

Empowering Sustainability Through Web-Based Information Sharing

Showcasing Campus Sustainability Initiatives Through Linked Data on the Web

The San Diego State University (SDSU) Library recently undertook a collaborative linked data project aimed to improve the discoverability of the university's sustainability efforts by integrating building data and sustainability features into Wikidata, an open-access knowledge base. Data were collected from the SDSU Facilities Department and the Office of Energy and Sustainability (OE&S) to create and enhance Wikidata entries for 119 campus buildings. This involved structuring unstructured data, developing a Wikidata data model, and linking building information to Google Maps. The information was then integrated into Wikidata, making it easier for researchers, students, and the public to explore and learn about sustainability initiatives and campus infrastructure.

In essence, the project improved the discoverability of SDSU's sustainability efforts and promoted data standardization and transparency to allow users to explore sustainability initiatives on campus. While challenges arose from inconsistent data formats and a lack of detailed information from OE&S, highlighting the need for standardized data management practices, the result is improved data transparency and a broader understanding of sustainability by allowing users to interact with building information and sustainability initiatives.

Linked Data and the Wikidata Platform

Linked data is a method for structuring and connecting information on the web to facilitate data sharing and improve data reliability and discoverability, whereas Wikidata is a collaborative, open-access knowledge base that stores structured data and information for diverse subjects.¹ Information on Wikidata is easily accessible and can be queried, analyzed, and visualized in various ways. Its extensive adoption within the Galleries, Libraries, Archives, and Museums (GLAM) community, along with its ability to facilitate wide data sharing and easy reuse by the global community, makes it an ideal platform for this project.² This accessibility allows for the development of tools and applications that can leverage building and campus data for research, decision-making, and public awareness.

Our library's Content Organization and Management team initiated a Wikidata institutional data project during the COVID-19 pandemic, laying the groundwork by creating

Sarah K. Tribelhorn is science and sustainability librarian at San Diego State University, email: stribelhorn@sdsu.edu. Greta Heng is cataloging and metadata strategies librarian at San Diego State University, email: gheng@sdsu.edu.

preliminary data for select SDSU buildings and a map visualization.³ This linked data project expanded on its previous scope by integrating sustainability efforts into the dataset, thereby providing a comprehensive resource for promoting transparency and sustainable practices across the SDSU campus.

Sustainability at CSU, SDSU, and the SDSU Library

This work is in line with the California State University (CSU) Sustainability Policy,⁴ to promote “the environmental sustainability of CSU’s operations” through awareness and actions, as well as the SDSU Senate Committee Policy on Sustainability,⁵ since SDSU is working toward achieving the goals of the 2017 Climate Action Plan (CAP) and working toward achieving carbon neutrality by 2040.⁶ This further ties into the role of SDSU being a leader in sustainability efforts across all campuses. This also aligns with the SDSU Library strategic plan, stating that this library “is an essential partner with campus colleagues in curricular and co-curricular endeavors.”⁷

Examples of sustainability efforts on campus include but are not limited to the establishment of the dedicated OE&S, fifteen Leadership in Energy and Environmental Design (LEED)-certified buildings, three on-campus gardens that grow produce for campus dining, more than 100 hydration stations, composting areas, recycling areas, bicycle lockers and racks, and a commuter hub including trolley and bus stops in the middle of campus. This does not include sustainability research projects that could also be included.

Sustainability and the Library

The SDSU Library is enrolled in the Sustainable Libraries Certification Program (SLCP) to benchmark sustainability.⁸ As part of this process, it is necessary for the library to illustrate collaborations with different campus partners on sustainability initiatives, so this project ties in well with these goals. The library is fundamental for this work with linked data, since it is core to organizing and managing information, and our expertise is fundamental for ensuring high-quality and accurate data. In addition, we understand the needs of users, and this will enhance the user experience in finding information on sustainability projects to enhance the awareness of and participation in these sustainability programs. As part of the SLCP, this project provided the library with relevant information about on-campus sustainability efforts and highlighted potential campus partnerships and collaborations to ultimately strengthen sustainability at SDSU.

Data Collection and Methods

The primary objective of the data collection effort was to compile comprehensive information on the sustainability features of SDSU buildings. Two main data sources have been used: the SDSU Facilities Department and OE&S.

The dataset from the Facilities Department includes structured information on 294 buildings, such as building codes, names (including alternative names), architectural style, architect, structural area, and other relevant details. The dataset comprises 69 data fields (columns)⁹ and is presented in a structured tabular format (CSV).

To link buildings to their Google Map locations, one student assistant was hired to manually search the building’s Google Maps Customer ID¹⁰ and add that information to the building dataset. GMBeverywhere chrome extension¹¹ was used for the ID search purpose.

Given that SDSU operates multiple campuses, this project focused exclusively on buildings at the main campus. The buildings were further categorized by campus location, with 119 buildings on the main campus selected for inclusion in this project.

In contrast, the data from OE&S is largely unstructured, comprising various forms such as graphs, reports, initiatives, and articles. Although this source includes critical information on sustainability features, it also contains unrelated content, such as general sustainability initiatives that do not pertain specifically to campus buildings. This project has only selected buildings with LEED certificates, buildings with green restaurants, and information related to OE&S and the SDSU Annual Sustainability Summit so far. By the time this article is published, no comprehensive dataset exists that consolidates all SDSU sustainability projects.

To describe buildings and their sustainable features on Wikidata, the following approaches were implemented: (1) data cleaning and filtering out irrelevant information, (2) building a Wikidata data model, and (3) creating or enhancing Wikidata entries for SDSU buildings, complete with their associated sustainability features.

Given the unstructured nature of the data from OE&S, a critical preliminary step was to isolate information specifically relevant to the sustainability features of SDSU buildings. For the structured building data, an evaluation was conducted to assess the relevance of each header, removing columns that contained sensitive or non-publicly relevant information, such as floor and room details. Cross-referencing entries with the structured dataset from the Facilities Department helped ensure alignment and accuracy, focusing only on the buildings listed therein.

Once the relevant data was identified, a standardized data model was constructed to effectively represent both building information and sustainability attributes on Wikidata. This model¹² was developed based on the previously published data model in 2021. For statistical purposes, a project ID¹³ was created and assigned to each building using the property 'on focus list of Wikimedia project' (P5008),¹⁴ which was created and enhanced by the team.

The final step involved creating or updating Wikidata entries for each SDSU building. To prevent duplicate entries, an initial duplication check was conducted. This process involved using SPARQL (acronym for SPARQL Protocol and RDF Query Language) queries to retrieve existing entries for SDSU buildings on Wikidata, followed by a manual comparison to identify Wikidata IDs of buildings already described on Wikidata. Those Wikidata IDs were then added to the dataset. Finally, QuickStatements¹⁵ was used to batch-create new entries or enhance existing ones for SDSU main campus buildings on Wikidata.

Building Entries on Wikidata

A total of 119 Wikidata entries were created and enhanced as part of this project. Figure 1 shows a partial screenshot of the buildings included in the project, with geo-coordinates added. The map was dynamically generated by Wikidata using a SPARQL query to identify all entities linked to the project through the 'on focus list of Wikimedia project' (P5008) statement. Each building is represented as a small red dot. Once clicked, the red dots can link to other Wikidata items via one or more statements, enabling users to explore additional related data. This interactive map offers an intuitive and visually engaging way to navigate SDSU building data, enhancing accessibility and illustrating the relationships among the buildings within the project. Figures 2, 3, and 4 show an example of Wikidata for a building with a LEED Gold certificate.

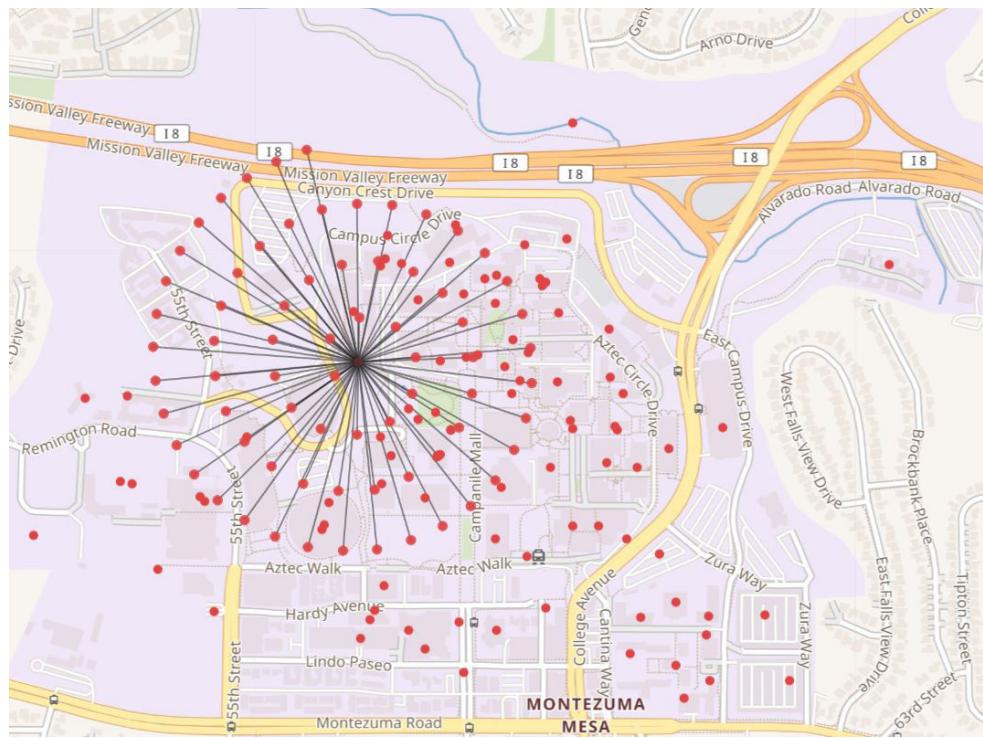


Figure 1. Partial screenshot of the buildings included in the project, with geo-coordinates added.

⧼A English ⓘ Not logged in Talk Contributions Create account Log in

Item [Discussion](#) Read [View history](#) Search Wikidata [edit](#)

Aztec Aquaplex (Q4832935)

Building of San Diego State University; outdoor swimming pool

▼ In more languages [Configure](#)

Language	Label	Description	Also known as
English	Aztec Aquaplex	Building of San Diego State University; outdoor swimming pool	
Spanish	No label defined	No description defined	
Traditional Chinese	No label defined	No description defined	
Chinese	No label defined	No description defined	

[All entered languages](#)

Statements

instance of	<ul style="list-style-type: none"> swimming center edit + add reference
	<ul style="list-style-type: none"> university building edit + add reference + add value
inception	<ul style="list-style-type: none"> 2007 edit + 1 reference + add value
country	<ul style="list-style-type: none"> United States of America edit + 0 references + add reference + add value
located in the administrative territorial entity	<ul style="list-style-type: none"> San Diego edit + 1 reference + add value

Figure 2. Screenshot of the Wikidata information of the Aztec Aquaplex, Wikidata ID Q4832935, accessed on December 13, 2024.

coordinate location

owned by

operator

date of official opening

occupant

on focus list of Wikimedia project

has certification

Figure 3. Screenshot of the Wikidata information of the Aztec Aquaplex, Wikidata ID Q4832935, accessed on December 13, 2024.

Identifiers

Freebase ID

Google Maps Customer ID

+ add statement

Figure 4. Screenshot of Aztec Aquaplex, Wikidata ID Q4832935, accessed on December 13, 2024.

Most information from the campus Facilities and OE&S can be translated into structured data. The screenshots illustrate the information added about the Aztec Aquaplex in Wikidata. Its type has been categorized as both an academic building and a swimming center. Location details, including geographic coordinates, the county, and the country where the building is situated, are also included. Aquaplex's relationship with SDSU is specified, identifying the university as both its owner and operator. Additionally, time-related details, such as the year of inception and the opening date, have been recorded. The building's certification, specifically its LEED status, is also documented. Finally, identifiers for the building, including its Freebase ID and Google Maps customer ID, are linked, providing direct access to its Freebase page and Google Maps page.

Discussion

Although the team successfully added most building-related information to Wikidata, the process revealed several significant challenges. A primary issue arose from the nature of the data provided by OE&S, which often consisted of free-text entries that lacked specificity and detail. For example, while the office supplied data on the locations of electric vehicle charging stations, the information did not specify their power sources. This lack of granularity made it difficult to convert the data into structured formats and identify appropriate Wikidata properties. Free-text entries further compounded these issues by requiring extensive interpretation and manual processing to extract meaningful, structured information.

Moreover, the absence of centralized data management for building-related information posed additional obstacles. The data provided to the team came in a variety of inconsistent formats—such as brochures, maps, and spreadsheets with descriptive details—making it challenging to integrate. Sustainability data at SDSU appeared to follow inconsistent recording and formatting practices, leading to fragmented and unstandardized datasets. These challenges underscore the need for more consistent data management practices and standardized recording procedures to improve sustainability-related data management at SDSU. Implementing such measures would significantly improve the quality and usability of the data.

Conclusion

By linking building data to Google Maps and integrating it into Wikidata, we made it easier for researchers, students, and the public to explore and learn about sustainability initiatives, architectural heritage, and campus infrastructure. With dynamic visualization through SPARQL, users can interact with the data in meaningful ways. This integration improves data transparency and standardization while also fostering a broader understanding of sustainability's impact in higher education and beyond. Additionally, it enables the development of innovative applications and tools that can support informed decision-making and promote sustainable practices in the construction and management of educational facilities. *»*

Notes

1. Theo Van Veen, “Wikidata: From ‘an’ Identifier to ‘the’ Identifier,” *Information Technology and Libraries* 38, no. 2 (2019): 72–81, <https://doi.org/10.6017/ital.v38i2.10886>.
2. Lihong Zhu, Amanda Xu, Sai Deng, Greta Heng, and Xiaoli Li, “Entity management using Wikidata for cultural heritage information.” *Cataloging & Classification Quarterly* 61, no. 1 (2023): 20–46, <https://doi.org/10.1080/01639374.2023.2188338>.

3. <https://w.wiki/3Xov>.
4. "Sustainability Policy," California State University, May 12, 2022, <https://calstate.policystat.com/policy/11699668/latest/>.
5. "Policy File AY 2023-2024," San Diego State University Senate, 2023, https://senate.sdsu.edu/06_policy-file/2023-08-25_policy-file.pdf.
6. "Climate Action," San Diego State University, 2024, <https://sustainable.sdsu.edu/climate-action>.
7. "Transcending borders: The SDSU Library Strategic Plan, 2022–2025," SDSU University Library, 2022, <https://library.sdsu.edu/about/strategic-plan/>.
8. "What is SLCP?" Sustainable Libraries Initiative, 2018, <https://sustainablelibrariesinitiative.org/about-us/program-faq>.
9. The 69 data fields are: Building_Key, Building_Name, Official/Long, Abbrev_Short, Other_Names, SFDB_Code, SFDB_Number, Facility_Code, Architect, Style, Region, Location, Affiliation, Category, Primary_Use, UBC_Code, Planning, Condition, Operations, Ownership, Financing, Address, City, County, State, Zip_Code, Address_Code, City_Code, County_Code, Floors, Height, Footprint, Perimeter, Total_Rooms, Assignable, Asgn_Spaces, Non-Assignable, Non-Asgn_Spaces, Rentable_SF, Net_Usable, Circulation, Custodial, Mechanical, Parking, Toilet, Special_Area, Basic_Gross, C/U_Gross, Outside_Gross, Outside_Gross_50, Structural_Area, Unrelated_Gross, Related_Gross, Maintained_Gross, Janitorized, Constructed, Occupied, Last_Reno, Vacated, Demolished, Construction, Capitalized, Replacement, X_Coordinate, Y_Coordinate, Latitude, Longitude, Comments, and Retired_Date. The cleaned dataset has the following columns: Building_Name, Abbrev_Short, Other_Names, Official/Long, City, State, County, Latitude, Longitude, Constructed (year), Address, Zip_Code, Architect, Demolished, Style, Google Maps Customer ID, LEED, and HasGreenRestaurants. OE&S and SDSU Annual Sustainability Summit were created and enhanced as separate Wikidata entries.
10. Google Maps Customer ID is customer identifier for a place. For more information, see <https://www.wikidata.org/wiki/Property:P3749>.
11. <https://www.gmbeverywhere.com/>
12. "SDSU Buildings Data Model", SDSU Library, 2024 <https://docs.google.com/spreadsheets/d/1TVpc016lmxUZl932wRnl63mXqwRbJk8LZjqXDPuAsPg/edit?usp=sharing>.
13. <https://www.wikidata.org/wiki/Q124258890>
14. <https://www.wikidata.org/wiki/Property:P5008>
15. <https://www.wikidata.org/wiki/Help:QuickStatements>