

Zen Library - Linked Open Data (ZL-LOD) Project Plan

Ryan Spellman

School of Information, Kent State University

Dr. Marcia Lei Zeng

LIS 61095: Linked Data

May 6, 2022

Table of Contents

Abstract	2
Introduction and Project Summary	3
Data Resources and Tools	3
Wikidata Knowledge Graph	3
Gephi Visualization Tool	4
Workflow and Tasks	4
Alignment and Consolidation: Toward 5-star LOD	5
Value Vocabularies	6
ZL-LoD Schema: Classes, Properties, URIs, and Namespaces	7
Wikidata Property and Object URIs	7
Standard Wikidata Namespaces Included in ZL-LOD	7
FOAF Namespace	
PREFIX foaf: <http://xmlns.com/foaf/spec/>	8
ZL-LOD Namespace	8
ZL-LOD Element Set: Non-Wikidata-Derived Classes and Properties in Current Use	8
Summary of Limitations and Future Enhancement	10
Long-term Plan for Extending the Data	10
References	12
Appendix A: SPARQL Query for WikiData Query Service	13
Appendix B: ZL-LOD Entity-Relationship Inference Diagram	14

Abstract

The following paper is a project summary for the ZenLibrary Linked Open Data (ZL-LOD) project. This plan summarizes the project's aims, methods, tools, workflow, and technical details such as namespaces, element set details, and the ZL-LOD entity-relationship inferences diagram. Through use of custom SPARQL queries, ZL-LOD will collect data from Wikidata to build a list of notable Zen Buddhists from the Sōtō and Rinzai schools. This data will establish connections between these entities and their students, additional religion or worldviews they have, as well as an image, link to their WorldCat Identity page, and list their notable works. ZL-LOD will supplement this data with lineage information for each entity, harvested from semi-structured source material and made freely available for all to explore through an intuitive front-end Sigma.JS data visualization. In addition to the front-end focus, ZL-LOD will hold to Sir Tim Berners-Lee's 5-star marker for LOD best practices. This means ZL-LOD will think strategically during each development stage so it can contribute to the greater LOD ecology, becoming "of the web" rather than simply "on the web".

Keywords: LOD, linked open data, SPARQL, Wikidata, zen, Sōtō, Rinzai

Introduction and Project Summary

This project, currently being called Zen Library Linked Open Data (ZL-LOD), will create a topically focused LOD resource that explores and visualizes relationships between Sōtō Zen and Renzai School Buddhist figures. This includes relationships such as their schools of thought, notable works, and connections to other entities that arise via LOD-powered inferences. What this project will contribute to preexisting data are the lineages of each figure. Ordained Zen Buddhists have ties to a lineage that purportedly runs all the way back to the original Buddha in ancient India (Burk, 2019). While working with Wikidata in the early pre-planning stages for ZL-LOD, it was discovered that lineage data is not available. DBpedia was also considered, as it does have a lineage property available in its ontology. However, the data values attributed to this property are very lacking. Therefore we will implement a custom RDF container `rdf:Seq` to allow ZL-LOD to populate lineage property values.

Inspiration for this project comes from the likes of the [Internet Philosophy Ontology](#) (InPho), the [Linked Jazz](#) project, and the [Zen Marrow](#) project. Though inspired by these projects, the first step in ZL-LOD will be mostly rooted in the planning phase. A simple entity-to-relationship model will focus on the person as the primary subject entity, from which data will infer connections to successors (person), students (person), lineages (person), and more to reveal new relationships.

Data Resources and Tools

Wikidata Knowledge Graph

The baseline data will come from the Wikidata Knowledge Graph via their SPARQL endpoint. To review the SPARQL query developed to retrieve the base wikidata information, refer to [Appendix A](#). The data rows are required to have certain property value pairs, which pulls a list of entities that match the project's scope (see [Table 1](#) and [Appendix B](#), respectively).

Table 1

Breakdown of Entity Requirements and Option Values of SPARQL Query in Figure 1

Entity	First Tier Property Requirements	Optional (Data to Seek) - Retrieve, if Exists	Filter - Require One Value
?person	<p>?person p:P31 wd:Q5. ... is instance of <i>human</i></p> <p>?person p:P140 wd:Q748. ... has worldview <i>Buddhism</i></p> <p>?person p:P140 wd:Q7953. ... has worldview <i>Zen</i></p>	p:P1066 <i>students of</i>	<p>?person p:P140 wd:Q639625. ... has worldview <i>Rinzai School</i></p> <p>?person p:P140 wd:Q749085. ... has worldview <i>Sōtō</i></p>
		p:P7859 <i>WorldCat Identities ID</i>	
		p:P800 <i>notable works by</i>	
		p:P140 <i>religion / worldview of</i>	
		p:P18 <i>image associated with</i>	

Gephi Visualization Tool

Gephi will be used to visualize RDF data offline. Gephi is a powerful open source data visualization tool that supports importing data via CSV, as well as many other formats (Gephi, n.d.). Once a visualization is ready, the [SigmaExporter plugin](#) will render an interactive graph that is almost suitable for the web.

Sigma.JS is a javascript framework for visualizing data that is published under an MIT license (Sigma.JS, n.d.). At the time of this writing, there is just a prototype available at <https://zenlibrary.org/visualization/>. This prototype is mostly a direct export to Sigma.JS from Gephi using the plugin mentioned above. In the finished project, a custom Sigma.JS instance will be built, tailored to the needs of this project.

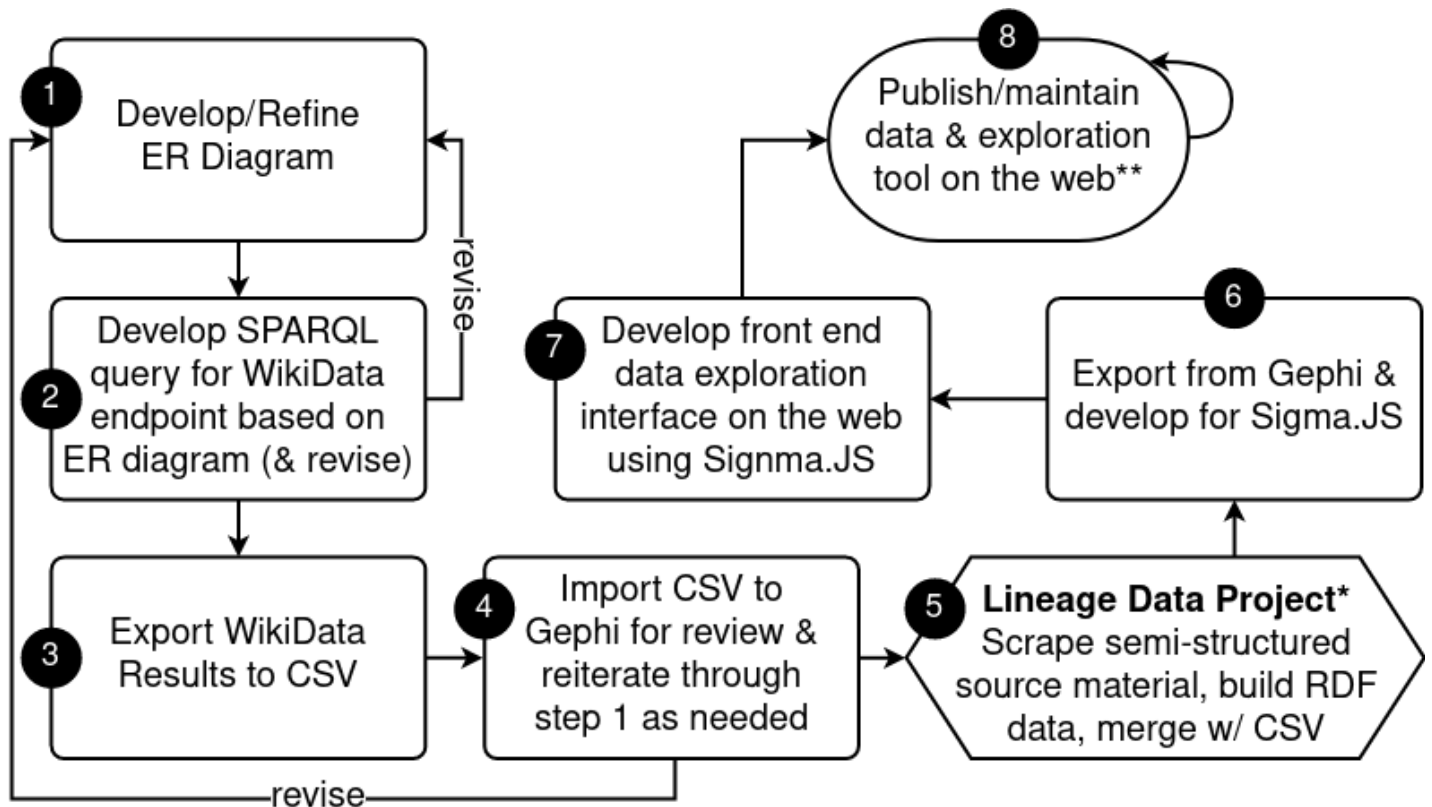
Workflow and Tasks

The general workflow will involve approximately eight phases. As you can see in [Figure 2](#), revision is going to be key as the project develops, particularly in phases two and five. Also take note of phase number 5. This phase is detailed enough to warrant its own workflow model. Once ZL-LOD

is to this stage, the first run of the project will build lineage data manually for one or two entities as a use case. Once this process is smoothed out, a workflow to either scrape and clean or manually enter lineage data will be established.

Figure 2

ZL-LOD Workflow and Tasks



* The Lineage Data Project is a separate subproject.

** Publishing the data will include reciprocal linking and uploading to DataHub.

Note. Diagram created by the author using Diagrams.net <https://www.diagrams.net>

Alignment and Consolidation: Toward 5-star LOD

Sir Tim Berners-Lee's five-tiered system helps guide best practices for open data. As data progresses up the tiers, it progressively becomes more beneficial to the world of LOD.

These tiers include:

- “★ Publish data on the Web in any format (e.g., PDF, JPEG) accompanied by an explicit Open License (expression of rights).
- ★★ Publish structured data on the Web in a machine-readable format (e.g., XML).
- ★★★ Publish structured data on the Web in a documented, non-proprietary data format (e.g., CSV, KML).
- ★★★★ Publish structured data on the Web as RDF (eg Turtle, RDFa, JSON-LD, SPARQL)
- ★★★★★ In your RDF, have the identifiers be links (URLs) to useful data sources” (Hyland et al., 2013).

This project aims to become a five-star LOD resource. As previously outlined, the project will organically work up to this benchmark as it develops. See [Table 2](#) below for a brief review of how these tiers will be met during the project’s development.

Table 2

Alignment of ZL-LOD with 5-Star LOD

Rating	Alignment with Rating
★	Data will be published on the web at zenlibrary.org
★★	The data is structured right after query
★★★	Data is exported to open format CSV
★★★★	After CSV builds interactive Gephi/Sigma.JS visualization, data will be published as JSON-LD and available at zenlibrary.org
★★★★★	The project will link to Wikidata entities, WorldCat Identities, and VIAF. Additionally, including sameAs.org co-references in RDF for entities.

Value Vocabularies

The property values for `zl:lineage` will adhere to the Virtual International Authority File (VIAF). VIAF is a resource that combines name authority files into a singular point of reference, and is provided by the OCLC (VIAF, n.d.). If no Wikidata entity for a person is available for lineage data,

VIAF will be consulted. If the name is not included in VIAF, `rdfs:comment` will be used to note that no authority file was used. The development of a local name authority file may be considered should the need arise.

ZL-LoD Schema: Classes, Properties, URIs, and Namespaces

The following is a list of the Wikidata URIs that are used explicitly in the project. Only one is a hard property-value-pair requirement (**R→**) for any entity that is returned. Three are of a group in which at least one is required (**||R→**) in addition to the hard requirement.

Wikidata Property and Object URIs

R→ = required property/value pair

||R→ = only one property/value pair out of all indicated is required

<https://www.wikidata.org/wiki/Property:P31> **R→** <https://www.wikidata.org/wiki/Q5>

<https://www.wikidata.org/wiki/Property:P140> **||R→** <https://www.wikidata.org/wiki/Q7953>

<https://www.wikidata.org/wiki/Property:P140> **||R→** <https://www.wikidata.org/wiki/Q639625>

<https://www.wikidata.org/wiki/Property:P140> **||R→** <https://www.wikidata.org/wiki/Q748>

<https://www.wikidata.org/wiki/Property:P7859>

<https://www.wikidata.org/wiki/Property:P18>

<https://www.wikidata.org/wiki/Property:P1066>

<https://www.wikidata.org/wiki/Property:P140>

<https://www.wikidata.org/wiki/Property:P800>

Standard Wikidata Namespaces Included in ZL-LOD

PREFIX wd: <<http://www.wikidata.org/entity/>>

PREFIX wdt: <<http://www.wikidata.org/prop/direct/>>

PREFIX wikibase: <<http://wikiba.se/ontology#>>

PREFIX p: <<http://www.wikidata.org/prop/>>

PREFIX ps: <<http://www.wikidata.org/prop/statement/>>

PREFIX pq: <<http://www.wikidata.org/prop/qualifier/>>

PREFIX rdfs: <<http://www.w3.org/2000/01/rdf-schema#>>

PREFIX bd: <<http://www.bigdata.com/rdf#>>

FOAF Namespace

PREFIX foaf: <<http://xmlns.com/foaf/spec/>>

ZL-LOD Namespace

PREFIX zl: <<https://www.zenlibrary.org/zl-lod/#>>

ZL-LOD Element Set: Non-Wikidata-Derived Classes and Properties in Current Use

A simple two-class, one-property element set is all that will be needed to record sequential lineage data. The purpose of the `zl:Lineage` hierarchical custom class structure is to provide extensibility. As the project develops, more subclasses and properties may be required to accurately represent more complex lineage information.

Class: Lineage	
Label	Lineage
Term Name	zl:Lineage
URI	https://www.zenlibrary.org/zl-lod/#Lineage
Definition	Class to denote lineage data.
Implementation	Optional Repeatable
Comment	Class is hierarchically implemented to point to the lineage of the subject. <pre><zl:Lineage> <rdf:Seq> <rdf:li> </rdf:li> </rd:Seq> </zl:Lineage></pre>
Value Space Control	—

Class: Seq	
Label	Seq
Term Name	rdf:Seq
URI	http://www.w3.org/2000/01/rdf-schema#Seq
Definition	Container class used to build a sequential list in RDF.
Implementation	Optional non-repeatable within same Lineage class
Comment	Class is used as a subclass of zl:Lineage. <pre><zl:Lineage> <rdf:Seq> <rdf:li> </rdf:li> </rd:Seq> </zl:Lineage></pre>
Value Space Control	—PREFIX foaf: < http://xmlns.com/foaf/spec/ >

Property: li (rdf:_n)	
Label	Ordered Item
Term Name	li (rdf:_n)
URI	http://www.w3.org/2000/01/rdf-schema#
Definition	RDF property used to identify entities of a lineage in sequential order.
Implementation	Optional Repeatable
Comment	Property is used to identify lineage members in sequential order. In RDF/XML, rdf:li is used. For N3 serialization, use rdf:_n. <pre><zl:Lineage> <rdf:Seq> <rdf:li rdf:about="http://www.wikidata.org/entity/Q318064"> </rdf:li> <rdf:li rdf:about="http://www.wikidata.org/entity/Q44847"> </rdf:li> </rd:Seq> </zl:Lineage></pre> <hr/> <pre>zl:Lineage rdf:_1 wd:Q318064 ; rdf:_2 wd:Q44847 .</pre>
Value Space Control	Literal values obtained from VIAF when necessary.

Summary of Limitations and Future Enhancement

This RDF data will be manually acquired in this first step by running queries through Wikidata's SPARQL endpoint. This is a major limitation of the project, as the graphs will not be dynamically generated. In future refinements, the Apache Jena Fuseki tool could be run on a Tomcat server to make dynamically generated SPARQL queries against locally stored RDF data sets based on a user's front-end input possible.

A future enhancement to address this might be employing a CSV spreadsheet to manage lineage information that is mapped to a RDF conversion tool such as Tarql. This would make it easy to create a user-editable spreadsheet through an online form interface that could create CSV data that can be converted to RDF and then incorporated into the ZL-LOD data. Some of this process could likely be automated as well.

Another future enhancement will involve refining how additional information about each entity can be obtained. Part of the data this project returns is a WorldCat Identities ID for each entity. WorldCat Identities is a resource that includes information from [WorldCat](#), [VIAF](#), [FAST](#), as well as details about the entity's works, and much more (OCLC Research, n.d.). It is a very rich resource. At a minimum, we have a goal of making the WorldCat Identities ID a clickable link to follow for more information. However, a deeper integration might be possible with a little research.

Long-term Plan for Extending the Data

The ZL-LOD project has already begun to spark new ideas on how it might grow in the future. For instance, birth and death places might be included and mapped using Getty's TGN. Similarly, the project could consider using this data to approximate eras, and then establish relationships between entities based on that calculation. This is a good example of how one extension of the data can lead to another. Unlike the lack of lineage information that spurred this project, both of these examples are achievable with data already available in Wikidata. There is also potential for establishing relationships based on topics of their talks and writings, notable practices (such as tea ceremonies, or

calligraphy), and much more. Establishing relationships like this could unearth unexplored connections between different entities. To find these relationships, semi-structured source material from [Sacred Texts](#), the [Gutenberg Project](#), and other open access resources will be run through PermID's Intelligent Tagging tool, which allows for exporting to RDF.

Once ZL-LOD gets to the stage of expressing new relationships based on processing semi-structured source material, it will be time to enlist help proofreading them. This will involve reaching out to subject specialists in the topic, asking them for feedback and revisions. A simple contact form could be constructed for this. Ideally, there would be a link in the Sigma.JS instance of the entity or relationship that needs revision that will take the user to the contact form while also logging the node that they came from.

References

- Burk, D. (2019, April 6). *Lineage in buddhism: The intersection between the individual and the collective tradition*. Zen Studies Podcast. <https://zenstudiespodcast.com/lineage-in-buddhism/>
- Gephi. (n.d.). *Features*. <https://gephi.org/features/>
- Hyland, B., Ateazing, G., Pendleton, M., Srivastava, B. (2013, June 27). *Linked data glossary*. W3C Working Group Note.
<https://dvcs.w3.org/hg/gld/raw-file/default/glossary/index.html>
- OCLC Research. (n.d.). *WorldCat identities*.
<https://www.oclc.org/research/areas/data-science/identities.html>
- Sigma.JS. (n.d.). *A JavaScript library aimed at visualizing graphs of thousands of nodes and edges*. <https://www.sigmajavascript.org/>
- VIAF. (n.d.). *VIAF: The virtual international authority file*. <http://viaf.org/>

Appendix A: SPARQL Query for WikiData Query Service

```

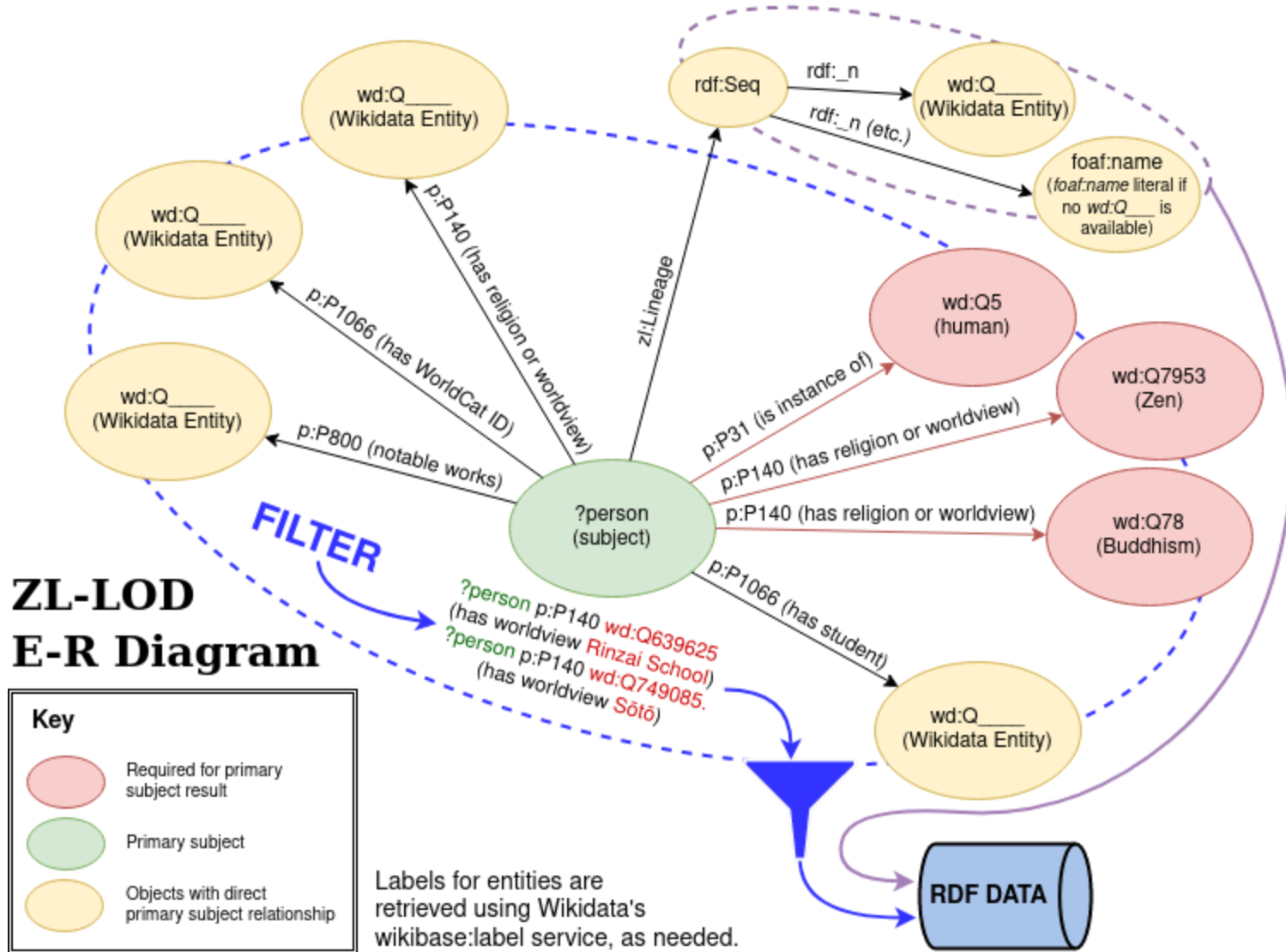
SELECT DISTINCT      ?personLabel ?person ?studentLabel ?student ?worldCatIDURL ?workLabel ?work
?worldViewLabel
                    ?worldView ?image
WHERE { SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". } #label
service
{
  SELECT DISTINCT ?person ?work ?student ?worldCatIDURL ?worldView ?image
  WHERE
  {
    ?person p:P140 ?statement0.
    ?statement0 (ps:P140/(wdt:P279*)) wd:Q748. # subject's religion or world view is Buddhism
    ?person p:P140 ?statement1.
    ?statement1 (ps:P140/(wdt:P279*)) wd:Q7953. # subject's religion or world view is Zen
    ?person p:P31 ?statement3.
    ?statement3 (ps:P31/(wdt:P279*)) wd:Q5. # subject is an instance of human
    OPTIONAL
    {
      ?person p:P1066 ?statement4.
      ?statement4 (ps:P1066/(wdt:P279*)) ?student. # optionally pull students of entity
      ?person p:P7859 ?statement5.
      ?statement5 (ps:P7859/(wdt:P279*)) ?worldCatID. # optionally pull WC Identities
      number
      ?person p:P800 ?statement6.
      ?statement6 (ps:P800/(wdt:P279*)) ?work. # optionally pull notable works
      ?person p:P140 ?statement7.
      ?statement7 (ps:P140/(wdt:P279*)) ?worldView. # optionally pull religion or worldview
      ?person p:P18 ?statement8.
      ?statement8 (ps:P18/(wdt:P279*)) ?image. #optionally pull the image for the entity }
    BIND(URI(CONCAT('https://www.worldcat.org/identities/', ?worldCatID)) AS ?worldCatIDURL)
    FILTER ( ?worldView = wd:Q748 || # worldview must be Buddhism OR
    ?worldView = wd:Q7953 || # worldview must be Zen OR
    ?worldView = wd:Q639625 || # worldview must be Rinzai School OR
    ?worldView = wd:Q749085). # worldview must be Sōtō
  }
}

```

}

Note: To see this query in action, follow this link: <https://zenlibrary.org/wikidata-query.php>

Appendix B: ZL-LOD Entity-Relationship Inference Diagram



Note. Diagram created by the author using Diagrams.net <https://www.diagrams.net>