

Publishing, Linking and Annotating Events via Interactive Timelines: an Earth Sciences Case Study

Lianli Gao¹, Jane Hunter¹

¹ School of ITEE, The University of Queensland,
Brisbane, Australia
{ l.gao1, j.hunter}@uq.edu.au

Abstract. Events are a critical entity for documenting information within many domains - and yet they are one class of information that, to date, has been relatively neglected with regard to both publishing on the Semantic Web and semantically annotating. In this paper we describe how we enable the interoperable integration, annotation and linking of information about major events from the earth sciences domain, by adopting a Linked Data approach to major events (earthquakes, tsunamis and volcanic eruptions) and the timelines and annotations that capture additional domain-expert knowledge. Firstly we describe the common *Event*, *Timeline*, *Annotation* and *TemporalRelation* ontologies that we use to enable interoperability and exchange of information about events and the relationships between them. We then harvest data describing major geological events from multiple authoritative sources, map it to our model(s) and publish it as RDF triples to the Web of Linked Data. We then describe the semantic annotation system that we have developed that enables the discovery, retrieval and ontology-based markup of such event data via interactive timelines. The resulting annotations significantly enhance the discovery and re-use of information about major geological events. More importantly these annotation tools enable scientists to document, share and discuss their hypotheses about the temporal relationships between such events.

Keywords: events, timelines, linked data, semantic annotations, geosciences

1 Introduction

Understanding the temporal relationships between historical events is often a critical step in predicting the occurrence of future events. The focus of this paper is on tools to assist scientists to improve their understanding of the temporal relationships between major geological events (earthquakes, tsunamis and volcanic eruptions). More specifically this paper describes the ontologies and semantic annotation system that we have developed that enables earth scientists to visualize, annotate and analyse temporal relationships between major geological events (earthquakes, tsunamis and volcanic eruptions) using interactive timelines. In the process of developing this system, we have also developed a Linked Data approach to such events that retrieves relevant data from a number of disparate authoritative

sources, integrates the datasets via a common *Event* model and publishes it as RDF triples (via Atom feeds or to a Linked Data Hub). Moreover to facilitate the interoperability, exchange and re-use of both the geological events and user-specified aggregations and annotations, we have also developed common, extensible ontologies to describe *Timelines*, *TemporalRelations*, and *Annotations*. The details and source of these ontologies are also described in this paper. The outcome is a set of tools for reasoning about the temporal relationships between major geological events that will hopefully lead to better models for predicting such potentially catastrophic events in the long run.

2 Objectives

Events are a critical information entity within many domains and yet they are overlooked when it comes to publishing as Linked Data. They are often hidden or encapsulated within databases, Web pages or timelines which prohibit their independent discovery, re-use, annotation or linking. The aim of the work described in this paper is to illustrate the benefits that are possible by treating events as first-class information objects and publishing them on the Semantic Web - where they can be annotated, interpreted and linked to other related datasets or events. More specifically, we demonstrate these benefits in the context of the earth sciences domain to enable scientists to analyse temporal relationships between past earthquake, tsunami and volcanic events (commonly known as “geochange” events) [1, 2].

Our first objective is to define a common data model for describing geochange events that enables interoperability between such events and a standardized model for publishing such events as Linked Data. Given this common model, we can then extract relevant geochange event data (about volcanic eruptions, earthquakes and tsunamis) from multiple authoritative Web sources (online databases, Web sites and timelines), represent it in our Event model and publish it as RDF triples, Atom feeds and/or to a Linked Data Hub.

One of the most common methods used to document, describe, aggregate, publish and visualize real world events on the Web is via *Timelines*. They provide a graphical representation of a chronology or sequence of events displayed along a time axis. One challenge to sharing event data is the multitude of timeline software systems and the lack of interoperability between them. When an event is published via a timeline (built using specific timeline software), the individual event data is not accessible, discoverable or re-usable – it is part of the “deep web”, locked inside the particular timeline software and format. It is necessary to decouple *events* from *timelines* and the timeline rendering software – so that both events and timelines are discoverable and re-usable independently. Hence our second objective is to describe a common, interoperable model for *timelines* – that incorporates links to the contained events and that can also be published to the Web of Linked Data. To evaluate our timeline model, we identify a number of existing geochange timelines and show how the encapsulated data can be mapped to our models, without loss of information.

Given the resulting availability of both geochange events and timelines on the Web as Linked Data, our third objective is to develop semantic annotation tools for events – that enable researchers or the general public to add semantic markup to critical events

to enable additional interpretations and knowledge to be captured and to facilitate further reasoning across the events. Such knowledge includes the identification of temporal or causal relationships between events, which will lead to better predictability and early warning systems. As part of this objective, we also aim to develop an ontology of temporal relationships between events and methods for annotating relationships between events within the same and different timelines. Finally we plan to evaluate these ontologies and services in the context of the earth sciences domain by applying them to *geochange* events.

3 Related Work

There has been significant past research within numerous domains that aims to develop a *common event model* to support information integration. For example, the ABC model [3] was defined to document events (primarily in the information domain) that capture the provenance of documents undergoing change across multiple systems and platforms. The CIDOC/CRM [4] focuses on an interoperable model to support metadata exchange within cultural institutions. The Event (and associated Time) Ontology [5] was originally defined to describe events in the music and performance domain but has since been applied more generally. Other upper ontologies in which events are key entities include DOLCE+DnS Ultralite [6], the F Event model [7] and OpenCYC [8]. Shaw et al [9] provides a comparison of some of these existing event models in an effort to provide an interlingua model – the LODE ontology. Our approach is to adopt a simplified version of the LODE ontology (which is described in Section 5.1) and to apply it to specific types of geochange events (earthquakes, tsunamis and volcanic eruptions).

There also exists a vast number of Web-based tools for authoring and editing timelines. Examples include: SIMILE¹, My Timelines, Timeline Builder, xtimeline, Time Morph, Timelinr, Preceden and TimeGlider². All of these systems rely on different sets of attributes and metadata to document the information contained within each timeline. There is little to no interoperability between the many different timeline software tools. The Timeline ontology developed by Raimond and Abdallah for the music domain is the most relevant previous work that aimed to develop a standardized RDF/OWL model for timelines [5]. However this timeline ontology is specifically aimed at the music domain to support mappings between time scales. In our Timeline model, we re-use a simplified sub-set of their classes and properties (described in Section 5.2). In addition, we define a “references/referencedBy” relationship between *timelines* and *events*, each of which are uniquely identifiable via persistent URIs. We also apply and evaluate the model using geochange data.

Existing approaches to annotating “events” have primarily involved proprietary approaches in which the annotations are locked inside the specific timeline tool or system. For example, Google’s Interactive Charts, enable users to attach annotations to interactive timelines/charts that are rendered using Flash³. The annotations are not Web resources and are only accessible through the Google javascript API used to

¹ <http://www.simile-widgets.org/timeline/>

² <http://www.shambles.net/pages/school/timelines/>

³ <http://code.google.com/apis/chart/interactive/docs/gallery/annotatedtimeline.html>

generate the timelines/charts. *RecordedFuture*⁴ is an example of a browser-based temporal analytics tool that enables users to explore and visualize time-based data. *RecordedFuture* enables users to add annotations to events in the Timeline view. It also enables users to share event visualizations through Facebook, Twitter or a newly generated URL. However it is a commercial product that only allows the sharing of timelines between users who have purchased RecordedFuture. Other examples of timeline-based tools that support annotations include ChronoViz [10] and Chronozoom⁵ – but again, neither system publishes the annotations as independent Web resources with unique persistent URIs that are discoverable, independent of the events. SemaTime [11] combines a timeline visualization interface with semantic annotation tools to annotate relationships between entities – but the focus is on visualizing semantic relationships that change over time (e.g., married_to).

4 Methodology

4.1 Case Study

Volcanic eruptions, earthquakes and tsunamis are strongly related to each other – both spatially and temporally. Both earthquakes and volcanoes occur at the boundaries of the tectonic plates that comprise the Earth’s surface. Earthquakes are caused by pressure built up when plates collide, move apart, or slide past each other or over each other. Volcanoes form when the magma that is generated at plate boundaries rises to the surface. The movement of magma within a volcano or the adjustment of plates under volcanoes, causes earthquakes. Tsunamis are caused by the occurrence of earthquakes in oceanic or coastal regions. Understanding the temporal relationships between these geochange events will help scientists to develop better predictive models and early warning systems that may save lives of communities living in endangered zones. Hence our objective is to provide geologists and earth scientists with Web-based tools that enable them to aggregate disparate data sets describing geochange events and to document, analyse and interpret the temporal relationships between such events.

4.2 Process

Our approach can be divided into the five stages:

1. Firstly we developed common ontologies/data models for describing *events*, *timelines*, *annotations* and *temporal relationships*;
2. Next we harvested geochange event data from multiple Web sites and timelines (NOAA’s National Geophysical Data Center’s (NGDC) Natural Hazards Data⁶, USGS Earthquake Database⁷ etc), and represented it in our event and timeline models. We stored this data in our own RDF triple store, generating HTTP URIs for each event and timeline – but we also generated an Atom feed and published the RDF triples to the the Comprehensive Knowledge Archive Network (CKAN) [12] Linked Data Hub;

⁴ <https://www.recordedfuture.com/>

⁵ <http://eps.berkeley.edu/~saekow/chronozoom/>

⁶ <http://www.ngdc.noaa.gov/hazard/>

⁷ <http://earthquake.usgs.gov/earthquakes/eqarchives/epic/database.php>

3. We then developed a SPARQL interface to our RDF triple store to enable users to search, retrieve and display events based on metadata fields and display them on a Simile Widget Timeline [13];
4. Next we developed the SAFE (Semantic Annotation For Events) Firefox plugin that enables users to: annotate a single event on single timeline; annotate multiple events on single timeline or within a time period; annotate multiple events on different timelines displayed simultaneously; annotate relationships between events on same timeline or different timelines. The annotations are stored on an annotation server – using our OAC-based annotation model [14] – but we can also publish/share them as RDF via HTTP URIs and Atom feed. Users can also search and retrieve events via the annotations.
5. Finally we evaluated the system through user feedback and performance measures.

5 Ontologies

This section describes the ontologies that we've developed to support the publishing, linking and annotation of geological events. We have drawn on existing vocabularies and terms from the namespaces listed in the table below.

Prefix	XML namespace	Description
dc	http://purl.org/dc/elements/1.1/	Dublin Core
event	http://purl.org/NET/c4dm/event.owl#	Event ontology
tl	http://purl.org/NET/c4dm/timeline.owl#	Timeline ontology
time	http://www.w3.org/2006/time#	OWL-Time
geo	http://www.w3.org/2003/01/geo/wgs84_pos#	WGS84 GeoPositioning

5.1 Event Ontology

There exists a wide variety of existing *Event* ontologies from which to draw on, including the ABC ontology, the CIDOC CRM ontology, the Event Ontology, DOLCE+DnS Ultralite, OpenCYC and LODDE [9]. Based on an analysis of these existing ontologies, as well as the requirements of our application, we defined a new *GeochangeEvent* class which is a subclass of the Event ontology's Event class [4] and defined the following properties in our GeochangeEvent class:

dc:identifier – HTTP URI for this event;

dc:title – title of the event;

dc:description – literal describing the event;

dc:source – HTTP URI of the source of the event data (e.g., USGS Web site);

event:time – range = time:TemporalEntity;

region – name of region where it occurred;

country – name of the country where it occurred;

geo:lat – coordinates in decimal degrees;

geo:long – coordinates in decimal degrees;

isReferencedBy – URIs of *Timelines* that reference this event.

We also defined a set of sub-classes of the *GeochangeEvent* class, that are specific to the geochange domain: *Earthquake*, *VolcanicEruption*, *Tsunami*. These three sub-classes each have additional specific properties. Earthquake events have the additional properties of: *magnitude* (0.0-9.9), *intensity* (0-12), *focalDepth* (0-700 km) and *numberOfDeaths*. Tsunamis have the additional properties of *waterHeight* (0-525 ms) and *numberOfDeaths*. VolcanicEruptions have the additional properties of *volcanoName*, *volcanoType* (Caldera, CinderCone, Lava, Mud, Pumice, Pyroclastic, Shield etc), *volcanicExplosivityIndex* (0-8) and *numberOfDeaths*.

5.2 Timeline Ontology

Our Timeline ontology was developed by analyzing the attributes used to describe existing timelines (e.g., the SIMILE widget/timeline) and also by drawing on terms from the Timeline ontology [5]. A critical addition to our model is the “references/referencedBy” relationship:

- dc:identifier** – HTTP URI for this timeline;
- dc:creator** – author of the timeline;
- dc:title** – literal title for the timeline;
- dc:description** – decription of the timeline;
- dc:date.created** – date the timeline was created;
- tl:beginsAtDateTime** – the start of the timeline;
- tl:endsAtDateTime** – the end of the timeline;
- intervalUnit** – the unit to be displayed on the axis intervals e.g. 1 hour, 1 year;
- references** – URIs of *Events* that are contained within this timeline..

5.3 Annotation Ontology

We chose to base our annotation ontology on the Open Annotation Collaboration (OAC) ontology [14] – which was specifically designed to enable the publishing and linking of annotations on the Web of Linked Data. The OAC ontology is ideal because: it is designed to support annotations in which the *body* and the *target* of the annotation may be of any media type (e.g. the body might be a seismograph); the annotation, body and target are all identifiable via HTTP URIs; and multiple targets are supported. This last aspect is particularly relevant as we want to support the annotation of temporal relationships between multiple events in different timelines.

Figure 1 below illustrates the OAC model corresponding to the annotation (A-1) of a “causal” relationship (B-1) between the Honshu Earthquake (E-1) in Japan in March 2011 contained in the Global Earthquakes timeline (TL-1) and the tsunami in the Miyako province (E-2) contained in the Global Tsunamis timeline (TL-2).

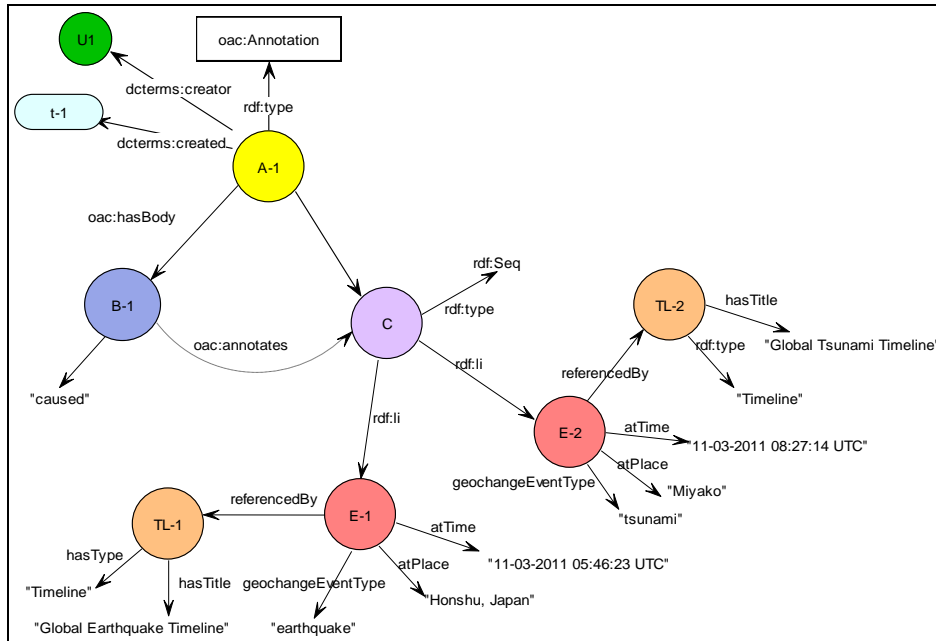


Fig 1: Using OAC to model the annotation of a causal relation between two events

5.4 TemporalRelation Ontology

The role of the TemporalRelation Ontology is to provide a set of controlled terms that can be used to tag relationships between different types of geochange events. The list of terms/properties that apply are listed below. This list is adapted from the list of temporal relations defined in the OWL Time ontology [15]:

- time:before/time:after** – one event precedes/follows another;
- time:intervalOverlaps** – the duration of two events overlaps;
- time:intervalEquals** – the start and end times of two events coincide;
- time:intervalMeets** – the end of one event coincides with the start of another event;
- time:intervalContains** – one event starts and finishes within the duration of a second event.

5.5 OtherRelations

Apart from the temporal relations described above, there were three other relationships that we defined:

- isRelatedTo** – one event is related to another – but the precise relationship is unclear;
- causes/causedBy** – one event causes/triggers another event (subPropertyOf isRelatedTo);
- requires/requiredBy** – one event cannot occur unless the other event has already occurred (subPropertyOf isRelatedTo).

6 Semantic Annotation Prototype

6.1 System Architecture

Figure 2 illustrates a high level view of the system architecture for our Semantic Annotation For Events (SAFE) service. A large collection of geochange event data is harvested from multiple authoritative Web sites and timelines including: USGS Earthquakes Database and the the NOAA NGDC natural hazards database. This data is mapped to our Event and Timeline models and stored as RDF triples in a Sesame RDF triple store. A user interface was developed that enables users to search and retrieve events from the Sesame triple store via titles, descriptions, dates/date ranges and/or keywords, using SPARQL. The retrieved events are dynamically displayed via a browser-based interactive timeline built using the SIMILE timeline widget. The SAFE annotation client is a Firefox sidebar built using XUL (XML User Interface Language), AJAX (Asynchronous JavaScript) and JavaScript. The annotation server is implemented using the Apache Tomcat. The various components of the system are accessible via the project Web Portal⁸.

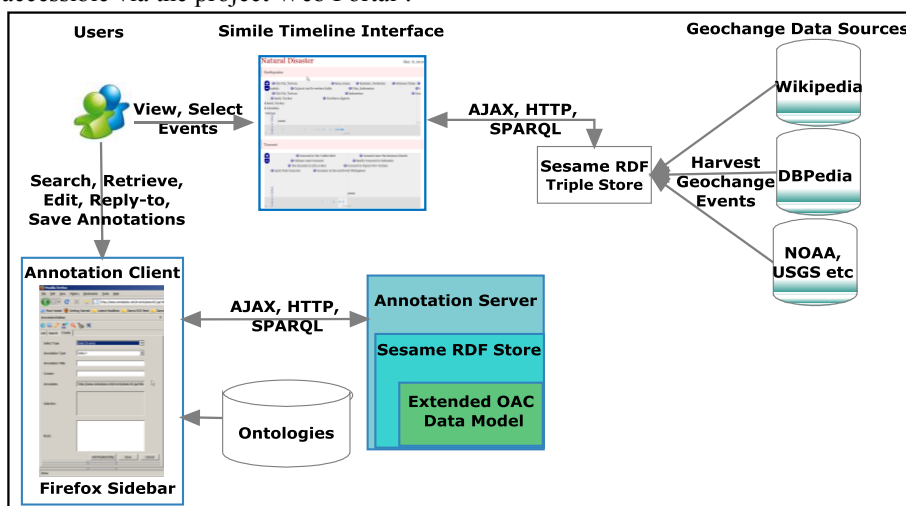


Fig. 2. High-level Architecture of the Semantic Annotation System

6.2 User Interface

The objective of the SAFE annotation system is to enable users to interactively:

- Annotate a single event on single timeline;
- Annotate multiple events on the same or multiple timelines, by selecting them individually or by interactively specifying a time period;
- Annotate relationships between events on the same or different timelines;
- Search and retrieve annotations and associated events based on the annotation metadata. Examples of is: "give me all causal relationships and associated events", "give me all of the Timelines that reference this event".

⁸ <http://seas.metadata.net/events/>

Figure 3 illustrates the SAFE Firefox user interface being used to annotate a relationship between two events on different timelines. In the top LHS of Fig 3, is the sidebar for creating a new annotation - users specify the annotation type, title, creator, and body (e.g., controlled term, free text or URI). Users can also search, browse, edit and delete annotations via this sidebar. On the RHS are displayed one or more Simile timelines. The annotation client communicates with the Simile timeline(s) to extract and record the time range and the selected events that are associated with each annotation. Clicking on an existing annotation in the sidebar, causes the timeline to jump to the annotated event, highlight the event and display the associated annotation. To annotate a relationship between multiple events, users open one or more timelines, select the events of interest, and create a new annotation – the body of which is chosen from a pull-down menu. The corresponding RDF graph is displayed in the lower left hand corner of the sidebar.

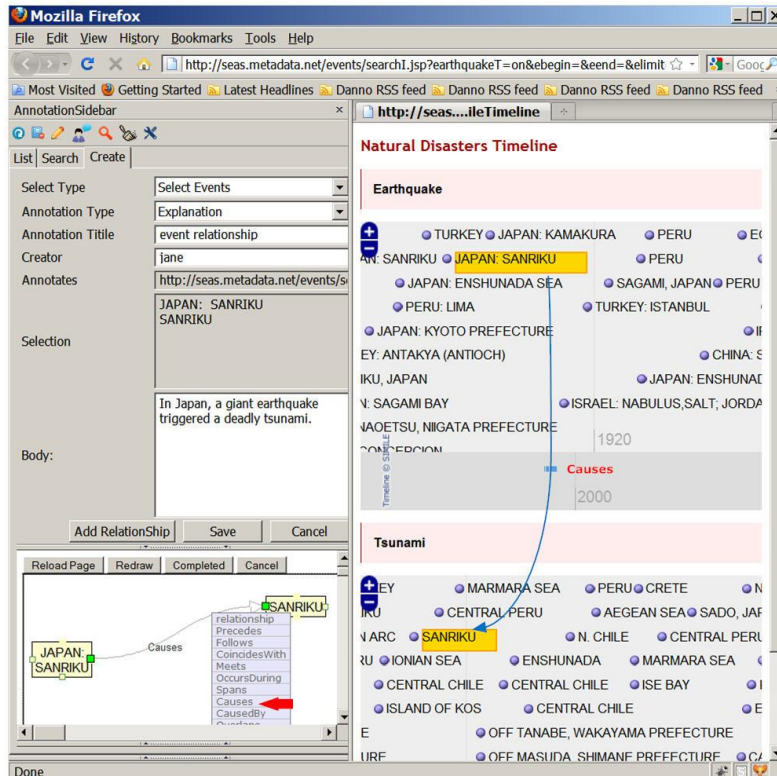


Fig 3: Using the SAFE Plugin with Firefox to annotate relations between events on 2 timelines

7 Evaluation and Discussion

The evaluation of the event, timeline and annotation ontologies described in Section 5 was based on their ability to support the mapping of harvested datasets and timelines from authoritative sources on earthquakes, tsunamis and volcanic eruptions. This mapping exercise illustrated that the property extensions to the sub-classes of geochange events enabled accurate descriptions to be captured e.g., the magnitude associated with an earthquake or the water height associate with a tsunami. An

additional sub-class of events that was identified as missing from the `geochangeEvent` class was “runup”. Runups are a consequence of tsunamis that occur when the water level rises onshore at multiple locations along the coastline, and of interest to geoscientists. The relationships ontology needs to be extended to support related rules and restrictions. For example, a “causal” relationship between two events is only possible if the start time of the first event occurs before the start time of the second event. Currently the system does not check for such prerequisite temporal relationships but this would be relatively easy to implement as a validation process within the client annotation tool before saving the annotation. The OAC model was ambiguous in the context of annotating relationships between multiple events. The OAC model recommends the use of `ore:aggregations` for annotating multiple targets – however if they are ordered (e.g., sequential/list) then perhaps a blank node which is an `rdf:Seq` or `rdf:List` is a better approach. The other potential disadvantage associated with the OAC approach is the need to generate URIs for the annotation, body and target. This may well lead to a URI management problem in the long term – as well as a scalability problem as the number of annotations becomes very large and SPARQL querying struggles with the size of the RDF triple store.

User feedback to the SAFE annotation service (via a questionnaire) was mixed. Users found the Firefox sidebar easy to download, install, configure and the annotation interface intuitive and user friendly. Users liked the integration of the sidebar and timeline within the single browser and the speed of synchronization between the two panels. Users requested the ability to open more than two timelines simultaneously and to tag relationships between events contained within three or more timelines. They also requested the ability to attach a *certainty* measure to relationship tags. For example, they might tag a particular tsunami event as being *causedBy* a particular earthquake event, but the author’s confidence in this assertion is only 75%. Finally, a significant number of users requested the ability to specify both geo-spatial and temporal relationships simultaneously via a combined mapping and timeline interface (such as TimeMap [16]). Related to this was the additional request to enable interactive specification of more sophisticated querying and inferencing rules. For example, “find all tsunami events that fall within a 1000 km radius and within 18 hours of a particular earthquake event and tag them as *causedBy* the earthquake”..

8 Future Work and Conclusions

We have identified a number of future work directions that we would like to pursue. Firstly, we plan to integrate the timeline with a mapping interface to enable the annotation and visualization of both spatial, temporal and spatio-temporal relationships between geochange events. We are also planning to implement (SWRL) inferencing rules that enable users to reason across the data based on the tagged or inferred temporal relationships. For example, if someone tags *earthquake-E1->causes->tsunami-T1*, and someone else tags *tsunami-T1->causes->runups-R1, R2, R3, R4*. Then because the *causal* property is transitive, the system can infer that *earthquake-E1->causes->runups-R1, R2, R3, R4*. Users can then ask queries such as “what is the total numberOfDeaths caused by earthquake-E1?”.

In conclusion, we have described a set of services that enable information about geological events (that was previously hidden in databases, Web sites and timelines) to be exposed on the Web as Linked Data. Given the availability of these rich datasets on earthquakes, tsunamis and volcanic eruptions, we then developed a set of timeline-based annotation services that enable users to document and share their ideas and hypotheses about the temporal relationships between such events. The outcome is an extensible framework and a robust foundation for future more advanced temporal reasoning - not only about geological events, but about events more generally, from many domains and disciplines.

References

1. Foerster, T., Trame, J., Remke, A.: Web-based GEONETCast Data for Geochange Research. In: Hennebohl, K., Vinhas, L., Pebesma, E., Camara, G. (eds.) GIScience for Environmental Change Symposium Proceedings, vol. 40, pp. 1-6. (2010)
2. Devaraju, A., Kauppinen, T.: Geo-Processes and Properties Observed by Sensors: Can We Relate Them? In: GeoChange 2010 – GIScience for Environmental Change. (2010)
3. Lagoze, C., Hunter, J.: The ABC Ontology and Model. Journal of Digital Information (JoDI) 2(2) (2001)
4. Doerr, M., The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. AI Magazine 24(3), 75–92 (2003)
5. Raimond, Y., Abdallah, S., Sandler, M., Giasson, F., Sandler, M.: The Music Ontology. The 8th International Conference on Music Information Retrieval (ISMIR 2007), pp. 417-422, Vienna, Austria (2007)
6. Gangemi, A., Mika, P.: Understanding the Semantic Web through Descriptions and Situations. In: International Conference on Ontologies Databases and Applications of SEMantics ODBASE, pp. 689-706. Springer, (2003)
7. Scherp, A., Franz, T., Saathoff, C., Staab, S.: F—A model of events based on the foundational ontology DOLCE+DnS ultralight. In: The Fifth International Conference on Knowledge Capture (K-CAP 2009), pp. 137-144. (2009)
8. Matuszek, C., Cabral, J., Witbrock, M., Deoliveira, J.: An introduction to the syntax and content of Cyc. Formalizing and Compiling Background Knowledge and Its Applications to Knowledge Representation and Question Answering, pp. 44--49 (2006)
9. Shaw, R., Troncy, R., Hardman, L.: LODÉ: Linking Open Descriptions of Events. In 4th Annual Asian Semantic Web Conference (ASWC'09), vol. LNCS 5926, pp. 153-167, Shanghai, China (2009)
10. Fouse, A., Weibel, N., Hutchins, E., Hollan, J.D.: ChronoViz: A system for supporting navigation of time-coded data. The 2011 annual conference extended abstracts on Human factors in computing systems, pp. 299-304. ACM, Vancouver, BC, Canada (2011)
11. Stab, C., Nazemi, K., Fellner, D.W.: SemaTime - Timeline Visualization of Time-Dependent Relations and Semantics. Lecture Notes in Computer Science, vol. 6455, pp. 514-523, Los Angeles, USA (2010)
12. Open Knowledge Foundation, CKAN - the Open Source Data Hub <http://ckan.net/>
13. Simile Widgets Timeline, <http://www.simile-widgets.org/timeline/>
14. Open Annotation Collaboration (OAC): Beta Data Model Guide, 10 August, 2011 <http://www.openannotation.org/spec/beta/>
15. W3C, Time Ontology in OWL, Eds. J.Hobbs, F.Pan, W3C Working Draft , 27 Sept 2006 <http://www.w3.org/TR/owl-time/>
16. TimeMap Javascript library <http://code.google.com/p/timemap/>