Earthquake LOD: Seismic Dataset Construction with Ontology Oriented Design Patterns

Hiroki Uematsu*
hiroki_u@nii.ac.jp
The Graduate University for Advanced Studies, SOKENDAI
Hayama, Kanagawa, Japan
National Institute of Informatics
Chiyoda-ku, Tokyo, Japan

Hideaki Takeda
takeda@nii.ac.jp
National Institute of Informatics
Chiyoda-ku, Tokyo, Japan
The Graduate University for Advanced Studies, SOKENDAI
Hayama, Kanagawa, Japan

ABSTRACT

This paper describes the construction of an Earthquake Linked Open Data (LOD) for seismological datasets using Ontology Oriented Design Patterns. In recent years, various studies have focused on utilizing machine learning technology to detect, classify, or predict seismic intensities based on vast amounts of observed seismic waveform data. Researchers need to collect hypocenter information, time of occurrence, and target observation stations related to seismic waveforms to compile data for machine learning purposes. Although seismic waveform datasets for machine learning are widely available worldwide, accessing waveform data observed by Japanese seismic networks is limited, and metadata retrieval is difficult. To address this, we developed the Earthquake LOD with Ontology Oriented Design Patterns to enhance the discovery and retrieval of seismic data. Ontology Oriented Design Patterns are used to construct interoperable knowledge graphs. This method categorizes tasks for humans and machines using a domain-specific knowledge usage checklist, enabling the efficient creation of valuable knowledge graphs. In this paper, we provide a detailed account of constructing the Earthquake LOD for seismic datasets from both Japan and overseas, utilizing Ontology Oriented Design Patterns.

KEYWORDS

Ontology, Linked Open Data, Ontology Oriented Design Patterns, Earthquake

ACM Reference Format:
Hiroki Uematsu and Hideaki Takeda. 2023. Earthquake LOD: Seismic Dataset Construction with Ontology Oriented Design Patterns. In *Proceedings of The 12th International Joint Conference on Knowledge Graphs (IJCKG 2023) (Conference IJCKG '23)*. ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Japan is one of the most earthquake-prone countries in the world. Around the Japanese archipelago, four plates collide with each other, and more than 100,000 earthquakes occur per year, averaging more than 300 earthquakes per day, including those that are not felt. Although Japan accounts for about 0.25% of the world's land area, earthquakes exceeding magnitude 6 have occurred nearly 200 times in 10 years, accounting for 20% of all earthquakes in the world.

Seismic motion is observed as waveform data of acceleration and is used in various research such as calculation of seismic intensity, determination of hypocenter, emergency earthquake warning, and predicted seismic intensity. In recent years, it has been used as training data for research using machine learning, such as predicting the seismic intensity at a specific observation station, whether the observed waveform is a seismic waveform, and identifying the P-wave/S-wave of an earthquake. Since machine learning requires a large amount of high-quality training data, seismic observation networks are useful. However, one of the networks K-NET[7] which was established by the National Research Institute for Earth Science and Disaster Resilience (NIED) waveform data acquisition site does not have an API, users need to specify the date and time, hypocenter, observation station, etc., and download the waveform data. In order to search for waveform data independently observed by researchers and observation networks of the Japan Meteorological Agency (JMA) and local governments, it is possible to create a database that aggregates waveform data. Although, since the waveform data cannot be republished and there is no URI that uniquely points to the waveform data, researchers will have their own databases, making it difficult to create a reusable open waveform database.

Therefore, we will collect metadata of "earthquakes" such as information on observation stations where earthquake ground motions were observed, seismic intensity, observation time, hypocenter position, and magnitude estimated from observed waveforms, and publish them in the form of Linked Data. In Linked Data, the link structure of things and concepts is represented by a model called triple using RDF, and data can be traced by following properties links in the same way as following links on websites. Here, website links indicate simple connections, but in Linked Data triples, properties that become links also have URIs and indicate link relationships.

In this research, we newly define the vocabulary related to seismic motion such as seismic source, seismic intensity, and observation time, and the vocabulary related to observation stations as properties specialized for earthquake data. For example, 'Earthquake motion' has vocabularies such as 'Seismic intensity', 'Magnitude', and 'Occurrence time' as properties. In addition, by using the term 'observed waveform' as a property, triples of waveforms

Conference IJCKG '23, Dec 08–09, 2023, Miraikan, Tokyo, Japan

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

obtained by observing a certain seismic motion are created. can be expressed.

2 EARTHQUAKE OBSERVATION

In general, the word "earthquake" refers to events such as tremors felt by people on their own, but in reality, it refers to the rapid displacement of the bedrock due to the pushing and pulling of the underground bedrock. Shaking occurs as a result of bedrock displacement and is recognized by us at the ground surface. The shaking is transmitted to the ground surface as waves, and the magnitude of the shaking observed by seismometers is expressed as the seismic intensity in Japan. The difference in the velocity of the P and S waves of the seismic waveform observed at each station is used to estimate the hypocenter of the earthquake. Because earthquakes occur underground, it is difficult to actually observe them. Therefore, information on the waveforms observed at each observation station is important, such as the hypocenter estimation and calculating the seismic intensity.

Observations of seismic activity are conducted in many countries. The International Federation of Digital Seismograph Networks (FDSN)[24] has 2196 registered seismograph networks with 24-bit resolution with data recorded in continuous time series at a sampling rate of at least 20 samples/second are registered. STEAD[12] registers approximately 1.2 million time series of seismic waveforms observed by seismometers, covering more than 19,000 hours of datasets. These datasets include data observed by seismic networks created by academic institutions such as the International Seismological Center (ISC)[1] and the United States Geological Surve, as well as data from seismometers installed at home by individuals such as Raspberry Shake[22]. The data sets contain a variety of data, including data observed by seismometers installed at home, such as Raspberry Shake.

In Japan, seismic waveforms observed by observation networks such as K-NET and Kik-net, which are based on data from observation stations established by NIED, JMA, and local governments, can be obtained. However, although the acquired data can be used for analysis and other purposes, it cannot be redistributed, and only some of the JMA's observation station and data are registered in the FDSN. The JMA releases observation data only for major earthquakes as strong-motion earthquake observation data[2]. On the other hand, the seismic intensity database search [3] allows users to search for earthquakes on a map by date and time of occurrence, seismic intensity, seismic intensity observed at each observation station from the observation station, hypocenter and depth, and magnitude. Other earthquake data is available in the Earthquake Monthly Report (Catalog Edition)[9]. In addition, the hypocenters of earthquakes up to several years before the hypocenters were determined are available on the hypocenter data [4] of the Earthquake Monthly Report (Catalog Edition). Although the observed waveforms themselves cannot be obtained, they can be considered to contain metadata on the observed seismic motions.

Using metadata of observed waveforms and seismic motions, various studies have been conducted not only to estimate seismic intensity and hypocenter but also to calculate predicted seismic intensity, and classify earthquakes and earthquake early warning systems. For example, for real-time seismic intensity estimation[10], data from stations that have observed the same earthquake are used as training data for learning. When downloading data using K-NET of NIED, it is necessary to find earthquake waveforms by using queries such as "search by observation station," "search by hypocenter," or "search by record start time. Figure 1 is a form for searching and downloading data from K-NET. You can retrieve the necessary data by specifying the observation stations and the period that includes the earthquake occurrence time.

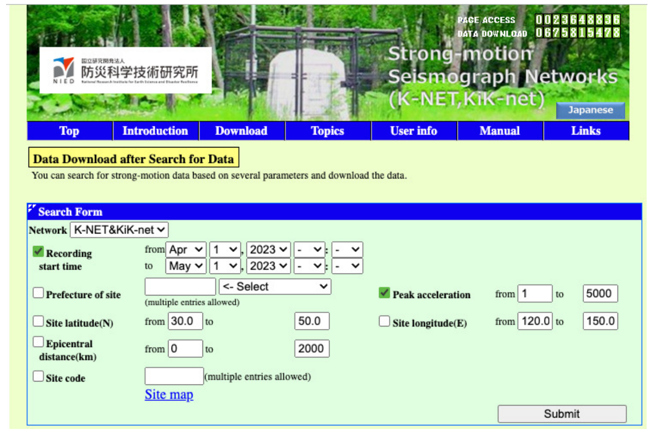

Figure 1: Data Download after Search for Data | K-NET

However, there is no list of which stations observe which earthquakes, although multiple stations must observe the same earthquake to be selected when searching by the station. In addition, when searching from the hypocenter, it is not known whether the observation stations observed the earthquake that occurred at that hypocenter or not without searching the data and making a list. Furthermore, since the observation stations are different from each other, it is difficult to retrieve the seismograms of earthquakes that occurred at the same hypocenter from multiple observation stations because the IDs are not assigned to each earthquake.

3 METHOD

To solve the problem in section 2, we aim to make the observed waveforms publicly available and searchable in the form of Linked Data, which links data together.

Uematsu et al.[27] showed the flow of determining the schema for improving data interoperability, transforming data, designing connections to other data, and registering the data in Wikidata. Using corporate databases distributed in Japan and insurance medical and pharmacy institutions as examples, they aim to make data distributed in XML, Excel, PDF, etc. reusable as a base registry.

To improve interoperability, it is important to facilitate reuse of frequently used schemas and connection with data from other domains. For example, it is easy to understand data and connect to Wikidata and DBpedia by using schemas defined in Wikidata, DBpedia, schema.org, and also as well as owl, rdf, and SKOS. In addition, by using LOV or DBpedia Archivo to search for ontologies

[1]http://www.isc.ac.uk/
[2]https://www.data.jma.go.jp/eqev/data/kyoshin/jishin/
[3]https://www.data.jma.go.jp/eqdb/data/shindo/
[4]https://www.data.jma.go.jp/eqev/data/bulletin/hypo.html

in the same domain, and if there are data or services that utilize those ontologies or schemas, the data can be interconnected and utilized by using a standard schema. It is natural to use a common vocabulary such as place names, latitude/longitude, time, etc. In addition, if there is a vocabulary such as earthquake or seismic intensity, it should be possible to combine and use data related to earthquakes that are already available to the public by using the same vocabulary.

On the other hand, it is important not only to use a common vocabulary but also to use a schema that is specific to that domain and can correctly represent the nature of the data. The same vocabulary may have different usages and meanings in various domains.

In such cases, it is necessary to create a schema tailored to the data to be converted to Linked Data, rather than using an existing ontology or vocabulary. Linked Data conversion is difficult without understanding the original data structure, ontology, and schema, and is often constructed empirically. In addition, simple methods are used, such as anyLink[5] and CSV2LOD[6], which first determine the mapping to an existing schema based on whether or not it matches the Resource or Property name.

Therefore, we organized these flows and created design patterns for reading data, checking domains, matching with existing schemas, and further data cleansing. In addition, by drawing attention to the ontology and vocabulary used to express the meaning and structure of the data, it is possible to understand the data and find the appropriate vocabulary.

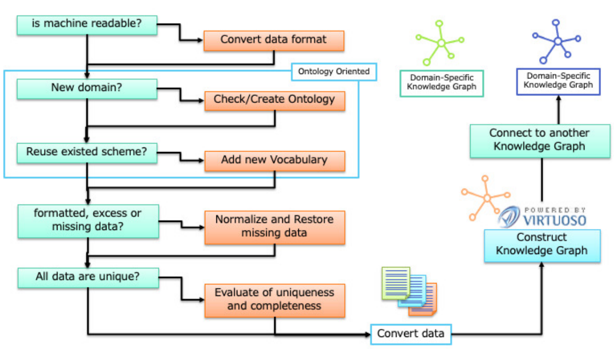

Figure 2: Ontology Oriented Design Patterns

Figure 2 shows simple data design flow using Ontology Oriented Design Patterns. The five checks indicated by the green boxes, "Is machine readable?", "New domain?", "Reuse existed schema?", "Formatted, excess or missing data?", "All data are unique?" If the answer is "yes," the data can be processed by the machine as is, but if the answer is "no," a human must check the contents of the data and process it. For example, in the case of machine-unreadable data created in PDF or Excel, the file format should be converted or the data contained in a single cell should be split. If the existing schema cannot be reused, a new schema should be created according to the ontology of the relevant data. Thus, the design pattern is largely a flow of reading the acquired data, checking and designing the domain of the data based on the ontology, correcting missing data or formatting, and checking for uniqueness to connect with external data. Each item and its task is itemized below.

(1) Collect original data
    - PDF, Excel, CSV/TSV, RDF, etc.
    - Convert to machine-readable format
(2) Design data structure with ontology orientation
    - Automatic or manual processing
    - Use machine support as much as possible for manual processing
    - check domain of data
        – find the same domain or suitable ontologies and vocabularies
        – create the suitable ontology
    - Existing domain or vocabularies
        – use existing vocabularies, schemas
(3) Normalize
    - Remove unnecessary characters
    - Convert full-width to half-width
(4) Restore missing data
    - Specify domain in Wikidata and Wikipedia and so on
    - Restore missing data
(5) Match columns between data of the same domain
    - Perform column matching between data to be unified
    - Create a mapping table manually
    - Evaluate the uniqueness of columns used for unification
        – Evaluate uniqueness and completeness

By using this design pattern, users who do not have empirical knowledge can follow the flow to create highly interoperable data. In particular, the flow clarifies which parts can be handled mechanically which need to be handled manually, and which tasks the user should focus on. Among these, focusing on ontology will help users understand the data and the domain of the data, and enable them to use vocabulary and schemas that have meaning appropriate to the data, rather than just schema mapping.

In the following sections, we present the procedure for converting seismic data to LOD using the ontology oriented design patterns.

4 EARTHQUAKE ONTOLOGY

First, we organized the vocabulary related to earthquakes. The JMA's Earthquake Monthly Report (Catalog Edition) does not provide data on observed waveforms, but it does provide metadata on observed earthquake ground motions.

On the other hand, an example of describing USGS Earthquake data utilizing the SOSA (Sensor, Observation, Sample, and Actuator) ontology of the SSN (Semantic Sensor Network)[5] has been published[7]. This description example shows the acceleration of the earth's surface observed by a sensor that is an instance of SOSA:Sensor when the earth is the observation target. Since the structure in which sensors detect tremors by observing the acceleration of the earth's surface and record the waveform is the structure of information used to measure seismic intensity and to estimate the hypocenter, the Earthquake Ontology to be constructed in this study will also be constructed utilizing SOSA. However, since there

[5] http://link.lodosaka.jp/
[6] https://koujikozaki.github.io/CSV2LOD/
[7] https://www.w3.org/TR/vocab-ssn/integrated/examples/seismograph.ttl

is no property that expresses the relationship between the observed waveforms and the seismic intensity or hypocenter, it is necessary to create a new property. In particular, the seismic intensity widely used in Japan is based on the JMA seismic intensity scale defined by the Japan Meteorological Agency, which requires the creation of unique properties.

Similarly, VEO (Volcano Event Ontology)[6, 23] is an ontology that utilizes SOSA and SSN. The VEO is intended to describe data observed using IoT systems for seismic motions caused by volcanic activity. In addition, it is used for machine learning-based classification to identify seismic events (underwater explosions, quarry blasts, and thunders).

SBEO (Smart Building Evacuation Ontology)[21] defines the urgency, severity, and intensity of natural disasters such as earthquakes and tsunamis, and human disasters such as terrorism and kidnapping as contexts, and builds an ontology for route recommendation systems to evacuate buildings.

As several examples show, "Earthquake" is defined as an ontology or vocabulary and applied to each case. However, it can be seen that what the term "Earthquake" expresses differs depending on the ontology, application, and user. The SOSA example and VEO indicate ground shaking, while SBEO indicates natural disasters. In Japan, the terms "jishin-dou", "jishin", and "shin-sai" are used interchangeably, although the English term would be "Earthquake". Earthquakes in Japan, as discussed in Section 2, refer to phenomena of bedrock displacement, and there is not always a single hypocenter or seismic motion that is estimated to be the source of the earthquake. Therefore, prominent seismic activity is summarized as major damaging earthquakes that occurred in Japan[8] . In addition, earthquake disaster is defined as "damage directly caused by seismic motion and damage caused by tsunamis, fires, explosions, and other unusual phenomena that occur as a result of such motion,"[9] and is distinguished from earthquake and seismic motion. However, Wikidata's Entity is listed as "Earthquake", as in the Great East Japan Earthquake[10] .

Since each of these concepts has a different meaning, existing ontologies and schemas with the vocabulary earthquake cannot correctly represent the concept. In this study, "jishin-dou" as seismic motion, "jishin" as Earthquake, and "shin-sai" as Earthquake disaster are used as separate concepts. However, the list of seismic motions that include earthquakes and earthquake disasters is not fully published in Japan, although the definitions of each concept are different. Therefore, we first created an ontology as an earthquake vocabulary based on the Earthquake Monthly Report (Catalog Edition) of JMA, which records metadata about seismic motions, such as hypocenters, intensity and observation station.

The data in the Earthquake Monthly Report (Catalog Edition) include source data, measured data, first motion mechanism solution data, CMT solution data, seismic intensity data, tsunami data, etc. In this paper, the seismic intensity data file was first selected as the target. The seismic intensity data contains a record called the hypocenter record, which contains information on the hypocenter, and information on the observation stations where the earthquake motion that occurred at the hypocenter was observed. Figure 3 is a part of the Earthquake Monthly Report (Catalog Edition).

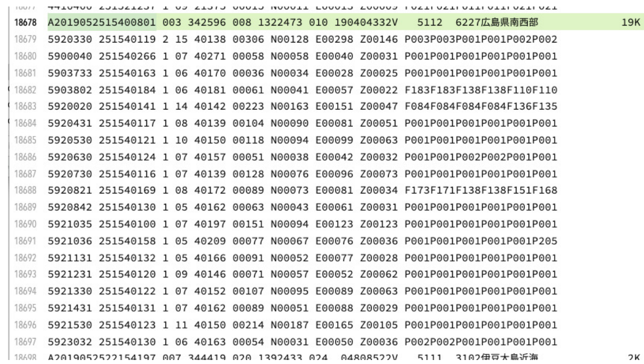
```
A2019052515400801 003 342596 008 1322473 010 190404332V   5112  6227広島県南西部          19K
5920330 251540119 2 15 40138 00306 N00128 E00298 Z00146 P003P003P001P001P002P002
5900040 251540266 1 07 40271 00058 N00058 E00040 Z00031 P001P001P001P001P001P001
5903733 251540163 1 06 40170 00036 N00034 E00028 Z00025 P001P001P001P001P001P001
5903802 251540184 1 06 40181 00061 N00041 E00057 Z00022 F183F183F138F138F110F110
5920020 251540141 1 14 40142 00223 N00163 E00151 Z00047 F084F084F084F084F136F135
5920431 251540117 1 08 40139 00104 N00090 E00081 Z00051 P001P001P001P001P001P001
5920530 251540121 1 10 40150 00118 N00094 E00099 Z00063 P001P001P001P001P001P001
5920630 251540124 1 07 40157 00051 N00038 E00042 Z00032 P001P001P002P002P001P001
5920730 251540116 1 07 40139 00128 N00076 E00096 Z00073 P001P001P001P001P001P001
5920821 251540169 1 08 40172 00089 N00073 E00081 Z00034 F173F171F138F138F151F168
5920842 251540130 1 05 40162 00063 N00043 E00061 Z00031 P001P001P001P001P001P001
5921035 251540100 1 07 40197 00151 N00094 E00123 Z00123 P001P001P001P001P001P001
5921036 251540158 1 05 40209 00077 N00067 E00076 Z00036 P001P001P001P001P001P205
5921131 251540132 1 05 40166 00091 N00052 E00077 Z00028 P001P001P001P001P001P001
5921231 251540120 1 09 40146 00071 N00057 E00052 Z00062 P001P001P001P001P001P001
5921330 251540122 1 07 40152 00107 N00095 E00089 Z00063 P001P001P001P001P001P001
5921431 251540131 1 07 40162 00089 N00051 E00088 Z00029 P001P001P001P001P001P001
5921530 251540123 1 11 40150 00214 N00187 E00165 Z00105 P001P001P001P001P001P001
5923032 251540130 1 06 40163 00054 N00031 E00050 Z00036 P002P002P001P001P001P001
```

Figure 3: Earthquake Monthly Report (Catalog Edition), 2019.05

First, the hypocenter record is listed, and the number at the end of this record is the number of stations that observed the earthquake. The number at the end of the record is the number of observation points that observed the earthquake. In the example, the number is 19, so 19 lines of information on the seismic intensity and acceleration observed at the observation points are listed from the next line of the seismic source record. Both the hypocenter, seismic intensity, and acceleration records are fixed-length data, and include the location and depth of the hypocenter, the magnitude, the time of occurrence (origin time), the observation point number, and the time when the earthquake was determined to have occurred at the observation point (trigger time). The Earthquake Monthly Report (catalog section) contains metadata recorded regarding tremors observed at hypocenters and observation stations. To convert metadata related to earthquakes into RDF, instead of focusing on the acceleration waveforms themselves, we constructed an ontology by utilizing the column names used in the seismic source and intensity records.

Listing 1: jp-earthquake.ttl

```
PREFIX xsd: <http://www.w3.org/2001/
    XMLSchema#>
PREFIX rdf: <http://www.w3.org
    /1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/
    rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/
    owl#>
PREFIX schema: <http://schema.org/>
PREFIX dcterms: <http://purl.org/dc/terms
    />
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX sosa: <http://www.w3.org/ns/sosa/>
PREFIX ssn:  <http://www.w3.org/ns/ssn/>

PREFIX jpe: <https://seismic.balog.jp/
    ontology/jp-earthquake.ttl#>
```

[8]https://www.data.jma.go.jp/eqev/data/higai/higai1996-new.html
[9]https://elaws.e-gov.go.jp/document?lawid=353AC0000000073&openerCode=1
[10]https://www.wikidata.org/wiki/Q36204

```
<https://seismic.balog.jp/ontology/jp-earthquake.ttl#>
    a   owl:Ontology ;
    rdfs:label  "earthquake ontology for seismic dataset"@en ;
    dcterms:creator [   foaf:homepage   <https://seismic.balog.jp/> ;
                        foaf:name       "hiroki_u" ;
                    ] ;
    dcterms:issued   "2022-10-16" ;
    dcterms:modified   "2023-09-18" .

jpe:earthquake  a   owl:Class ;
    rdfs:label  "earthquake"@en ;
    rdfs:subClassOf schema:Event .

jpe:observer    a   owl:Class ;
    rdfs:label  "Station"@en ;
    rdfs:subClassOf sosa:Sensor .

jpe:seismicMotion   a   sosa:Observation ;
    rdfs:label  "seismic motion"@en ;
    jpe:shindo rdf:literal;
    jpe:hasHypocenter jpe:hypocenter ;
    sosa:isObservedBy jpe:observer .

jpe:observedWave   a   sosa:Observation ;
    rdfs:label  "observed wave"@en ;
    sosa:isObservedBy jpe:observer .

jpe:hypocenter   a   owl:Class ;
    rdfs:label  "hypocenter"@en ;
    rdfs:subClassOf schema:Event .

jpe:hypocenter   a   owl:Class ;
    rdfs:label  "hypocenter"@en ;
    rdfs:subClassOf schema:Event .

jpe:shindo  a   sosa:ObservableProperty ;
    rdfs:label  "intensity"@en .

jpe:originTime  a   rdf:Property ;
    rdfs:label  "origin time"@en .

jpe:magnitude  a   rdf:Property ;
    rdfs:label   "magnitude"@en ;
    rdfs:subPropertyOf schema:value .

jpe:magnitudeType  a   rdf:Property ;
    rdfs:label   "magnitude type"@en ;
    rdfs:subPropertyOf schema:value .

jpe:depth  a   rdf:Property ;
    rdfs:label  "depth"@en ;
    rdfs:subPropertyOf schema:value .

jpe:seismicIntensity  a   sosa:ObservableProperty ;
    rdfs:label  "Instrumental Seismic Intensity"@en ;
    rdf:subPropertyOf schema:value .
...
```

Figure 4: Earthquake Ontology

A graph created based on the vocabulary related to earthquakes, which is mainly used, is shown in Figure 4.

Since the earthquake itself cannot be observed, it is important to show the relationship between the waveform information actually observed at the observation station and the hypocenter and magnitude estimated from the observed waveform as the semantics of the earthquake. The earthquake ontology was constructed based on the hypocenter, seismic motion, observed waveforms, and observation station that constitute an earthquake. The seismic motion and the observation station that observes the waveforms at the ground surface were described using the SOSA and SSN. SSN is an ontology for describing and sharing sensor data, focusing on modeling SSN sensor networks and providing terms and relationships to describe sensor characteristics, capabilities, and observed data. SOSA provides terms and relationships to describe sensor data and related elements such as sensors, observations, samples, and actuators. The earthquake ontology observation station class inherits from the

Figure 5: Earthquake data in Hiroshima on May 25, 2019

SOSA:Sensor class, and seismicMotion and observedWave are set to the properties observed from the observation station. This is because seismic motion and observed waveform are two different classes, and observation stations, especially seismometers installed at home by individuals such as CSN or Raspberry Shake, may falsely detect non-seismic motion such as movement of people or pets, truck or bus traffic, etc., resulting in non-seismic waveforms being observed. In the earthquake ontology, the hypocenter is identified from the seismic waveforms observed by the stations, and the data set summarizing these three relationships is intended to be captured as an earthquake.

Figure 5 shows the hypocenter and observed waveforms graphically.

It can be seen that one earthquake is indicated by the waveform observed at the observation point and the hypocenter estimated from the observed waveform.

5 EARTHQUAKE LOD

We converted the available data from the JMA's Earthquake Monthly Report (Catalog Edition) and FDSN earthquake events into Linked Data based on Earthquake Ontology. The data in the JMA's Earthquake Monthly Report (Catalog Edition) is written in a fixed-length format with the character code Shift-JIS. In addition, information such as date-time, seismic intensity, magnitude, etc. is recorded in a fixed-length format for each digit in the format, as described on the JMA's Earthquake Monthly Report (Catalog Edition) website[11]. Therefore, according to the design pattern, the LOD is created through the process of data normalization including format conversion and character code change.

Since FDSN includes observation networks registered with ISC and STEAD, data outside Japan are using FDSN. The FDSN has an API[12] that can retrieve a list of stations and a list of hypocenters for each registered network. Since the data of seismic waveforms observed at each station must be collected within the site of the network, this time the data of hypocenters and stations were collected from the API and converted to turtle format using the Earthquake Ontology.

The earthquake ontology and the converted data from the JMA earthquake monthly report (catalog) and FDSN earthquake event are available at Citizen Seismology Network (CSN), which is an earthquake observation network for Yokohama citizens operated by Yokohama City University[13], and data retrieval through SPARQL endpoints (https://seismic.balog.jp/sparql).

For example, among the earthquakes that occurred after 2018 with a maximum seismic intensity of 5-upper observed, the observed waveforms with an instrumental seismic intensity of 4 or higher and the hypocenters of the earthquakes were converted to LOD data using the earthquake ontology. An example query is shown below 2. In the case of JMA's Earthquake Monthly Report and K-NET's download site, it is possible to search for observed waveforms from earthquakes and data observed by stations, but they do not support complex queries. To perform complex searches, it was necessary to download the data by specifying the earthquake, read the file, and create a local database for the search. However, using Earthquake LOD and SPARQL Query, flexible searches are now possible.

Listing 2: The earthquakes that occurred after 2018 with a maximum seismic intensity of 5-upper observed

```
PREFIX jpe: <https://seismic.balog.jp/
    ontology/jp-earthquake.ttl#>

SELECT DISTINCT * WHERE {
  ?s a <https://seismic.balog.jp/ontology
      /jp-earthquake.ttl#hypocenter> .
  ?s jpe:originTime ?origin .
  ?s jpe:shindo ?shindo.
    FILTER(xsd:dateTime(?origin) >
        "2018-01-01T00:00:00"^^xsd:
        dateTime)
    FILTER CONTAINS(xsd:string(?shindo),
        5)
  ?obs_wave jpe:hasHypocenter ?s ;
    jpe:observedBy ?sta .
  ?sta rdfs:labal ?name .
} LIMIT 100
```

5.1 Dataset in Earthquake LOD

This section describes the Earthquake LOD datasets. The Earthquake LOD was created using various seismological network data registered on FDSN and data from the JMA.

- United States National Seismic Network[2]
- Hawaiian Volcano Observatory Network[8]
- Montana Regional Seismic Network[15]
- Southern California Seismic Network[17]

[11] https://www.data.jma.go.jp/eqev/data/bulletin/data/shindo/format_j.txt

[12] https://www.fdsn.org/webservices/

[13] https://seismic.balog.jp/

- Nevada Seismic Network[28]
- Pacific Northwest Seismic Network - University of Washington[19]
- USGS Northern California Seismic Network[20]
- Alaska Geophysical Network[1]
- Oklahoma Seismic Network[25]
- University of Utah Regional Seismic Network[18]
- Raspberry Shake[22]
- Alaska Volcano Observatory[13]
- Texas Seismological Network[4]
- Puerto Rico Seismic Network & Puerto Rico Strong Motion Program[16]
- US Geological Survey Networks[3]
- Lamont-Doherty Cooperative Seismographic Network[11]
- Geological Survey Networks[26]
- National Tsunami Warning Center Alaska Seismic Network[14]

Our earthquake LOD is available at seismic.balog.jp and can be searched through the SPARQL endpoint (https://seismic.balog.jp/sparql).

FDSN has data registered until 2022, including stations not currently in operation, while the JMA has data from 1919 to 2019. This is because the JMA data is based on the Earthquake Monthly Report (catalog version), which is published only after the hypocenter is determined and other processes are completed, and the data is not updated. In this paper, FDSN data since 1970 and JMA data from 1919 to 2019 were collected and converted to LOD. Table 1 shows statistics of the Earthquake LOD.

Table 1: Number of Hypocenters and Stations

Organization	Hypocenters	Stations
FDSN	1602972	60646
JMA	100740	6795

Figure 6: all Stations

All Stations are plotted on the map in Figure 6. STEAD, ISC, and FDSN only display data outside of Japan, but using Earthquake LOD makes it possible to use more than 65,000 observation stations, including those in Japan.

Figure 7 shows the locations of hypocenters of magnitude 7 or greater that have occurred since 2010. It can be seen that the hypocenters are located along the plate and that they are concentrated in Japan.

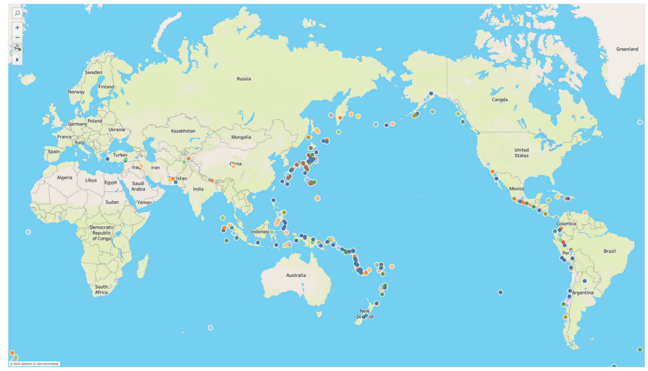

Figure 7: Hypocenter map from 2010

Table 2: Number of hypocenters over magnitude 7

Year	Hypocenters
2023	10
2022	10
2021	18
2020	9
2019	18
2018	33
2017	31
2016	42
2015	23
2014	19
2013	23
2012	24
2011	44
2010	27

Table 2 shows the estimated number of hypocenters observed since 2010 with a magnitude of 7 or more. Although there are variations from year to year, we can see that 2011, when the Great East Japan Earthquake occurred in Japan, had the largest number of hypocenters with a magnitude of 7 or higher.

Listing 3: All hypocenters from 2010

```
PREFIX jpe: <https://seismic.balog.jp/ontology/jp-earthquake.ttl#>

SELECT year(xsd:dateTime(?origin)) COUNT(*) AS ?cnt WHERE {
    ?s a jpe:hypocenter ;
        jpe:originTime ?origin ;
        jpe:magnitude ?mag .
    FILTER(xsd:dateTime(?origin) > "2010-01-01T00:00:00"^^xsd:dateTime)
    FILTER(?mag >= 7)
} GROUP BY year(xsd:dateTime(?origin))
  ORDER BY DESC(year(xsd:dateTime(?origin)))
```

Plotting hypocenters by their own latitude and longitude on the map, we can be seen that are located along the plate and that they are concentrated in Japan. These data can be retrieved with the following SPARQL Query3.

6 CONCLUSION

In this paper, we organized earthquake data sourced from the Japan Meteorological Agency's Earthquake Monthly Report and the FDSN dataset. Additionally, we developed the Earthquake Ontology using Ontology Oriented Design Patterns. The Earthquake Ontology consists of a vocabulary related to earthquakes and has been published in the form of Linked Open Data. This was achieved by assigning URIs to earthquakes based on information from observed waveforms and hypocenters.

At first, utilizing the ontology oriented design patterns, we defined the phenomenon of earthquakes, which is challenging to observe directly. This led to the creation of an earthquake LOD. After discovering data that should be converted to improve interoperability, it is useful to reuse existing ontologies and schemas in order to easily convert the data into LOD. Comprehending the original data is crucial for ontology discovery, validation, and the conversion of existing data. This complexity makes automation challenging and requires careful consideration of the workflow. In this paper, we show that it is possible to create seismic ontologies and seismic LODs by applying design patterns to the example of earthquake data that occur frequently in Japan. This paper demonstrates the feasibility of creating Earthquake Ontology and Earthquake LODs through the application of design patterns, using the example of frequently occurring earthquake data in Japan. We also aim to build graphical platforms and graphical applications for using design patterns.

Next steps, we aim to create an infrastructure for multiple observation networks by converting the latest data published in the JMA's seismic intensity database, data in NIED's seismic observation network, and seismic data in FDSN into LOD. Furthermore, the dataset will be extended by discovering the observed waveforms that observed the hypocenters registered in the FDSN and linking them with the source data and the observation station data. Observed seismic waveforms can be obtained from the data catalog of the network registered in the FDSN and the site of NIED, where observation data is publicly available, but it is necessary to construct a query based on the information of the hypocenter. Therefore, making it possible to generate parameter settings for acquisition from the Earthquake LOD created this time would promote data distribution for seismic research. By using the Earthquake Ontology to describe data from proprietary observation networks such as CSN, seismometers developed by individuals, and data from seismometers installed in smart homes, which are difficult to release in normal times, it will be possible to reuse many data in case of emergency. Furthermore, we will promote the availability of an earthquake catalog format that can be used as earthquake data and learning data, and LOD conversion of data observed by our own observation network. Benchmarking based on the same dataset is important for source determination, calculation of predicted seismic intensity, and training data for machine learning, but it is believed that datasets for reproduction are not distributed due to the fact that Japanese data cannot be redistributed and IDs are not assigned. By using the Earthquake ontology created in this paper to describe the datasets used in earthquake research in LOD, it is expected that the availability of datasets for reconstruction will be improved.

On the other hand, while the Earthquake Ontology and Earthquake LOD developed in this study have been made available to the Japanese seismic community, their practical implementation is still pending. One contributing factor is the prevalent use of older systems, such as fixed-length data in Shift-JIS format. In the future, we aim to gain acceptance in the seismic community by developing a system to convert various seismic data into LOD and other utilization infrastructures in conjunction with the dissemination of the Earthquake LOD developed in this study.

REFERENCES

[1] Univ. of Alaska Fairbanks Alaska Earthquake Center. Alaska Geophysical Network [Data set]. 1987. https://doi.org/10.7914/SN/AK
[2] Albuquerque Seismological Laboratory (ASL)/USGS. United States National Seismic Network [Data set]. 1990. https://doi.org/10.7914/SN/US
[3] Albuquerque Seismological Laboratory (ASL)/USGS. US Geological Survey Networks [Data set]. 1980. https://doi.org/10.7914/SN/GS
[4] The University of Texas at Austin Bureau of Economic Geology. Texas Seismological Network [Data set]. 2016. https://doi.org/10.7914/SN/TX
[5] Michael Compton, Payam Barnaghi, Luis Bermudez, Raúl García-Castro, Oscar Corcho, Simon Cox, John Graybeal, Manfred Hauswirth, Cory Henson, Arthur Herzog, Vincent Huang, Krzysztof Janowicz, W. David Kelsey, Danh Le Phuoc, Laurent Lefort, Myriam Leggieri, Holger Neuhaus, Andriy Nikolov, Kevin Page, Alexandre Passant, Amit Sheth, and Kerry Taylor. 2012. The SSN ontology of the W3C semantic sensor network incubator group. *Journal of Web Semantics* 17 (2012), 25–32. https://doi.org/10.1016/j.websem.2012.05.003
[6] Mariarosaria Falanga, Enza De Lauro, Simona Petrosino, Diego Rincon-Yanez, and Sabrina Senatore. 2022. Semantically Enhanced IoT-Oriented Seismic Event Detection: An Application to Colima and Vesuvius Volcanoes. *IEEE Internet of Things Journal* 9, 12 (June 2022), 9789–9803. https://doi.org/10.1109/JIOT.2022.3148786
[7] National Research Institute for Earth Science and Disaster Resilience (NIED). 2019. NIED K-NET, KiK-net,National Research Institute for Earth Science and Disaster Resilience. https://doi.org/10.17598/NIED.0004
[8] USGS Hawaiian Volcano Observatory (HVO). Hawaiian Volcano Observatory Network [Data set]. 1956. https://doi.org/10.7914/SN/HV
[9] Japan Meteorological Agency (JMA). 2023. Earthquake Monthly Report (Catalog Edition). Retrieved May 8, 2023 from https://www.data.jma.go.jp/eqev/data/bulletin/
[10] Hisahiko KUBO and Takashi KUNUGI. 2022. PREDICTION APPROACH OF RISE PART OF REAL-TIME SEISMIC INTENSITY. *Journal of Japan Association for Earthquake Engineering* 22, 6 (2022), 6:22–6:38. https://doi.org/10.5610/jaee.22.6_22
[11] Columbia University Lamont Doherty Earth Observatory (LDEO). Lamont-Doherty Cooperative Seismographic Network [Data set]. 1970. https://doi.org/10.7914/SN/LD
[12] S Mostafa Mousavi, Yixiao Sheng, Weiqiang Zhu, and Gregory C Beroza. 2019. STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. *IEEE Access* (2019). https://doi.org/10.1109/ACCESS.2019.2947848
[13] Alaska Volcano Observatory/USGS. Alaska Volcano Observatory [Data set]. 1988. https://doi.org/10.7914/SN/AV
[14] NOAA National Oceanic and Atmospheric Administration (USA). 1967. National Tsunami Warning Center Alaska Seismic Network [Data set]. https://doi.org/10.7914/SN/AT
[15] Montana Bureau of Mines and MT USA) Geology/Montana Tech (MBMG. Montana Regional Seismic Network [Data set]. 1982. https://doi.org/10.7914/SN/MB
[16] University of Puerto Rico. Puerto Rico Seismic Network & Puerto Rico Strong Motion Program [Data set]. 1986. https://doi.org/10.7914/SN/PR
[17] California Institute of Technology and United States Geological Survey Pasadena. Southern California Seismic Network [Data set]. 1926. https://doi.org/10.7914/SN/CI
[18] University of Utah. University of Utah Regional Seismic Network [Data set]. 1962. https://doi.org/10.7914/SN/UU
[19] University of Washington. Pacific Northwest Seismic Network - University of Washington [Data set]. 1963. https://doi.org/10.7914/SN/UW
[20] USGS Menlo Park. USGS Northern California Seismic Network [Data set]. 1966. https://doi.org/10.7914/SN/NC

[21] Khalid Qasim, Fernandez Alberto, Lujak Marin, and Doniec Arnaud. 2023. SBEO:Smart Building Evacuation Ontology. *Computer Science and Information Systems* 20, 1 (2023), 51–76. https://doi.org/10.2298/CSIS220118046K

[22] S.A Raspberry Shake. 2016. Raspberry Shake [Data set]. https://doi.org/10.7914/SN/AM

[23] Diego Rincon-Yanez, Enza De Lauro, Mariarosaria Falanga, Sabrina Senatore, and Simona Petrosino. 2020. Towards a semantic model for IoT-based seismic event detection and classification. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. 189–196. https://doi.org/10.1109/SSCI47803.2020.9308329

[24] Gerardo Suarez, Torild van Eck, Domenico Giardini, Tim Ahern, Rhett Butler, and Seiji Tsuboi. 2008. The International Federation of Digital Seismograph Networks (FDSN): An Integrated System of Seismological Observatories. *IEEE Systems Journal* 2, 3 (2008), 431–438. https://doi.org/10.1109/JSYST.2008.2003294

[25] Oklahoma Geological Survey. Oklahoma Seismic Network [Data set]. 1978. https://doi.org/10.7914/SN/OK

[26] U.S. Geological Survey. U.S. Geological Survey Networks [Data set]. 2016. https://doi.org/10.7914/SN/GM

[27] Hiroki Uematsu, Phuc Nguyen, and Hideaki Takeda. 2022. Design for Data Structures: Data Unification and Federation with Wikibase. In *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 6169–6178. https://doi.org/10.1109/BigData55660.2022.10020686

[28] Reno University of Nevada. Nevada Seismic Network [Data set]. 1971. https://doi.org/10.7914/SN/NN