

# Smithsonian Libraries and Archives Wikidata & Wikibase Projects

2021 LD4



July 21 2021



[bit.ly/2021LD4SMI0721](https://bit.ly/2021LD4SMI0721)



Smithsonian  
*Libraries and Archives*

# How do we do this?

GLAMs - Modeling out data from the *Artists Files*  
to the Zs: *Zebra Genomes*, 中國人 (*Zhongguo ren*), *Zulu Kingdom*, in the Smithsonian Institution



Smithsonian  
*Libraries and Archives*

# Who we are



**Jackie Shieh**

Librarian  
[\*ShiehJ\*](#)



**Anne Evenhaugen**

Librarian  
[\*aceven\*](#)



**Kathryn Phillips**

Librarian  
[\*JuliaAgricola\*](#)



**Diane Shaw**

Special Collections Cataloger  
[\*Uncommon fritillary\*](#)



**Heidy Berthoud**

Head of Resource Description  
[\*Berthoudh\*](#)



**Richard Naples**

Technical Information Specialist  
[\*Drastician\*](#)



**Suzanne Pilsk**

Librarian  
[\*PilskS\*](#)

*...plus many other staff members at the  
Smithsonian Libraries and Archives!*



**Smithsonian**  
*Libraries and Archives*

# Topics that we will cover today:

- Project genesis: Smithsonian and its world of data
- Why Wikidata platform
- Our PCC Projects
- Wikidata challenges and solutions
- Wikibase journey

# How did we get started?

- Authority control need
- Starting point
  - Artists files—Libraries and museums
  - Non-Latin collections
  - Research outputs



Smithsonian  
*Libraries and Archives*

# PCC Wikidata Smithsonian Projects

- African Ethnic Groups
- Artists' Files
- Chinese Ancestors Portraits
- Dibner Scientific Portraits
- Smithsonian Research Online

*These projects all centered around dealing with **names** (people, organizational units), but also have challenges about **multilingual scripts**, concerns about **ethics**, and questions on how we can make **inferences***

# This is what we did

## Property

- Application & constraint
- Policies
- Best practices
- Ethical question
- Political issue
- SPARQL impact

## Issues and resolutions

- ...

# Properties



# Example 1: Chinese Ancestor Portraits

## What's in a Name?

- family (P53)
- family name (P734)

## Issues and resolutions

- Given name constraints
- Nominal kinship
- Naming conventions for women
- Lost to history

# Example 1: Chinese Ancestor Portraits



Gong Family (Q103537674)



Smithsonian  
*Libraries and Archives*

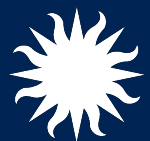
# What would you have done?

## Slack questions

- Inference issue, same clan, not of same family name
- Women's name, <wife of>, <Mrs. \*\*\*>



# Sex or Gender



Smithsonian  
*Libraries and Archives*

# Example 2: Sex or Gender

## Sex or gender

- ...

## Issues and resolutions

- ...



# What would you have done?

## Slack questions

- How to get around the problem of retrieving an item based on data at birth and publicly available, but the named subject objects to this approach.
- Should the property reflect the latest gender status of the subject for living vs historical person?
- In the absence of known gender, if this is a <core> property, how will you reconcile the requirement and practice?
- How non-human, e.g., Lassie? A <she> character but in reality a <he> actor?!

# Ethnic Groups

# Example 3: Ethnic Groups

## Ethnic group

- ...

## Issues and resolutions

- ...





# What would you have done?

## Slack questions

- How do we decide the name of an ethnic group when it has changed across the history?
- What people call themselves vs. how others call them? That could be also a multilingual concern especially during the colonial period.
- Describing people of variety of ethnic groups, e.g., Qing dynasty from Han ethnic to becoming Manchu ethnic due to honorific promotion; or self-promotion for social status, changing from Christian, Jewish to Muslim, vice versa?

# Floruit (floru-what?!?)



Smithsonian  
*Libraries and Archives*

# Example 4: Active Period of an Artist

## Active period of an artist

- floruit
- work period start & work period end

## Issues and resolutions

- ...

# What would you have done?

## Slack questions

- What is the best practices for recording an active period of an artist, writer, etc.?
- How does the period of “best known period” of an artist aside from the start of their work?
- What is the criteria of an editor to use in terms of the research for adding a property statement?

# Wikibase: an SI-centric Wikidata



Smithsonian  
*Libraries and Archives*

# Wikibase: an SI-centric Wikidata

- What to emulate from Wikidata?
- Data modeling
  - Greater local control over data description and quality, but at the cost of interrelatability?
  - PII concerns
- Documentation: best practices
- System customization
  - Wikidata: Schemas in place of <property—constraint>
  - Hierarchical user account setup
  - UI customization

# What will you do?

- Slack questions
  - Data modeling needs
  - User account for various access/edit needs

# References and Credits

## PCC Wikidata Pilot: Smithsonian Libraries

[wikidata.org/wiki/Wikidata:WikiProject\\_PCC\\_Wikidata\\_Pilot/Smithsonian\\_Libraries](https://wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/Smithsonian_Libraries)

## Wikibase yearly summary 2021

[lehir.net/wikibase-yearly-summary-2021/](https://lehir.net/wikibase-yearly-summary-2021/)

.....Images used ....



Smithsonian  
*Libraries and Archives*



# Thank you!



Smithsonian  
*Libraries*