# Tsunami-Related Data: A Review of Available Repositories Used in Scientific Literature

**Tomáš Nacházel [1], František Babič [2], Marco Baiguera [3], Pavel Čech [1], Martina Husáková [1], Peter Mikulecký [1], Karel Mls [1], Daniela Ponce [1], Dimitra Salmanidou [3], Kamila Štekerová [1], Ioanna Triantafyllou [4], Petr Tučník [1], Marek Zanker [1] and Vladimír Bureš [1,*]**

[1] Faculty of Informatics and Management, University of Hradec Kralove, 500 03 Hradec Kralove, Czech Republic; tomas.nachazel@uhk.cz (T.N.); pavel.cech@uhk.cz (P.Č.); martina.husakova.2@uhk.cz (M.H.); peter.mikulecky@uhk.cz (P.M.); karel.mls@uhk.cz (K.M.); daniela.ponce@uhk.cz (D.P.); kamila.stekerova@uhk.cz (K.Š.); petr.tucnik@uhk.cz (P.T.); marek.zanker@uhk.cz (M.Z.)

[2] Faculty of Electrical Engineering and Informatics, Technical University of Košice, 042 00 Košice, Slovakia; frantisek.babic@tuke.sk

[3] UCL Department of Civil, Environmental and Geomatic Engineering, and Department of Statistical Science, University College London, London WC1E 6BT, UK; m.baiguera@ucl.ac.uk (M.B.); d.salmanidou.12@ucl.ac.uk (D.S.)

[4] Department of Geology and Geoenviroment, National & Kapodistrian University of Athens, 106 79 Athens, Greece; ioannatriantafyllou@yahoo.gr

\* Correspondence: vladimir.bures@uhk.cz; Tel.: +420-493332259

**Abstract:** Various organizations and institutions store large volumes of tsunami-related data, whose availability and quality should benefit society, as it improves decision making before the tsunami occurrence, during the tsunami impact, and when coping with the aftermath. However, the existing digital ecosystem surrounding tsunami research prevents us from extracting the maximum benefit from our research investments. The main objective of this study is to explore the field of data repositories providing secondary data associated with tsunami research and analyze the current situation. We analyze the mutual interconnections of references in scientific studies published in the Web of Science database, governmental bodies, commercial organizations, and research agencies. A set of criteria was used to evaluate content and searchability. We identified 60 data repositories with records used in tsunami research. The heterogeneity of data formats, deactivated or nonfunctional web pages, the generality of data repositories, or poor dataset arrangement represent the most significant weak points. We outline the potential contribution of ontology engineering as an example of computer science methods that enable improvements in tsunami-related data management.

**Keywords:** tsunami; repository; data set; data format; review; evaluation; ontology

## 1. Introduction

Tsunamis are long waves with periods ranging from a few minutes to about an hour and wavelengths from tenths to hundreds of kilometers depending on the type and dimensions of the causative source [1]. Various sources can produce tsunamis as long-propagating waves. Seismically triggered tsunamis represent approximately 80% of all tsunamis worldwide [2]. This means that most of the sudden displacements of the water column are associated with earthquakes as the main trigger. Furthermore, volcanic activity, submarine and subaerial mass wasting, atmospheric disturbances, and cosmic impacts can generate tsunamis [3,4]. Once generated, tsunamis travel at high speed and spread over a large area of water. In deep water, the tsunami wave amplitude may remain small, typically ranging up to a few meters. The waves become higher and shorter in shallow water and may have run-up heights exceeding several tens of meters. After reaching coastal areas, waves inundate land up to several kilometers in the case of large tsunamis.

Consequently, casualties and damage or ruining of infrastructure and built-up areas occur. Moreover, environmental issues, such as the destruction of the alongshore topography, erosion of the sea bottom, and devaluation of terrestrial soil, are associated with this event [5]. Figure 1 depicts the three major tsunami occurrences that played a significant role in tsunami research. The devastation of coastal areas during the 2004 Indian Ocean tsunami can be considered the first event, which has activated intensive research on the phenomenon. Then, an increase in scientific publications can be identified in connection with the 2009 South Pacific Tsunami, 2010 Maule Tsunami, and especially with the 2011 Tohoku-oki tsunami, which triggered the other wave of interest [6]. The last occurrence is associated with tsunamis on the island of Sulawesi in Indonesia in 2018.
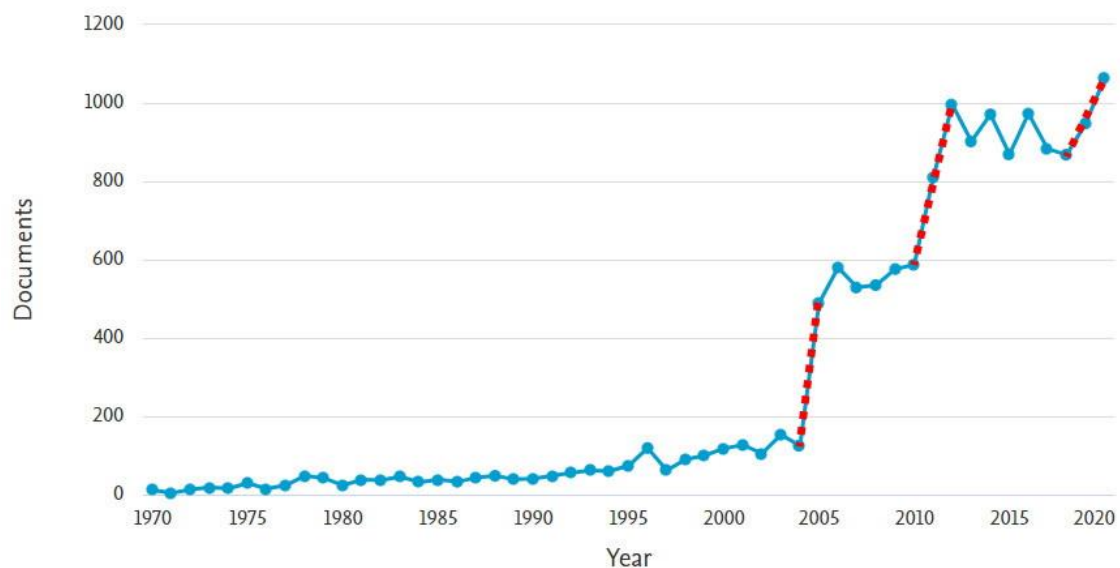


**Figure 1.** Tsunami-focused publications (authors' analysis in the Scopus database).

Anpalagan and Woungang [7] stated that almost 90% of the geological records of tsunamis in the Mediterranean Sea area were incorrectly interpreted. Because historical records in that area were accurately recorded [8], they were used to analyze the three types of predictions focused on the probability of tsunami occurrence, fatal casualties, and financial losses. The National Oceanic and Atmospheric Administration (NOAA) datasets served as sources of data. Studies of this type confirm how the availability of proper datasets and processable formats represents the primary assumption of research. Additionally, various theoretical models must be quantified to obtain the expected added value (e.g., [9–11]). The topics and use cases range from the analysis of impacts in the case of prevalent ex-post assessment of tsunami hazards and related damages [12] to the maximum earthquake magnitude scenario [13], analysis of historical events such as the 1755 CE Lisbon earthquake and the largest historical tsunami ever impacting the Europe's Atlantic coasts [14], or confirming the existence of low-level tsunamis [15].

The remainder of this paper proceeds as follows. Section 2 introduces the general context of this study and the identification of the main research issue. Section 3 describes the methodology applied in this study. Section 4 presents the achieved results with emphasis on the identified organizations and attributes of the provided repositories. Moreover, a focus on the existing data formats is provided. Section 5 discusses the main findings and outlines possible solutions in the form of an ontology. The final section concludes the study.

## 2. The Issue Description

There are two types of data: primary and secondary. The former is linked to existing monitoring or surveillance systems such as the Global Navigation Satellite System (GNSS), which is used in seismology to study ground displacements [16], or the Seafloor

Observation Network for Earthquakes and tsunamis along the Japan Trench (S-net), which is currently the world's largest network of ocean bottom pressure sensors for real-time tsunami monitoring [17]. These data are considered primary, as they are original and collected from the primary resources with the help of sensors. Although web-based services provide these data in real time, the availability and ability to process them are not straightforward and easy for researchers. Therefore, secondary data, that is, data collected by someone else and stored in a repository, are used. These data can be either experimental or empirical. The former is connected to the experiments and acquired results. For instance, Mulia and Satake [17] analyzed the efficacy of tsunami forecasting through exhaustive synthetic experiments. They considered 1500 hypothetical tsunami scenarios from megathrust earthquakes with magnitudes ranging from 7.7 to 9.1. These types of data are associated with published papers and studies. They enable the testing of various scenarios without the necessity of possessing empirical data. Empirical data are collected in the environment of interest, for instance, in the form of field surveys.

For certain types of digital objects, well-curated, deeply integrated, special-purpose datasets such as those provided by the NOAA. Various organizations and institutions store large volumes of data, the availability of which should benefit society, as it can improve decision-making before the tsunami occurrence, during the tsunami impact, and after coping with the aftermath. Nevertheless, the existing digital ecosystem surrounding tsunami research prevents researchers from extracting the maximum benefit from their research investments. We see the emergence of numerous general-purpose data repositories, at scales ranging from institutional to open globally scoped datasets. Furthermore, other specifics such as geographical location, data formats, or applied data models make the situation even more complicated from a technical perspective. The wide scale and multipurpose nature of repositories is understandable. Multidisciplinary research has been perceived as a mode of exploration or investigation with great potential to uncover new knowledge, understanding, and insight for a long time. In tsunami research, profit is expected from bridging different disciplines, helping advance disaster-related science. Tsunami research is multidisciplinary, as it is explored from the perspectives of not only natural science disciplines (e.g., geology, geomorphology, volcanology, meteorology, seismology, and geochemistry), but also technical disciplines (e.g., civil engineering or computer science) or social science disciplines (e.g., psychology or decision science). Thus, repositories associated with tsunamis are plentiful from various perspectives, such as research domains, data format, access mode, and type of institution.

This study emphasizes the existence of data-related issues in tsunami research. Pararas-Carayannis [18] provides a brief insight into the history of tsunami research, showing a significant role of data generation, storage, and sharing. Regardless of the existing volume of data, several studies stress the existing limitations in data availability and the effectiveness of their handling. To provide two illustrations, studies written by Behrens et al. [3] and Trinaistich, Mulligan, and Take [19] are reviewed. Behrens et al. [3] explored existing research gaps in the field of probabilistic tsunami hazard and risk analysis. They prioritized research gaps and evaluated whether closing a gap is a data-related issue or a problem of missing theoretical understanding. Several findings have been reported. For instance, we believe that the lack of tsunami exposure data is just as important as is modeling complicated aspects of inundation, but the former is assumed to be easier to achieve. Several other similar examples are found in this study. The second example is associated with the run-up of landslide-generated waves, which can significantly damage the environment. Although data focused on the run-up of non-breaking waves are available, there is a lack of data on the run-up of waves at the point of breaking before interaction with the opposing shore [19].

The availability of sufficient data in the required quality remains a principal bottleneck in tsunami research. There is an urgent need to improve the infrastructure to support data reuse [2]. What constitutes "good data management" is, however, largely undefined, and is left as a decision for the data or repository owner. The main objective of this study is to

explore the environment of datasets and data repositories associated with tsunami research, analyze the current situation in associated data management, and propose possible ways of coping with identified issues.

## 3. Methodology

### 3.1. Process of Selection

As this study intends to review data repositories and their datasets used in scientific research, papers published in the Web of Science database were analyzed. Using the keyword "tsunami" the database returns an initial set of papers that can be further filtered. First, the search term "water" had to be added to avoid papers dealing with tsunamis in different contexts, for example, the tsunami of obesity among children. Further, filters including language (English), document type (article), and publishing date (last five years from the search date, i.e., 1 August 2020) were applied. This procedure returned 1047 research papers from the Web of Science categories, such as Geosciences Multidisciplinary, Engineering Multidisciplinary, Civil Engineering, Meteorology, Atmospheric Sciences, Oceanography, Engineering Ocean, Geochemistry Geophysics, Engineering Marine, and Multidisciplinary Sciences. References or acknowledgments of any data repository were identified and collected. A data source was not used when the found data were no longer available, a dataset was stored on a private server without any further description, or a citation led only to another article. When a found citation pointed at a dataset in a data repository, this source was added, explored, and described by the criteria presented in the section below. When the paper did not provide a direct link to a data repository, information was searched in the relevant resources (e.g., government agencies, national and international institutions, including universities or NGOs).

### 3.2. Evaluation of Data Repositories

Different repositories provide different sets of features for browsing and searching datasets. This section lists and explains all the parameters used to describe and compare repositories. Since the availability and general overview of repositories vary from well-structured catalogs to obsolete web pages without any searchability, some details about repositories could not have been conclusively acquired.

The parameters used for the repository comparison were selected based on several aspects. First, parameters were collected either from databases focusing on data repositories or our experience with found repositories. Some of these characteristics were excluded as they were unhelpful to users in searching datasets (e.g., institution type and funding). Second, additional parameters were added during the search. Some repositories offer features specific to this field of research. Filters based on the time or location of the event that a dataset describes can serve as an example. A few repositories help users by offering a preview of data before downloading or even by the manual rating of dataset content regarding its openness and usability.

After identifying features and criteria for comparing dataset repositories, some were removed because they are either unobtainable or getting their values would cost an excessive amount of time. Comparing the overlap of datasets in all repositories is such an example. Because various formats and availability of metadata hinder automatic processing of most repositories, manual searching and comparisons of individual datasets are required. The evaluation of datasets and data repositories was based on the attributes presented in the following subsections.

#### 3.2.1. Repository Content

*Datasets total:* The total number of datasets in a repository. The number is shown when accessing the repository or is often listed as the number of results for an empty search. (Note that the number of datasets differs from the number of files as datasets usually contain more than one file.)

*Tsunami datasets:* The number of datasets that the repository lists as search results of the term *tsunami*. Therefore, only datasets with this word in the title, description, or keywords were counted. Note that many valuable datasets for tsunami research are not tagged by the *tsunami* keyword (e.g., general bathymetry data). Hence, this number is not conclusive because none of the repositories offer a *usability-for-tsunami* filter. It serves solely as an approximate indicator of the orientation of the repository. If a repository lacks a search feature, then the number is not investigated unless it contains only units of datasets allowing manual count.

*Last update:* The year of the last update or addition of dataset directly related to tsunami (i.e., set of datasets described in the previous paragraph). It provides a rough estimate of the activity of content creators on a repository of tsunamis. The condition for acquiring this repository parameter is a sorting feature based on the last updates or additions. Alternatively, it was collected for repositories with a limited number of datasets, thus allowing a manual check.

*Repository domain:* Repositories have varying breadth of focus. This characteristic describes a preference for a specific area of interest if there is one. Generally, catalog repositories are either completely general or have broad topics covering various research areas to encourage public contributors to share data on their platforms. Databases and web presentations typically keep the narrow focus of the organization responsible for the data, which is also allowed by the unified format of all presented data.

### 3.2.2. General Availability

*Availability:* Online availability of repositories. Most of them had free access without restrictions. Some repositories may be behind a paywall or require logging. For restricted repositories, there may be different levels of access to the data. While the search feature and viewing metadata of datasets may be free, it requires a paid account to download data.

*Downloadable data:* Datasets usually need to be downloaded to obtain data. However, some listed repositories are structured collections of links to datasets on other platforms. Here, the repositories contained only metadata. To download data, a user should acquire data from the linked storage.

*Data usability rating:* Repositories rate the usability of stored datasets based on various criteria. The objective of this rating is to indicate the level of potential obstacles during data processing. Standardized rating schemas exist (e.g., Tim Berners-Lee's Five Stars of Openness), but repositories create their own system to incorporate their specific features.

*Metadata:* Metadata is a structured description of datasets. It may be downloadable as a JSON or XML file or listed in a table on the dataset profile page. This feature enables the automatic processing of datasets outside a repository or reading of additional details that are not explicitly stated in a repository's dataset profile.

*Dataset preview:* Some repositories allow users to preview data before downloading it. This feature enables users to investigate the structure of data and make the selection of appropriate data easier. However, this is impossible in every format.

### 3.2.3. Filter/Search Options

*Search:* The ability to search for terms in title or metadata is an essential feature of all databases or online catalogs. However, some included repositories were in the form of a simple list of links to the datasets. These repositories usually contain only a few datasets and focuses on presenting a project or organization rather than offering a structured catalog of datasets.

*Dataset filter:* If the repository contains items other than datasets, this parameter describes whether it is possible to filter results to datasets only. This feature simplifies the search for users interested in data. Repositories with sections dedicated to datasets and repositories containing only datasets are appropriately marked (e.g., *datasets only*), as they are missing this filter.

***The location filter:*** Unlike repositories with geological focus, general repositories usually miss the location filter because data validity is not usually limited to a location. However, this recognition is important for searching for historical data about tsunamis in a specific region.

***Field/topic filter:*** The ability of filter-selected topics or research fields is useful for general repositories with a great variety of datasets. It allows users to browse among possibly related datasets without necessarily knowing the exact keywords that label the required data.

***Format filter:*** Datasets contain data in various formats; some are proprietary, and some are not suitable for automatic processing. Therefore, this filter helps to narrow a set of datasets to one fitting an intended use.

***License filter:*** Datasets are shared under various licenses. Research or educational use is usually not limited and requires only a citation, but some datasets may be restricted for noncommercial use. Thus, some repositories offer a license filter along with a license statement in the dataset details.

***Year/Date filter:*** Whether the user searches for data from a specific historical event or prefers to look through recent data only, a time-based filter is a useful tool to narrow datasets to the most relevant set. Note that only repositories focusing on historical events offer this filter based on the date of an event. The vast majority of repositories that allow this filtering consider only dates of addition or updates of datasets, which does not correspond to the date of the event that the dataset may be describing.

***Tsunami magnitude filter:*** When browsing tsunami data, some tsunami databases (other types of repositories do not offer such specific features) allow users to filter results based on the magnitude of tsunamis. All magnitude and intensity scales were included in this characteristic because this feature is rare, and repositories in some occurrences do not specify the exact scale they used.

*3.3. Ontological Contribution into the Data Repositories and Datasets*

To outline possible solutions to the identified data-related issues, an approach based on ontological engineering as a subarea of artificial intelligence is provided. Ontology represents a well-defined collection of concepts that describe a specific domain. Concepts are the abstractions of a particular set of instances, that is, in ontology engineering, concepts are regarded as classes and instances as individuals. An example of a class can be a data repository, and an example of an individual is the Japan Tsunami Trace Database. Ontology also encompasses the links between individuals that are described as special types of concept properties. A dataset containing wave parameters can thus be linked to a dataset containing geospatial parameters. Even more advanced hierarchical relations can be represented, allowing concepts to be generalized or aggregated into other concepts to decompose even the most complex domain. Concepts and properties can be associated with various types of logical constraints that enable the inference of facts not explicitly stated in the ontology. Ontology is expressed as a graph-based structure in which the nodes represent concepts, and the edges represent relations. In this respect, ontology can be viewed as a semantic map of a given domain that can serve to navigate in that domain using complex queries. For instance, we have a statement: "*Oceanography is Earth and Environmental Science.*". This statement can be easily expressed in the resource description framework (RDF), a data model, which is the actual standard for semantic graph database development. Oceanography (a subject/an instance) and Earth and Environmental Science (a parental class) are uniquely identifiable "resources." Natural science is the parental class for Earth and Environmental Science (a child class). Various methodologies can be used to develop formal ontologies [20–22]. The Noy and McGuiness methodology [23] was used for the ontology development in this study. It is based on an iterative developmental approach and can be used for any kind of application domain and developmental tool. This methodology is influenced in part by the Protégé environment, which was also used in our study for ontology development. It provides the following seven developmen-

tal phases for ontology building: domain and scope specification, the reuse of existing ontologies, enumeration of important terms with their properties, class definitions and class hierarchy development, modeling properties of classes, and inclusion of details for properties and instance modeling. The categories are based on the International Disaster Database EM-DAT [24], which is slightly customized for tsunami research, and the Library of Congress Recommended Formats Statement (2020–2021), providing categories of creative content [25].

## 4. Results

This section presents the presentation of required results. A list of identified repositories and data resources is provided, and an evaluation of a single resource is provided on a pre-defined scale of alternatives. Then, an analysis of existing formats is presented.

### 4.1. Repositories

Altogether, 60 repositories with tsunami-relevant datasets were identified. Table A1 in Appendix A provides basic information on the resources found. The acquired list represents the number of datasets that the repository lists as the search results of the term tsunami. Therefore, only datasets with this word in the title, description, or keywords were counted. Many more useful datasets for tsunami research exist; however, they are not tagged with the tsunami keyword (e.g., general bathymetry data). Hence, this number is not conclusive because none of the repositories offer a usability-for-tsunami filter. It serves solely as an approximate indicator of the orientation of the repository. If a repository lacks a search feature, then the number is not investigated unless it contains only units of datasets that allow manual counting.

There are three types of repositories: catalog, database, and presentation. Catalog offers a sortable list of items usually accompanied by search and filter features. The items in the catalog represent individual datasets. Here, datasets are usually uploaded by multiple organizations. A database is also a sortable list of items, but these items already represent individual records. A database can be perceived as a single-structured dataset. They are usually focused on a narrow topic (e.g., database of tsunamis and water level). The organization operating the database is responsible for inserting data into the database. In this context, the presentation refers to a static web page presenting a single dataset or a non-sortable list of links that lead to projects or datasets. There are usually no searching or filtering features, as the presentation pages show few items. Its purpose is to share results from a project or organization; therefore, the organization running the presentation page is responsible for the data as well.

From a general perspective, repositories incorporate distinct volumes of datasets ranging from hundreds of thousands (global multidomain resources such as Data.gov, Mendeley, or OSF share) to single units (e.g., the Novosibirsk Tsunami Laboratory or the Japan Tsunami Trace database). From the perspective of tsunami research, the volume of datasets is significantly lower, ranging from thousands (e.g., Pangea) or hundreds (e.g., Data.gov or Data World) to single units (e.g., the Queensland Government database or the Humanitarian Data Exchange). Repositories were created and maintained by private organizations, public institutions, or governmental bodies. Twenty-nine data repositories have been updated during the last three years, which indicates the general usability of the current research. It is possible to find data at a global scale, that is, data are associated with various geographical locations from Australia to the Mediterranean Sea and the United States. Apparently, data related to tsunami-jeopardized regions with advanced technologies such as Japan (IRIDeS), European countries (EMODnet), and the United States (NOAA, NASA) are available in large volumes. Multinational organizations such as the World Bank Group or the World Health Organization support tsunami-related research with data repositories. This is not the case for regions with less developed countries. Therefore, global technologically intensive initiatives and activities are crucial for tsunami research. NOAA or the Japan Tsunami Trace Database with tens of thousands of records can serve as

examples. Unsurprisingly, considerable heterogeneity is the main attribute of the generated list. From the domain perspective, repositories contain data related to various disciplines, such as seismology, meteorology, hydrology, and bathymetry. This makes the list of data repositories difficult to compare, and evaluation bias is almost inevitable.

Most of the datasets had free access without any restrictions. Some repositories may be hidden behind a paywall or require logging in, for example, the European Marine Observation and Data Network. For restricted repositories, there may be different levels of access to the data. Moreover, websites such as Study of the Tsunami Aftermath and Recovery (STAR) does not always respond. While the search feature and viewing metadata of datasets may be free, it requires a paid account to download data. Datasets need to be downloaded to obtain the data. However, some listed repositories do not enable downloads (e.g., the Japan Tsunami Trace database) or are structured collections of links to datasets on other platforms. Here, the repositories contained only metadata. To download data, the user must acquire data from linked storages. Data usability is likely to represent the most serious issue associated with datasets. Some repositories rate the usability of stored datasets based on various criteria. The objective of this rating is to indicate the level of potential obstacles during data processing. Standardized rating schemas exist (e.g., Tim Berners-Lee's Five Stars of Openness), but repositories often create their own system to incorporate their specific features. Metadata represent the structured description of datasets. It may be downloadable as JSON or XML files (see the discussion section) or listed in a table on the dataset profile page. This feature allows the automatic processing of dataset outside a repository or reading additional details that are not explicitly stated in the dataset profile in a repository. Only a few datasets enable metadata download (e.g., Queensland Government or the Spanish National Center of Geographic Information). However, these repositories are mostly general data repositories with a small fragment of tsunami-focused datasets. This is also the case for datasets that only enable the view of metadata. As for the search or filtering abilities, the capability to search for terms in the title or metadata and use filter queries are essential features of all databases or online catalogs. We identified only one repository that enabled the filtration of a search query by location, topic, file format, license, and time. However, most datasets were only in Portuguese. However, some of the identified repositories were in the form of a simple list of links to datasets. These repositories usually contain a few datasets and focuses on presenting a project or organization rather than offering a structured catalog of many datasets. The evaluation of repositories is presented in Table A2.

*4.2. Data Formats*

Repositories and datasets are associated with dozens of data formats. In this section, we introduce and explain the main characteristics. Table 1 presents all identified formats available for the tsunami topic as they occur in the two biggest and most established data repositories, NOAA and Data.gov. The frequency of occurrence of tsunamis was introduced.

**Table 1.** General overview of frequencies of dataset formats in the most important repositories NOAA and DATA.GOV.

| Format | NOAA_Tsunami | DATA.GOV_Tsunami |
|---|---|---|
| application/x-netcdf | 643 | 0 |
| HTML | 486 | 0 |
| WMS | 407 | 262 |
| WCS | 375 | 182 |
| CSV | 275 | 282 |
| Esri REST | 192 | 240 |
| PDF | 181 | 199 |
| KML | 89 | 93 |
| ZIP | 10 | 78 |
| XML | 9 | 6 |
| WFS | 7 | 7 |
| KMZ | 3 | 0 |
| TIFF | 0 | 10 |
| JSON | 0 | 2 |
| TAR | 0 | 2 |
| RDF | 0 | 1 |

Of this number, eight formats proved to be the most common in the "tsunami" category (they represent more than 90% of records in particular databases). We provide examples of datasets using these data formats (all accessed on 8 August 2021).

4.2.1. General Formats
NetCDF

The Network Common Data Form (NetCDF) represents a community standard for sharing scientific data. It is a set of software libraries and machine-independent data formats. These formats support the creation, access, and sharing of array-oriented scientific data.

- https://data.noaa.gov/metaview/page?xml=NOAA/NESDIS/NGDC/Collection//iso/xml/NTWC_Waterlevel_Collection.xml&view=getDataView&header=none

HTML

Unlike other formats, this is usually not a downloadable file and links to project web pages, metadata, or even file downloads from different sources.

- https://catalog.data.gov/dataset/developing-solid-earth-modeling-and-simulation-capabilities-for-future-gravity-satellite-m
- https://data.doi.gov/dataset/radiocarbon-age-dates-for-sections-of-19-sediment-cores-from-offshore-puerto-rico-and-the-u-s-v
- https://data.europa.eu/data/datasets/0195b79b-e28f-4de9-ba1a-a6744a4bf0a9?locale=en

CSV

This is a general table format used over similar ones because it is not dependent on a specific application to read. It often contains historically recorded values or lists of items with an array of parameters.

- https://catalog.data.gov/dataset/port-of-los-angeles-tsunami-evacuation-routes-signs
- https://data.doi.gov/dataset/sediment-grain-size-distributions-of-three-carbonate-sand-layers-in-anahola-valley-kauai-hawaii

ZIP

This is a general archive file containing files of any other format (including formats unsupported by the repository). It includes batches of supplementary files for specialized software, an additional description of how to use data, or all other files connected to a dataset for easy download.

- https://catalog.data.gov/dataset/port-of-los-angeles-tsunami-evacuation-routes-signs
- https://data.doi.gov/dataset/radiocarbon-age-dates-for-sections-of-19-sediment-cores-from-offshore-puerto-rico-and-the-u-s-v

PDF

This is a printable text/presentation file. Because of the structure of this format, its content is not designed for automatic processing. It is a paper or manual explanation dataset and other complementary or legal information.

- https://catalog.data.gov/dataset/probzones-generalized-100-and-500-year-flood-zones-for-seaside-oregon-determined-by-probabilis0
- https://catalog.data.gov/dataset/noaa-wds-paleoclimatology-lmdz-inca-climate-chemistry-model

### 4.2.2. Mapping Formats

These formats contain a data layer meant to be placed over a map to mark points/areas of interest with additional information. Some formats may be downloadable and viewable as XML, but they need to be viewed over a map with the appropriate tool for human readability. It provides an overview of historical tsunamis, sea levels, or other geospatial data on a map and offers additional information.

### KML

Keyhole Markup Language (KML) is an XML notation made for 2D and 3D Earth browsers (e.g., Google Earth). It is one of the international standards of the Open Geospatial Consortium.

- https://catalog.data.gov/dataset/port-of-los-angeles-tsunami-evacuation-routes-signs
- https://data.humdata.org/dataset/hotosm_idn_sunda_strait_tsunami_waterways
- https://catalog.data.gov/dataset/misr-level-2-aerosol-parameters-v002

### WMS

Additionally, localized data, the Web Mapping Service contain map images. The Open Geospatial Consortium developed it.

- https://catalog.data.gov/dataset/california-state-waters-map-series-offshore-of-pigeon-point-web-services88b54
- https://data.doi.gov/dataset/california-state-waters-map-series-salt-point-to-drakes-bay-web-services
- https://www.geonode-gfdrrlab.org/layers/hazard:ts_mih_rp10

### WCS

Web Coverage Service is a more dynamic map format than WMS, as it can represent time-varying phenomena. The Open Geospatial Consortium also contributed to this creation.

- https://catalog.data.gov/dataset/chignik-alaska-1-arc-second-coastal-digital-elevation-model
- https://catalog.data.gov/dataset/1-9th-arc-second-digital-elevation-models-dems-usgs-national-map-3dep-downloadable-data-collect
- https://catalog.data.gov/dataset/1-arc-second-digital-elevation-models-dems-usgs-national-map-3dep-downloadable-data-collection

### Esri REST

The map format developed by Esri. It can be viewed in the ArcGIS application.

- https://catalog.data.gov/dataset/california-state-waters-map-series-offshore-of-pigeon-point-web-services88b54
- https://data.doi.gov/dataset/california-state-waters-map-series-salt-point-to-drakes-bay-web-services

## 5. Discussion

This study builds on the work of Gusiakov, Dunbar, and Arcos [26], who outlined and discussed the existing issues with data compilation, cataloging, and distribution, as well as the incompleteness of certain types of data. Hence, we intend to support the improvement of data management in tsunami research. The importance of archiving data in this domain is in fact the same as in other disciplines: verification of published results, better meta-analysis, new questions, increased citation and credit, new opportunities for teaching and learning, and reducing the risk of loss [27].

*5.1. Identified Issues and Perspectives*

Sharing data usable in tsunami research has several advantages. Different interpretations or approaches to existing data contribute to scientific progress, especially in a multidisciplinary setting characteristic of tsunami research. Proper management and long-term preservation help retain data integrity. Furthermore, when data are available, re/collection of data is minimized, optimizing resource use. Finally, the availability of data enables replication studies, which can be used as training tools for new tsunami researchers [28]. While sharing data is the first step toward reuse, it is also critical that the data be simple to understand and use [29]. However, proper data management is not a goal, but rather is the key conduit leading to knowledge discovery and innovation [30], and subsequent data and knowledge integration and reuse by the community after the data publication process.

White et al. [29] suggested nine recommendations for improving data management in research: sharing your data, providing metadata, providing an unprocessed form of the data, using standard data formats, using good null values, making it easy to combine your data with other datasets, perform basic quality control, use an established repository, and use an established and open license. Unfortunately, this study reveals that most recommendations are not met in the case of tsunami-related data repositories. This is somewhat in contradiction to the opinions of scientists within specific disciplines [28].

The analysis of data repositories reveals issues with which researchers searching for reliable data should cope with. First, the most common technique for data management in the form of metadata description is insufficient for many datasets. For our analysis, their existence was not as necessary as it was for researchers who needed data for their experiments or decision-makers for their decisions. There is also an overlap among various databases, which may seem to be an advantage. However, in some cases, redundancy can lead to confusion because research may need the latest version of the dataset or work in distributed teams. Thus, coordination or synchronization might be a more significant issue than expected. Initially, there were other evaluation criteria on the list, which, in the end, remained unused. To give two examples, the detailed orientation of the repository would be an interesting piece of information. The problem is that it is not usually specified, and it is necessary to go inside and through datasets to determine whether there is no topic filter available. Furthermore, licenses associated with repositories are rarely specified, as they are usually a property of individual datasets. This criterion would be applicable if an organization running the repository is also the author of its content.

We encountered issues during the evaluation process, which exhibited weak points of existing repositories:

1. Data resources are heterogeneous and poorly arranged, which prevents automatic machine processing. Moreover, in some cases, even searching or filtering tools are missing, which significantly reduces the effectiveness of manual work with the source repository.
2. Even the most significant actors in the field, such as NOAA or data.gov, change the form of presentation or search in their repositories from time to time [31]. Although this issue seems minor, user interface or interface usability plays a significant role when a huge volume of data needs to be searched and processed.
3. Research papers and studies refer to datasets that are not directly associated with tsunamis (e.g., general geography), but their data can be used, and it is impossible to identify them when searching with relevant search terms. This reveals that the demarcation line between the tsunami and non-tsunami fields of study is difficult to define. The multidisciplinary nature of tsunami-oriented research makes the analysis of datasets and repositories more complicated.
4. The semantic differences among concepts of datasets, data, resources, and repositories generate confusion. These concepts are used in various contexts. The development of a virtual data collection system can help improve the organization of tsunami-related datasets.

5. There are many deactivated, nonfunctional, or unavailable files, even found during the search in the dataset. This issue is typical of the outcomes of research projects. Project documents or data are available only within the sustainability period, after which websites or interfaces are not managed or maintained.

6. Although there are datasets offering one or more formats of the same data, there are specific formats of data associated with specific software applications unreadable for standard available SW solutions. Typically, old data prepared for obsolete applications are impossible to run in existing operating systems.

7. Noise is often present in the data that must be filtered out, and void data that need to be dealt with (at least from the modeling side).

8. Not all datasets are the primary resources and only contain a reference. However, their features can be used as catalogs or guideposts as they work with datasets more appropriately than pages in which datasets are originally uploaded.

*5.2. Demonstration of Ontology-Engineering Help*

Various efforts leading to improved data have already been made. For instance, Murnane et al. [32] considered the lack of a consistent data structure, which hinders the development of tools that can be used with more than one set of data. They report on an effort to solve these problems through the development of extensible, internally consistent schemas for risk-related data. This study contributes to these endeavors and outlines possible solutions in the form of ontology-based systems. In the domains of natural hazards, natural disasters, disaster management, or emergency management, ontologies are mentioned in two lines of research:

1. The first line of research is focused on the usage of ontologies for categorization of concepts related to the above-mentioned domains, sharing of these ontological structures between interested parties (humans, humans and computers, or between computers), and system interoperability.

2. The second line illustrates how ontologies can be directly integrated or connected to the designed system.

As for the first line, the Wikipedia project provides a huge collection of information related to different application domains, including disaster or emergency management, categorization of natural hazards, and natural disasters. The issue is that the information presented by Wikipedia is not well machine processable. More specific queries defined by the user often fail. The DBPedia project [33] solves this issue by encoding facts (found in the Wikipedia info-boxes) into more formal structures expressed in RDF. Wikidata is a complementary project to DBPedia, which is continuously updated by users and bots (computational autonomous robots). The content of the DBPedia was automatically extracted from Wikipedia. Vocabularies related most to natural hazards or disasters are cross-domain or geography-related. If we are interested in vocabularies related to emergency management, we can visit the Linked Open Data (LOV) web. The vocabulary used for the annotation dataset repositories expressed in RDF is available in [34]. The authors of the paper did not find any ontologies (vocabularies) used for annotating datasets expressed in other formats (or datasets repositories).

As for the second research line, an ontology-based conceptual framework is proposed for improving shared situation awareness among teams of rescuers in case of emergency incidents. Mass evacuation during a tsunami event is a case study for framework demonstration [35]. Infrastructure failures caused by natural hazard events were modeled using the InfraRisk ontology. A software prototype using infrared ontology was introduced in [36]. It also provides a visualization of the data published using the ontology. Zhong et al. [37] presented a meteorological disaster system in which an ontological approach was used to model the domain knowledge of meteorological events, emergency management, disaster-specific knowledge, and geographical (geospatial) characteristics. Sermet and Demir [38] presented a different system called the flood artificial intelligence system. It is based on the flood ontology, which covers geological hazards, meteorological haz-

ards, diseases, wildfires, floods, monitoring devices, and environmental concepts. It is a question-answering and decision support system that can provide factual responses using domain-specific ontological knowledge in case of flood-related events. Voice-based and text-based communication channels are available to users. The unified knowledge-based Crisis Response Ontology (CROnto) was introduced in [39]. This ontology provides a sharable vocabulary for facilitating communication and problem-solving between emergency response organizations during disaster events. It is obvious that during disaster hazards, such as earthquakes, fast reactions are inevitable for mitigating damage to life and property. Formal ontology has been developed and integrated into a rule-based and case-based reasoning system that develops recommendations based on similar cases (disaster events) from the past. Ontology has been used to manage earthquake data in intelligent systems [40]. Liu et al. [41] presented a knowledge model called the geologic hazard emergency response (GHER) used for modeling emergency knowledge, which is inevitable for providing a fast emergency response during geological hazards. This model has been implemented in the GHERS system.

The main purpose of the ontology is to provide a semantics-based structure that can help the user.

- To receive fundamental insights into tsunami-related and tsunami-not-directly related data repositories.
- To discover which characteristics are shared by more data repositories.
- To explore the backbone of the ontology consisting of core ontological classes together with the relationships between them.
- To ask concrete questions on data repositories and related facts.

We have developed a semantic graph database according to the Noy and McGuiness methodology [23] (Section 3.3). All phases of the methodology are fulfilled, except for the modeling definitions of the classes. Only descriptions of the selected classes were modeled. The semantic graph has 105 classes. These classes model type of access into data repositories (access class), application domains of datasets (domain class), types of formats of datasets (Format class), languages in which datasets are expressed (LanguageFamily class), owners of data repositories (owner class), categories of data repositories (repository class), possible datasets (dataset class), locations from which the datasets are received (location class), and various types of disasters (disaster class). The domain class, format class, dataset class, and disaster class are the most structured parts of the class-based layer. Therefore, these parts are shown in Figures 2–5. The domain class is extended by more specific application domains, where the main attention is paid to the domains related to tsunamis. Because of the readability of the whole domain class taxonomy, only classes without a data-based layer (instances) are presented.

More specific formats are added to the format class. This categorization of formats is based on the Library of Congress Recommended Formats Statement (2020–2021) [25]. Because of the readability of the entire format class structure, only classes without the data-based layer (instances) are visible. The dataset class categorizes datasets according to the data, that is, what is the content of the datasets and what the user can directly find in them. The dataset class is "divided into" SeaDataset and AboveSeaLevelDataset classes. SeaDataset class model kinds of data received from the sea-level and below sea-level measurements. AboveSeaLevelDataset class model kinds of data measured above sea-level, for example, from the atmosphere. The disaster class extends the ontology by various disaster types, including tsunamis and meteotsunami. The user can receive wider insight into how the tsunami can be classified next to other disasters and which relationships exist between them. Categorization of natural disasters is based on the International Disaster Database (EM-DAT) [24], which is slightly customized for tsunami research.

Formal statements are modeled using predicates (properties). The following are distinguished in the semantic graph database (Table 2): Object property hasTsunamiDataset is a general relationship indicating whether it makes sense to visit a dataset repository for browsing tsunami-related datasets. If the user wants to know which dataset repositories,

including datasets provide e. g., data collected by tsunami detection buoys or tide gauges, the rdf:type (is-a/instanceOf) relationship is used for this purpose. Datasets are categorized into more specific classes, based on which it is possible to determine the nature of the dataset, i.e., which types of data the datasets include. Figure 6 depicts the relationships between one concrete dataset repository (Kaggle) and one tsunami-related dataset (data related to the volcano-induced tsunami) volcano tsunami.csv. Figure 6 is prepared in the OntoGraf Protégé plugin, which does not provide datatype properties. Datatype property containsDataOfTimeScaleMin and containsDataOfTimeScaleMax (see Table 2 for a deeper explanation) are an inherent part of the datasets modeling in the ontology.
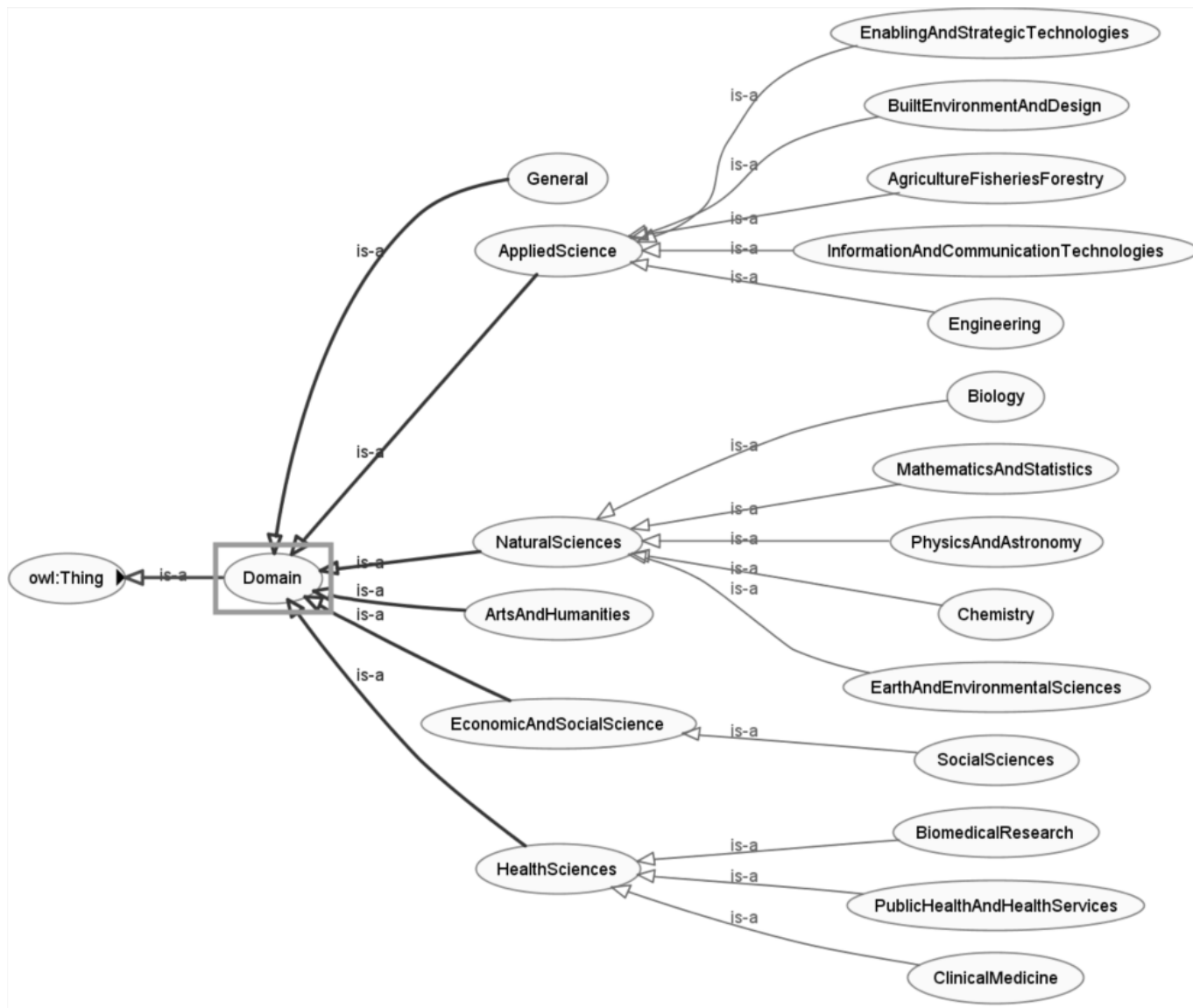
**Figure 2.** Categorization of the tsunami-related application domains of data repositories.

**Figure 3.** Categorization of the tsunami-focused data formats provided by data repositories.



**Figure 4.** Categorization of the datasets related to the tsunami research.
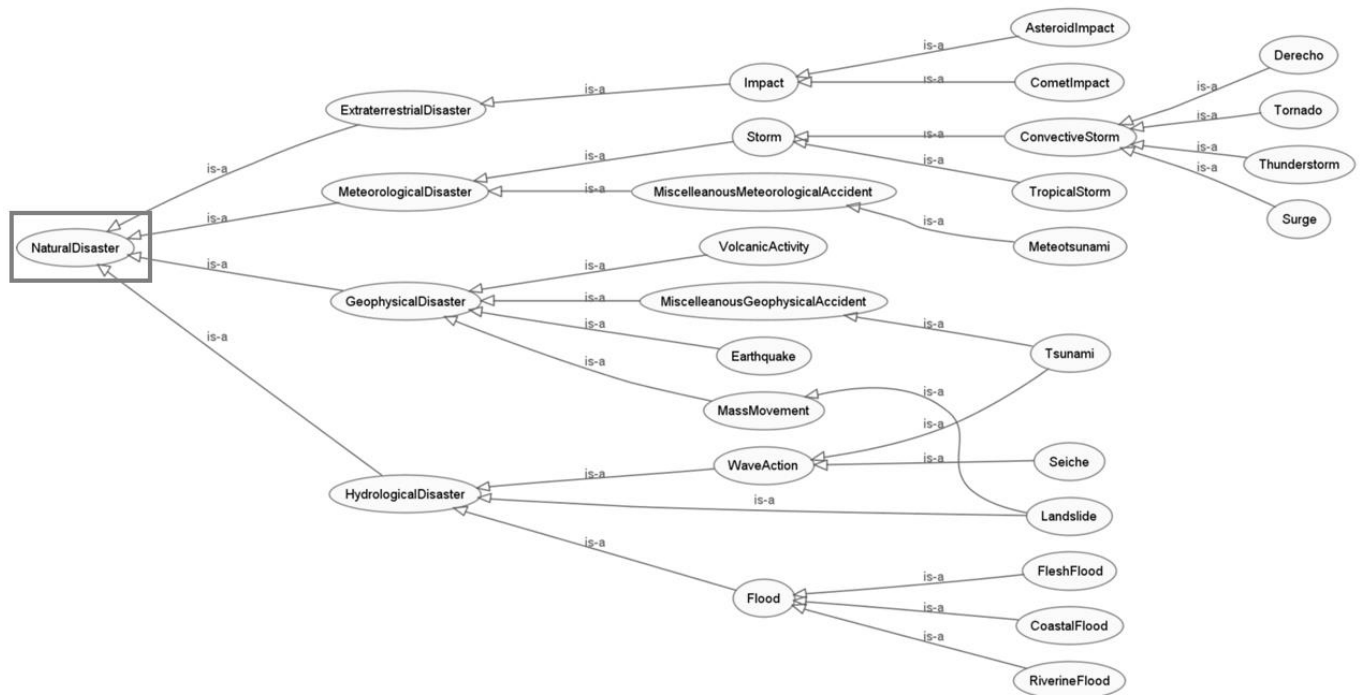
**Figure 5.** Categorization of natural disasters (adjusted according to [24]).

**Table 2.** Domain-specific predicates (properties).

| Domain-Specific Property | Type of Property | Purpose |
|---|---|---|
| hasAccess | object property | How is a data repository available (paid, free, or under registration)? |
| hasDomain | object property | Which application domain is a data repository interested in? |
| hasOwner | object property | Who is the owner of the data repository? |
| hasPart/isPartOf | object property | Relationship between whole and its parts. |
| providesFormatOfDataset | object property | Which data formats are available in the data repository? |
| providesLanguageOfDataset | object property | Which language is data sets expressed in? |
| areUsedForStudy | object property | Which data are used for the investigation of which disasters? |
| containsDataFrom | object property | Which location data are come from? |
| hasDataForDownloading | datatype property | Does a data repository provide datasets for downloading? |
| Types of properties for datasets filtering: hasDatasetFilter hasDomainFilter hasLicenseFilter hasLocationFilter | datatype property | Does a data repository provide filters for datasets, their domains, licenses, or locations? |
| hasMetadataForDownloading | datatype property | Is it possible to download the metadata of datasets? |
| hasSearchField | datatype property | Is there any search functionality? |
| hasTimeScaleFilter | datatype property | Is it possible to filter datasets according to a timescale? |
| hasUsabilityOfRating | datatype property | Is there information about the rating of usability? |
| offersPreviewOfDataset | datatype property | Is it possible to preview the datasets? |
| hasTotalDatasets | datatype property | How many datasets are in the data repository? |
| hasTsunamiDataset | datatype property | How many datasets related to the tsunami are in the data repository? |

**Table 2.** *Cont.*

| Domain-Specific Property | Type of Property | Purpose |
|---|---|---|
| containsDataOfTimeScaleMin | datatype property | When were the data of datasets measured (min. year-month-day)? |
| containsDataOfTimeScaleMax | datatype property | When were the data of datasets measured (max. year-month-day)? |
| alternativeName | annotation property | Expression of an alternative name for the data repository. |
| description | annotation property | Specification of more details of data repository. |
| identifier | annotation property | Identifier of the data repository (if it is available). |
| url | annotation property | URL link to the data repository. |
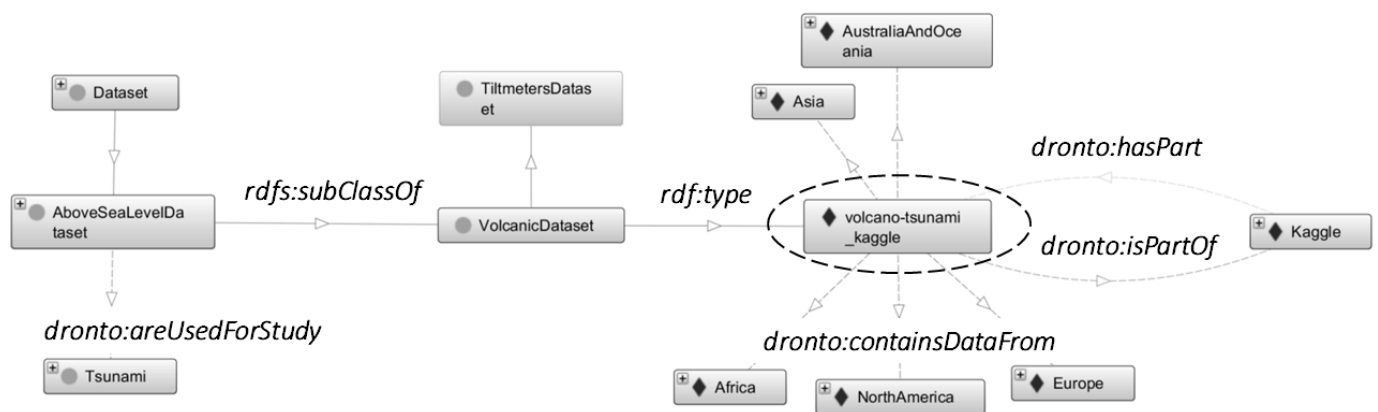


**Figure 6.** Relationships between datasets repositories and datasets (the example).

The RDF-based semantic graph database in Figure 7 depicts the fragment of the semantic graph database, where 15 web-based data repositories are visible: Data.gov Catalog, US gov–Department of the Interior, OSF Share, OSF Home, Japan Tsunami Trace database, NCEI (formerly NGDC), Queensland Government, EM-DAT Public, Fishare, Science Data Bank, Kaggle, Data World, Harvard Dataverse, Google—Dataset Search, and World Bank Water data—Data catalog. This data repository has links to various resources (individuals), which provide more detail about the repository, e.g.,:

- hasDomain; MeteorologyAndAtmosphericSciences, EnvironmentalSciences, or Oceanography.
- hasOwner; U.S.GeneralServicesAdministration.
- hasAccess; free.
- providesFormatOfDataset; PDF, XML, HTML, or TIFF (most cited data formats).
- providesLanguageOfDataset; English.

Consequently, Protocol and RDF Query Language (SPARQL), a W3C standard for querying RDF graphs, can be used for "mining" the content of the semantic graph database with data repositories. Examples follow.

Figure 8 shows how SPARQL-based querying is realized in the Protégé ontological editor. The user is the author of the SPARQL queries encoded in the SPARQL Query tab available in Protégé. The query is the input for the query engine, which searches the content of ontological and data layers. Specific results are provided in the resulting table based on the SPARQL-based query structure.

**RDFS-based layer**
(TBOX – Terminological part)

**RDF-based layer**
(ABOX – Assertional part)
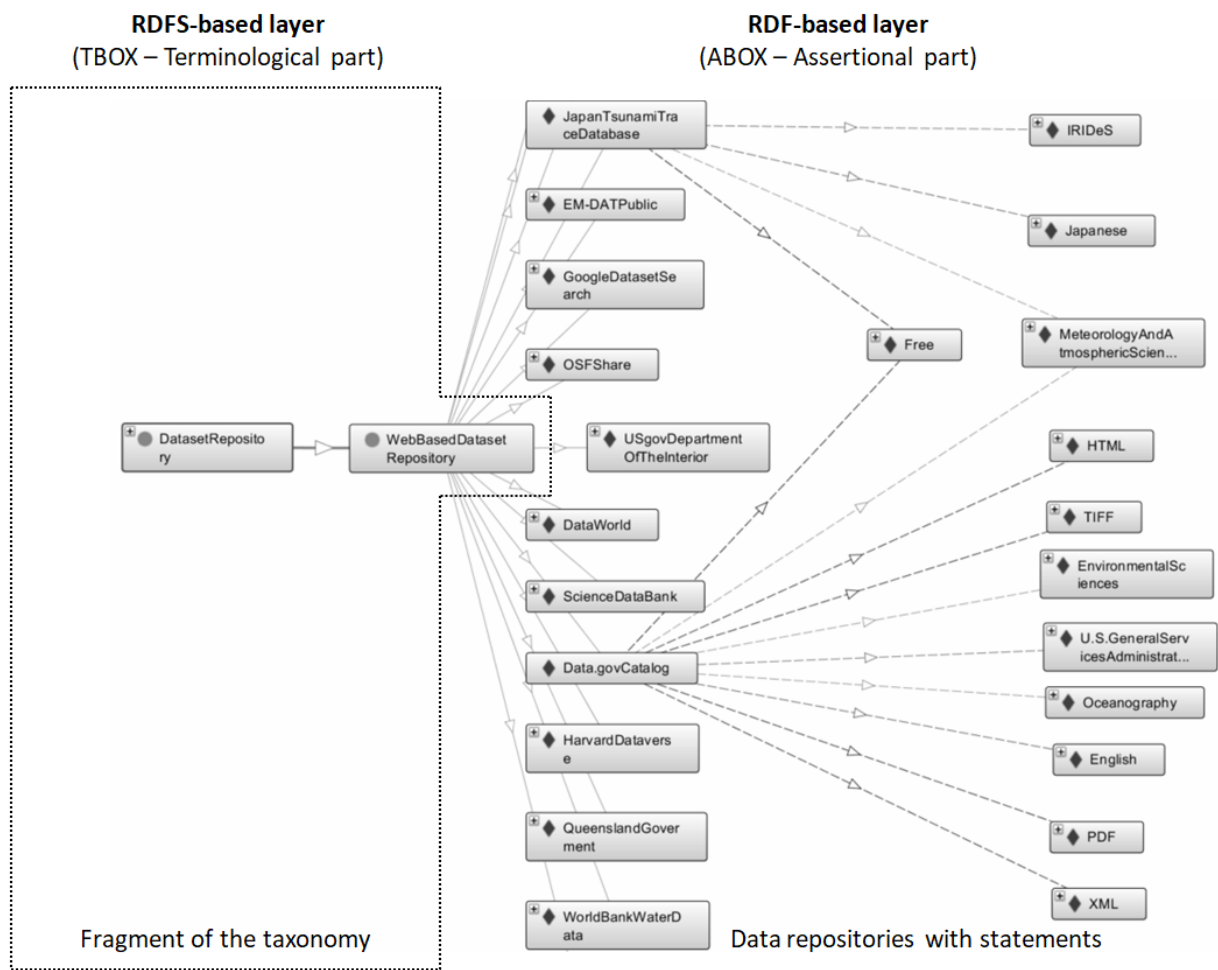
**Figure 7.** The structure of the semantic graph database (the fragment).

**Figure 8.** Philosophy of SPARQL querying.

We can zoom the SPARQL query part in Figure 8 and uncover a hidden structure. As an example, we would like to know which dataset repositories can provide datasets related to tsunamis. Additionally, ordering results according to the number of tsunami datasets in descending order are required. The first three results are presented in the output section. The complete query written in SPARQL is described below.

SELECT ?repository ?countOfTsunamiDatasets
WHERE {
? Repository rdf:type dronto:WebBasedDatasetRepository.
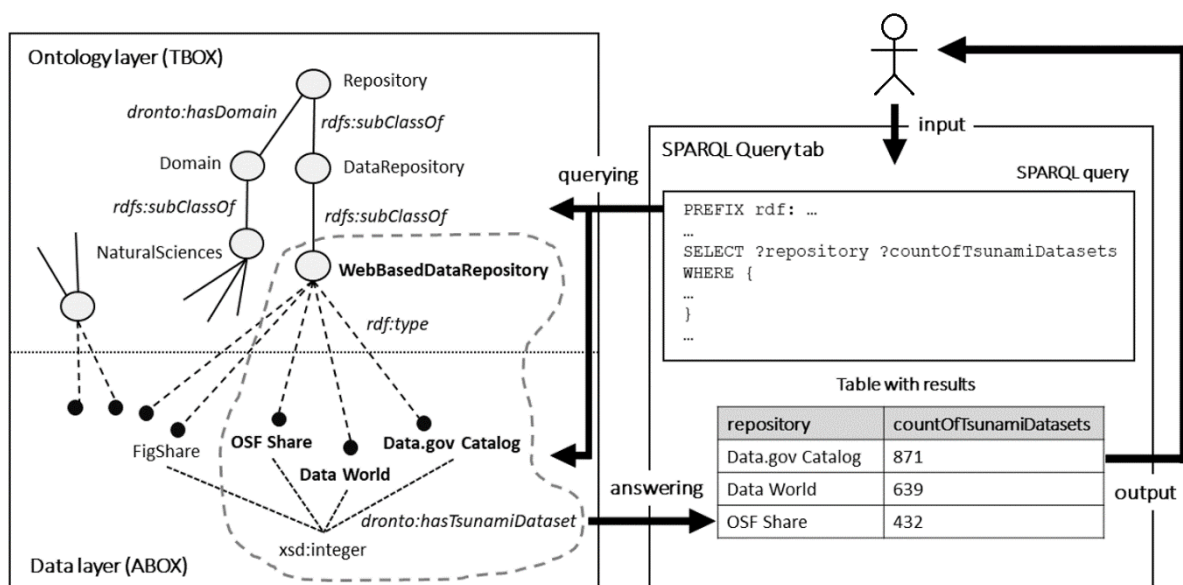? repository dronto:hasTsunamiDataset ?countOfTsunamiDatasets
}
ORDER BY DESC (?countOfTsunamiDatasets)
LIMIT 3

Output:

- Data.gov Catalog: 871.
- Data World: 639.
- OSF Share: 432.

The SPARQL query begins with the name spaces identified by prefixes that tell us which external vocabularies must be used in the query. Our ontological structure was identified using the dronto prefix. Results of the query are stored in the variables (identified by "?" symbols) mentioned in the "select" section (?repository, ?countOfTsunamiDatasets). The required content is written in the "where" section. Additional modifiers can be mentioned below, such as ordering (ORDER BY) or visualization limited amount of results (LIMIT). Figure 8 shows that only a fragment of the knowledge structure corresponds to the SPARQL query. This fragment is depicted as a gray dashed part on the left side of Figure 8. This fragment is visualized as the resulting table on the right side of Figure 8.

The second example demonstrates how to answer the question of which data repositories are focused on meteorology and atmospheric sciences. The philosophy of querying is the same as the aforementioned example.

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX dronto: <http://www.semanticweb.org/husáková/ontologies/2021/DROntology#>
SELECT ?repository
WHERE
{? Repository rdf:type dronto:WebBasedDatasetRepository.
?repository dronto:hasDomain dronto:MeteorologyAndAtmosphericSciences.}

Output: OSF Share, Figshare, Data.gov Catalog, Japan Tsunami Trace Database, Data World, Google Dataset Search, World Bank Water Data, NCEI, and Science Data Bank.

This semantic graph database does not need to be interpreted only as a pure collection of statements about data repositories. The inference process can be used to discover hidden facts inside the database. As an example, the inference engine can "ingest" the statements about data repositories and cluster them according to the application area of the data which are provided by these repositories. The semantically richer language is inevitable for this process and has been slightly introduced above. The classes-based layer is general, so it can be used as a backbone for projects not only aimed at data repositories. It can be customized freely for its own purposes.

## 6. Conclusions

Increased connectivity has accelerated progress in global research, and estimates indicate that the scientific output is doubling approximately every decade [42]. As science is becoming data-intensive and collaborative, this rise in research activity increases research data output [28,43]. However, efficient work with data is increasingly deemed an important part of the scientific process because approximately 80% of research data are inaccessible or

unpublished [42]. Making data publicly available allows original results to be reproduced and new analyses to be conducted [27,29]. Nevertheless, there is a growing debate about how quickly scientific findings can and should influence disaster mitigation policies [44]. Although a relatively new research area, tsunami science depends on data from various disciplines falling within the scope of geosciences, oceanography, engineering, physics, mathematics, and disaster management, including politics, media, communication, and education [4]. Furthermore, tsunami hazard assessment and mitigation plans based on numerical modeling and simulation of tsunamis have gained increasing importance.

In the realm of tsunami research, a growing number of scientists are trained in surveying techniques. Thus, more data will be collected [45]. Because data are the infrastructure of tsunami research, this study provides a unique, extensive review of tsunami-related datasets and repositories. Existing data repositories have several issues, ranging from missing updates or dysfunctional webpages to limited search filtering, metadata downloadability, and data usability. Although the list of repositories presented in this study is not exhaustive because of the applied methodological approach, diverse sets of experts and practitioners can take advantage of the repositories identified and evaluated in this study as datasets contain data related to volcanology, geoscience, water research, or civil engineering. By using data, their teams and institutions will provide various types of datasets and repositories, which can be used for analysis, modeling, simulation, or prediction of tsunami occurrence. Thus, multidisciplinary research, suggested by the latest research [46], to design and propose practical solutions can be supported. Hopefully, the primary outcome of this study will catalyze the data lifecycle in tsunami research.

Indeed, there are ways to mitigate these issues. Computer science techniques and methods from artificial intelligence, such as k-Nearest Neighbors (k-NN) models [47], C4.5 algorithm [48], the Random Forest method [49], Bayes classifier [50], or non-linear parametric model [51] are worth mentioning. They all process data to acquire new knowledge and insights. We briefly outline ontology engineering as an approach that we believe can help the global research community reduce the analytical workload of all stakeholders and enhance the quality of the recorded data. It plays the role of a metadata provider that can improve orientation in existing tsunami-related data repositories. The proposed solution is presented only as an outline. It shows viability, usability, and feasibility of solving the issues identified in this review. Its development to a full version with all suggested functionalities can improve work with tsunami-related data repositories.

**Appendix A**

**Table A1.** Data sources (Notes are partial or full direct citations from available links).

| # | Name | Organization | Link |
|---|------|--------------|------|
| 1. | Data.gov Catalog | U.S. General Services Administration | https://catalog.data.gov/dataset |
| | Note: Data.gov Catalog was launched in 2009 and is managed and hosted by the U.S. General Services Administration, Technology Transformation Service. It is an online repository of policies, tools, case studies, and other resources to support data governance, management, exchange, and use throughout the federal government. Data.gov follows the DCAT-US Schema v1.1. | | |
| 2. | Data.gov Dataset | U.S. General Services Administration | https://data.doi.gov/dataset |
| | Note: was launched in 2009 and is managed and hosted by the U.S. General Services Administration, Technology Transformation Service, and includes data from several U.S. departments and services. This database contains more than 28,000 datasets. The majority of these datasets are geospatial types. The dataset finder offers filtering based on tags of datasets, dataset formats, origin organization, publisher, and bureaus. | | |
| 3. | OSF Home | Center for Open Science | https://osf.io/ |
| | Note: OSF Home is a free, open platform to support your research and enable collaboration. Part of OFS called OSF Share. | | |
| 4. | OSF Share | Center for Open Science | https://share.osf.io/ |
| | Note: OSF Share was founded in 2013 by the Association of American Universities and the Association of Public and Land-grant Universities. It is a community open-source initiative developing tools and services to connect related, yet distributed, research outputs, enabling new kinds of scholarly discovery. The National Endowment for the Humanities (NEH) is currently supporting SHARE in a project to integrate digital humanities into the scholarly web. | | |
| 5. | Japan Tsunami Trace database | IRIDeS | https://tsunami3.civil.tohoku.ac.jp/ |
| | Note: Japan Tsunami Trace database contains information about destroyed buildings and drifting objects, traces indicating the inundation limit. This database includes more than 30,000 records. The majority of data in this database are collected from Japan. This database also contains a Japan map with tsunami incidents and pictures from these incidents. | | |
| 6. | National Centers for Environmental Information | National Centers for Environmental Information | https://www.ngdc.noaa.gov/hazard/ |
| | Note: NCEI (formerly NGDC)—Natural Hazards Data, Images, and Education is part of NOAA. NCEI archives and assimilates tsunami, earthquake, and volcano data to support research, planning, response, and mitigation. Long-term data, including photographs, can be used to establish the history of natural hazard occurrences and help mitigate against future events. Related groups of datasets to a tsunami are Hazards Data: Map Search, Tsunami Events, Tsunami Runups, and Recent/Significant Tsunami Events. | | |
| 7. | Queensland Goverment—Open Data Portal | The State of Queensland | https://www.data.qld.gov.au/dataset |
| | Note: Queensland Government offers Open Data Portal with more than 2700 datasets and 11,200 resources across all fields related to Queensland under the Right to Information Act 2009 (RTI Act) and the Information Privacy Act 2009 (IP Act). Six of these datasets are related to tsunamis. | | |
| 8. | EM-DAT Public | EM-DAT | https://public.emdat.be/ |
| | Note: EM-DAT (Emergency Events Database) Public was launched in 1988 by the Centre for Research on the Epidemiology of Disasters (CRED) with the initial support of the World Health Organization (WHO) and the Belgian Government. The main objective of the database is to serve the purposes of humanitarian action at national and international levels. The initiative aims to rationalize decision making for disaster preparedness, as well as provide an objective base for vulnerability assessment and priority setting. EM-DAT contains essential core data on the occurrence and effects of over 22,000 mass disasters in the world from 1900 to the present day. | | |
| 9. | Figshare | Figshare | https://figshare.com/ |
| | Note: Figshare was launched at the beginning of the year 2012. It is an online open-access repository with aim of preserving and share research outputs, including datasets, figures, and videos. Figshare offers almost one and a half million datasets; 254 of them are related to tsunamis. Furthermore, it provides nearly 200 figures and more than 150 journal contributions related to tsunamis. | | |
| 10. | Science Data Bank | Computer Network Information Center, Chinese Academy of Sciences | https://www.scidb.cn/en |
| | Note: Science Data Bank (ScienceDB) is a public, general-purpose data repository aiming to provide data services (e.g., data acquisition, long-term preservation, publishing, sharing, and access). ScienceDB is devoted to becoming a repository of long-term data sharing and data publishing in China. According to authors key features of ScienceDB, for example are data findability, open and sharing data, data traceability, and permanent accessibility. ScienceDB offers 484 datasets; one of them is related to tsunamis. | | |
| 11. | Kaggle | Kaggle | https://www.kaggle.com/datasets |
| | Note: Kaggle was founded in 2010, and it is a subsidiary of Google LLC. Kaggle is an online platform for a community of data scientists. Focus on finding and publishing datasets and models. Kaggle offers more than 70,000 datasets; 14 of them are related to tsunamis. | | |
| 12. | Data World | data.world, Inc. | https://data.world/ |
| | Note: Data World was founded in 2016, it is public benefit corporation, which aims at makes data easily understandable for the public. This company has three main goals: (1) build the most meaningful, collaborative and abundant data resource in the world in order to maximize data's societal problem-solving utility. (2) advocate publicly for improving the adoption, usability, and proliferation of open data and linked data. (3) serve as an accessible historical repository of the world's data. It offers more than 600 datasets related to the tsunami. | | |
| 13. | Harvard Dataverse | The President & Fellows of Harvard College | https://dataverse.harvard.edu/ |
| | Note: The Harvard Dataverse Repository is a free data repository open to all researchers from any discipline, both inside and outside of the Harvard community, where you can share, archive, cite, access, and explore research data. Each individual Dataverse collection is a customizable collection of datasets (or a virtual repository) for organizing, managing, and showcasing datasets. Harvard Dataverse offers more than 110,000 datasets; 33 of them are related to tsunamis. | | |
| 14. | Google—Dataset Search | Google | https://datasetsearch.research.google.com/ |
| | Note: Google—Dataset Search is a search engine for data sets created by Google, it was launched in 2018. Data Search has two main goal: (1) foster a data sharing ecosystem that will encourage data publishers to follow best practices for data storage and publication. (2) give scientists a way to show the impact of their work through citation of data sets that they have produced. | | |

**Table A1.** *Cont.*

| # | Name | Organization | Link |
|---|------|-------------|------|
| 15. | The World Bank Water Data catalog | The World Bank Group | https://wbwaterdata.org/organization/worldbank-data-catalog |
| | Note: World Bank—Data catalog offers different types of data, e.g., geospatial, microdata, time series, and other types of datasets. Part of World Bank is World Bank Water data. | | |
| 16. | The World Bank Data catalog | The World Bank Group | https://datacatalog.worldbank.org/ |
| | Note: World Bank Water data—Data catalog is a source for all water-related open data at the World Bank. It contains datasets and applications generated or compiled by the Water Global Practice. In to total it contains 2654 datasets, 26 of them are related to tsunamis. | | |
| 17. | PANGAEA | PANGAEA | https://www.pangaea.de/ |
| | Note: PANGAEA is an information system, which operated as an Open Access library with aim on archive, publish and distribute georeferenced data from earth system research. PANGAEA is member of the World Data System (WDS) of the International Science Council (ISC). The search engine is powered by the open-source software Elasticsearch and metadata processing is provided by panFMP. PANGAEA offers more than 400,000 datasets; more than 3500 are related to tsunamis. | | |
| 18. | WHO data collections | WHO | https://www.who.int/data/collections |
| | Note: WHO manages and maintains a wide range of data collections related to global health and well-being. WHO has 194 Member States across six regions, and from more than 150 offices, WHO staff are united in a shared commitment to achieve better health for everyone, everywhere. WHO offers 80 datasets; none of them is related to the tsunami. | | |
| 19. | Sendai Framework for Disaster Risk Reduction | United Nations Office for Disaster Risk Reduction | https://www.desinventar.net/ |
| | Note: Sendai Framework for Disaster Risk Reduction is built on two modules. The first one is Administration and Data Entry module, which is a relational and structural database through which the database is fed by filling in predefined fields (space and temporal data, types of events and causes, and sources) and by both direct and indirect effects (deaths, houses, infrastructure, and economic sectors). The second is The Analysis module, which allows access to the database by queries that may include relations among the diverse variables of effects, types of events, causes, sites, dates, etc. This module allows at the same time to represent those queries with tables, graphics, and thematic maps. | | |
| 20. | The Humanitarian Data Exchange | United Nations Office for the Coordination of Humanitarian Affairs | https://data.humdata.org/ |
| | Note: The Humanitarian Data Exchange (HDX) is an open platform for sharing data across crises and organizations. Launched in July 2014, the goal of HDX is to make humanitarian data easy to find and use for analysis. HDX growing collection of datasets has been accessed by users in over 200 countries and territories. | | |
| 21. | Earth Online | European Space Agency | https://earth.esa.int/eogateway/ |
| | Note: European Space Agency (ESA)—Earth Online. Earth Online is the entry point for scientific-technical information on Earth Observation activities by the European Space Agency (ESA). The web portal provides a vast amount of content, grown and collected over more than a decade: Detailed technical information on Earth Observation (EO) missions; satellites and sensors; EO data products and services; online resources such as catalogs and library; applications of satellite data; access to promotional satellite imagery. | | |
| 22. | EU Open Data Portal | Publications Office of the European Union in Luxembourg | https://data.europa.eu/data/datasets |
| | Note: EU Open Data Portal was funded by the European Union, and it is managed by Publications Office of the European Union. This portal provides access to open data from international, EU, national, regional, local, and geo data portals. It replaces the EU Open Data Portal and the European Data Portal. This portal has four main sections: searching data, providing data, using data, and training and library. The portal offers more than 15,000 datasets; 9 of them are related to tsunamis. | | |
| 23. | Novosibirsk Tsunami Laboratory | Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Tsunami Laboratory, Novosibirsk, Russia | http://tsun.sscc.ru/nh/list.html |
| | Note: Novosibirsk Tsunami Laboratory was founded in 2004. It contains databases related to tsunamis, earthquakes, impacts events, volcanic activities, bolides and asteroids, and hurricanes. | | |
| 24. | InnovationLab GeoNode | Labs GeoNode | https://www.geonode-gfdrrlab.org/ |
| | Note: InnovationLab GeoNode is a geospatial content management system, a platform for the management and publication of geospatial data. It brings together mature and stable open-source software projects under a consistent and easy-to-use interface allowing non-specialized users to share data and create interactive maps. Data management tools built into GeoNode allow for integrated creation of data, metadata, and map visualizations. Each dataset in the system can be shared publicly or restricted to allow access to only specific users. Social features such as user profiles and commenting and rating systems allow for the development of communities around each platform to facilitate the use, management, and quality control of the data the GeoNode instance contains. | | |
| 25. | STAR—Study of the Tsunami Aftermath and Recovery | None * | http://stardata.org/data.html |
| | Note: The Study of the Tsunami Aftermath and Recovery (STAR) is a longitudinal survey of individuals, households, communities, and facilities in the provinces of Aceh and North Sumatra, Indonesia. The study is designed to provide evidence on the immediate and longer-term consequences of the 2004 Sumatran-Andaman earthquake and tsunami and recovery efforts. * STAR is a collaborative project involving investigators at Duke University; SurveyMETER (Indonesia); the University of North Carolina, Chapel Hill; the University of California, Los Angeles (UCLA); the University of Pennsylvania; the University of Southern California; the World Bank; Statistics Indonesia. | | |
| 26. | Mendeley Data | Elsevier Inc. | https://data.mendeley.com/ |
| | Note: Mendeley Data was launched in 2015 by Elsevier. It is an open, cloud-based research data management (RDM) platform that empowers research institutions to manage the entire lifecycle of research data, and enables researchers to discover, collect, and share research data. It enables librarians and administrators to moderate, manage, report on, and showcase research data output regardless of which data repository researchers use. It contains more than seven million datasets; more than 2000 of them are related to the tsunami. | | |

**Table A1.** *Cont.*

| # | Name | Organization | Link |
|---|------|--------------|------|
| 27. | Dryad | Dryad | https://datadryad.org/search?q= |

Note: Dryad is an open source, community driven project that takes a unique approach to data publication and digital preservation. Dryad focuses on search, presentation, and discovery and delegates the responsibility for the data preservation function to the underlying repository with which it is integrated. Dryad's original iteration launched in 2009, in 2019, Dryad merged with Dash.

| # | Name | Organization | Link |
|---|------|--------------|------|
| 28. | National Center for Biotechnology Information Support Center | National Library of Medicine | https://www.ncbi.nlm.nih.gov/search/ |

Note: NCBI (National Center for Biotechnology Information Support Center) was established in 1998. NCBI is now a leading source for public biomedical databases, software tools for analyzing molecular and genomic data, and research in computational biology. Today NCBI creates and maintains over 40 integrated databases for the medical and scientific communities as well as the general public.

| # | Name | Organization | Link |
|---|------|--------------|------|
| 29. | Qualitative Data Repository | Qualitative Data Repository | https://data.qdr.syr.edu/ |

Note: Qualitative Data Repository (QDR) is a dedicated archive for storing and sharing digital data (and accompanying documentation) generated or collected through qualitative and multi-method research in the social sciences. QDR provides search tools to facilitate the discovery of data, and, also, serves as a portal to material beyond its own holdings, with links to U.S. and international archives. The repository's initial emphasis is on political science.

| # | Name | Organization | Link |
|---|------|--------------|------|
| 30. | DataCite | DataCite | https://search.datacite.org/ |

Note: Datacite is a leading global non-profit organization that provides persistent identifiers (DOIs) for research data and other research outputs. Organizations within the research community join DataCite as members to be able to assign DOIs to all their research outputs. This way, their outputs become discoverable and associated metadata is made available to the community. DataCite then develops additional services to improve the DOI management experience, making it easier for our members to connect and share their DOIs with the broader research ecosystem and to assess the use of their DOIs within that ecosystem. DataCite is an active participant in the research community and promotes data sharing and citation through community-building efforts and outreach activities.

| # | Name | Organization | Link |
|---|------|--------------|------|
| 31. | Open data initiative of the Government of Spain | Government of Spain | https://datos.gob.es/en/catalogo |

Note: The Aporta Initiative was launched in 2009 to promote the opening of public information and development of advanced services based on data. It is backed by the Ministry of Economy and Business, the Ministry of Territorial Policy and Civil Service, and the Public Corporate Entity Red.es. The main goal of the Aporta initiative, a key element in the Spanish government's open data policy, is to harmonize and efficiently take advantage of the synergies between ongoing open data projects. It seeks to always drive and coordinate actions being carried out by different levels the administration, the private sector, and academic field, according to an integrating governance model. It does all of this in order to promote new products and services from the private sector and civil society to benefit society.

| # | Name | Organization | Link |
|---|------|--------------|------|
| 32. | Malta Data Portal | Government of Malta | https://open.data.gov.mt/dashboard.html |

Note: Malta Data Portal was founded by the Republic of Malta in 2014. It was financed by the Malta government and European Union. It offers 205 datasets related to the republic of Malta.

| # | Name | Organization | Link |
|---|------|--------------|------|
| 33. | National Opendata Portal | Republic of Cyprus | https://www.data.gov.cy/ |

Note: National Opendata Portal was founded by the Ministry of Finance of the Republic of Cyprus in 2014, and it offers datasets related to the Republic of Cyprus. In total, it offers more than 1100 datasets. Most of them are related to the environment, economy and finance, population, government, and health.

| # | Name | Organization | Link |
|---|------|--------------|------|
| 34. | Open Data From Public Administration | Agenzia per l'Italia Digitale | https://dati.gov.it/ |

Note: Open Data From Public Administration was born as a project promoted in 2011 by the Italian government and since 2015 it has been managed by the Agency for Digital Italy. The data.gov.it Portal is the national catalog of metadata relating to open type data released by public administrations and constitutes the search tool and the access point to the data made available according to the open data paradigm, in accordance with the provisions from art. 9 of the legislative decree n. 36/2006 (transposition of the European Directive on the reuse of public sector information). Dati.gov.it is also the tool with which the Agency for Digital Italy promotes the policies for the enhancement of national public information assets. To this end, the portal makes available to administrations and developers a series of useful resources to deepen the topic of open data, to improve the quality of the data exposed and, ultimately, to encourage their reuse.

| # | Name | Organization | Link |
|---|------|--------------|------|
| 35. | O catálogo central de dados abertos em Portugal | Agência para a Modernização Administrativa | https://dados.gov.pt/en/ |

Note: O catálogo central de dados abertos em Portugal is the Portuguese Public Administration's open data portal. Its function is to aggregate, reference and store open data from different Public Administration's bodies and sectors, therefore creating the central catalogue of open data in Portugal. Besides working as a shared data storing and publication service, it may be used by any public body, also working as an indexing portal of contents in other open data portals/catalogues. It also provides several interaction mechanisms between data suppliers and re-users, such as the possibility to commend, submit complementary data versions and suggest improvements to the platform.

| # | Name | Organization | Link |
|---|------|--------------|------|
| 36. | OECD Data | Organisation for Economic Co-operation and Development | https://data.oecd.org/ |

Note: The Organisation for Economic Co-operation and Development (OECD) is an international organization that works to build better policies for better lives. OECD's goal is to shape policies that foster prosperity, equality, opportunity, and well-being for all. OECD draws on 60 years of experience and insights to better prepare the world of tomorrow. Data OECD is part of OECD and offers more than 7500 datasets, two of them are related to tsunamis. Another part of OECD, which focuses on datasets, is OECD iLibrary.

| # | Name | Organization | Link |
|---|------|--------------|------|
| 37. | OECD iLibrary | Organisation for Economic Co-operation and Development | https://www.oecd-ilibrary.org/ |

Note: OECD iLibrary is the online library of the Organisation for Economic Cooperation and Development (OECD) featuring its books, papers, and statistics and is the knowledge base of OECD's analysis and data.

**Table A1.** *Cont.*

| # | Name | Organization | Link |
|---|------|--------------|------|
| 38. | DesignSafe-CI | The Natural Hazards Engineering Research Infrastructure | https://www.designsafe-ci.org/data/browser/public/ |

Note: DesignSafe-CI is supported by the National Science Foundation (NSF) is an independent federal agency created by U.S. Congress in 1950. DesignSafe-CI is the web-based research platform of the NHERI Network that provides the computational tools needed to manage, analyze, and understand critical data for natural hazards research.

| 39. | The official portal for European data | Publications Office of the European Union | https://data.europa.eu/en |

Note: European Data Portal—the portal provides access to open data from international, EU, national, regional, local, and geo data portals. It replaces the EU Open Data Portal and the European Data Portal. The portal addresses the whole data value chain, from data publishing to data reuse. Going beyond collecting metadata (data about data), the strategic objective of the portal is to improve accessibility and increase the value of open data. The portal is divided into four sections: (1) searching data, (2) providing data, (3) using data, and (4) training and library.

| 40. | SAGE Research methods | SAGE Publications | https://methods.sagepub.com/Datasets |

Note: SAGE Research methods supports research at all levels by providing material to guide users through every step of the research process. Nearly everyone at a university is involved in research, from students learning how to conduct research to faculty conducting research for publication to librarians delivering research skills training and doing research on the efficacy of library services. SAGE Research Methods has the answer for each of these user groups, from a quick dictionary definition, a case study example from a researcher in the field, a downloadable teaching dataset, a full-text title from the Quantitative Applications in the Social Sciences series, or a video tutorial showing research in action. SAGE Research Methods is the ultimate methods library with more than 1000 books, reference works, journal articles, and instructional videos by world-leading academics from across the social sciences, including the largest collection of qualitative methods books available online from any scholarly publisher. The site is designed to guide users to the content they need to learn a little or a lot about their method. The Methods Map can help those less familiar with research methods to find the best technique to use in their research. Built upon SAGE's researchers.

| 41. | Gebco | British Oceanographic Data Centre | https://www.gebco.net/ |

Note: Gebco is a non-profit making organization which relies largely on the voluntary contributions of an enthusiastic international team of geoscientists and hydrographers. GEBCO is continually working to improve its gridded data sets with the aim of providing the most authoritative publicly available bathymetric grids for the world's oceans.

| 42. | Irish National Seabed Survey Data Access | GSI Seabed Mapping | http://www.gsiseabed.ie/data.htm |

Note: INFOMAR /Irish National Seabed Survey (INNS)—INFOMAR is a successor to the Irish National Seabed Survey (INSS) and concentrates on creating integrated mapping products of the physical, chemical and biological features of the seabed in the near-shore area. The INSS (1999–2005) involved mapping of more than 80% of Ireland's seabed territory, supporting the delineation of the exclusive economic zone, and extending inshore coverage to the 200 m depth contour overall, with extension in to 50 m depth offshore. The INFOMAR program which was tasked with mapping the remaining coastal areas. This is being undertaken in two phases; Phase one (2006–2016) focusing on 26 inshore priority bays and 3 priority coastal areas, and Phase two (2016–2026) mapping the remaining unsurveyed coastal and shelf areas.

| 43. | University of Hawaii Sea Level Center | University of Hawaii Sea Level Center | http://uhslc.soest.hawaii.edu/data/ |

Note: University of Hawaii Sea Level Center maintains one of the largest global networks of tide gauges that feed the FD stream, but there are numerous other international agencies that contribute. These FD data are received from partner agencies on a monthly basis and incorporated into the FD stream.

| 44. | European Marine Observation and Data Network | The European Marine Observation and Data Network | https://emodnet.eu/en/portals |

Note: European Marine Observation and Data Network (EMODnet) is a network of organizations supported by the EU's integrated maritime policy. These organizations work together to observe the sea, process the data according to international standards and make that information freely available as interoperable data layers and data products. EMODnet provides access to European marine data across seven discipline-based themes. For each of these themes, EMODnet has created a gateway to a range of data archives managed by local, national, regional, and international organizations. Through these gateways, users have access to standardized observations, data quality indicators, and processed data products, such as basin-scale maps. These data products are free to access and use.

| 45. | Spanish National Center of Geographic Information | Centro Nacional de Información Geográfica | http://centrodedescargas.cnig.es/CentroDescargas/ |

Note: Spanish National Center of Geographic Information—The Download Center (CdD) is a web site created by the National Center for Geographic Information (CNIG), aimed at serving users as a free tool for downloading geographic digital files generated by the Directorate General for the National Geographic Institute (IGN). It offers Geo-referenced images of maps with various scales of representation, to display on the computer screen or on mobile devices. These images do not contain neither marginal information (captions) nor framework of coordinates, Altimetric information that represents the landform of the national territory, and, in the case of Lidar data, of the elements that are found on it as well, and other geographic information.

| 46. | The Centre for Environmental Data Analysis | Science and Technology Facilities Council | https://catalogue.ceda.ac.uk/ |

Note: The original "CEDA" group followed the merger of two of NERC's data centers—the BADC and NEODC—in 2005, originally being called the Centre for Environmental Data Archival. However, with greater support for users by analyzing the data with a slight name change from the A in CEDA from Archival to Analysis occurred in 2015 to reflect this growing and important role for CEDA. CEDA aims to support environmental science, further environmental data archival practices, and develop and deploy new technologies to enhance access to data. Additionally, they provide services to aid large scale data analysis.

| 47. | Nasa—MODIS | NASA | https://modis.gsfc.nasa.gov/data/dataprod/ |

Note: Nasa—MODIS—Moderate Resolution Imaging Spectroradiometer is a key instrument aboard the Terra and Aqua satellites. Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every 1 to 2 days, acquiring data in 36 spectral bands, or groups of wavelengths (see MODIS Technical Specifications). These data will improve our understanding of global dynamics and processes occurring on the land, in the oceans, and in the lower atmosphere. MODIS is playing a vital role in the development of validated, global, interactive Earth system models able to predict global change accurately enough to assist policy makers in making sound decisions concerning the protection of our environment.

**Table A1.** *Cont.*

| # | Name | Organization | Link |
|---|------|-------------|------|
| 48. | NASA—ASTER | NASA | https://asterweb.jpl.nasa.gov/gdem.asp |

Note: ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) was founded by the Ministry of Economy, Trade, and Industry (METI) of Japan and the United States National Aeronautics and Space Administration (NASA), and focus on spaceborne thermal emission and reflection. The first version of the ASTER GDEM, released in June 2009, was generated using stereo-pair images collected by the ASTER instrument onboard Terra. ASTER GDEM coverage spans from 83 degrees north latitude to 83 degrees south, encompassing 99 percent of Earth's landmass.

| # | Name | Organization | Link |
|---|------|-------------|------|
| 49. | Geospatial Information Authority of Japan | Geospatial Information Authority of Japan | https://fgd.gsi.go.jp/download/menu.php |

Note: The Geospatial Information Authority of Japan is Japan's national mapping organization and a special organization of the Ministry of Land, Infrastructure, Transport, and Tourism. Part of The Geospatial Information Authority of Japan focus on collecting and providing disaster prevention information using the latest technology, and information about past disasters.

| # | Name | Organization | Link |
|---|------|-------------|------|
| 50. | RadioCarbon—IntCal13 Supplemental Data | Radiocarbon | http://www.radiocarbon.org/IntCal13.htm |

Note: RadioCarbon—IntCal13 Supplemental Data was published in Radiocarbon journal in 55th volume in 2013. Radiocarbon is the main international journal of record for research articles and date lists relevant to 14C and other radioisotopes and techniques used in archaeological, geophysical, oceanographic, and related dating.

| # | Name | Organization | Link |
|---|------|-------------|------|
| 51. | Satellite Imaging Corporation | Satellite Imaging Corporation | https://www.satimagingcorp.com/gallery/ |

Note: Satellite Imaging Corporation Satellite Imaging Corporation (SIC) was formed in the early 1990's as a response to increasing demand for medium and high resolution 2D and 3D satellite image data. Management has over 40 years of experience in on/offshore survey, satellite remote sensing, and GIS industry. SIC has processed satellite image data for clients belonging to a variety of industries for their domestic and international mapping, GIS, environmental and design project needs. SIC have performed projects throughout Africa, Europe, Russia, North America, South America, the Middle East, and Southeast Asia. SICs portfolio includes cadastre and GIS projects for the United States Agency for International Development (USAID), precision agriculture mapping, transportation and pipeline corridor surveys, near shore bathymetry, support for 2D/3D seismic data acquisition, and the planning of well locations and access roads located in rural, and remote areas around the world.

| # | Name | Organization | Link |
|---|------|-------------|------|
| 52. | OpenTopography | https://portal.opentopography.org/datasets | OpenTopography Facility, San Diego Supercomputer Center, University of California San Diego |

Note: OpenTopography Facility is based at the San Diego Supercomputer Center at the University of California, San Diego. Main missions of OpenTopography Facility is to democratize online access to high-resolution (meter to sub-meter scale), Earth science-oriented, topography data acquired with lidar and other technologies. Harness cutting edge cyberinfrastructure to provide Web service-based data access, processing, and analysis capabilities that are scalable, extensible, and innovative. Promote discovery of data and software tools through community populated metadata catalogs. Partner with public domain data holders to leverage OpenTopography infrastructure for data discovery, hosting and processing. Provide professional training and expert guidance in data management, processing, and analysis. Foster interaction and knowledge exchange in the Earth science lidar user community.

| # | Name | Organization | Link |
|---|------|-------------|------|
| 53. | Sea Level Station Monitoring Facility | Flanders Marine Institute | http://www.ioc-sealevelmonitoring.org/list.php |

Note: Sea Level Station Monitoring—The Global Sea Level Observing System (GLOSS) was established by the UNESCO Intergovernmental Oceanographic Commission (IOC) in 1985 to establish a well-designed, high-quality in situ sea level observing network to support a broad research and operational user base. The backbone of the global tide gauge network is the GLOSS Core Network (GCN), a global set of 300 tide gauge stations that provide optimal sampling of the global ocean. GCN gauges were allocated to each island or group of islands at intervals not closer than 500 km, and along continental coasts at intervals generally not less than 1000 km. Preference was given to islands in order to maximize exposure to the open ocean.

| # | Name | Organization | Link |
|---|------|-------------|------|
| 54. | Open platform for French public data | Government of France | https://www.data.gouv.fr/en/datasets/ |

Note: Open platform for French public data is a portal governed by the French government. Part of Data.Gouv.FR is Etalab, which supports the opening up of public data for the State and administrations. As such, Etalab develops and manages the open platform for public data data.gouv.fr, a platform which hosts the datasets and lists their reuse.

| # | Name | Organization | Link |
|---|------|-------------|------|
| 55. | AVISO | AVISO CNES Data Center | https://aviso-data-center.cnes.fr/ |

Note: AVISO from Centre National d'Etudes Spatiales (Cnes), which is the government agency responsible for shaping and implementing France's space policy in Europe, was founded in 1998. In recent years, AVISO has become a reference in international oceanographic and altimetric communities. In 2014 AVISO opens to wider applications than the ocean themes. Thus, becoming AVISO +, the portal opens to hydrology/coastal/ice and merges with the CTOH website to provide users with more operational and demonstration products, and expertise in an intuitive and modern web site.

| # | Name | Organization | Link |
|---|------|-------------|------|
| 56. | Italian Tsunami Effects Database | Istituto Nazionale di Geofisica e Vulcanologia | https://tsunamiarchive.ingv.it/ited.1.0/ |

Note: The Italian Tsunami Effects Database (ITED), the first database dealing with the tsunami effects observed along the Italian coasts from historical times. ITED was compiled starting from the Euro Mediterranean Tsunami Catalogue. ITED focuses on the propagation effects observed along the Italian coasts providing information on how each locality was interested by tsunamis effects over time. Currently ITED contains about 300 observations of tsunami effects referred to 184 localities of the Italian coasts and related to the 70 Italian tsunami events present in EMTC v2. Whenever a place experienced a tsunami effects more than once, details of each observation is supplied to allow the user to build the tsunami-history of the locality

| # | Name | Organization | Link |
|---|------|-------------|------|
| 57. | WEBRITEC | European Commission | https://webritech.jrc.ec.europa.eu |

Note: There are three section of the repository. TAD show the theorical Sea Levels Tide calculated by an algorithm and compare them with a real Measurements for each buoys in Database. TAT provides the tsunami analysis tool dealing with tsunami public calculations, a list of user calculations and ability to submit a new calculation. Sea levels at specific points on the Globe are also available.

**Table A1.** *Cont.*

| # | Name | Organization | Link |
|---|------|--------------|------|
| 58. | European/NEAMTWS Tsunami Catalogue | UNESCO/IOC Project Office | http://www.ioc-tsunami.org/ |
| Note: A unified catalogue containing 290 tsunamis generated in the European and Mediterranean seas since 6150 B.C. to current days was developed based on the GITEC, GITEC-II, and TRANSFER projects. | | | |
| 59. | Euro-Mediterranean Paleotsunami Database | Istituto Nazionale di Geofisi-ca e Vulcanologia | http://paleotsunami.rm.ingv.it/index.php |
| Note: Database was developed within the frame of the EC TRANSFER project with the aim to collect data on tsunami inundations occurred in the past. Evidence of paleotsunamis is derived from coastal stratigraphy because of the presence of peculiar sediments or boulders. Dating of the paleotsunami deposits helps in correlating events with historical tsunamis or previous ones. This Database provide mainly two types of information of use for developing tsunami scenarios and time dependent hazard calculations: locations of past inundations and their frequency. | | | |
| 60 | Tsunami Measurement Data | IUGG Tsunami Commission | http://www.nda.ac.jp/~fujima/TMD/index.html |
| Note: IUGG Tsunami Commission requested to all researchers to provide the data of tsunami traces of Indian Ocean Tsunami. Provided data are written by the common definition with available tables and figures. IUGG Tsunami Commission collects and authorizes the data, makes small-scale maps and large-scale surveyed-area maps and distributes them to tsunami community. | | | |

**Table A2.** Evaluation of data resources (selected criteria).

| | Description | | General Availability | | | | | | Filter/Search Options | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Last Update | Repository Domain | Availability | Downloadable Data | Data Usability Rating | Metadata | Dataset Preview | Search | Dataset Filter | Location Filter | Field/Topic Filter | Format Filter | License Filter | Year/Date Filter |
| 1. | 2021 | General | Free | Yes | No | Downloada | Yes | Yes | Datasets only | Yes | Yes | Yes | Yes | No |
| 2. | 2020 | General | Free | Yes | No | Downloadable | Yes | Yes | Datasets only | Yes | No | Yes | No | No |
| 3. | 2018 | General | Free | No | No | Downloada | No | Yes | Yes | No | No | No | No | Yes |
| 4. | 2018 | General | Free | Yes | No | View only | Yes | Yes | No | No | No | No | No | No |
| 5. | 2013 | Tsunami | Free | No | No | No | N/A | Yes | Datasets only | Yes | No | No | No | Yes |
| 6. | N/A | Natural Hazards | Free | No | No | No | N/A | No | Datasets only section | No | No | No | No | No |
| 7. | 2021 | General | Free | Yes | Yes | Downloada | No | Yes | Datasets only | No | Yes | Yes | Yes | No |
| 8. | 2020 | Disasters | Registration needed | Yes | N/A | N/A | Yes | N/A | N/A | Yes | Yes | N/A | N/A | Yes |
| 9. | 2021 | General | Free | Yes | No | View only | Yes | Yes | Yes | No | Yes | No | Yes | No |
| 10. | 2018 | General | Free | Yes | No | Downloadable | No | Yes | Datasets only | No | Yes | No | No | Yes |
| 11. | 2021 | General | Free | Yes | Yes | View only | Yes | Yes | Datasets only section | No | No | Yes | Yes | No |
| 12. | 2021 | General | Paid | N/A | N/A | View only | Yes | Yes | Yes | No | Yes | No | No | No |
| 13. | 2020 | General | Free | Yes | No | Downloada | Yes | Yes | Yes | No | Yes | Yes | Limited | Yes |
| 14. | 2021 | General | Free | No | No | View only | No | Yes | Datasets only | No | Yes | Yes | Limited | Limited |
| 15. | 2020 | Water-related | Free | Yes | Yes | View only | No | Yes | Datasets only | Yes | No | Yes | Yes | No |
| 16. | 2017 | General | Free | Yes | Not rated | Downloadable | No | Yes | Datasets only section | Yes | No | No | Yes | No |
| 17. | 2021 | Environmei | Free | Yes | No | View only | Yes | Yes | Datasets only | Yes | Yes | No | No | Yes |
| 18. | N/A | Health | Free | No | No | No | No | Yes | Datasets only section | No | No | No | No | No |
| 19. | N/A | Disasters | Free | Yes | No | No | No | No | Datasets only | Yes | No | No | No | No |
| 20. | 2021 | General | Free | Yes | No | View only | No | Yes | Datasets only | Yes | No | Yes | Yes | No |
| 21. | N/A | General | Free | No | No | No | No | Yes | Datasets only section | No | Yes | No | No | No |

<p align="center">**Table A2.** *Cont.*</p>

| | Description | | | | | | | | Filter/Search Options | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Last Update | Repository Domain | Availability | Downloadable Data | Data Usability Rating | Metadata | Dataset Preview | Search | Dataset Filter | Location Filter | Field/Topic Filter | Format Filter | License Filter | Year/Date Filter |
| 22. | 2021 | General | Free | Yes | No | View only | No | Yes | Datasets only section | Yes | Yes | Yes | No | No |
| 23. | 2020 | Natural Hazards | Free | No | N/A | N/A | N/A | N/A | N/A | Yes | Yes | N/A | N/A | Yes |
| 24. | 2017 | Natural Hazards | Free | Yes | No | View only | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes |
| 25. | 2015 | Tsunami | Registration | Yes | N/A | N/A | N/A | No | Datasets only | No | No | No | No | No |
| 26. | 2021 | General | Free | Partially | No | N/A | Yes | Yes | Yes | No | No | No | No | No |
| 27. | 2020 | General | Free | Yes | No | View only | No | Yes | Datasets only | Yes | Yes | No | No | No |
| 28. | 2021 | Biotechnology | Free | Partially | No | Downloadable | No | Yes | No | No | Yes | No | No | Limited |
| 29. | N/A | General | Free | Yes | No | Downloada | Yes | Yes | Yes | No | Yes | Limited | Limited | Yes |
| 30. | 2021 | General | Free | No | No | Downloadable | No | Yes | Yes | No | No | No | No | Yes |
| 31. | 2015 | General | Free | Yes | No | Downloada | No | Yes | Datasets only section | No | Yes | Yes | No | No |
| 32. | N/A | General | Free | No | No | View only | No | Yes | Datasets only section | No | Yes | No | No | No |
| 33. | N/A | General | Free | Yes | No | Downloada | Yes | Yes | Yes | No | Yes | Yes | Yes | No |
| 34. | N/A | General | Free | Yes | No | View only | No | Yes | Datasets only | No | Yes | No | No | No |
| 35. | 2019 | General | Free | Yes | No | View only | No | Yes | Datasets only | Yes | Yes | Yes | Yes | Yes |
| 36. | 2010 | General | Free | No | No | No | No | Yes | No | No | Yes | No | No | No |
| 37. | N/A | General | Limited | Yes | No | No | Yes | Yes | Yes | Yes | Yes | No | No | Yes |
| 38. | 2021 | Natural Hazards | Free | Yes | No | View only | Yes | Yes | Datasets only | No | Yes | Yes | No | No |
| 39. | 2020 | General | Free | Yes | Yes | Downloada | Yes | Yes | Datasets only section | Yes | Yes | Yes | Yes | No |
| 40. | N/A | General | Paid | N/A | No | View only | N/A | Yes | Yes | No | Yes | Yes | No | Yes |
| 41. | 2020 | Bathymetry | Free | Yes | N/A | N/A | N/A | No | Datasets only section | No | No | No | No | No |
| 42. | 2010 | Bathymetry | Free | Yes | N/A | N/A | N/A | No | Datasets only | No | No | No | No | No |

**Table A2.** *Cont.*

| | Description | | | General Availability | | | | | Filter/Search Options | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Last Update | Repository Domain | Availability | Downloadable Data | Data Usability Rating | Metadata | Dataset Preview | Search | Dataset Filter | Location Filter | Field/Topic Filter | Format Filter | License Filter | Year/Date Filter |
| 43. | 2021 | Sea Level | Free | Yes | N/A | N/A | N/A | No | Datasets only | No | No | No | No | No |
| 44. | N/A | Water-related | Paid | N/A | N/A | N/A | N/A | No | Datasets only section | No | Yes | No | No | No |
| 45. | 2020 | Geography | Free | Yes | No | Downloada | Yes | No | Datasets only | No | No | No | No | No |
| 46. | 2019 | Environment | Registration needed | Yes | No | Downloadable | No | Yes | Yes | No | No | No | No | No |
| 47. | N/A | Satellite Imaging | Free | N/A | N/A | N/A | N/A | No | Datasets only section | No | Yes | No | No | No |
| 48. | N/A | Geography | Free | Yes | N/A | N/A | Yes | No | Datasets only section | No | No | No | No | No |
| 49. | N/A | Geography | Registration needed | Yes | N/A | N/A | N/A | No | Datasets only | Yes | Yes | No | No | No |
| 50. | N/A | Environment | Free | Yes | N/A | N/A | N/A | No | Datasets only | No | No | No | No | No |
| 51. | N/A | Satellite Imaging | Free | Yes | N/A | N/A | N/A | No | Datasets only | No | No | No | No | No |
| 52. | 2011 | Topography | Limited | Yes | No | Downloadable | Yes | Yes | Datasets only | Yes | Yes | Limited | No | No |
| 53. | 2021 | Sea Level | Free | Yes | N/A | N/A | Yes | No | Datasets only | No | No | No | No | No |
| 54. | 2020 | General | Free | Yes | Yes | View only | Yes | Yes | Datasets only | Yes | No | Yes | Yes | Yes |
| 55. | 2021 | Geography | Limited | N/A | No | View only | No | Yes | Datasets only section | No | No | No | No | No |
| 56. | 2019 | Tsunami | Free | Yes | N/A | N/A | Yes | Yes | Datasets only | Yes | No | N/A | N/A | Yes |
| 57. | 2021 | Tsunami | Free | Yes | N/A | N/A | Yes | Yes | Datasets only | Yes | No | N/A | N/A | Yes |
| 58. | 2014 | Tsunami | Free | No | N/A | N/A | Yes | No | Datasets only | No | No | N/A | N/A | No |
| 59. | N/A | Tsunami | Free | Yes | N/A | N/A | No | Yes | Datasets only | No | No | N/A | N/A | No |
| 60. | 2010 | Tsunami | Free | Yes | No | No | No | No | Datasets only | No | No | No | No | No |

## References

1. Papadopoulos, G.; Lorito, S.; Løvholt, F.; Rudloff, A.; Schindele, F. Geophysical risk: Tsunami. In *Science for Disaster Risk Management 2017: Knowing Better and Losing Less*; Poljanšek, K., Marín Ferrer, M., De Groeve, T., Clark, I., Eds.; Publications Office of the European Union: Luxembourg, 2017; pp. 162–176. ISBN 978-92-79-60678-6.

2. Harbitz, C.B.; Løvholt, F.; Bungum, H. Submarine Landslide Tsunamis: How Extreme and How Likely? *Nat. Hazards* **2014**, *72*, 1341–1374. [CrossRef]

3. Behrens, J.; Løvholt, F.; Jalayer, F.; Lorito, S.; Salgado-Gálvez, M.A.; Sørensen, M.; Abadie, S.; Aguirre-Ayerbe, I.; Aniel-Quiroga, I.; Babeyko, A.; et al. Probabilistic Tsunami Hazard and Risk Analysis—A Review of Research Gaps. *Front. Earth Sci.* **2021**, *9*. [CrossRef]

4. Röbke, B.R.; Vött, A. The Tsunami Phenomenon. *Prog. Oceanogr.* **2017**, *159*, 296–322. [CrossRef]

5. Shinozaki, T. Geochemical Approaches in Tsunami Research: Current Knowledge and Challenges. *Geosci. Lett.* **2021**, *8*, 6. [CrossRef]

6. Chagué-Goff, C.; Szczuciński, W.; Shinozaki, T. Applications of Geochemistry in Tsunami Research: A Review. *Earth Sci. Rev.* **2017**, *165*, 203–244. [CrossRef]

7. Anpalagan, A.; Woungang, I. Tsunami Prediction and Impact Estimation Using Classifiers on Historical Data. In Proceedings of the 2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA), Valencia, Spain, 19–22 October 2020; pp. 119–126.

8. Papadopoulos, G.A.; Gràcia, E.; Urgeles, R.; Sallares, V.; De Martini, P.M.; Pantosti, D.; González, M.; Yalciner, A.C.; Mascle, J.; Sakellariou, D.; et al. Historical and Pre-Historical Tsunamis in the Mediterranean and Its Connected Seas: Geological Signatures, Generation Mechanisms and Coastal Impacts. *Mar. Geol.* **2014**, *354*, 81–109. [CrossRef]

9. Ai, C.; Ma, Y.; Yuan, C.; Xie, Z.; Dong, G. A Three-Dimensional Non-Hydrostatic Model for Tsunami Waves Generated by Submarine Landslides. *Appl. Math. Model.* **2021**, *96*, 1–19. [CrossRef]

10. Titov, V.; Moore, C. Meteotsunami Model Forecast: Can Coastal Hazard Be Quantified in Real Time? *Nat. Hazards* **2021**, *106*, 1545–1561. [CrossRef]

11. Macías, J.; Escalante, C.; Castro, M.J. Multilayer-HySEA Model Validation for Landslide-Generated Tsunamis-Part 2: Granular Slides. *Nat. Hazards Earth Syst. Sci.* **2021**, *21*, 791–805. [CrossRef]

12. Sugawara, D. Numerical Modeling of Tsunami: Advances and Future Challenges after the 2011 Tohoku Earthquake and Tsunami. *Earth Sci. Rev.* **2021**, *214*. [CrossRef]

13. Kurniawan, T.; Yuliatmoko, R.S.; Sunardi, B.; Prayogo, A.S.; Muzli, M.; Rohadi, S. Tsunami Simulation for Disaster Mitigation Based on Earthquake Scenarios in the Molucca Subduction Zone (Case Study of the Molucca Sea Earthquake on 7 July 2019). *AIP Conf. Proc.* **2021**, *2320*, 040026-1–04006-7.

14. Bosnic, I.; Costa, P.J.M.; Dourado, F.; La Selle, S.; Gelfenbaum, G. Onshore Flow Characteristics of the 1755 CE Lisbon Tsunami: Linking Forward and Inverse Numerical Modeling. *Mar. Geol.* **2021**, *434*, 106432. [CrossRef]

15. Kim, J.; Omira, R. The 6–7 July 2010 Meteotsunami along the Coast of Portugal: Insights from Data Analysis and Numerical Modelling. *Nat. Hazards* **2021**, *106*, 1397–1419. [CrossRef]

16. Ravanelli, M.; Occhipinti, G.; Savastano, G.; Komjathy, A.; Shume, E.B.; Crespi, M. GNSS Total Variometric Approach: First Demonstration of a Tool for Real-Time Tsunami Genesis Estimation. *Sci. Rep.* **2021**, *11*, 3114. [CrossRef]

17. Mulia, I.E.; Satake, K. Synthetic Analysis of the Efficacy of the S-Net System in Tsunami Forecasting. *Earth Planets Space* **2021**, *73*. [CrossRef]

18. Pararas-Carayannis, G. Brief History of Early Pioneering Tsunami Research—Part A. *Sci. Tsunami Hazards* **2018**, *37*, 49–129.

19. Trinaistich, W.C.; Mulligan, R.P.; Take, W.A. Runup of Landslide-Generated Waves Breaking on Steep Slopes Captured Using Digital Imagery and Hydrochromic Paint. *Coast. Eng.* **2021**, *166*, 103888. [CrossRef]

20. Keet, M. Methodologies for Ontology Development. Available online: https://eng.libretexts.org/Bookshelves/Computer_Scienc e/Programming_and_Computation_Fundamentals/Book%3A_An_Introduction_to_Ontology_Engineering_(Keet)/06%3A_ Methods_and_Methodologies/6.01%3A_Methodologies_for_Ontology_Development (accessed on 14 July 2021).

21. Contreras, M.C.B.; Reyes, L.F.H.; Ortiz, J.A.R. Methodology for Ontology Design and Construction. *Contad. Adm.* **2019**, *64*, 134. [CrossRef]

22. Gómez-Pérez, A.; Fernández, M.; Vicente, A. de Towards a Method to Conceptualize Domain Ontologies. In *Proceedings Workshop: Ontological Engineering, Proceedings of the 12th European Conference on Artificial Intelligence (ECAI'96), Budapest, Rumanía, 13 August 1996*; Facultad de Informática (UPM): Budapest, Rumanía, 1996.

23. Noy, N.F.; McGuinness, D.L. Ontology Development 101: A Guide to Creating Your First Ontology 2001. Available online: https://protege.stanford.edu/publications/ontology_development/ontology101.pdf (accessed on 8 August 2021).

24. CRED EM-DAT: The International Disaster Database. Available online: https://www.emdat.be/classification (accessed on 14 July 2021).

25. Library of Congress Recommended Formats Statement. Available online: https://www.loc.gov/preservation/resources/rfs/T OC.html (accessed on 26 April 2021).

26. Gusiakov, V.K.; Dunbar, P.K.; Arcos, N. Twenty-Five Years (1992–2016) of Global Tsunamis: Statistical and Analytical Overview. *Pure Appl. Geophys.* **2019**, *176*, 2795–2807. [CrossRef]

27. Whitlock, M.C. Data Archiving in Ecology and Evolution: Best Practices. *Trends Ecol. Evol.* **2011**, *26*, 61–65. [CrossRef] [PubMed]

28. Tenopir, C.; Allard, S.; Douglass, K.; Aydinoglu, A.U.; Wu, L.; Read, E.; Manoff, M.; Frame, M. Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE* **2011**, *6*, e21101. [CrossRef]
29. White, E.; Baldridge, E.; Brym, Z.; Locey, K.; McGlinn, D.; Supp, S. Nine Simple Ways to Make It Easier to (Re)Use Your Data. *Ideas Ecol. Evol.* **2013**, *6*. [CrossRef]
30. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef]
31. Wernet, G.; Bauer, C.; Steubing, B.; Reinhard, J.; Moreno-Ruiz, E.; Weidema, B. The Ecoinvent Database Version 3 (Part I): Overview and Methodology. *Int. J. Life Cycle Assess.* **2016**, *21*, 1218–1230. [CrossRef]
32. Murnane, R.J.; Allegri, G.; Bushi, A.; Dabbeek, J.; de Moel, H.; Duncan, M.; Fraser, S.; Galasso, C.; Giovando, C.; Henshaw, P.; et al. Data Schemas for Multiple Hazards, Exposure and Vulnerability. *Disaster Prev. Manag. Int. J.* **2019**, *28*, 752–763. [CrossRef]
33. Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; Ives, Z. DBpedia: A Nucleus for a Web of Open Data. In Proceedings of the The Semantic Web, Busan, Korea, 11–15 November 2007; Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., et al., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 722–735.
34. Alexander, K.; Cyganiak, R.; Hausenbals, M.; Zhao, J. Describing Linked Datasets with the VoID Vocabulary. Available online: https://www.w3.org/TR/void/ (accessed on 31 May 2021).
35. Javed, Y.; Norris, T.; Johnston, D. Ontology-Based Inference to Enhance Team Situation Awareness in Emergency Management. In Proceedings of the 8th International ISCRAM Conference, Lisbon, Portugal, 8–11 May 2011; pp. 1–9.
36. Roman, D.; Sukhobok, D.; Nikolov, N.; Elvesæter, B.; Pultier, A. The InfraRisk Ontology: Enabling Semantic Interoperability for Critical Infrastructures at Risk from Natural Hazards. In Proceedings of the On the Move to Meaningful Internet Systems. OTM 2017 Conferences, Rhodes, Greece, 23–28 October 2017; Panetto, H., Debruyne, C., Gaaloul, W., Papazoglou, M., Paschke, A., Ardagna, C.A., Meersman, R., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 463–479.
37. Zhong, S.; Fang, Z.; Zhu, M.; Huang, Q. A Geo-Ontology-Based Approach to Decision-Making in Emergency Management of Meteorological Disasters. *Nat. Hazards* **2017**, *89*, 531–554. [CrossRef]
38. Sermet, Y.; Demir, I. Towards an Information Centric Flood Ontology for Information Management and Communication. *Earth Sci. Inform.* **2019**, *12*, 541–551. [CrossRef]
39. Bannour, W.; Maalel, A.; Ben Ghezala, H.H. Ontology-Based Representation of Crisis Response Situations. In Proceedings of the Computational Collective Intelligence, Hendaye, France, 4–6 September 2019; Nguyen, N.T., Chbeir, R., Exposito, E., Aniorté, P., Trawiński, B., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 417–427.
40. Jain, S.; Mehla, S.; Agarwal, A.G. An Ontology Based Earthquake Recommendation System. In Proceedings of the Advanced Informatics for Computing Research, Shimla, India, 15–16 June 2019; Luhach, A.K., Singh, D., Hsiung, P.-A., Hawari, K.B.G., Lingras, P., Singh, P.K., Eds.; Springer: Singapore, 2019; pp. 331–340.
41. Liu, X.; Liu, Z.; Liu, Y.; Tian, J. Integration of a Geo-Ontology-Based Knowledge Model and Spatial Analysis into Emergency Response for Geologic Hazards. *Nat. Hazards* **2021**. [CrossRef]
42. Gonzalez, A.; Peres-Neto, P.R. Act to Staunch Loss of Research Data. *Nature* **2015**, *520*, 436. [CrossRef]
43. Perrier, L.; Blondal, E.; Ayala, A.P.; Dearborn, D.; Kenny, T.; Lightfoot, D.; Reka, R.; Thuna, M.; Trimble, L.; MacDonald, H. Research Data Management in Academic Institutions: A Scoping Review. *PLoS ONE* **2017**, *12*, e0178261. [CrossRef]
44. Normile, D. Scientific Consensus on Great Quake Came Too Late. *Science* **2011**, *332*, 22–23. [CrossRef] [PubMed]
45. Arcos, N.P.; Dunbar, P.K.; Stroker, K.J.; Kong, L.S.L. The Impact of Post-Tsunami Surveys on the NCEI/WDS Global Historical Tsunami Database. *Pure Appl. Geophys.* **2019**, *176*, 2809–2829. [CrossRef]
46. Jain, N.; Virmani, D.; Abraham, A. Tsunami in the Last 15 Years: A Bibliometric Analysis with a Detailed Overview and Future Directions. *Nat. Hazards* **2021**, *106*, 139–172. [CrossRef]
47. Dilectin, H.D.; Mercy, R.B.V. Classification and Dynamic Class Detection of Real Time Data for Tsunami Warning System. In Proceedings of the 2012 International Conference on Recent Advances in Computing and Software Systems, Chennai, India, 27–27 April 2012; pp. 124–129.
48. Kusumah, Y.; Irawan, B.; Setianingsih, C. Sea Wave Detection System Using Web-Based Decision Tree Algorithm. In Proceedings of the 2020 10th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS), Malang, Indonesia, 26–28 August 2020; pp. 231–236.
49. Pughazhendhi, G.; Raja, A.; Ramalingam, P.; Elumalai, D.K. Earthosys—Tsunami Prediction and Warning System Using Machine Learning and IoT. In Proceedings of the International Conference on Computational Intelligence and Data Engineering, Chennai, India, 21–23 February 2019; Chaki, N., Devarakonda, N., Sarkar, A., Debnath, N.C., Eds.; Springer: Singapore, 2019; pp. 103–113.
50. Liliana, D.Y.; Priharsari, D. Tsunami Early Warning Detection Using Bayesian Classifier. In Proceedings of the 2019 2nd International Conference of Computer and Informatics Engineering (IC2IE), Banyuwangi, Indonesia, 10–11 September 2019; pp. 44–48.
51. Yoshikawa, M.; Igarashi, Y.; Murata, S.; Baba, T.; Hori, T.; Okada, M. A Nonlinear Parametric Model Based on a Power Law Relationship for Predicting the Coastal Tsunami Height. *Mar. Geophys. Res.* **2019**, *40*, 467–477. [CrossRef]