

Earthquake Ontology and LOD

Hiroki Uematsu^{1,2,*}, Hideaki Takeda^{1,2}

¹National Institute of Informatics, Japan

²The Graduate University for Advanced Studies, SOKENDAI, Japan

Abstract

In this paper, we constructed an earthquake ontology and Linked Open Data for seismology. Earthquake ontology defines data on seismic waveforms such as seismic intensity and occurred time, data on observation stations where seismic waveforms were observed, and classes and properties such as the size and depth of the hypocenter of the observed waveforms. By using the earthquake ontology, it is possible to assign URIs to "earthquakes" that cannot be observed unlike seismic sources and observed waveforms, making it possible to use the necessary earthquake data based on observation information. We developed the seismic dataset not only the world's seismic data for machine learning represented by STEAD but also data publicly available in Japan on a limited basis, such as JMA and NIED prevention, as Linked Open Data using earthquake ontology.

Keywords

Ontology, Linked Open Data, Earthquake, Seismology

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.

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[1] 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

ISWC 2023 Posters and Demos: 22nd International Semantic Web Conference, November 6–10, 2023, Athens, Greece

*Corresponding author.

✉ hiroki_u@nii.ac.jp (H. Uematsu); takeda@nii.ac.jp (H. Takeda)

🆔 0000-0003-4215-3112 (H. Uematsu); 0000-0002-2909-7163 (H. Takeda)



© 2023 Copyright © 2023 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

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.

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. 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)[2] 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[3] registers approximately 1.2 million time series of seismic waveforms observed by seismometers, covering more than 19,000 hours of datasets.

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. Other earthquake data is available in the Earthquake Monthly Report (Catalog Edition)[4]. Although the observed waveforms themselves cannot be obtained, they can be considered to contain metadata on the observed seismic motions.

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 observe 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. 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. 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 points where the earthquake motion that occurred at the hypocenter was observed. Therefore, we created an ontology as an

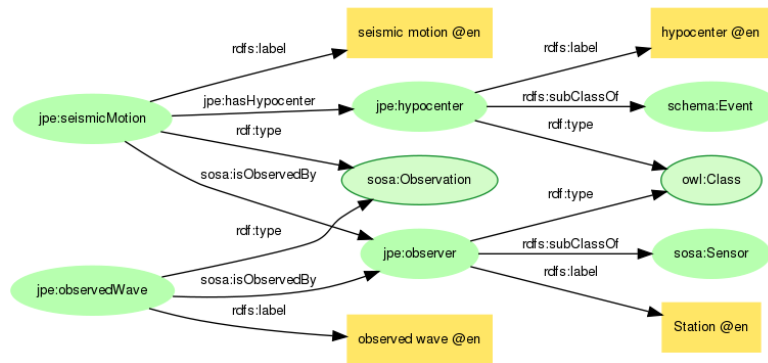


Figure 1: Earthquake Ontology (observedWave and hypocenter)

earthquake vocabulary based on the data contained in the Earthquake Monthly Report (Catalog Edition) of JMA. 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.

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. Figure 1 shows the hypocenter and observed waveforms graphically.

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 observe the waveforms at the ground surface were described using the SOSA (Sensor, Observation, Sample, and Actuator) ontology of the SSN (Semantic Sensor Network)[5]. The earthquake ontology observation station class inherits from the SOSA:Sensor class, and seismicMotion and observedWave are set to the properties observed from the observation station. 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.

4. 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. Since FDSN includes observation networks registered with ISC and STEAD, data outside Japan are using FDSN. Networks in FDSN are *United States National Seismic Network, Hawaiian Volcano Observatory Network, Montana Regional Seismic Network, Southern California Seismic Network, Nevada Seismic Network, Pacific Northwest Seismic Network - University of Washington, USGS Northern California Seismic Network, Alaska Geophysical Network, Oklahoma Seismic Network, University of Utah Regional Seismic Network, Raspberry Shake, Alaska Volcano Observatory, Texas Seismological Network, Puerto Rico Seismic Network & Puerto Rico Strong Motion Program, US Geological Survey Networks, Lamont-Doherty Cooperative Seismographic Network, Geological*

Table 1
Number of Hypocenters and Stations

Organization	Hypocenters	Stations
FDSN	1602972	60646
JMA	100740	6795

Our earthquake LOD is available at seismic.balog.jp and can be searched through the SPARQL endpoint (<https://seismic.balog.jp/sparql>). 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 2
Number of hypocenters over magnitude 7

Year	2019	2018	2017	2016	2015	2014	2013	2012	2011	2010
Number	10	33	31	42	23	19	23	24	44	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.

Figure 2 shows the location of hypocenters of magnitude 7 or greater that have occurred since 2010. It can be seen the hypocenters are located along the plate and that they are concentrated in Japan. These data can be retrieved with the following SPARQL Query (Listing 1).

Listing 1: SPARQL Query

```

1  PREFIX jpe: <https://seismic.balog.jp/ontology/jp-earthquake.ttl#>
2
3  SELECT year(xsd:dateTime(?origin)) COUNT(*) AS ?cnt WHERE {
4    ?s a jpe:hypocenter ;
5       jpe:originTime ?origin ;
6       jpe:magnitude ?mag .
7    FILTER(xsd:dateTime(?origin) > "2010-01-01T00:00:00"^^xsd:dateTime)
8    FILTER(?mag >= 7)
9  } GROUP BY year(xsd:dateTime(?origin))
10 ORDER BY DESC(year(xsd:dateTime(?origin)))

```

5. Conclusion

In this paper, we organized earthquake data registered in the Earthquake Monthly Report from JMA and data from FDSN published. We also created an ontology by organizing vocabulary related to earthquakes and published it in the form of Linked Open Data by assigning URIs to earthquakes based on information on observed waveforms and hypocenters.

In the future, 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, we will promote



Figure 2: Hypocenter map from 2010

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.

References

- [1] N. R. I. for Earth Science, D. R. (NIED), NIED k-net, kik-net, national research institute for earth science and disaster resilience, 2019. URL: <https://doi.org/10.17598/NIED.0004>.
- [2] G. Suarez, T. van Eck, D. Giardini, T. Ahern, R. Butler, S. Tsuboi, The international federation of digital seismograph networks (fdsn): An integrated system of seismological observatories, *IEEE Systems Journal* 2 (2008) 431–438. doi:10.1109/JSYST.2008.2003294.
- [3] S. M. Mousavi, Y. Sheng, W. Zhu, G. C. Beroza, Stanford earthquake dataset (stead): A global data set of seismic signals for ai, *IEEE Access* (2019). doi:doi:10.1109/ACCESS.2019.2947848.
- [4] J. M. A. (JMA), Earthquake monthly report (catalog edition), 2023. URL: <https://www.data.jma.go.jp/eqev/data/bulletin/>.
- [5] M. Compton, P. Barnaghi, L. Bermudez, R. García-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, V. Huang, K. Janowicz, W. D. Kelsey, D. Le Phuoc, L. Lefort, M. Leggieri, H. Neuhaus, A. Nikolov, K. Page, A. Passant, A. Sheth, K. Taylor, The ssn ontology of the w3c semantic sensor network incubator group, *Journal of Web Semantics* 17 (2012) 25–32. URL: <https://www.sciencedirect.com/science/article/pii/S1570826812000571>. doi:10.1016/j.websem.2012.05.003.