

Final Report

Review Beyond Notability

University of London

Table of contents

Introduction	3
Recommendations	4
1. Record	5
2. Link	8
3. Query	9
4. Engage	12
5. Export	14
Conclusions	15
Colofon	16
Appendix 1 - User Stories	17
Appendix 2 - Useful queries	19

Introduction

Beyond Notability is a re-evaluation of women's work in archaeology, history and heritage between 1870 and 1950. The main datastore of the project is a Wikibase instance hosted on Wikibase.cloud. This datastore currently contains research data on over 500 women that have been identified as notable by the project in this period. This growing number of women are described using an extensive data model¹ that consists of over a 100 fields.

The project has gathered an impressive amount of information on the women and their lives by closely researching the sparse sources that are available. The project is contributing to the notability of these people and the deserved attention and respect for this group.

In order for the research to have a lasting impact on the other data sources, and to increase the value of its data output, the project investigators have asked IP Squared and Digitaal Werktuig to review the Beyond Notability platform and its Wikibase implementation, that holds linked data on the researched persons, and their lives in a structured manner.

This is the final report of the review. The review consisted of a Quick Scan² that provided a baseline for discussions, an interview that provided an overview of the needs of the project and a workshop that combined the findings with the needs and added prioritisation to these.

¹ The project's data model is described on beyond-notability.wikibase.cloud/wiki/Project:Meta

² The quick scan can be found [here](#).

Recommendations

The recommendations presented in this report are based on an investigation using User Stories and Epics, a common method to identify needed functionality, priorities and champions for projects that rely on technological implementations.

The user story analysis is a tool that helps to structure the thinking on how to work with the linked data within the Beyond Notability project. The stories are based on an interview session and a workshop with the project team. In addendum 1 you will find an overview of the user stories and a more detailed description of the theory behind them.

The report uses the main categories (epics) of the user stories to structure its recommendation. These categories are:

1. Record
2. Link
3. Query
4. Engage
5. Export

The report describes recommendations for each of the categories that are introduced in the user stories.

1. Record

Recording data is fundamental to creating a linked datastore. Three **must-have** user stories were identified.

to quickly enter items, based on the research, so the data from the archives of the SAL and the RAI can be modelled and catalogued. To improve this goal we recommend the following:

1. Adopt Cradle³.

This speeds up the recording process of adding new persons, organisations and events to the system. Cradle allows you to create forms to more easily add new information. We've already set up a basic Cradle form and added a description of how to add and edit these forms on Project:Cradle⁴

2. Adopt QuickStatements.

When researchers already have a structured set of information that they want to add to the system (e.g. a spreadsheet with names of relevant persons), then adopt QuickStatements. QuickStatements is a tool that allows you to import a preformatted spreadsheet into Beyond Notability.⁵

to express uncertainty in information (e.g. datetime) so (nuanced, conflicting, etc.) findings of the archival research are best represented we recommend:

1. Investigate whether the adoption of the EDTF datetime format

This is a supported data type in Wikibase Cloud, that gives more opportunity to express uncertainty in temporal information. E.g. when you know the day of the month, but not the month of a certain event, or that an event was recurring for a couple of years.

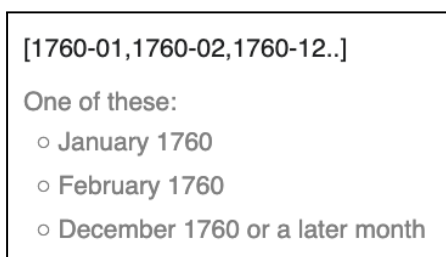


Figure. Demonstration of some of the expression possibilities of EDTF.⁶

³ It has to be mentioned that Cradle, at the moment of writing, is experiencing issues on Wikibase Cloud. These need to be resolved first. Please see the technical ticket to check whether these have been resolved: phabricator.wikimedia.org/T317109

⁴ As demonstrated using beyond-notability.wikibase.cloud/tools/cradle/#/subject/person_of_interest

⁵ See also beyond-notability.wikibase.cloud/wiki/User:Mzeinstra/sandbox/QS#Basic_working for a detailed guide on how to work with QuickStatements.

⁶ edtf.wikibase.wiki/wiki/Property:P1 provides many more examples of the expression scope of EDTF.

Existing information can be transformed and re-ingested using QuickStatements (see above).

to develop a means of recording complex and nuanced data points so the complexity of women's lives are appropriately represented in the data and provoke users to think differently about how linked data can be used. To satisfy this user story, a clear data model and data model documentation is needed. It is therefore recommended to:

1. Add more structural documentation and define a minimal required set of statements per modality (person, event, organisation, etc.).

Enforcing this minimal required set of statements can go hand-in-hand with adopting Cradle for data entry.

2. Remove unused properties to make data entry clear.

Properties that are not used can confuse the data contributor and the general public who want to use queries to explore the data. It is recommended to remove the properties from the platform that are not used. This can be done with the administrator account.

3. Unambiguously mark each item as being part of one of those modalities to make them machine readable.

E.g. record that an organisation is an instance of an organisation, that a person is a person as to not rely on (optional) information like gender or the date that the organisation was founded on.

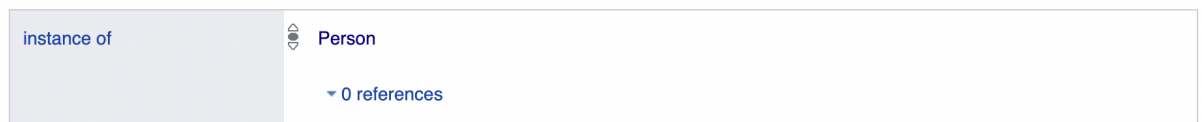


Figure. Example of stating an instance of property with the value "Person"

4. Add additional statements to the items in Beyond Notability, based on values that can be extracted from (part of) the current labels.

Labels are not easily machine readable, as they have no clear definition. A label of an event has a different meaning than a label of a person. Consider adding the labels as statements that clearly indicate their meaning (first name, last name, name of organisation, location, etc.) as well.

This can be done semi-automatically by exporting all current labels of a certain modality and transformed into QuickStatement commands (see above).

5. In addition to the previous recommendation, avoid items that only have labels.

Labels should not be considered as machine readable information, as the information stored within is not qualified. Currently over 300 items and properties do not have descriptive information.

```

10 SELECT ?item ?itemLabel ?statements WHERE {
11     ?item wikibase:statements ?statements .
12     FILTER (?statements <1)
13     SERVICE wikibase:label {
14         bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en,en-gb".
15     }
16 }

```

Figure. Query to find 'empty items' ([source](#))

6. Consider introducing a new property based on assigned gender from Homosaurus.

The project is almost exclusively dealing with assigned genders, which conflicts with the definition of gender that is used on Wikidata, which is currently linked with your definition of gender.

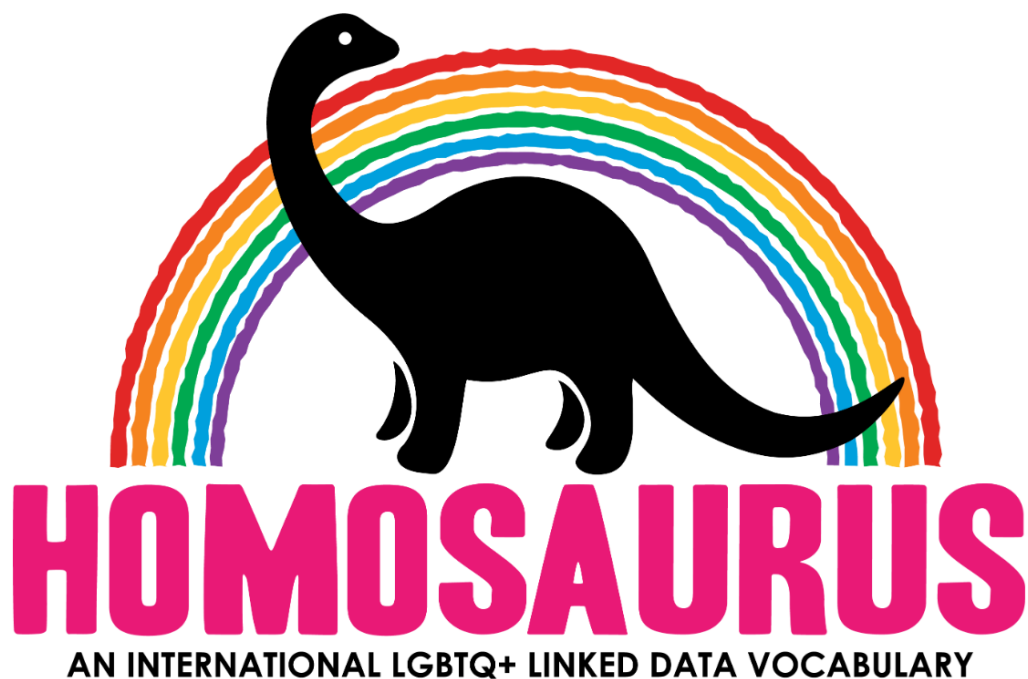


Figure. The HOMOSAURUS!

If you so desire, you can still relate this property to gender in Wikidata with a relation from the Similarity Ontology.⁷

⁷ Please refer to this paper on the Similarity Ontology, providing alternatives to a sameAs relation for LOD: [When owl:sameAs isn't the Same: An Analysis of Identity in Linked Data](#)

2. Link

Currently, datasets exist and thrive in an ecosystem of linked data. The more they reference other sources of information, the more they can find and exchange information. This is also reflected in the user stories.

There is one **must-have** user story **to link data to cultural heritage institutions so (example) it provides an opportunity to look at 'blind spots' together with other cultural heritage organisations.** To turn more information in the datastore into linked data, we recommend to:

1. Research whether persons are registered in Beyond Notability appear in other databases, like VIAF and ISNI

If they already appear, include a statement in your existing person record. If they don't appear, register for a new ISNI through the British Library (where possible) so that other institutions can also refer to the same identifier, increasing the collective knowledge and findability of the persons.

2. Where possible, link properties to (an) external data model(s).

Links between the Beyond Notability data model and external data model increase the chances of interoperability and possible future reuse of the collected information. Currently, Beyond Notability links to properties on Wikibase to identify that the property is the same as a property on Wikidata. At the time of writing there are still 70 properties that do not have a link to an external data model, Wikidata or otherwise. Consider adding these additional links to the Wikibase data model or another datamodel like Schema.org.

3. Make 'Described at URL' an unambiguous property.

A field within a data model should have a clear and unambiguous function. It is advised to split up fields that have ambiguous function or contain unclear data. For example, the field 'Described at URL' can currently either mean that a property is an exact match, that the value is an external identifier or that this is a main source of an item. Consider splitting up the values in this field into statements that are specific to the goal (exact matches, external identifiers and sources). By using queries and QuickStatement (see above) it becomes possible to move the contents of this field to other properties.

3. Query

Gathering and linking information is the start of the creating value with the Beyond Notability project. Being able to successfully query the data is the next step in creating this value. The project identified two **must-have** user stories for project members that are connected to querying.

First is the ability **to ask interesting questions about the data so the project team can better understand women's work in archaeology, history, and heritage**. Here we recommend to:

1. Clean up the data model.

This avoids deprecated properties etc. to be taken into account in queries, resulting in misleading results, errors and unintended machine-driven exploration of the datastore.

As well as **to use data analysis and visualisation of the entire dataset so the project can generate new insights into large-scale, longitudinal changes in the nature of women's archaeological, historical and heritage work**.

2. Visualise the data using SPARQL queries.

Visualisation also provides a quick overview of the relations between people that are described in Beyond Notability. Wikidata provides a good overview of all the options that are available to create these types of visualisations.

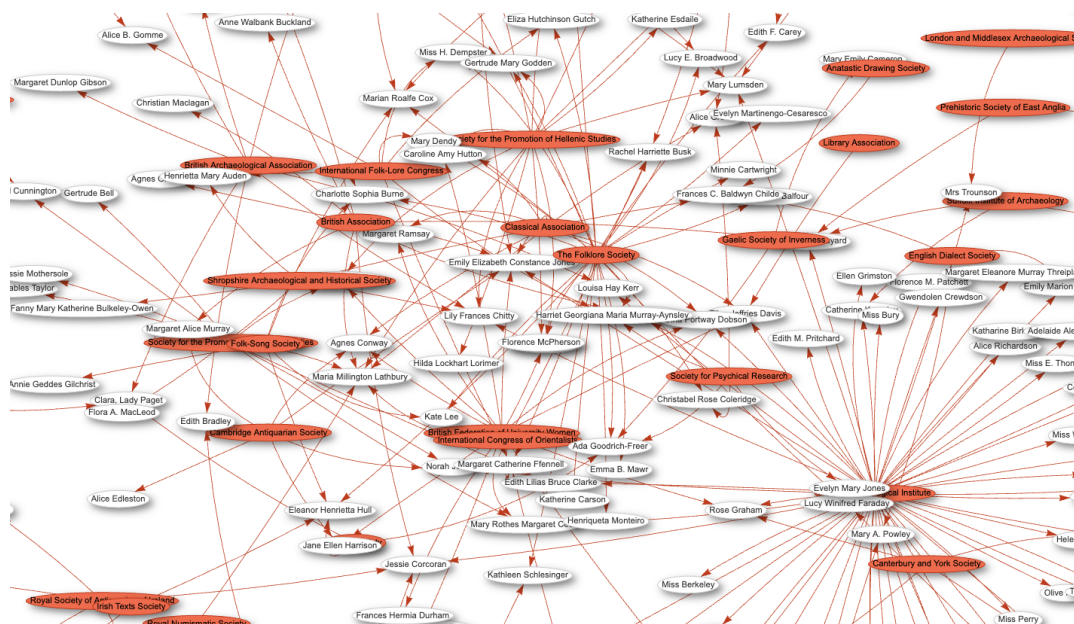


Figure. Example visualisation showing which person was a member of which organisation ([source](#)).

3. Use federated queries to enhance data visualisation and to add information to the datastore.

Wikibase allows you to display information from multiple Wikibase together in one overview. This is one of the powerful options of linked data. For example it is possible to

display information that is stored in Wikidata on Beyond Notability items, based on a match on the Wikidata identifier, ISNI, VIAF, birth date, etc. Here QuickStatements could again be used as a tool to potentially enrich the items in your datastore, with information retrieved from other datastores through federated queries.

```
SELECT ?woman ?womanLabel ?WDitem ?VIAFlink ?ISNI ?LibraryOfCongres ?Oxford ?Gutenberg ?GND ?BNF
WHERE {
  #Select all women
  ?woman bnwdt:P3 bnwd:Q3 .
  #select all women that have a Wikidata link
  ?woman bnwdt:P117 ?Qnumber .

  #create reference to Wikibase entity
  BIND(IRI(concat("http://www.wikidata.org/entity/", ?Qnumber)) as ?WDitem )
  #on Wikibase do
  SERVICE <https://query.wikidata.org/sparql> {
    #get VIAF P214
    OPTIONAL{?WDitem wdt:P214 ?VIAF
      BIND(IRI(CONCAT("https://viaf.org/viaf/", ?VIAF)) as ?VIAFlink)}
    #get ISNI P213
    OPTIONAL{?WDitem wdt:P213 ?ISNI}
    #get Library of Congres P244
    OPTIONAL{?WDitem wdt:P244 ?LibraryOfCongres}
    #get Oxford Dictionary of National Biography ID P1415
    OPTIONAL{?WDitem wdt:P1415 ?Oxford}
    #get Project Gutenberg author ID P1938
    OPTIONAL{?WDitem wdt:P1938 ?Gutenberg}
    #get Deutsche Biographie (GND) ID P7902
    OPTIONAL{?WDitem wdt:P7902 ?GND}
    #get Bibliothèque nationale de France ID P268
    OPTIONAL{?WDitem wdt:P268 ?BNF}
  }

  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en-gb". }
}
ORDER BY ?womanLabel
```

Figure. Example of a federated query ([source](#)).

The project also identified one lower priority (should-have) users stories for project members **to calculate percentages for external coverage of BN items, so the representation of BN subjects in other databases can be gauged at, to see where BN can contribute to an improvement.** Here we recommend to:

4. Explicitly link to additional sources, like WorldCat.

This will help find the person in other databases to see what the external coverage in relevant/reputable resources is of the persons that are described in the data store. Having more sources attached to the items will also create more notability for the persons to be adopted in e.g. Wikidata further down the line.

Finally the project identified another **should-have** user story for end-users, they should also be able **to ask interesting questions of the data so people's understanding of what women's work in archaeology, history, heritage can be enhanced (new notions of visibility)**. For easy exploration of the datastore, compatibility with common example queries for Wikibase datastores and general settings is of great importance. In the Engagement section we also provide recommendations on serving end users in more general terms. Here specific to querying we at minimum recommend to:

5. Adopt the general English language instead of British English to record information.

Most tools use English and do not present British English. This means that the querying functionality of the project is diminished. This recommendation has already been put into practice.

To enforce this setting it is recommended to go to the user preferences page and under 'internationalisation' select "en - english".

Internationalisation

Language:

Figure. Screenshot of language switch in account preferences.

4. Engage

The project wants to build relationships with other groups and organisations, not only in this project but also for possible future projects. Two user stories relate to this need of the project;

A **must-have** user story **to build relationships with cultural heritage institutions so that I can connect national institutions to more local heritage institutions to enhance local sites and institutions and to work with not necessary local professionals.** And the lower priority (**should-have**) user story **“to build relationships with women's history groups the project can have follow-up research projects.**

For both of these user stories it is important to clearly communicate and introduce the technical aspects of the project. We recommend that to:

1. Clearly communicate the goals of the project on the Wikibase main page.

This can be augmented with concrete calls to action. The project Wikibase is currently mostly being used as an internal communication tool. This means that clear concise information aimed at end-users on the main page is missing.

2. Add links to explainers on how to interact with the data.

This increases user engagement with the data and better description of example queries. Also consider adding a wish list for future data integrations and/or queries. The community might be able to help.

3. Configure Wikibase-SortedProperties.

The SortedProperties functionality is a means to pre-organise the way statements are presented on a Wikibase. This allows the data contributors and the public to quickly find information on an item page. With the additional benefit that recording data will become quicker, as the contributor can quickly see which data is present.

4. Explore additional skins to MediaWiki.

This can make the Wikibase more attractive and readable for the general audience. The tab 'Appearance', under the user preferences section gives an overview of the existing skins that are available. The administrator can set a default skin for all users. The screenshot below shows Beyond Notability in the 'timeless' skin.

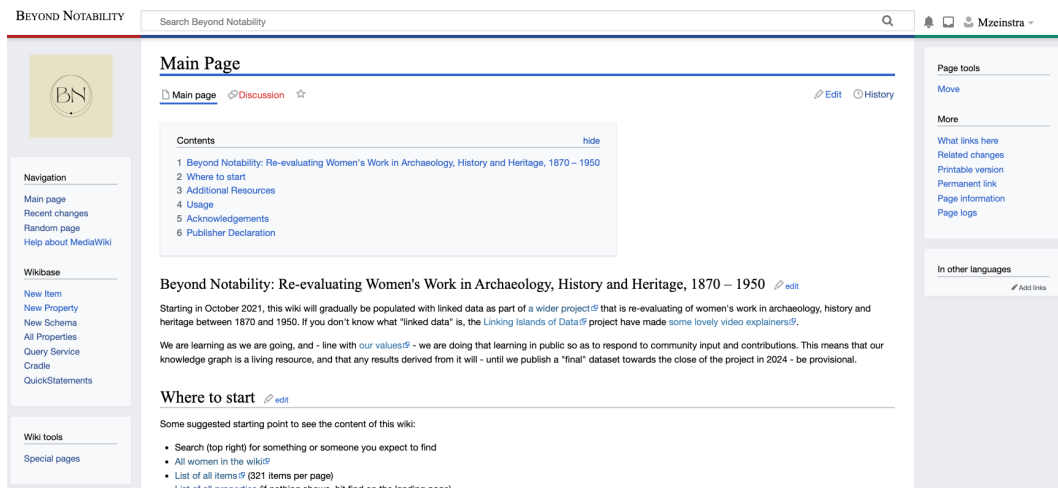


Figure. Screenshot of an alternative skin for Beyond Notability.

5. Write a privacy policy.

Users of the Beyond Notability platform need to be aware that the project deals with person information and applicable law requires that a proper privacy statement needs to be in place. The current privacy policy is not about privacy but about intellectual property.

6. Add a CC0 licence to the project.

The project currently licences its data as a Creative Commons Attribution 4.0 International License, this licence makes the data incompatible with Wikidata, meaning that after the project the data cannot be ingested into Wikidata. We recommend using the CC0 1.0 Universal (CC0 1.0) Public Domain Dedication instead.

5. Export

Thinking about the data after the project is a good way to ensure operability and lasting value creation. This is also reflected in the in two **must-have** users stories **to export the statements/triples (from Wikibase.cloud) so the data can be preserved for potential reuse post-project** and for **to export the data in raw form so it can be made available in a academic data repository.**

To further these two user stories it is recommended to:

1. Research and test exporting all data from the Wikibase.

Additionally a lower priority **should-have** user stories was added that has some connection to the 'engage' epic. Project members indicate that the general should have a way **to programmatically retrieve BN data so that the data can be reused in other projects (silos).** In order to further this user story we recommend to:

1. Create example queries to easily export some data.

To facilitate exporting of data, it is recommended to create some queries that allow users to quickly export some key data from the Beyond Notability datastore.

```
1 #All triples in the wiki, limited to 1000
2 SELECT ?a ?aLabel ?b ?c WHERE {
3   ?a ?b ?c
4   SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en-gb". }
5 }
6 LIMIT 1000 #please don't make this number too big as it will slow down the site!
```

Figure. All triples in the Wikibase, limited to 1000 ([source](#)).

Finally a lower priority **could-have** user story was developed where project members need a way **to export the history, discussion, etc (from Wikibase.cloud) so that the data can be preserved for potential reuse post-project.** We recommend to:

1. Contact Wikibase.cloud for this specific request, as the execution of this depends on the functionality that they offer in their services.⁸

⁸ This recommendation has already been put into action at the time of writing this report. See phabricator.wikimedia.org/T312131

Conclusions

The Beyond Notability datastore already embraces many of the possibilities Wikibase offers to record research findings into structured data and design a custom data model. In the remaining years of the project, this strong foundation can be further exploited by introducing auxiliary tools and methods like Cradle, QuickStatements and Federated Queries. This will allow the project to become more efficient, and less reliant on manual import.

The review also identified additional opportunities to embrace Linked Data principles and best-practices by connecting the datastore to additional (trusted) sources and data models. This enhances the machine-readability of the datastore and makes the data more interoperable for future data reuse.

Although the project evidently already developed (complex) query examples at the time of review, additional possibilities with methods to federate queries on multiple datastores and to visualise query results in the Query Service were developed.

The project has the potential to seek engagement with other initiatives and domains, well beyond the lifespan of the project. Some quick wins were identified to foster these possibilities and to ensure data reuse can actually happen in a Linked Open Data context.

Finally, the review identified some current challenges and work-arounds for exporting the (final) result of the project, as stored and contextualised in your datastore.

Colofon

About the authors

Maarten Zeinstra is a strategic information consultant and intellectual property lawyer.

Maarten Brinkerink is an expert in digital culture and information processing.

Licence



This report is licensed under a [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Go to <https://creativecommons.org/licenses/by/4.0/> to obtain a full copy of this licence. You are free to copy, distribute, modify this report, as long as you mention the authors above and link to the licence.

Appendix 1 - User Stories

A user story is an informal, general explanation of a feature written from the perspective of the end user. Its purpose is to articulate how a feature will provide value. User stories are expressed in a simple three part sentence, structured as follows:

As a [user], I want [functionality], so [benefit]

Priority in this analysis is defined as followed:

Must-have: Required to allow Beyond Notability to succeed.

Should-have: Required to allow Beyond Notability to excel.

Could-have: Required to go beyond expectations for Beyond Notability.

Nice-to-have: Not required, but makes Beyond Notability more pleasant to work on.

Epic	As a [user]	I want [functionality]	so [benefit]	Priority	Champion
Engage	Project member	to build relationship with cultural heritage institutions	I can connect national institutions to more local heritage institutions to enhance local sites and institutions and to work with not necessary local professionals	Must-have	Katherine
Engage	Project member	to build relationships with women's history groups	the project can have follow-up research projects	Should-have	Amara
Export	Project member	to export the statements/triples (from Wikibase.cloud)	the data can be preserved for potential reuse post-project	Must-have	James
Export	Funder	BN to export the data in raw form	it can be made available in a academic data repository	Must-have	Katherine
Export	Data scraper	to programmatically retrieve BN data	the data can be reused in other projects (silos)	Should-have	James

Epic	As a [user]	I want [functionality]	so [benefit]	Priority	Champion
Export	Project member	to export the history, discussion, etc (from Wikibase.cloud)	the data can be preserved for potential reuse post-project	Could-have	James
Link	Project member	to link data to cultural heritage institutions	(example) it provides an opportunity to look at 'blind spots' together with cultural heritage organisations	Must-have	Amara
Query	Project member	to ask interesting questions about the data	the project team can better understand women's work in archaeology, history, heritage	Must-have	James
Query	Project member	to use data analysis and visualisation of the entire dataset	the project can generate new insights into large-scale, longitudinal changes in the nature of women's archaeological, historical and heritage work	Must-have	James
Query	Project member	to calculate percentages for external coverage of BN items	the representation of BN subjects in other databases can be gauged at, to see where BN can contribute to an improvement	Should-have	James
Query	End user	to ask interesting questions of the data	people's understanding of what women's work in archaeology, history, heritage can be enhanced (new notions of visibility)	Should-have	Amara
Record	Project member	to quickly enter items, based on the research	the data from the archives of the SAL and the RAI can be modelled and catalogued	Must-have	Ammandeep
Record	Project member	to express uncertainty in information (e.g. datetime)	the (nuanced, conflicting, etc.) findings of the archival research are best represented	Must-have	Katherine
Record	Project member	to develop a means of recording complex and nuanced data points	the complexity of women's lives are appropriately represented in the data and provoke users to think differently about how linked data can be used	Must-have	James
Record	Project member	to record complex and nuanced data points	BN pushes the boundaries of Linked Data in a digital humanities context	Could-have	Amara

Appendix 2 - Useful queries

Some useful queries were developed to research the data in the project. For documentation purposes they are reproduced below.

Properties without Wikidata links

```
SELECT ?property ?propertyLabel ?wikidata WHERE {
  ?property wikibase:statementProperty ?ps .
  FILTER NOT EXISTS {?property bnwdt:P117 ?wikidata. }
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language "en" .
  }
}
```

Fetch all 'Described at' with a Wikidata link

```
PREFIX bnwdt: <https://beyond-notability.wikibase.cloud/prop/direct/>
SELECT ?id ?link
WHERE {
  ?id bnwdt:P14 ?link .
  FILTER regex(?link, ".*?wikidata.*", "i")
  SERVICE wikibase:label { bd:serviceParam wikibase:language
"[AUTO_LANGUAGE],en-gb". }
}
```

Get all birth and death dates with their references from Wikidata

```
PREFIX bnwd: <https://beyond-notability.wikibase.cloud/entity/>
PREFIX bnwds: <https://beyond-notability.wikibase.cloud/entity/statement/>
PREFIX bnwdv: <https://beyond-notability.wikibase.cloud/value/>
PREFIX bnwdt: <https://beyond-notability.wikibase.cloud/prop/direct/>
PREFIX bnp: <https://beyond-notability.wikibase.cloud/prop/>
PREFIX bnps: <https://beyond-notability.wikibase.cloud/prop/statement/>
PREFIX bnpq: <https://beyond-notability.wikibase.cloud/prop/qualifier/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX p: <http://www.wikidata.org/prop/>
PREFIX pr: <http://www.wikidata.org/prop/reference/>
PREFIX prov: <http://www.w3.org/ns/prov#>
```

```
SELECT ?woman ?womanLabel ?WDitem ?WDdateOfBirth
(GROUP_CONCAT(?BirthStatedInEnglish;separator=" | ") AS ?sources)
WHERE {
  #Select all women
  ?woman bnwdt:P3 bnwd:Q3 .
  #select all women that have a Wikidata link
  ?woman bnwdt:P117 ?Qnumber .
```

```

#create reference to Wikibase entity
BIND(IRI(concat("http://www.wikidata.org/entity/", ?Qnumber)) as ?WDitem )
#on Wikibase do
SERVICE <https://query.wikidata.org/sparql> {
  #get the date of birth
  ?WDitem wdt:P569 ?WDdateOfBirth .
  #get the statement of birth
  ?WDitem p:P569 ?Birthstatement .
  #get its node
  ?Birthstatement prov:wasDerivedFrom ?BirthRefnode.
  #get the value of 'Stated in'
  ?BirthRefnode pr:P248 ?BirthStatedIn.
  #get the english label of the referece 'stated in'
  OPTIONAL {?BirthStatedIn rdfs:label ?BirthStatedInEnglish
    FILTER (LANG(?BirthStatedInEnglish) = "en")}.
}

SERVICE wikibase:label { bd:serviceParam wikibase:language
"[AUTO_LANGUAGE],en-gb". }
}
GROUP BY ?woman ?womanLabel ?WDitem ?WDdateOfBirth
ORDER BY ?womanLabel

```