhance the amount of high-quality knowledge that can be made available to the general public through the Wikidata platform. The project also seeks to facilitate the production of a foundational set of globally reusable materials that can be further developed, utilised as content for webinars and to brief the Wikimedia Australia community presenting to GLAMs about Wikidata and improve relationships with Australian GLAM institutions.

It is also desired that data from 5 new institutions will be identified as target datasets to be available via Wikidata, and pitches made to these GLAMs.

## SECTION 2 CURRENT CONTEXT

Pru Mitchell, of Wikimedia Australia, provided the following analysis of current GLAM participation in Wikidata.

# 1. Map of Australian GLAMs and Wikidata July 2020

Candidate collection      In discussion      Data only      Wikidata partnership/project

Type of collection	Galleries	Libraries	Archives	Museums
Small GLAM collection, collection metadata not online, and/or minimal internal IT capacity	Women's Art Register			
Mid-size GLAM collection in a regional location				
Mid-size GLAM collection in a capital-city location with onsite IT capacity				Dictionary of Sydney- Person ID; Organisations; Buildings; Places; Structures; Events  Sydney Living Museums (NSW Archives)
Large state-level GLAM collection with open, professionally-managed collection metadata and onsite IT capacity	Art Gallery of South Australia – Work ID  National Gallery Victoria -Artwork ID; People ID	State Library NSW  State Library Queensland	NSW State Archives – Government Agencies ID	
National-level GLAM collection with open, professionally-managed collection metadata and onsite IT capacity		National Library of Australia – People ID; Newspaper ID		Australian Centre for the Moving Image – Object ID  Australian National Maritime Museum – Person ID  Australian War Memorial – People ID; Art ID; Place ID; Military units; Military events

## 2. Wikidata and GLAM partnership pathways

GLAM Wikidata status	Description	Action
Candidate collection	A GLAM collection identified as a desirable or likely candidate to approach for data and/or to discuss a Wikidata partnership or project.	<ol style="list-style-type: none"> <li>1. Complete a <a href="#">Candidate Collection Preparation Checklist</a></li> <li>2. Prepare a <a href="#">Wikidata Value Proposition</a> specific to this GLAM Collection</li> <li>3. Customise the <a href="#">Preparing your Data for Wikidata</a> flyer for this GLAM [see Toby's Step 1]</li> </ol>
Data only	A GLAM collection that provides open data that has been imported into Wikidata, (for example mix n match), without reference to, or involvement by, the GLAM itself.	<ol style="list-style-type: none"> <li>1. Complete a <a href="#">Wikidata mapping of the GLAM dataset</a> noting areas for discussion</li> <li>2. Prepare a <a href="#">Wikidata Value Proposition</a> specific to this GLAM dataset</li> <li>3. Set up some <a href="#">Sample Queries</a> that include this GLAM dataset</li> </ol>
In discussion	WMAU has approached this GLAM with a <a href="#">Wikidata Value Proposition</a> and <a href="#">Preparing your Data for Wikidata</a> flyer	<ol style="list-style-type: none"> <li>1. Document the outcomes of the discussion in a <a href="#">Response to Value Proposition</a></li> <li>2. Develop a <a href="#">Next Steps Workflow</a> document appropriate to this GLAM</li> <li>3. Assign appropriate people as ongoing <a href="#">Liaison Team</a></li> </ol>
Wikidata partnership/project	<p>GLAM is contributing to and/or using Wikidata in its work.</p> <p>This may involve a formal or informal agreement with Wikimedia Australia, and/or a staff member with Wikidata expertise/responsibilities and/or a Wikidatan-in-Residence</p>	<ol style="list-style-type: none"> <li>1. Document this project/partnership in a <a href="#">Case Study</a> and/or <a href="#">This Month in GLAM</a> article</li> <li>2. Discuss <a href="#">Roundtripping Investigation Checklist</a> with GLAM to identify possible enrichment and re-use projects</li> <li>3. Celebrate within the GLAM and Wikidata communities</li> <li>4. Evaluate project</li> </ol>

### 3. Support documentation

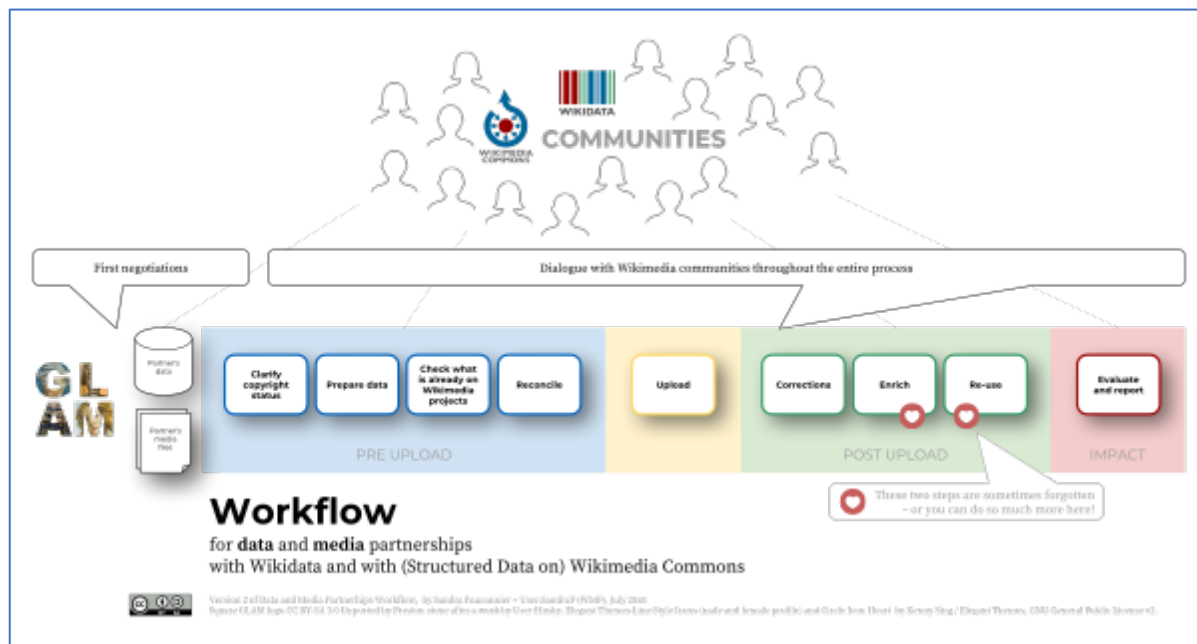
Resource	Content
<b>Candidate Collection Preparation Checklist</b>	Detail reasons for approaching this GLAM, and any research that can be undertaken prior to meeting that will improve our understanding of their needs and help prepare a Wikidata Value Proposition specific to this GLAM
<b>Wikidata Value Proposition</b> Slides and script	What is the value proposition specific to this GLAM Collection and to Wikimedia Australia and Wikidata? What are we hoping to achieve? Why? What are the alternatives?
<b>Wikidata Mapping of the GLAM Dataset</b> Metadata/Wikidata table/sheet	Once a candidate GLAM dataset has been identified, select the most appropriate Wikidata project/template/properties and map the dataset in " <u>a standard/useful format</u> ". Note areas for discussion with GLAM to ensure a practical/clean import
<b>Response to Value Proposition</b> Short report template	Document the outcomes of an initial discussion with a GLAM. <ul style="list-style-type: none"> <li>• What was the level of interest?</li> <li>• What were they most interested in?</li> <li>• What were the areas of concern, and how could we overcome these?</li> </ul>
<b>Sample Queries</b> URLs for SPARQL queries	Develop a small number of queries that are highly relevant to this GLAM. Explain in words what the query is doing, and sources it uses. Include this GLAM's own dataset where possible
<b>Next Steps Workflow</b> Text checklist with space to record decision points, and may include links to existing online resources	This is a checklist document recording discussion and decisions to move a data donation, project or partnership to the next stage. Possible items include: <ul style="list-style-type: none"> <li>• Desired outcome / Success factors</li> <li>• Dependencies</li> <li>• What will each party do?</li> <li>• Who will do this– specify backup contacts from both GLAM and WMAU</li> <li>• By when, or escalation process</li> <li>• Any paperwork required</li> <li>• Technical steps, Tools required</li> <li>• Documentation, data quality and constraints</li> <li>• Metrics, Networks and Celebration</li> </ul>
<b>Preparing your Data for Wikidata</b> A4 brochure that can be readily edited and printed on demand	Designed A4 brochure that can be handed to GLAMs outlining <ul style="list-style-type: none"> <li>• Value proposition</li> <li>• Minimum requirements for data [see Toby's Stage 1]</li> <li>• Mention copyright</li> <li>• High level workflow [customise Toby's &amp; Sandra's workflow diagrams below]</li> <li>• WMAU contact names, email addresses</li> <li>• Image and short blurb + URLs of two relevant Wikidata projects</li> </ul>
<b>Roundtripping Investigation Checklist</b> Text checklist or script for discussion. Possible slides with stories from other GLAMs (if available)	Checklist to tease out options for re-use and enrichment using Wikidata Likely to require understanding of the GLAM's own internal workflows and data management systems/infrastructure Likely to require expertise beyond WMAU Wikidata people <ul style="list-style-type: none"> <li>• Tools required</li> <li>• Highlight current examples of GLAMs and linked data using Wikidata</li> <li>• Contacts able to provide advanced level of advice</li> </ul>

#### 4. Visual version of our partnership pathway

**TASK:** To integrate our pathway improvements into existing Wikidata community workflow diagrams.

The focus of our project is

1. “Pre-upload” stages of GLAM institutional engagement (a value proposition, early stage planning, data identification and prioritisation); and
2. “Impact” stages (measurement, benefits realisation)



[https://outreach.wikimedia.org/wiki/GLAM/Resources/Data\\_and\\_media\\_partnerships\\_workflow](https://outreach.wikimedia.org/wiki/GLAM/Resources/Data_and_media_partnerships_workflow) by SandraF



[https://outreach.wikimedia.org/wiki/GLAM/Resources/Data\\_and\\_media\\_partnerships\\_workflow](https://outreach.wikimedia.org/wiki/GLAM/Resources/Data_and_media_partnerships_workflow) by Toby Hudson



# SECTION 3 INTERVIEWS WITH WIKIDATA USERS

## 1. Introduction

During July 2020, 6 interviews were conducted with existing Wikidata users to better understand the issues GLAMs face in joining and using Wikidata. The individuals are not identified here beyond their fields and experiences. The following questions were asked of the subjects and summaries of their responses are reproduced in **Appendix A**. These interviews have been used to identify the key issues below.

## 2. Questions:

- What successes have you had?
- What facilitated these relationships?
- Did face-to-face events like conferences (which ones?) or showcase linked data projects help bring the GLAM on board?
- What kind of data was it and how did the GLAM want it used?
- (What data is suitable for Wikidata? If it can be linked to other items.)
- Were there licensing issues or concerns? How were they handled?
- Did the data need work to be structured in a special way before being added to Wikidata?
- Did the data donation result in standout projects to demonstrate value?
- Have the GLAMs used Wikidata tools? (Eg Crotos)
- Value expectations or how was value measured?
- Were there advantages for the institution internally- eg: increased access to their own data?
- What has been the best tool available to you in gaining GLAM donations?
- What are the main barriers to institutions using Wikidata?
- What were the barriers for you (the interviewee) in terms of having these discussions with institutions?
- What materials would you like to make it easier?
- Who else would be good to talk to on these matters?

## 3. Overarching principles

The following issues came up in a number of the interviews and provide generalised direction for the next stage of the project.

### 1. All GLAMs are different

All GLAMs are different and specifically libraries and archives differ from museums and galleries in their collections, metadata structures and likely uptake of the platform. Additionally, the size of an institution will have a significant impact on its ability to dedicate time to onboarding data to Wikidata. MWAU should target a particular GLAM type and develop resources to support them rather than taking a one size fits all approach.

### 2. Tailor content for users

It's not new content that's required but rather a tailoring of the existing content for different institutions and GLAM types, and formats of communication. This should include looking at producing things like printed guides as well as presenting at GLAM events and having active user workshops.

## KEY ISSUES

The following are the key issues that emerged during the interviews:

## 1. Develop and refine reasons why GLAMs should be involved in Wikidata.

This is not about generating new information, but collating and refining what is there to a coherent and accessible format. Possibly a series of different formats targeted to different GLAM types.

Reasons that have been important in establishing GLAM relationships in the past include:

- A global platform- reaching wider audiences- including outside their home countries or regions
- A free, multilingual platform with a worldwide readership
- Link your dataset to other major data sets
- Gives smaller institutions or groups a level playing field to be in dialogues

## 2. Clarify Wikidata's role as a data repository and the benefits/freedoms compared to other platforms.

Wikidata as a platform where GLAMs can add to their own data by linking to other data sets, data round tripping and through greater visibility. Acknowledges that no institutional dataset is whole-gains from linking to other data sets. Promise of LOD is scale and visibility, where data is actively edited on a massive scale: "life gets breathed into it".

## 3. Provide examples of successful collaborations.

As with other issues these examples should be tailored to the type and scope of the GLAM. An example of this can be found in the White Paper, Finding GLAMs:

*"The Metropolitan Museum of Art in the United States is an example of a successful partnership between a high-profile cultural institution and Wikimedia. In 2017, the museum launched their Open Access Policy, putting over 406,000 images of public-domain artworks under the CC0 waiver.[8] This is the most generous of the Creative Commons tools, enabling anyone to freely copy, modify and distribute the materials, including for commercial use.[9] Those images were then uploaded to Wikimedia Commons by a Wikimedian in Residence.[10]*

*A year later, it was clear that the resources shared by the museum were very popular.[11] The nearly 4,000 images included in Wikipedia articles were reaching 10 million viewers per month. What was particularly interesting is that the most popular artworks reached a much larger audience through Wikipedia than through the museum's own website; a third of that audience came from Wikipedia articles in languages other than English. One of the world's most famous museums still benefited tremendously from sharing their materials on open platforms."*

Finding GLAMs, White Paper, Alicia Fagervig

<[https://meta.wikimedia.org/wiki/FindingGLAMs/White\\_Paper](https://meta.wikimedia.org/wiki/FindingGLAMs/White_Paper)>

See also: [https://outreach.wikimedia.org/wiki/GLAM/Model\\_projects](https://outreach.wikimedia.org/wiki/GLAM/Model_projects)

#### 4. Provide clear channels of support and liaison between the GLAM and the open knowledge movement.

Wikimedians in Residence is one example, but others include the Project Chat page on Wikidata and @Wikidata on Twitter.

Some respondents indicated GLAMs may need to be able to point to formal support in order to get approval for Wikidata projects.

[https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)

Twitter conversations- @Wikidata

#### 5. Identify tools required for use of Wikidata and direction to clear and simple training

Provide clear onboarding instructions and identify avenues of support, if required. The role and functions of processes such as Mix'n'match, OpenRefine and Sparql Query are central to Wikidata functions. These should be demystified early and resources for further training indicated (review to see if existing training can be used or if new material should be produced).

Existing examples include: <https://www.wikidata.org/wiki/Wikidata:Tours>

#### 6. Establish learning communities for Wikimedians in libraries, cultural heritage, and research institutions

One respondent at a major state library identified the importance of having a Wikidata working group within the library. NZ respondents discussed the impact of presenting at conferences and other events within the small GLAM community of NZ. Other dispersed groups are active in the communication and facilitation of Wikidata knowledge and practices. This represents a significant way GLAMs can be encouraged to use Wikidata and also supported in their ongoing engagement with it.

Investigate the best methods for nurturing learning communities across GLAMs for Wikidata knowledge sharing recognising that different institutions will require different levels of support.

#### 7. Develop the value proposition.

As identified previously the impact of GLAM Wikidata involvement needs to be better identified and presented. This includes measurements, benefits and realisations as in successful outcomes.

This might be related to institutional changes post COVID-19 or it might be readily available tools that are driven by Wikidata that enable GLAMs to either examine their own data or provide public facing collection exploration. Examples include:

- Resonator, which allows Wikidata to be formatted differently and easily change the language, interface etc. <https://reasonator.toolforge.org/?&q=12418>
- 'Crotos'- project example- Extra institutional – a search and display engine for visual artworks based on Wikidata and using Wikimedia Commons files. <http://zone47.com/crotos/?l=en>
- 'Sum of all paintings' visualisation of Wikidata painting information [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_sum\\_of\\_all\\_paintings/Property\\_statistics/Sandbox](https://www.wikidata.org/wiki/Wikidata:WikiProject_sum_of_all_paintings/Property_statistics/Sandbox)

Identify others or create a central place where examples are promoted.

## SECTION 4 DEVELOPMENT OF CONTENT

As identified in Section 2, the focus of the second part of the current project is to develop specific content for the following stages of the Wikidata workflow:

1. “Pre-upload” stages of GLAM institutional engagement (basic knowledge of capabilities, early stage planning, data identification and prioritisation); and
2. “Impact” stages (measurement, benefits realisation or the value proposition)

### TARGET GLAMS

Understanding that GLAMs and their collections are all different, with different needs, this content will be developed for specific glam types. The following GLAM types have been identified as preliminary targets. This is based on both current gaps and perceived values propositions. Onboarding content is to be developed and tested as part of this stage of the project. Target GLAMs are:

GLAM type	Example
Mid-size GLAM collection in a capital-city location with onsite IT capacity	Sydney Living Museums
Large state-level GLAM collection with open, professionally-managed collection metadata and onsite IT capacity	NSW State Archives

### DRAFT CONTENT

The following is developed from the existing introduction material located at:

[https://outreach.wikimedia.org/wiki/GLAM/Resources/Data\\_and\\_media\\_partnerships\\_workflow](https://outreach.wikimedia.org/wiki/GLAM/Resources/Data_and_media_partnerships_workflow)

The draft material below is to be developed following consultations with the target GLAMs. Following this it is to be reviewed and edited with contributions from the Wikimedia Australia team. That should follow this stage of the project, and precede user testing.

## 1. Pre-upload- Preparing your data for Wikidata

**Approach:** Detail reasons for approaching this GLAM, and any research that can be undertaken prior to meeting that will improve our understanding of their needs and help prepare a Wikidata Value Proposition specific to this GLAM.

### Target GLAM 1 (Mid-size with onsite IT)

Stage	Tips	Tools
Clarify copyright status	<p><b>Make sure that copyright of the data and media files is compatible with Wikimedia projects.</b></p> <ul style="list-style-type: none"> <li>• Data for Wikidata must be CC0. (See <a href="#">Wikidata and copyright</a>)</li> <li>• Media files' copyright must be compatible with Commons policy. (See <a href="#">Commons:Licensing</a> for comprehensive information, and <a href="#">this infographic</a> for a brief overview of how it works)</li> </ul>	<p>If permissions and licenses for copyrighted media files aren't published in a public place: make sure the permissions are clarified via an e-mail to <a href="#">OTRS, the platform used by the Wikimedia projects to manage and archive e-mail conversations</a>. (<a href="#">Licensing images: when do I contact OTRS?</a>)</p>
Preparing data	<p><b>Clean up the data to be consistent and compatible with Wikimedia Commons and/or Wikidata.</b></p> <ul style="list-style-type: none"> <li>• Look at similar media or data items on Wikimedia Commons or Wikidata for inspiration how to model the data.</li> <li>• <a href="#">Wikidata's WikiProjects</a>– the 'groups' where volunteers work together on common interests – often have recommendations on data modelling for specific subjects.</li> </ul>	<p>Spreadsheet software - allows non-programmers to run checks against existing Wikimedia content</p> <ul style="list-style-type: none"> <li>○ <a href="#">Google Sheets</a> - free spreadsheet software that can be collaborative</li> <li>○ <a href="#">Wikipedia and Wikidata tools for Google Spreadsheets</a> is a free add-on for Google Sheets that adds functions for querying Wikipedia and Wikidata.</li> <li>○ <a href="#">OpenRefine</a> (formerly Google Refine) - popular tool for advanced data cleaning, transformation and matching against Wikidata content. Its <a href="#">homepage</a> includes video tutorials and a <a href="#">guide</a> on how to use version 3.0 and higher for Wikidata manipulation and uploading.</li> <li>○ <a href="#">PAWS</a> and <a href="#">Pywikibot</a> - for those with some programming experience allows for large scale querying and advanced actions.</li> </ul>

Stage	Tips	Tools
Check what is already on Wikimedia projects	<p><b>Always check which data and media items are already present on Wikidata and Wikimedia Commons.</b></p> <ul style="list-style-type: none"> <li>• Volunteers have often already autonomously uploaded quite a few images from GLAM collections.</li> <li>• Wikidata will probably already contain quite a few data items about creative works, people and topics related to specific GLAM collections.</li> <li>• On Wikimedia Commons, it is considered good practice to upload new (higher-quality) media files. Don't overwrite existing files.</li> <li>• On Wikidata, duplicate items must be avoided and merged when they are discovered. It is OK (and even highly recommended) to add extra sources and statements to existing items though.</li> </ul>	<ul style="list-style-type: none"> <li>• Search function of Commons and Wikidata</li> <li>• <a href="#">Wikidata Query Engine</a></li> <li>• <a href="#">PetScan</a>, an advanced search and query tool for Wikimedia projects</li> </ul>
Reconcile	<p><b>Reconciliation is the step where data items from a source dataset are matched with their corresponding Wikidata items.</b></p> <ul style="list-style-type: none"> <li>• Be thorough during this step. Creating many duplicate Wikidata items must be avoided, as these cause a lot of cleanup work for the Wikidata community!</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">OpenRefine</a> and its Wikidata reconciliation feature</li> <li>• <a href="#">Mix'n'match</a>, an online tool for crowdsourced reconciliation (<a href="#">Manual</a>)</li> </ul>

## Target GLAM 2 (Large GLAM collection with open metadata)

Stage	Tips	Tools
Clarify copyright status	<p><b>Make sure that copyright of the data and media files is compatible with Wikimedia projects.</b></p> <ul style="list-style-type: none"> <li>• Data for Wikidata must be CCO. (See <a href="#">Wikidata and copyright</a>)</li> <li>• Media files' copyright must be compatible with Commons policy. (See <a href="#">Commons:Licensing</a> for comprehensive information, and <a href="#">this infographic</a> for a brief overview of how it works)</li> </ul>	<p>If permissions and licenses for copyrighted media files aren't published in a public place: make sure the permissions are clarified via an e-mail to <a href="#">OTRS, the platform used by the Wikimedia projects to manage and archive e-mail conversations</a>. (<a href="#">Licensing images: when do I contact OTRS?</a>)</p>
Preparing data	<p><b>Clean up the data to be consistent and compatible with Wikimedia Commons and/or Wikidata.</b></p> <ul style="list-style-type: none"> <li>• Look at similar media or data items on Wikimedia Commons or Wikidata for inspiration how to model the data.</li> <li>• <a href="#">Wikidata's WikiProjects</a>—the 'groups' where volunteers work together on common interests – often have recommendations on data modelling for specific subjects.</li> </ul>	<p>Spreadsheet software - allows non-programmers to run checks against existing Wikimedia content</p> <ul style="list-style-type: none"> <li>○ <a href="#">Google Sheets</a> - free spreadsheet software that can be collaborative</li> <li>○ <a href="#">Wikipedia and Wikidata tools for Google Spreadsheets</a> is a free add-on for Google Sheets that adds functions for querying Wikipedia and Wikidata.</li> <li>• <a href="#">OpenRefine</a> (formerly Google Refine) - popular tool for advanced data cleaning, transformation and matching against Wikidata content. Its <a href="#">homepage</a> includes video tutorials and a <a href="#">guide</a> on how to use version 3.0 and higher for Wikidata manipulation and uploading.</li> <li>• <a href="#">PAWS</a> and <a href="#">Pywikibot</a> - for those with some programming experience allows for large scale querying and advanced actions.</li> </ul>
Check what is already on Wikimedia projects	<p><b>Always check which data and media items are already present</b></p>	<ul style="list-style-type: none"> <li>• Search function of Commons and Wikidata</li> </ul>

Stage	Tips	Tools
	<p><b>on Wikidata and Wikimedia Commons.</b></p> <ul style="list-style-type: none"> <li>• Volunteers have often already autonomously uploaded quite a few images from GLAM collections.</li> <li>• Wikidata will probably already contain quite a few data items about creative works, people and topics related to specific GLAM collections.</li> <li>• On Wikimedia Commons, it is considered good practice to upload new (higher-quality) media files. Don't overwrite existing files.</li> <li>• On Wikidata, duplicate items must be avoided and merged when they are discovered. It is OK (and even highly recommended) to add extra sources and statements to existing items though.</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Wikidata Query Engine</a></li> <li>• <a href="#">PetScan</a>, an advanced search and query tool for Wikimedia projects</li> </ul>
Reconcile	<p><b>Reconciliation is the step where data items from a source dataset are matched with their corresponding Wikidata items.</b></p> <ul style="list-style-type: none"> <li>• Be thorough during this step. Creating many duplicate Wikidata items must be avoided, as these cause a lot of cleanup work for the Wikidata community!</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">OpenRefine</a> and its Wikidata reconciliation feature</li> <li>• <a href="#">Mix'n'match</a>, an online tool for crowdsourced reconciliation (<a href="#">Manual</a>)</li> </ul>



## 2. Wikidata Value Proposition

**Approach:** What is the value proposition specific to this GLAM Collection and to Wikimedia Australia and Wikidata? What are we hoping to achieve? Why? What are the alternatives?

**Draft Content:** This could take the format of slides and script that are able to be edited and adjusted for different GLAMs and collection types. A number of existing precedents are listed below.

Interviews with target GLAMs will identify which resonate the most and which ones answer their specific queries.

**Examples projects to be discussed with targets:**

### ARL WHITE PAPER ON WIKIDATA

<https://www.arl.org/wp-content/uploads/2019/04/2019.04.18-ARL-white-paper-on-Wikidata.pdf>

See for examples for use of Wikidata for improving the diversity, equity, and inclusion and community outreach.

### WHY DATA PARTNERS SHOULD LINK THEIR VOCABULARY TO WIKIDATA: A NEW CASE STUDY- Europeana

<https://pro.europeana.eu/post/why-data-partners-should-link-their-vocabulary-to-wikidata-a-new-case-study>

Arguments for creating rich linking data for cultural heritage material – linking it to link it to semantically rich and multilingual resources. Practical steps data providers can take to upload and align their vocabularies with Wikidata.

### WIKIDATA AND CULTURAL HERITAGE COLLECTIONS- Science Museum Group

<https://www.sciencemuseumgroup.org.uk/wikidata-and-cultural-heritage-collections/>

Webinar with set of short case studies from practitioners who have worked with Wikidata, presenting their work and the opportunities and challenges as they saw them.

Sample impact table, to be developed specifically for each target GLAM:

Impact	Measurable aspects	Examples
Evaluate and report	<p><b>Evaluate the impact of the media files and/or data by measuring improvements and (re-)use</b></p> <p>Measurable aspects may include</p> <ul style="list-style-type: none"> <li>(Number of) people who worked on the data and media</li> <li>Types of enrichment</li> <li>Inclusions in Wikimedia project</li> <li>Pageviews of pages where data/media is used</li> </ul>	<p>Wikimedia Commons:</p> <ul style="list-style-type: none"> <li><a href="#">GLAMorous</a> shows how often media files from a Commons category (or uploaded by Commons user) are used in other Wikimedia projects</li> <li><a href="#">BaGLAMa</a> shows Wikimedia page views over time, for specific categories of media files on Wikimedia Commons. <a href="#">Get in touch with its maintainer, Magnus Manske</a>, who can add your own category/ies.</li> <li><a href="#">GLAMorgan</a> shows Wikimedia page views for a specific Wikimedia Commons category for a specific month.</li> </ul>

Impact	Measurable aspects	Examples
		<ul style="list-style-type: none"> <li>• <a href="#">Fae's GLAM Dashboard</a>, a set of templates that show interesting data about a Commons category, including the most edited files and the most active volunteers who have contributed to them.</li> </ul> <p>Wikidata:</p> <ul style="list-style-type: none"> <li>• <a href="#">SPARQL Recent Changes</a>, shows changes to items from a Wikidata query over a given period of time.</li> </ul>

## SECTION 5 NEXT STEPS

### CONTENT TESTING

The draft material developed in Section 4 is to be tested by GLAM collection managers with responses recorded via follow up interviews. It is anticipated that this research will form the foundation of a series of training materials able to be adapted to a range of experience levels and formats for further development by WMAU.

### FINAL OUTCOMES

This material should form part a suite of Wikidata material that is able to be adapted for use by a variety of GLAM types. It is anticipated that GLAM specific content for the following stages of Wikidata usage will also be developed, as required support documentation identified in Section 2.

Phase	Description
<b>Wikidata Mapping of the GLAM Dataset</b> Metadata/Wikidata table/sheet	Once a candidate GLAM dataset has been identified, select the most appropriate Wikidata project/template/properties and map the dataset in “ <u>a standard/useful format</u> ”.  Note areas for discussion with GLAM to ensure a practical/clean import
<b>Response to Value Proposition</b> Short report template	Document the outcomes of an initial discussion with a GLAM. <ul style="list-style-type: none"> <li>• What was the level of interest?</li> <li>• What were they most interested in?</li> <li>• What were the areas of concern, and how could we overcome these?</li> </ul>
<b>Sample Queries</b> URLs for SPARQL queries	Develop a small number of queries that are highly relevant to this GLAM. Explain in words what the query is doing, and sources it uses. Include this GLAM’s own dataset where possible
<b>Next Steps Workflow</b> Text checklist with space to record decision points, and may include links to existing online resources	This is a checklist document recording discussion and decisions to move a data donation, project or partnership to the next stage. Possible items include: <ul style="list-style-type: none"> <li>• Desired outcome / Success factors</li> <li>• Dependencies</li> <li>• What will each party do?</li> <li>• Who will do this– specify backup contacts from both GLAM and WMAU</li> <li>• By when, or escalation process</li> <li>• Any paperwork required</li> <li>• Technical steps, Tools required</li> <li>• Documentation, data quality and constraints</li> <li>• Metrics, Networks and Celebration</li> </ul>
<b>Preparing your Data for Wikidata</b>	Designed A4 brochure that can be handed to GLAMs outlining <ul style="list-style-type: none"> <li>• Value proposition</li> </ul>

Phase	Description
A4 brochure that can be readily edited and printed on demand	<ul style="list-style-type: none"> <li>• Minimum requirements for data [see Toby's Stage 1]</li> <li>• Mention copyright</li> <li>• High level workflow [customise Toby's &amp; Sandra's workflow diagrams below]</li> <li>• WMAU contact names, email addresses</li> <li>• Image and short blurb + URLs of two relevant Wikidata projects</li> </ul>
<p><b>Roundtripping Investigation Checklist</b></p> <p>Text checklist or script for discussion. Possible slides with stories from other GLAMs (if available)</p>	<p>Checklist to tease out options for re-use and enrichment using Wikidata</p> <p>Likely to require understanding of the GLAM's own internal workflows and data management systems/infrastructure</p> <p>Likely to require expertise beyond WMAU Wikidata people</p> <ul style="list-style-type: none"> <li>• Tools required</li> <li>• Highlight current examples of GLAMs and linked data using Wikidata</li> <li>• Contacts able to provide advanced level of advice</li> </ul>

---

<sup>i</sup> [https://meta.wikimedia.org/wiki/FindingGLAMs/GLAM\\_statistics](https://meta.wikimedia.org/wiki/FindingGLAMs/GLAM_statistics)