

# Ethical and Inclusive Data Modeling for Gender Representation in Wikidata

Lessons and Recommendations from the Personal Pronouns Project



Queering Wikipedia Conference  
May 12, 14, and 17, 2023

## Presenters:

Alex Jiyun Jung, University of Toronto Libraries, [jya.jung@utoronto.ca](mailto:jya.jung@utoronto.ca)

Arielle Rodriguez, University of Washington Information School, [rodria7@uw.edu](mailto:rodria7@uw.edu)

Crystal Yragui, University of Washington Libraries, [cec23@uw.edu](mailto:cec23@uw.edu)

## Acknowledgements:

Clair Kronk, maestra for [LGBTdb](#) (Wikibase project) and key collaborator

LD4 Wikidata Affinity Group

Wikimedia Group Discussion Participants on Gender in Wikidata

# Gender in Wikidata: An Overview

# Why Inclusive and Accurate Gender Representation Matters

- Wikidata has great influence and reach
  - Gender information about individuals that is sought by Internet users almost always comes from Wikidata
- Wikimedia content and projects related to gender issues
- “Gender is the most popular yet personal detail..”
- Including gender properties has been and continues to happen
  - Shaping how this is done and creating best practices and documentation can help mitigate or eliminate harm to individuals

[Wikidata:WikiProject LGBT/gender](#)

# Identified Problems & Ethical Considerations

- Used for humans (living and deceased; fictional and non-fictional) and non-human animals
- Conflation of distinct properties in humans
- No clear reference standards and no references needed for value added
- Overreaching bots & Quickstatements
  - Misgendering based on name, appearance, etc.
  - Sometimes inferred from pronouns
- Lacks consent & violates the privacy of living people


# Personal Pronouns Project

# Scope

- Identified as an area of concern which would be relatively simple to tackle
  - It didn't end up being as simple as we assumed!
- Implementation would be limited
  - Incomplete sets of lexemes and Wikidata items for pronouns
  - Small number of existing statements linking items for people to their personal pronoun(s)
  - [Feasibility section of project documentation](#) shows that this is a manageable dataset
- Focused on English language with hope for sparking similar work across various language communities in Wikidata

# Areas of Concern

- [Documentation](#) for Personal Pronouns Project
- [Summary of Concerns](#) about Personal Pronouns in Wikidata

**personal pronoun** (P6553)  edit

personal pronoun(s) this person goes by  
pronoun | preferred pronoun | preferred gender pronoun | personal gender pronoun | gender pronoun | pronouns | personal pronouns | PGP | PGPs

[In more languages](#)

Language	Label	Description	Also known as
English	personal pronoun	personal pronoun(s) this person goes by	pronoun preferred pronoun preferred gender pronoun personal gender pronoun gender pronoun pronouns personal pronouns PGP PGPs
Spanish	pronombre preferido	pronombre personal con el que esta persona prefiere ser aludida	



# Accuracy

- Assigning incorrect information, using it to infer further incorrect information
  - For example, bots using personal pronouns to make P21 “sex or gender” statements about living people
  - A person’s gender is not connected to their personal pronouns

## personal pronouns (Q65067284)

third-person pronouns that an individual wants to be used when referring to them

personal gender pronoun | PGP | preferred gender pronoun | personal gender pronouns | preferred gender pronouns | PGPs

▸ Reconcin: Most relevant properties which are absent

▼ In more languages

Language	Label	Description	Also known as
English	personal pronouns	third-person pronouns that an individual wants to be used when referring to them	personal gender pronoun PGP preferred gender pronoun personal gender pronouns preferred gender pronouns PGPs
Spanish	pronombre preferido	pronombre personal en tercera persona que describe el género de una persona	

## Wikipedia (7 entries) edit [move]

ca [Pronom de preferència](#) [ref]

en [Preferred gender pronoun](#) [ref]

eo [Preferata genra pronomo](#) [ref]

he [כינוי גוף מגדרי](#) [ref]

ru [Предпочитаемые гендерные местоимения](#) [ref]

sr [Префериране заменице](#) [ref]

uk [Особисті гендерні займенники](#) [ref]

[English Wikipedia entry](#) for Wikidata item for [personal pronouns](#). [See talk page](#)

# Consent & Privacy

- Tension between:
  - Honoring identities and feelings and privacy of living people
  - “Quantity over Quality” editing practices
  - Facilitating automation of natural language content for languages which rely on grammatical gender
- Lack of guidance:
  - When to include personal pronouns
  - What constitutes an acceptable reference
  - What to do when pronouns change

# Guidance: How Personal Pronouns Work

## Pronouns-- A How To Guide

**Subject:** 1 laughed at the notion of a gender binary.

**Object:** They tried to convince 2 that asexuality does not exist.

**Possessive:** 3 favorite color is unknown.

**Possessive Pronoun:** The pronoun card is 4.

**Reflexive:** 1 think(s) highly of 5.

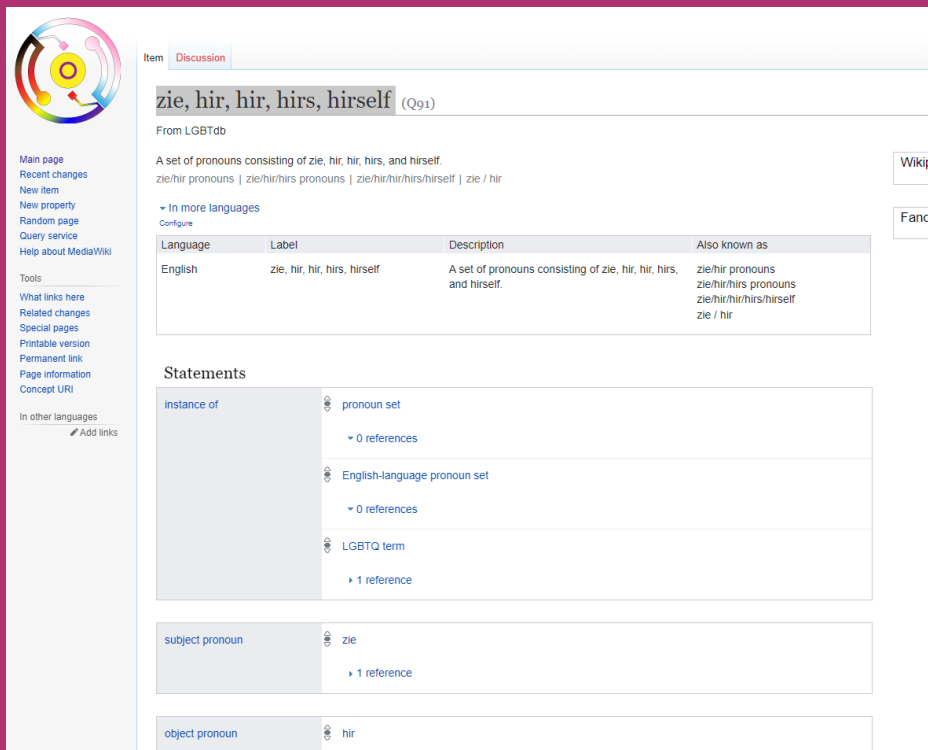
The pronoun list on the reverse is not an exhaustive list. It is good practice to ask which pronouns a person uses.

© 2011, 2016 UW-Milwaukee LGBT Resource Center

1	2	3	4	5
(f)ae	(f)aer	(f)aer	(f)aers	(f)aerself
e/ey	em	eir	eirs	eirself
he	him	his	his	himself
per	per	pers	pers	perself
she	her	her	hers	herself
they	them	their	theirs	themself
ve	ver	vis	vis	verself
xe	xem	xyr	xyrs	xemself
ze/zie	hir	hir	hirs	hirself

# Data Modeling

- Individual pronouns as lexemes and/or items, and pronoun sets as items
  - Individual pronouns as values for P6553 don't account for cases where a principal pronoun isn't enough to go on (for instance, “ze/zir” vs. “ze/hir”)
- How items are related to lexemes is inconsistent
- Example of how personal pronouns can be modeled well:
  - **LGBTDB** by Clair Kronk, “the free LGBTQIA+ database created for and by LGBTQIA+ people”
  - Example: [zie, hir, hir, hirs, himself](#)



The screenshot shows a Wikidata page for the item "zie, hir, hir, hirs, himself" (Q91). The page includes a navigation menu on the left, a description of the item, a table of properties, and a list of statements.

**Item** [Discussion](#)

## zie, hir, hir, hirs, himself (Q91)

From [LGBTdb](#)

A set of pronouns consisting of zie, hir, hir, hirs, and himself.  
[zie/hir pronouns](#) | [zie/hir/hirs pronouns](#) | [zie/hir/hirs/hirself](#) | [zie / hir](#)

[In more languages](#)  
[Configure](#)

Language	Label	Description	Also known as
English	zie, hir, hir, hirs, himself	A set of pronouns consisting of zie, hir, hirs, and himself.	<a href="#">zie/hir pronouns</a> <a href="#">zie/hir/hirs pronouns</a> <a href="#">zie/hir/hirs/hirself</a> <a href="#">zie / hir</a>

### Statements

instance of	<a href="#">pronoun set</a> 0 references
	<a href="#">English-language pronoun set</a> 0 references
	<a href="#">LGBTQ term</a> 1 reference
subject pronoun	<a href="#">zie</a> 1 reference
object pronoun	<a href="#">hir</a>

# Proposed changes

- Disassociate **personal pronouns** from sex and gender (P21)
- Move to **pronoun sets**, as implemented in *LGBTdb* (we thank Clair Kronk!)
  - We are exploring feasibility of updating personal pronouns (P6553)
  - If we can't, we may need to propose a new property
- Require **active consent** to add pronoun set, unless subject has a **Wikipedia page**
- **Ban overreaching bot activity**
- Establish + steward **best practice** for all pronoun additions (a Wikiproject?)

# How to apply pronoun sets

## Steps to implementation

- [Implementation plan here!](#)

## Logic model (i.e. a flowchart, to guide decision-making)

- [Full flowchart \(v2.0\) here!](#)
- A visual, step-by-step guide to our **proposed best practice**
- Given limited time, here are some key ideas and progress (next slide)
- This will stay a living document – Suggestions are always welcome!

# Key ideas + progress

Key ideas	v1.0 (Feb 2023)	v2.0 (Apr 2023)
Account for <b>multiple pronoun sets</b> + variations across pronoun sets	Develop logic on the basis of pronoun sets. Logic handles one pronoun set at a time. Subject with multiple pronoun sets may use them differently for different circumstances. Each set is evaluated individually.	
Comprehensively accommodate <b>diverse use cases</b>	Account separately for fictional personalities and real-life human beings, for whom privacy and sourcing authority work differently. Other differentiating traits include deceased/disappeared and historical personalities	
<b>Require consent</b> for this data to mitigate privacy violation and open data reuse harms	Subject's expression of non-consent prohibits data addition	<b>Stronger criteria for data inclusion.</b> Active consent required unless subject is a public figure (has a <b>Wikipedia page</b> ).
Clear <b>sourcing</b> and account for <b>change over time</b>	Require clear, self-stated sources	Clarify + refine applicable sourcing circumstances and info retrieval date

# Pronouns in Wikidata

ver 2.0 (April 19, 2023)  
Queering Wikipedia Conference 2023

\*PN: Each set of pronouns (e.g. "they, them, their, theirs, themself" as a set) used by or attributable to, the personality.

Repeat flowchart as many times as necessary to account for all sets.

To enter PN values ("Add PN"), add corresponding pronoun sets OR undisclosed by choice (Q117788244).



[Full flowchart here!](#)



# Next steps + how you can help

## Personal pronouns (P6553)

- **Task force** (Wikiproject) to document + implement pronouns best practices
- Logic model/Flowchart will stay a living document – **Suggestions welcome!**
- Continue testing use cases to challenge + improve proposed best practices
- Please contact us if you are keen to help with the above!

## Sex or gender (P21), more broadly

- Email list is now active! We can use it to discuss + collaborate on P21-related work.
- Arielle and Crystal will be hosting a **roundtable discussion** – see you there!

Q&A

# Q&A

Thanks for listening!

We are curious to hear your questions + feedback.

We can be reached at:

- Alex Jiyun Jung, University of Toronto Libraries, [jya.jung@utoronto.ca](mailto:jya.jung@utoronto.ca)
- Arielle Rodriguez, University of Washington Information School, [rodria7@uw.edu](mailto:rodria7@uw.edu)
- Crystal Yragui, University of Washington Libraries, [cec23@uw.edu](mailto:cec23@uw.edu)