

Smithsonian Libraries and Archives Wikidata & Wikibase Projects

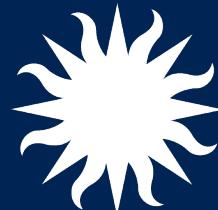
2021 LD4



July 21 2021



bit.ly/2021LD4SMI0721



Smithsonian
Libraries and Archives

*We gratefully acknowledge
the Native Peoples on
whose ancestral
homelands we gather, as
well as the diverse and
vibrant Native
communities who make
their home here today.*

—NMAI Land Acknowledgement

NATIONAL MUSEUM of the AMERICAN INDIAN

Smithsonian

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

Who we are



Jackie Shieh

Descriptive Data Management
Librarian
ShiehJ



Anne Evenhaugen

Head, Art Libraries
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

Head of Metadata Department
Pilks

*...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: It's already here!
- Our PCC Projects & Wikidata: issues and challenges
- Wikibase journey

How did we get started?

- Authority control need
- Starting point
 - Artists files—Libraries and museums
 - Languages and scripts for non-Latin collections
 - Researchers and published outputs



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

Wikidata Properties

- Understanding data model in the Wikidata ecosystem
- Application & constraints
- Smithsonian internal and Wikidata community policies
- Standards, best practices and patterns
- Ethical questions
- Political issues
- Impacts from SPARQL queries



Names

Example 1: Names

- Family (P53)
- Family name (P734)
- Birth name (P1477)
- Temple name (P1785)
- Posthumous name (P1786)
- Given name (P735)

Issues and resolutions

- Naming conventions
- Nominal kinship
- Given name constraints
- Inference

Names: Chinese Ancestor Portraits



Qianlong Emperor (Q19133)



Consort Qian (Q7675445)



Li clan of Longxi (Q15914504)

Gong Family (Q103537674)

Xing (Q61942409) and Shi (Q11128762)

What would you have done?

- How do you model a noble Chinese family? same clan, not of same family name, nominal kinship
- Do given name constraints suggest an appeal for change?
- Is there a way to "build in" inference for family relationships?

Sex or Gender



Example 2: Sex or Gender

Sex or gender (P21)

Issues and resolutions

- Ethical questions
- Conceptualization
- How to operationalize

Example 2: Sex or Gender

Ethical Considerations:

- Is Gender necessary in our Core Data Model?
- Counterpoint: if you can't count it, does it count?
- Informed consent, privacy, and dead people

Example 2: Sex or Gender

The screenshot shows the homepage of the Smithsonian American Women's History Initiative. At the top left is a logo for "BECAUSE OF HER STORY". The main title "Because of Her Story" is prominently displayed. Below it, a subtitle reads: "To create a more equitable and just American society, the Smithsonian American Women's History Initiative will create, disseminate, and amplify the historical record of the accomplishments of American women." A large image of a woman in military uniform is centered, with the text "I shed my camouflage." overlaid. To the right of the image is a sidebar with links to "Main page", "Contents", "Current events", "Random article", "About Wikipedia", "Contact us", "Donate", "Contribute", "Help", "Learn to edit", "Community portal", "Recent changes", "Upload file", "Tools", "What links here", "Related changes", "Special pages", "Permanent link", "Page information", "Print/export", "Download as PDF", "Printable version", "Languages", and "Add links". A "Discover Her Story" button is located at the bottom left.

<https://womenshistory.si.edu/>

The screenshot shows a Wikipedia page titled "Wikipedia:Meetup/DC_30/To-do list". The page header includes links for "Project page", "Talk", "Read", "Edit source", "View history", and "More". The main content area starts with a thank you message for expressing interest in improving coverage related to women in science on Wikipedia. It states that every woman listed on the to-do list is included in the Smithsonian Institution Archives, with a link to an article that needs improvement or creation. A "To-do list" section follows, containing a numbered list of names and their professions, each with a link to their Wikipedia page. On the right side of the page is a small image of Mary Agnes Chase working at a desk.

Project page Talk

Read Edit source View history More Search Wikipedia

Wikipedia:Meetup/DC_30/To-do list

From Wikipedia, the free encyclopedia
< Wikipedia:Meetup | DC 30

Thank you for expressing interest in improving coverage related to women in science on Wikipedia!

Every woman listed on this to-do list is included in the [Smithsonian Institution Archives](#). Here you will find a link to an article that needs improvement or an article that needs to be created, and resources for use.

To-do list [edit source]

- Annie Montague Alexander - Bontanist, needs improvement [1] ↗
- Florence Augusta Merriam Bailey - Ornithologist[2] ↗[3] ↗ [Images on Commons!](#)
- Carole Baldwin - Ichthyologist[4] ↗[5] ↗[6] ↗
- Florence Bascom - Geologist[7] ↗ [Images on Commons!](#)
- Doris Holmes Blake - Entomologist[8] ↗[9] ↗
- Emma Lucy Braun - Botanist[10] ↗[11] ↗
- Denise Breitburg - Marine Ecologist[12] ↗[13] ↗photo[14] ↗
- Janine Brown - Research Scientist[15] ↗
- Nellie A. Brown - Botanist[16] ↗
- Kari Bruwelheide - Forensic Anthropologist, article needs clean-up[17] ↗[18] ↗[19] ↗
- Mary Van Rensselaer Buell - Chemist[20] ↗ [Images on Commons!](#)
- Charlotte Gower Chapman - Anthropologist [21] ↗
- Mary Agnes Chase - Botanist[22] ↗[23] ↗[24] ↗
- Mary Strong Clemens - needs improved inline citations
- Elzada Clover - Botanist, Curator of the University of Michigan Botanical Gardens, One of 2 first women to complete entire run of the Grand Canyon [25] ↗[26] ↗
- Doris Mable Cochran - Herpetologist[http://siarchives.si.edu/collections/siris_arc_217308][27] ↗[28] ↗
- Wimmatte Porter Cockerell - Botanist[29] ↗[30] ↗[31] ↗[32] ↗

Mary Agnes Chase

https://en.wikipedia.org/wiki/Wikipedia:Meetup/DC_30/To-do_list

Example 2: Sex or Gender

Discussion [edit source]

Doesn't the conflation between sex and gender into a single category ruin the entire point of wikidata? [edit source]

It seems obvious to me that sex and gender (and possibly gender identity) should be distinct properties, if they are indeed different things. It'd be like combining a 'height or weight' property. It just... doesn't make much sense. Likewise, the related male/female/man/wo conflated, vague, and unclear. Which again seems like it ruins the entire point of the dataset. If "male" can refer to someone who is bi/sex. Ultimately it seems the intended usage is for gender identity, and then assuming that unless otherwise stated, a person's sex is gender. avoid acknowledging that some people have a gender identity that differs from their sex. But... that's the entire definition of transgender, differs from their sex? Yet instead we see a solution being "transgender male" and "transgender female" which makes the distinction ob/ also see that there's a distinction between human males and non-human males. This doesn't make much sense. It's not like the concep humans vs other animals.

It's obvious that all of this gender identity stuff should have its own property and values, which should ideally be named distinctly and di differentiate them. I think it's obvious that some people are intentionally trying to conflate them, which causes problems with unambiguou attempting to achieve.

Also noteworthy: Transsexualism should be distinct from Gender Identity Disorder. It appears that "transgender" typically refers to peopl don't match this. Instead it seems that a fictitious concept of gender identity is coined and applied to everyone (despite it not being a thi justification for how GID and Gender Dysphoria result, when this is untrue. There is then further attempts to conflate Transgender/GID/C Gender Identity with Sex. And that is what has lead to the confusion and horribly unclear labeling. Each concept should have a distinct & definitions.

Kafke (talk) 11:14, 2 February 2021 (UTC)

Akava'ine - can this be added? [edit source]

Hi all, I was adding someone to Wikidata, who is Akava'ine - which closely corresponds to the Samoan fa'afafine, a third gender in Pacific culture. [Wikidata](#) lets you add fa'afafine, but doesn't recognise Akava'ine as a constraint for gender. Please could it be added? Lajmooore (talk) 15:06, 3 June 2021 (UTC)

https://www.wikidata.org/wiki/Property_talk:P21

Wikidata:WikiProject Properties/Wikidata properties that may violate privacy

< Wikidata:WikiProject Properties

Manually update list

This list is periodically updated by a bot. Manual changes to the list will be removed on the next update!

WDQS | PetScan | TABernacle | Find images | Recent changes | Query: `SELECT ?item ?property ?wikilink ?cnt ?cnt2 ?cnt3 WHERE { { SELECT ?property ?cnt ?cnt2 ?cnt3 WHERE { { SELECT ?property (COUNT(?p) AS ?cnt) WHERE { wd:Q44601380 ^wdt:P8274 ?property . ?property wikibase:claim ?p . [] ?p [] . } GROUP BY ?property } } BIND(IRI(CONCAT('http://www.wikidata.org/entity/Q', SUBSTR(STR(?property), 33))) AS ?item) . BIND(CONCAT('\u007B\u007BP\u007C', SUBSTR(STR(?property), 32), '\u007D\u007D') AS ?wikilink) . } ORDER BY ?property`

#	Property	Number of claims
1	occupation (P106)	8517896
2	employer (P108)	1522254
3	religion (P140)	379906
4	ethnic group (P172)	150739
5	place of birth (P19)	2804912
6	sex or gender (P21)	7291197
7	position held (P39)	1308223
8	educated at (P69)	2154712
9	sexual orientation (P91)	3744
10	medical condition (P1050)	239239
11	phone number (P1329)	127807
12	political alignment (P1387)	3822
13	convicted of (P1399)	9820
14	birth name (P1477)	126232

https://www.wikidata.org/wiki/Wikidata:WikiProject_ProProperties/Wikidata_properties_that_may_violate_privacy

Example 2: Sex or Gender

Conceptualization:

- Conflation of Sex and Gender in one Property
- Cultural and Temporal Specificity of Gender
- Homosaurus?

Example 2: Sex or Gender

Operationalization:

- Mechanisms and Procedures to Collect Gender – VIVO?
- Implications of Sharing Data that Could be Edited

What would you have done?

- Should the Wikidata property “sex or gender” be in our core set of properties? What are the main reasons? Are they enough to justify the privacy issues involved?
- How would you model the data differently than what is in Wikidata? And if so, what would the consequences be?
- How should we weigh the lack of control of our data in the Wikidata ecosystem versus the responsibility of being an “authoritative” data source?

Ethnic Groups

Example 3: Ethnic Groups

ethnic group as **item**
(Q41710)

Zulu people (Q129962)
instance of (P31)
ethnic group (Q41710)

ethnic group as **property**
(P172)

Nelson Mandela (Q8023)
ethnic group (P172)
Thembu tribe (Q1779499)

Example 3: Ethnic groups

- Geography of ethnic groups
 - Location (P276) or Country (P17) and complexity of borders
 - Indigenous to (P2341) is for language, dance, food, culture, etc.
- Contextualizing histories of ethnic groups
- Contextualizing variant names of ethnic groups

Place holder for AKA: images

What would you have done?

- How can we decide the name of an ethnic group when it has changed across history?
- How can we contextualize variant forms of names within the same Wikidata record?
- How can we describe people belonging to a variety of ethnic groups, e.g., Han ethnic becoming Manchu ethnic due to honorific promotion; or self-promotion for social status, changing from Christian, Jewish to Muslim (Al-Andalus)?
- How can we surface colonial relationships in Wikidata? Is this helpful?

Floruit (floru-what?!?)

Example 4: Active Period of an Artist

floruit: /'flôr(y)əwət/ (Oxford Languages)

- Verb: used in conjunction with a specified period or set of dates to indicate when a particular historical figure lived, worked, or was most active. Noun: the period during which a historical figure lived or worked.
“The artist’s floruit was in the early of the 20th century”

floruit: (P1317) (Wikidata)

- date when the person was known to be active or alive, when birth or death not documented

A	B	C	D	E	F	G	H	I	J
lname	firstname	lifedates	nationality	country	atAAPG	atAFA	atANAC	atNMAI	VIAFID
Suk	Latcholassie	1919-2000	Inuit	Canada				NMAI	
boui	Leila	1982-2016		Morocco		AFA			
n	Louis	1868-1913			AAPG				3.2E
n	Margo	ca. 1894-1988			AAPG				4573
ndue	Monday	-1952		Nigeria		AFA			
ment	Nicolas	fl. 15th century			AAPG				
xander	Pat	-1946			AAPG				
ojojotor	Patrick	-1982		Nigeria		AFA			
ctise	Peju	-1975		Nigeria		AFA			
ee	Percy F.	1883-1959			AAPG				
situs	Petrus	ca 1410-1472 or 3			AAPG				
Simitiere	Pierre Eugene	ca 1736-1784			AAPG				9.6E
pe	Pieter	ca. b. 1613 - ca. d. 1653			AAPG				
issa	Raja	-1958		Tunisia		AFA			
enda	Ricci	-1966			AAPG				1.2E
drews	Robert	d. 1856 or 7			AAPG				6.4E
	S. J. (Sunday								
an	Jack)	approximately 1940-		Nigeria		AFA			
s	Sarah C.	ca b. 1839			AAPG				
ophiare	Sidney	-1960		Nigeria		AFA			
ambre	Stephen D.	ca. b. 1841-			AAPG				
oomuje	Tabuno	-1957		Nigeria		AFA			
oade	Uko	-1930		Nigeria		AFA			
xander	W. (William)	-1915			AAPG				
rrett	W. H.	fl 19-20th Century			AAPG				
a	Willem Claesz	ca. b. 1593-1680			AAPG				
g	William	ca. b. 1841			AAPG				
erger	William	b. 1846			AAPG				
ers	William	b. 1840			AAPG				
ter	William H.	ca. b. 1841			AAPG				

Potential options in Wikidata:

- date of birth (P569)
- date of death (P570)
- floruit (P1317)
- work period (start) (P2031)
- work period (end) (P2032)

Issues we are grappling with:

- How *wikidata* defines these properties vs. How *we* define in the field
- Who is right, wikidata or “the field”
- Lack of Consistency of Use
- How do you SPARQL these?

Use case in action:

“I want a list of Black American artists active in the 19th century.”

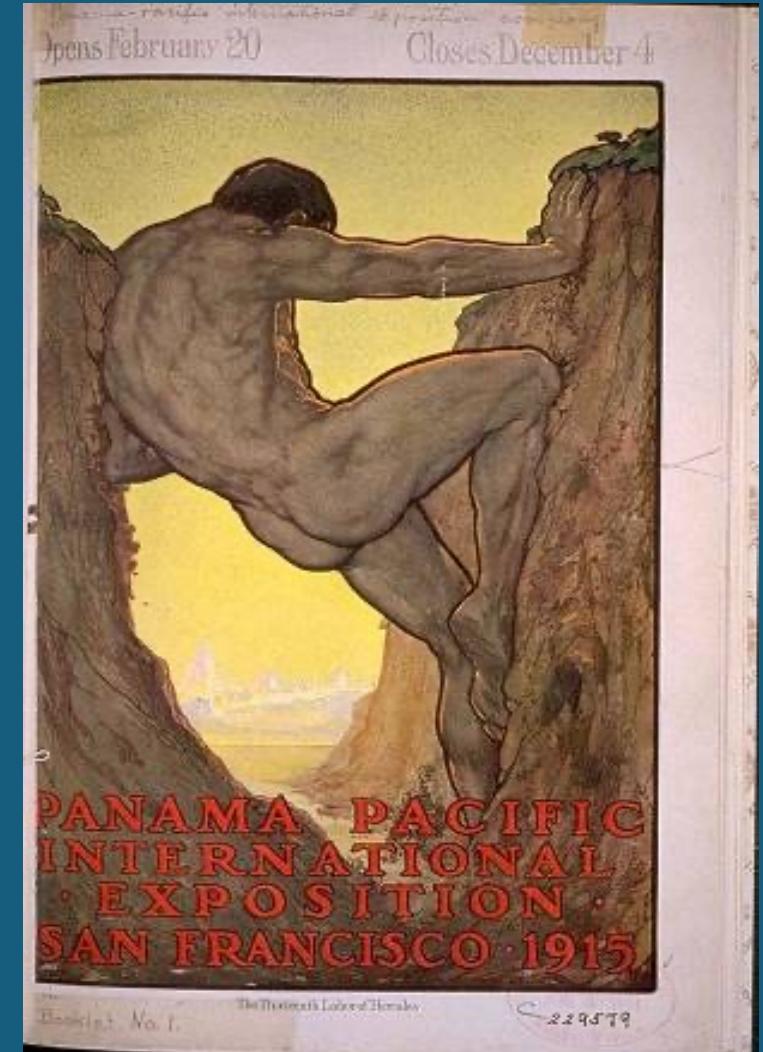
Artists of a certain race and citizenship whose work period started OR ended within the 19th century.

- But because artists born in the 18th century *may* have been active in the desired time period, we can't use birth dates alone.
- The property "floruit" has a specific use: "date when the person was known to be active or alive, when birth or death not documented."

- But wait, what about "work period (start)" and "work period (end)"? They are also indicated for use if they are active for a period.
- Sadly, many artist records lack any work period information, whether "work period (end/start)" or "floruit".

What would you have done?

Ever feel like you are caught between a rock and a hard space?



Wikibase

a decentralized perspective for Wikidata



Wikibase: an SI-centric Wikidata

- What are the descriptions that will matter to an entity?
- How much should be said about them?

Where do we start?

Where do we stop?

Should there be different treatments for opinions and interpretation?

What form should the description take?

Who will do the work?

oclc.org/research/presentations/2019/111319-godby-NISO-what-are-entities-that-matter.html

Wikibase: augmented and enriched with Wikidata

- Independent of, with a federation option to, Wikidata
- Technology stack, potential viable for developing linked data store
- Ability to develop specialized data model
- Local control over data (e.g., determining which claims to remain private)
- System customization, e.g., Wikidata:Schemas in place of <property—constraint>

Installing Wikibase: a reality check

- What is the question: Docker with or without
- APIs: unbundled without Docker
- UI customization for user preferences

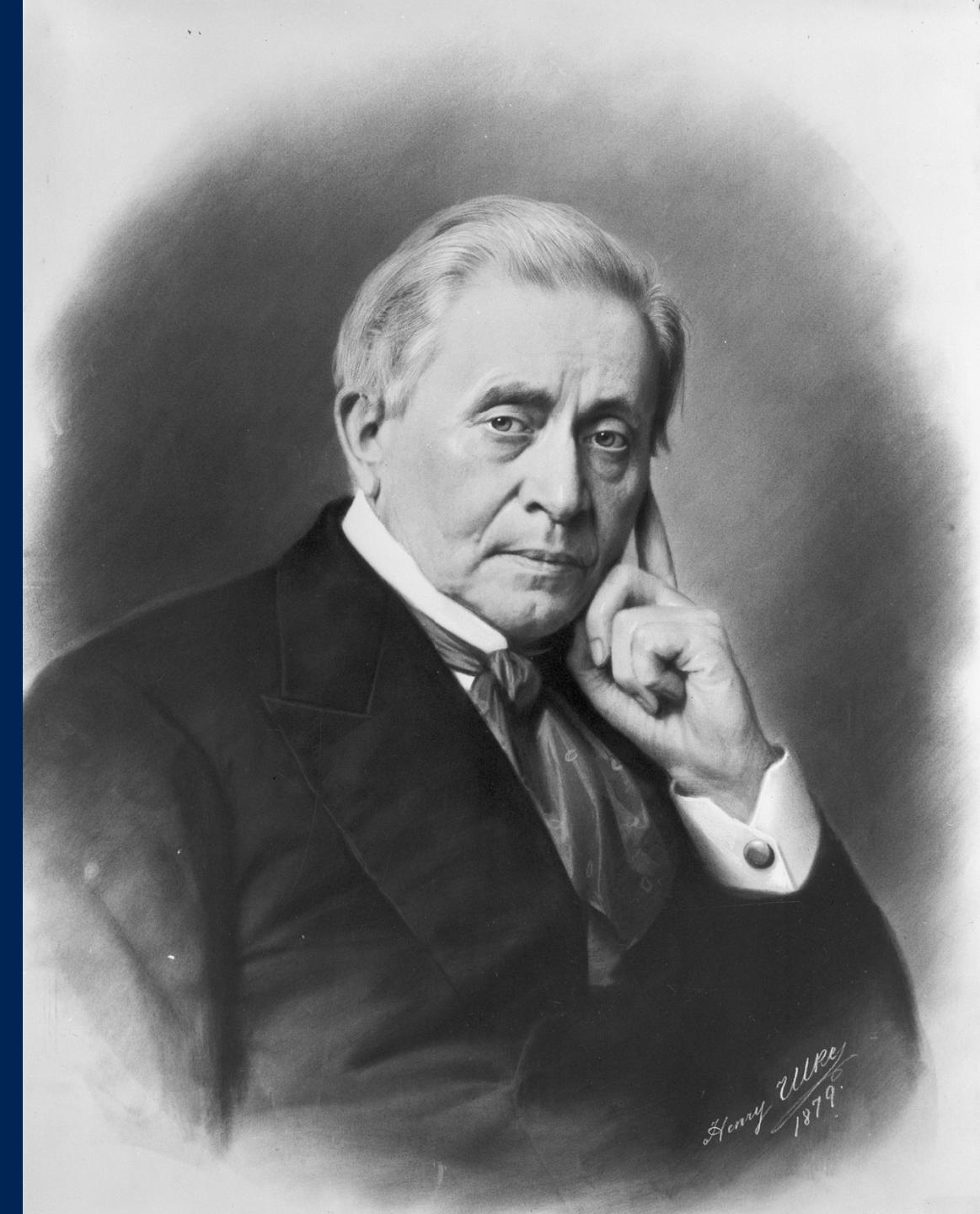
What will you do in your linked data platform?

- What will you do for data modeling needs for SPARQL queries to return the expected and avoid the unexpected?
- What will you do for user accounts to allow various access levels and editing capabilities?
- Are you creating policies and documenting constraints in your linked data project?
- Are you leveraging constraints in other linked data systems when you query? Please share your experience with constraint wins, frustrations, desires

“The worth and importance of the Institution is not to be estimated by what it accumulates within the walls of its building, but by what it **sends forth** to the world.”

—Joseph Henry, 1852

First Secretary of the Smithsonian Institution
(1846-1878)



Our Full Team

Heidy Berthoud, Head of Resource Description. [Berthoudh](#)

Rebecca Bruner, Library Technician, Hirshhorn Museum and Sculpture Gallery Library. [Rebeccabruner](#)

Anne Evenhaugen, Head Librarian, Smithsonian American Art and Portrait Gallery Library. [aceven](#)

Nilda Lopez, Library Technician, Cooper Hewitt Smithsonian Design Library. [NeoNilz](#)

Richard Naples, Technical Information Specialist, Metadata Services. [Drastician](#)

Kathryn Phillips, Librarian for South, Southeast and Middle Eastern Art and Architecture. [JuliaAgricola](#)

Suzanne Pilsk, Head, Metadata Services. [Pilsks](#)

Alexandra Reigle, Librarian, Smithsonian American Art and Portrait Gallery Library. [AReigle](#)

Keala Richard, Conservation Technician, Smithsonian Libraries and Archives Preservation Department. [Km31284](#)

Diane Shaw, Special Collections Cataloger. [Uncommon fritillary](#)

Jackie Shieh, Descriptive Data Management Librarian. [ShiehJ](#)

Mike Smith, Librarian for Korean art, American art, Conservation and Scientific Research, and serials. [78smith](#)

Janet Stanley, African Art Museum Librarian.

Amy Watson, Resource Description.

Yue Shu, Librarian for Chinese language. [Shuyuex0320](#)

References

PCC Wikidata Pilot: Smithsonian Libraries

wikidata.org/wiki/Wikidata:WikiProject_PCC_Wikidata_Pilot/Smithsonian_Libraries

Wikibase yearly summary 2021

lehir.net/wikibase-yearly-summary-2021/

Image Credits



Landsdowne Portrait by Gilbert Stuart (1796) National
Portrait Gallery, Washington D.C.
commons.wikimedia.org/entity/M33579

The Hope Diamond
Photo courtesy of the Smithsonian Institution
www.si.edu/newsdesk/photos/hope-diamond-0

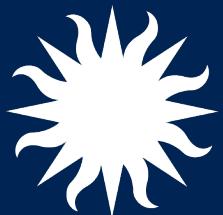
Dorothy's Ruby Slippers
americanhistory.si.edu/press/fact-sheets/ruby-slippers

Mei Xiang (Meghan Murphy photos)
insider.si.edu/2010/04/national-zoos-giant-panda-mei-xiang-is-not-pregnant/

1909 Wright Military Flyer
airandspace.si.edu/collection-objects/1909-wright-military-flyer/nasm_A19120001000

Smithsonian Institution Building (aka the Castle)
commons.wikimedia.org/entity/M96312914

Thank you!



Smithsonian
Libraries and Archives