## Guide to Enhancing SDSU Faculty Profiles on Wikidata

If you’re reading this Guide, you’ve probably been asked to help add information to faculty profiles by a librarian or a faculty member. Maybe you’ve seen some profiles already and think there’s more helpful information to add. Either way, this Guide will be essential for navigating some of the trickier aspects of profile enhancements.

This Guide does not provide the step-by-step instructions for data entry. They can be found [here](https://docs.google.com/document/d/1QK0MfltVvIkOYeG05CQXINyFpla-gqfKF6IkJUMtxK8/edit?usp=sharing).

### What is Wikidata?

[Wikidata](https://wikidata.org) is a public repository for structured data – like in a database or spreadsheet – that people and institutions around the world use to share important information about anything – people, places, institutions, and so forth.

The Wikimedia Foundation has designed Wikidata to allow for links to other databases or for links between different types of information within Wikidata itself. This “linked data” approach, as you’ll see, will allow you to link data to information about research that a faculty member has conducted with other researchers at other institutions, to give just one example.

When someone wants to retrieve information about a person, place, or thing entered in Wikidata, it’s easy to pull on all other links to create information-rich reports about something.

### Why do we use it for faculty profiles?

We use Wikidata because its method of data storage lets us create profiles on our faculty members from authoritative data found all over the internet. We don’t reenter a lot of information; we link together what’s already there.

### Why are there special terms for Wikidata data entry?

If you’ve done data entry before, you’re familiar with the terms, “record”, “field”, “row”, and “column”. There are comparable terms for Wikidata elements that have slightly different meanings. When you read about Wikidata, authors will talk about “entities”, “properties,” and “objects” or “literals”.

These terms can be confusing at first. They indicate how data are stored in Wikidata. Rather than database tables with records made of rows and columns, Wikidata entries are stored as *statements*, that is, in the form <item>+<property>+<value>. When you add a statement, you choose from a valid list of items (person, place, thing) (e.g., Dr. Anthony Fauci) and then properties (Place of Birth) and then add the valid value (Brooklyn, NY). That statement gets saved to a giant heap of statements in Wikidata. This method makes linking and retrieving different types of data easier.

**Important**: things in Wikidata have unique codes that start with “Q” (**Qnum**), and properties have unique codes that start with “P” (**Pnum**). That’s because multiple distinct things may use the same name, like “Anthony Fauci.” The same goes for properties, like “member.” You’ll occasionally encounter this situation with faculty names.

For different entities with the same name (referred to in Wikidata as a **label**), it’s usually easy to select the right Q code profile record by looking at the “label” and “description” values for the individual, which are required for all Wikidata entries. For instance, all SDSU profiles created by the automation tool use the description, “academic”, so if you searched Wikidata on the name, “Roberto Hernandez,” who is a SDSU faculty member in Chicana Studies, you’ll find many results. However, only the [Roberto Hernandez you’re looking for](https://www.wikidata.org/wiki/Q112194437) has the description, “academic.”

 Other Preparatory Reading

1. For more discussion of Wikidata elements see [Wikidata:Introduction - Wikidata](https://www.wikidata.org/wiki/Wikidata%3AIntroduction)

### What You Will Need Before You Start

1. [A Wikidata account](https://www.wikidata.org/w/index.php?title=Special:CreateAccount&returnto=Wikidata%3AMain+Page). It’s free and easy to set up.
2. A list of SDSU faculty who need profile enhancement ([here’s an example](https://docs.google.com/spreadsheets/d/1hi6GHCnK2R7WWLBHIH8lKJ81xk4AXXbmRonMsrsKKvE/edit?usp=sharing) from the College of Arts and Letters). You obtain one of these lists from the Cataloging and Metadata Strategies Librarian or from a task manager working with the librarian.
3. The allowable [profile properties list](https://docs.google.com/spreadsheets/d/1-cBO1rqnusTD5WRbVSrkM8ZjIitwzW965s6UenMN6vI/edit?usp=sharing). Make sure to check with the librarian to ensure you have the most recent official list.
4. The [process steps](https://docs.google.com/document/d/1QK0MfltVvIkOYeG05CQXINyFpla-gqfKF6IkJUMtxK8/edit?usp=sharing) for basic faculty profile enhancement.

### Getting Started

The faculty profile enhancement process is made easier by some computer automation that creates the basic entries for the faculty member before manual enhancement starts. The spreadsheet of faculty names you receive will already have links to the basic profile entry for each faculty.

Once you have the Wikidata page loaded, scroll down to the “official website” entry, which will take you to the SDSU faculty profile web page for that faculty member. Much of your enhancement information will come from this web page – specifically, information that’s hard or impossible for our automated tool to extract.

Sometimes, the faculty webpage is less helpful than you expect, and you have to use other sources of information for the [allowed enhancement properties](https://docs.google.com/spreadsheets/d/1-cBO1rqnusTD5WRbVSrkM8ZjIitwzW965s6UenMN6vI/edit?usp=sharing). These alternate sources are discussed below.

### Using the SDSU Profile Property List

Your goal as a profile cataloger is to ***enhance*** the profile, not complete the profile. You should enter as many easy-to-find properties from [the master list](https://docs.google.com/spreadsheets/d/1-cBO1rqnusTD5WRbVSrkM8ZjIitwzW965s6UenMN6vI/edit?usp=sharing) as possible in a reasonable amount of time. For instance, if you can’t find “notable work” for a faculty member on Wikidata, simply move on to the next property.

Based on our prior enhancement work, we provide the property groups below according to how frequently you’ll find values for them. Note that all the required properties are typically added beforehand by the automated tool, though you may find additional data to add for some of them, especially occupation, field of work, and affiliation (see the section on these below).

#### Required (automated entry)

* Label (basic info)
* Description (basic info)
* instance of (P31)
* occupation (P106)
* employer (P108)
* field of work (P101)
* affiliation (P1416)
* official website (P856)

#### Common (manual entry)

* given name (P735) **\***
* family name (P734) **\***
* notable work (P800)
* field of work (P101) (additional)
* educated at (P69) **\***
* educated at (P69) subproperty: academic degree (P512) **\***
* educated at (P69) subproperty: end time (P582) **\***
* ORCID iD (P496)
* Library of Congress authority ID (P244)
* VIAF ID (P1687)
* Google Scholar author ID (P1960)

**\*** *found nearly all the time*

#### Occasional (manual entry)

* member of (P463)
* award received (P166)
* ResearchGate ID (2038)

The remaining properties on the official list are rare for SDSU faculty. It is a good practice to enter them only if you discover them in the course of reading a source for more commonly found properties. It’s fairly common, for example, for faculty to put a link to their personal website on their SDSU biography webpage. These personal sites usually offer a rich source of additional information. **Note**: if you do encounter a personal website for a faculty member, add that URL to a second “official website” statement in Wikidata, and use the “rank” control to set the personal website as the “preferred” official website. Using sources besides the official faculty webpage takes you into a gray area: you must use your judgment to decide if a particular element – say, “educated at” – is important enough to retrieve from an alternative source that may or may not have the needed information (see the section on alternate sources below).

### Important Variations in Enhancement Workflow

When enhancing faculty profiles you will sometimes find that you can obtain the enhancement data you need from the faculty profile webpage and enter it without complications into Wikidata. However, it will often be the case that you will need to do extra work in Wikidata to successfully enter key enhancement values. The sections below provide guidance on how to deal with some of the more complex data management tasks you may need to perform to complete your enhancement.

#### Given Name and Family Name

Entering values for names can be complicated by given or family names that haven’t yet been entered into Wikidata. In Wikidata users can only enter values for types of things that have a pre-existing entry. For example, the family name Gonzalez-Rivera did not exist in Wikidata prior to our cataloger entering it so he could complete the enhancement record for the professor Victoria Gonzalez-Rivera.

#### Handling Hyphenated Names

Wikidata does not provide a way to indicate if a name is hyphenated or not, so one must 1) create a new family name entry for the entire hyphenated name, or 2) add two family name entries in the profile and indicate which name comes first by using the “series ordinal” subproperty. When in doubt use the second option. This is an imperfect solution but is common for hyphenated names, since they tend to be an outgrowth of particular family circumstances rather than a multi-generational name that happens to be hyphenated.

Here is some contextual information for name hyphenation practices in various cultures:

***Chinese***: Given names are sometimes represented with two characters. When names like this are Romanized, a hyphen is sometimes used between name parts, such as Ming-Hsiang.

***North American/European***: People getting married sometimes attach their spouse’s family name to their own family name with a hyphen to create a compound family name. This name is sometimes given to the children of the couple as well.

***Hispanic***: Because Hispanic family names are often made up of one’s father’s and one’s mother’s family name, Hispanic people emigrating to North American often hyphenate their family names together to avoid confusion with North American middle names. Most Hispanic cultures don’t use middle names.

You will also encounter single given names and family names that do not exist in Wikidata. Before you create the profile entry for a professor’s name(s), you’ll have to [create new name items first](https://www.wikidata.org/wiki/Special%3ANewItem). You will be required to add an “instance of” statement (e.g. “family name”), “native label” statement (e.g., “Hernandez”), and a writing system statement (“Latin script”). The “native label” statement requires that you provide a language of origin for names; it may be clear to you that the name is, for example, of Spanish origin or Chinese origin, etc. Assigning foreign language origins is difficult unless you are familiar with that particular language. In these cases we recommend applying the “multiple languages (Q20923490)” value for this sub-property. This indicates that the name is used in multiple language contexts, e.g. Chinese and English.

Once you enter the name entity data, you can then complete your entries in the profile.

#### Occupation, Field of Work, and Affiliation

Occupation, Field of Work, and Affiliation statements are all created automatically prior to enhancement. However, entering additional statements using these properties is sometimes necessary.

The automated tool enters “researcher” and “university teacher” as occupation statements. If you encounter descriptions of other current positions, such as “director” or “dean”, enter them as occupation statements. Past occupations do not need to be entered.

A field of work statement is automatically entered for each academic that corresponds to their department affiliation, e.g. an “anthropology” field of work for everyone belonging to the SDSU Department of Anthropology. Any relevant specialties such as “medical anthropology” or “women’s studies” should also be entered as a field of work statement.

New affiliation statements are rare, but sometimes an academic will be affiliated with more than one department. Note that there is no “department” concept in the Wikidata system. Rather, department entities and name values are recorded using the “affiliation” property.

#### Education

All education data should be entered within the “educated at” property. “Educated at” stores the university where the academic earned their highest degree, i.e., PhD, EdD, MD, MS, MA, etc. Enter the actual degree distinction and the year received by using the “academic degree” and “end time” sub-properties.

Occasionally, the academic’s biography won’t list the university, just the degree. In these cases, if the university information can’t be found using an alternate information source (see below), use the “academic degree” property to enter the degree data.

#### Notable Work

The notable work property links to scholarly work that is already stored in Wikidata. Linking notable work to the academic’s profile has high value because it associates biographical data with the academic’s principal professional output. It can become a way to showcase someone’s best or most influential work, though determining the noteworthiness of a given article can be difficult (looking up the total citations for a paper on Google Scholar is one way). Generally speaking, you should enter up to three scholarly works for an academic if that many are available. Often, there will be no entries to add to notable work for a given academic. In this case, simply move on to the next property.

The best way to find scholarly work on Wikidata is to enter the academic’s name in the search bar and then select “search for pages containing [scholar’s name]” from the popup list. Scholarship entries usually will be at the top of the results list. Sometimes you may have to click back to a second or third results page, but that is rare. Once you’ve located some articles, add them to the academic profile page using notable work as described in the [data entry document](https://docs.google.com/document/d/1QK0MfltVvIkOYeG05CQXINyFpla-gqfKF6IkJUMtxK8/edit?usp=sharing).

#### External Name Authority and Profile Identifiers

Name Authority Identifiers are a way of definitively identifying someone via cross reference. The principal name authority data stored in Wikidata profiles is the Library of Congress Name Authority ID. If the automation tool has not already entered this ID, search for it on the [Library of Congress website](https://id.loc.gov/authorities/names.html). Most faculty have a LOC ID, but not all. You should also add the VIAF authority ID found at VIAF.org if there is one.

Profile identifiers in Wikidata link an academic’s Wikidata profile to external sources. The SDSU Library is conducting an ongoing effort to have researchers create profiles in the [ORCID academic database](https://orcid.org/). Wikidata catalogers can then link an academic profile to ORCID article entries using the academic’s ORCID ID.

However, you will find that some academics don’t yet have an ORCID profile. Luckily, there are a number of other sites that house academic works that should also be linked to the Wikidata profile if they are available. These include the Google Author ID, the Academia.edu profile URL, and the ResearchGate profile ID.

The Google Scholar ID is somewhat complex to capture.

1. Go to the [Google Scholar site](https://scholar.google.com/schhp?hl=en&as_sdt=0,38) and search on the academic’s name.
2. If they have a profile page, it will show up at the top of the search results; or, article results will have the academic’s name hyperlinked. Click on the hyperlink to go to the profile.
3. Look at the URL in the address bar. It will contain a section that starts with “user=”. Copy the characters that come after user=. DON’T copy any text that follows the ID value, if any exists. Once you reach an “&” character, stop selecting and copy the text between “=” and “&”.
4. Add a Google Scholar Author ID statement to the Wikidata profile and paste the code in the value text field.

For Academia.edu, you will have to navigate to [the base site](https://www.academia.edu/) and then click on a focus area of the day, such as “History.” The results page will give you the search bar you need to find the academic in question. If you find the academic’s profile, copy the entire URL address string to use as the value in Wikidata.

For [ResearchGate](https://www.researchgate.net/), scroll down the landing page to find the search bar. If you find a profile, copy only the text in the profile URL that comes after the last “/”. This text is usually in the form, [given name]-[family name].

#### Duplicate profile statements

Occasionally, you will find a second Wikidata profile for an academic, probably created separately by the academic or by someone working with them. You will most likely uncover these during your search for notable work on Wikidata. In this event, use Wikidata’s [Merge two Items page](https://www.wikidata.org/wiki/Special%3AMergeItems) to merge the second record into the SDSU record, per Library policy. Merging items is difficult to undo, so make sure to follow these steps carefully.

1. Review both profile pages to ensure you know which one is the SDSU profile. The best way to determine this is to look at the description. If it says “academic,” it’s an SDSU item. If it’s still unclear which is from SDSU, click the [Item History link](https://www.wikidata.org/w/index.php?title=Q112129156&action=history) at the top of the page and look for entries by GretaHengbot18, which is the account used by the automated tool.
2. Copy down the Qnums of the two profiles you want to merge.
3. Enter the non-SDSU Qnum into the ID to Merge From field.
4. Enter the SDSU Qnum into the ID to Merge To field.
5. Click Merge Items
6. You will now find the merged record under the SDSU profile entry. Note that duplicate values can occur; review the profile for these and delete any exact duplicates.

#### Alternate Information Sources

You will find that many of the common and occasionally found properties aren’t on the faculty member’s webpage, but in other sources such as personal websites mentioned above. It’s possible to find key data, such as education history, in MARC records at VIAF.org (as in [this example](https://viaf.org/processed/LC%7Cn%20%2000109505); examine the fourth 670 entry line), academic profiles at Academia.edu, Library of Congress, Google Scholar, or ResearchGate. Entering statements for external academic profiles themselves can become very important if no notable works are entered in Wikidata for a given academic.

If these sources don’t yield any enhancement values to add, it’s worth trying to search using an internet search engine and see if you can quickly discover relevant information. However, you should only allow a minute or two to find enhancement values, since internet searching frequently yields lower value information.

Ultimately you will have to use your own judgment on when to keep searching. Ask yourself: do I already have several enhancements added to the profile; Am I spending much longer than my average time to complete a profile; have I spent several minutes looking for a hard-to-find property, such as “student of?” If the answer to any of these questions is “yes,” it’s best to move on to the next profile.