Project funded by the EUROPEAN UNION

SciNet NatHaz Prevention

Black Sea CROSS BORDER COOPERATION

Common Borders. Common Solutions.

A Scientific Network for Earthquake, Landslide & Flood Hazard Prevention

Geodatabase development

Deliverable No.: D.02.02

GA 2. GIS Development, activities 2.3 & 2.5

RESPONSIBLE: TEI of Central Macedonia (LP)

INVOLVED PARTNERS: ALL

Project Details

Programme	**Black Sea JOP**
Priority and Measure	Priority 2 (Sharing resources and competencies for environmental protection and conservation), Measure 2.1. (Strengthening the joint knowledge and information base needed to address common challenges in the environmental protection of river and maritime systems)
Objective	Development of a Scientific Network
Project Title	**A Scientific Network for Earthquake, Landslide and Flood Hazard Prevention**
Project Acronym	**SCInet NatHaz**
Contract No	MIS-ETC 2614
Lead Partner	TEI OF KENTRIKI MAKEDONIA, GREECE
Total Budget	700.000,00 Euro (€)
Time Frame	
Start Date – End Date	01/05/2013 – 30/04/2015

Book Captain: **K. PAPATHEODOROU (TEI KENTRIKI MAKEDONIA)**
Contributing Authors: **D. Kotzinos, K. Papatheodorou, K. Ntouros, K. Evangelidis, K. Konstantinidou (EPPO)**

Document Release Sheet

Book captain:	**K. PAPATHEODOROU (TEI CENTRAL MACEDONIA)**		**11.07.2014**
Approval	**K. PAPATHEODOROU (TEI CENTRAL MACEDONIA)**		**11.07.2014**
Approval	**A.ANSAL (IPA – BU-KOERI)**	M.Atilla Ansal	**11.07.2014**
Approval	**N. KLIMIS (Partner 1 - DUTH)**		**11.07.2014**
Approval	**V. MARGARIS (Partner 2 - EPPO)**	Basil N. Margaris	**11.07.2014**
Approval	**V. NENOV (Partner 3 - Burgas Assen Zlatarov University)**		**11.07.2014**
Approval	**L. TOFAN (Partner 4 - Ovidius University of Constanta)**		**11.07.2014**
Approval	**A.SIDORENKO (Partner 5 – IEEN "D.GHITU" Of Academy Of**		**11.07.2014**
Approval	**K. STEPANOVA (Partner 6 - Ukranian Environmental Academy Of Sciences, Black Sea**		**11.07.2014**
Approved by the Project Coordinator:	**K. PAPATHEODOROU**		**11.07.2014**
Distribution:	**ALL PARTNERS**		

RECORD of REVISIONS

Issue/Rev	Date	Page(s)	Description of Change	Release
	11.07.2014	33	Final version	0.5
	11.07.2014	30	Release of version 0.4 –internal, partners	0.4
	31.05.2014	15	Release of version 0.3 (internal – More Models)	0.3
	31.03.2014	10	Release of version 0.2 (internal – Core Model)	0.2
	31.12.2013	3	Release of version 0.1 (internal - TOC)	0.1

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 BACKGROUND OF THE DOCUMENT

1.1 SUMMARY

This document provides an overview of the data modeling activities in the project. The project adopted and extended substantially an already existing data and metadata model that is used to describe geospatial events and the geoscientific observations and measurements around them.

This is an event driven model, which allows for the description of natural geospatial phenomena as events and associates them with the place and the time span of their occurrence. This allows for an exact and rich modeling of the information. Additionally the model is built in a way that facilitates information integration, which allows the integration of data coming from the different partners, which can follow a different structure and a different format.

Finally this conceptual model is extensible (which was proven since we extended it to cover concepts not existing in the original version) and can be easily updated to cover additional areas around natural hazards.

1.2 SCOPE AND OBJECTIVES

The document aims at describing information integration issues with emphasis on geospatial information and specifically the area of natural hazards' management. This is a critical are since the proper management of such events highly affects the lives of individuals. The management of such information in an integrated way is critical for the successful reaction and management of such phenomena. So the main objective of this document is to describe an integrated conceptual model that allows researchers to integrate different kind of information that might need to manage together, i.e. the management of tsunami events after an earthquake becomes more and more common; for the successful management of such integrated events, integrated models as the proposed one are needed.

The proposed model builds on existing models that have become standards in the corresponding communities. But with the exception of INSPIRE there is no other data and metadata model that handles in an integrated way such diverse and rich information. So this document aims at describing the new concepts introduced by the model around the three main themes (floods, earthquakes, landslides) that are relevant to the project's scope and not to reintroduce all the concepts that have been defined in all other models that are referenced. When needed the reader is referred to these models descriptions to acquire a more in deep understanding of certain concepts.

Final the document's scope is limited to provide a high level description and understanding for the model's user of the different concepts, their hierarchy and the relations among them. The document does not aim at providing in its core a detailed description of each class and property introduced; this is for the document's Annex (Annex I), which is delivered with the document and where all the introduced concepts (both classes and properties) are numbered, explained conceptually and their position in the hierarchy along with their relations with other concepts in the model are explained. The reader is urged to refer to the Annex if some descriptions in the text are deemed not clear. Giving a high level view has been chosen given the number of diverse scientific disciplines covered by the model and the fact that describing everything in the outmost detail would be useless for the usual reader who looks to understand the main concepts of the model and then probably focus on the part that he needs to use most.

So this document is a high level description of the main (core) concepts of the model along with the core concepts of the three main themes described by it.

1.3 RELATED DOCUMENTS

1.3.1 Input

List of former deliverables acting as inputs to this document

Document ID	Descriptor
D.2.01	Data collection and data processing

1.3.2 Output

List of other deliverables for which this document is an input.

Document ID	Descriptor
D.2.03	WebGIS Development

2 INTRODUCTION

2.1 PRINCIPLES OF INFORMATION INTEGRATION

SCInetNatHaz project deals with publishing diverse data that can extend at a large-scale, which means that many different and diverse datasets will be made available through the project's infrastructure. Users of these datasets are usually interested in combining information crossing through the available datasets regardless of origin, purpose or format. This means that we need eventually to deal with data integration issues at various levels and with various degrees of coupling.

Data integration means, among other things, that we need to provide end users with the notion that a single common well-managed data source exists instead of many – distributed or not – individual ones. The database community has introduced the concept of *integrated data-spaces* ([FHM05], [HFM06], [HFM06]) as a response to this problem. Integrated data-spaces should make use of techniques that infer relationships between data sources and that refine them in the light of additional information extracted from related data sources. Although there do exist a wide range of proposals either for specific components (e.g. [SDH08, JFH08]) or for complete data-space management systems (e.g. [DS2006], [MCD+07]), there is a lack of major commonalities since existing proposals stem from quite different application contexts and data paradigms.

Overcoming this diversity is the objective of the recent W3C ***Linked Open Data***[1] (LOD) initiative for a unifying, machine-readable data representation that makes it possible to semantically access and interlink *heterogeneous* but *(semi-)structured* resources at data level – no matter what the structure of the data is, who created it, or where it comes from. Exhibiting a higher degree of interoperability (than documents) and ease of reuse, Linked Open Data has become a prominent choice for sharing (semi-) structured data on the web originating from heterogeneous sources (Enterprises, Scientific communities, Press institutions, but also emerging Social Media applications). Linked Open Data offers great potential for building innovative applications that create new value from the already collected data. Thus Linked Open Data is an effort of integrating semantically described data on RDF and thus, can be used for the creation of *semantically interoperable data-spaces*.

Using Linked Open Data principles and practices allow easy integration of datasets of the project with other datasets and the broader Semantic Web. In order to do that we might need to express our data using tools and languages used by the semantic web. These include but are not limited to, using RDF or OWL as the main languages of communication and using their underlying models for the semantic descriptions and the interoperability among the different datasets and using SPARQL as the main query language for managing the data and the metadata. Current efforts in the area of geospatial information like the INSPIRE Directive, although provide a comprehensive coverage of many and diverse domains of information, do not provide yet a layer of semantic descriptions that can be directly used and exploited on the

1 http://linkeddata.org/

semantic web and for linking with other datasets. Based on that the current document describes an effort to provide an information integration layer over the existing datasets that would allow them to retain their existing formats while allowing for seamless information and query integration for data and metadata models. Similar efforts have been reported in the literature, e.g. [TSS2011] but usually address the issue at a different (syntactic) level or assume that our datasets are already INSIRE compliant, which was not the case in our project.

Finally an important part of publishing data as Linked Open Data is the ability to employ URIs as the unique (and hopefully persistent) identifiers (IDs) for each data object published. URIs (Universal Resource Identifiers) have been proposed early enough in the semantic web initiative [BFM05] as the main mean for resource identification. A constant problem, that the semantic web is facing, is the lack of a global authority that would be responsible to assign such URIs. Since such a task would definitely have a sequence of implications on the operation of the semantic web the chosen path of action in this area has been the assignment of URIs at the highest possible level. A discussion on URI creation issues can be found in the next subsection..

Thus this document covers the necessary issues to allow the project partners to publish their datasets and supports publication in a semantic web and especially Linked Open Data compliant way, which is becoming a de facto requirement in many cases of public data publication. This document is based on and extends Deliverable 2.2 of the InGeoClouds Project (www.ingeoclouds.eu), which firstly introduced the described model for their needs. This model was evaluated and adopted by the SCInet NatHaz projects as it covers the project's needs in terms of modeling data regarding earthquakes and landslides and has been extended to also cover data regarding floods [ING12]. The project believes that extending work produced by other EU supported projects and reusing existing conceptual data models contributed to the uniformity of the data production in the scientific areas covered by the project and provides a more extended validation and visibility of the work. The extensions provided by SCInet NatHaz project are clearly identified in the text.

2.2 URI CONSTRUCTION

The need for unique identification of data objects has a long tradition stemming from the relational databases tradition of unique keys. Producing unique ways to identify data allows for:

- Matching information for the data at hand with information declared elsewhere for the same data
- Differentiate a specific data point from others that might have similar characteristics.

In the Semantic Web and Linked Open Data tradition the unique keys have been replaced by Universal Resource Identifiers (URIs). As also described in [IGC12], a Universal Resource Identifier (URI) [BFM05, DS06] serves to uniquely identify information objects or real world entities, be they local or published on the Internet ("Semantic Web"). URLs are a special case of URIs, with a particular syntax. They normally identify the location of a document on a server visible on the Internet. Therefore, it is a good practice to identify the document itself by a location independent URI (more good practices on URI construction can be found at: http://www.w3.org/2011/gld/wiki/223_Best_Practices_URI_Construction and other generic suggestions from EU bodies like EU commission regulation No 1089/2010 and EU commission regulation No 1205/2008 do exist[2]).

Other relevant forms of URIs are:

- Universal Resource Names (URN). Organizations can request a domain name at IANA[3] to be used following the "urn:" prefix, and define their own syntax and resolution rules.
- Digital Object Identifiers (DOI), managed by http://www.doi.org, a central authority hierarchically delegating subdomains to institutions, but only for digital objects.
- Universally Unique IDentifier (UUID): With the help of a random algorithm, a UUID generator can generate a most likely new identifier without a central authority.

The worst case for information integration is, when everybody declares a new identifier for something, without others knowing. From a use point of view, we have to consider three basic questions: (a) how will I know that an object has gotten an identifier at all; (b) who will be able to confirm that the object is indeed identified by this identifier, and what that means; (c) how will I be able to find the object if you give me only an identifier. After this, the question that might appear is what to do if an object has more than one identifier, or if an identifier is used wrongly in a description.

For the needs of this project, we recommend a mixed approach:

- For all items only locally known to a particular dataset, a UUID is the best solution. It is generated locally, and anybody to use it in the future will learn about the item only in connection with this identifier, and hence together with the knowledge of the item, will also have the knowledge of the identifier.
- For all items publicly known to belong to a certain institution, the institution can maintain LOD identifiers, such as for boreholes, measurement equipment, laboratories, because people will search there for "official identifiers" of those items. Institutions can apply their own policy (according to LOD guidelines) or follow a policy specified within their domain or country of origin.

[2] In the COMMISSION REGULATION (EU) No 1089/2010 of 23 November 2010 about "Interoperability of spatial data sets and services", the definition is given for a 'identifier' [of an object]: means a linguistically independent sequence of characters capable of uniquely and permanently identifying that with which it is associated, in accordance with EN ISO 19135". The annex I of the directive defines a more precise definition of the "External unique object identifier published by the responsible body, which may be used by external applications " and describes the attributes of each identifier:

- localID : identifier in the namespace
- namespace : Namespace uniquely identifying the data source of the spatial objet
- versionID : The identifier of the particular version of the spatial object with a maximum length of 25 character

[3] www.iana.org

- For industrial products of companies providing scientific equipment, we recommend the creation of "pseudo LOD identifiers" by a suitable combination of the company's web domain, the company's product identifier (and serial number), as long as they do not offer LOD identifiers themselves already.
- For place names we recommend to use the identifiers of a suitable gazetteer service (e.g.GeoNames[4]). If the gazetteer service does not provide URIs, a "pseudo LOD identifier" can be created from the domain name of the service and the respective place name identifier.

It is to be expected that organizations introducing identifiers at a later time, will follow a method, which will allow for automatically matching our "pseudo LOD identifiers" with the future correct LOD identifiers. It should also be noted that the generation of unique URIs is not a requirement when the data are kept at a local level and are not directly shared on the web or not plan to be extensively mixed with a lot of external datasets.

3 THE (GEO) SCIENTIFIC OBSERVATION MODEL

3.1 PURPOSE

This model is built (as it is also stated in [ING12]) as a common specification or an ontology model that supports data sources integration from different data providers (geospatial databases) in order to share data. The data providers are all public sector or research bodies and thus the project is supporting Community policy in the Re-use of Public Sector Information Directive. In addition to this it will enable a more complete and coherent implementation of Groundwater and Soils protection Directives. It will directly support GMES and GEOSS.

The Global Earth Observation System of Systems (GEOSS) is being built by the Group on Earth Observations (GEO). GEOSS seeks to connect the producers of environmental data and decision-support tools with the end users of these products, with the aim of enhancing the relevance of Earth observations to global issues. The end result is to be a global public infrastructure that generates comprehensive, near-real-time environmental data, information and analyses for a wide range of users. The earth observation systems consist of instruments and models designed to measure, monitor and predict the physical, chemical and biological aspects of the Earth system. Our conceptual model has been created or extended based on standards in order to answer the needs of this project. Regardless of the fact whether the project plans or not to publish on the web directly the available datasets we decided that it will beneficial for all to provide a more state of the art description of the data models and thus we adopted a paradigm based on the semantic web principles.

[4] www.geonames.org

3.1.1 Enabling of information interoperability and integration

Mediation tools and Semantic Web activities require an integrated, shared ontology for the information accumulated by different scientific domains. In addition, such interrelationships should be either integrated in common information storage, or at least virtually integrated through mediation devices that allow a query to be simultaneously launched on distinct information depositories, which in all cases requires common semantic tools.

The model needs a standard way of publishing the data as RDF and especially, ways of providing unique URIs are required for the data used within the project. The model is designed as a generic, extendable and interoperable model, which relies not only to geographic data but also for the documentation of scientific observation and measurement in general. It integrates different datasets created from sources w.r.t. their semantics. It is an event-centric model that captures complex historical contexts where time, place and participants are associated; so, it is capable to describe complex relationships and to integrate or connect rich information.

3.1.2 An extension of the CIDOC CRM

CIDOC CRM is explicitly compatible in formalism with the World Wide Web Consortium's Resource Description Framework (RDF), which can only be beneficial for our model. This is also an opportunity to extend the scope of the CIDOC CRM to geo-scientific information. Consequently, it opens the way for a model to benefit from further extensions of the scope of CIDOC CRM, such as the scientific heritage of observations and experiments.

3.1.2.1 Sources

CIDOC CRM

The CIDOC CRM5 model was being developed from 1996 under the auspices of the ICOM-CIDOC (International Council for Museums – International Committee on Documentation) Documentation Standards Working Group. The definition of the CIDOC CRM model has now become ISO standard 21127 [CID11].

INSPIRE

INSPIRE Directive is based on the infrastructures for spatial information established and operated by the 27 Member States of the European Union. The Directive addresses 34 spatial data themes needed for environmental applications, with key components specified through technical implementing rules. One of the main tasks of the INSPIRE program is to enable the interoperability and, where practicable, harmonization of spatial data sets and services within Europe.

[5] "CIDOC CRM" stands for "Comité international de documentation [= International Committee on Documentation] Conceptual Reference Model," which, when isolated from any context, is not particularly meaningful (CIDOC is affiliated to ICOM, the International Council of Museums).

The text of the INSPIRE Directive is available from the INSPIRE web site (http://inspire.jrc.ec.europa.eu/) and (http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32007L0002:EN:NOT)

DUBLIN CORE

The fifteen-element "Dublin Core" achieved wide dissemination as part of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and has been ratified as IETF RFC 5013, ANSI/NISO Standard Z39.85-2007, and ISO Standard 15836:2009.. The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description. The name "Dublin" is due to its origin at a 1995 invitational workshop in Dublin, Ohio; "core" because its elements are broad and generic, usable for describing a wide range of resources.

The fifteen-element "Dublin Core" described in this standard is part of a larger set of metadata vocabularies and technical specifications maintained by the Dublin Core Metadata Initiative (DCMI). The full set of vocabularies, DCMI Metadata Terms [DCMI-TERMS], also includes sets of resource classes (including the DCMI Type Vocabulary [DCMI-TYPE]), vocabulary encoding schemes, and syntax encoding schemes. Available online at http://dublincore.org/metadata-basics/

3.2 CORE MODEL DESCRIPTION

The core model is extensively described in [ING12], here we include for completeness of the text the main elements of its description and connect them with the extensions made; the reader is referred to this document if further reading is deemed necessary. As we can see in Figure 1, the core view of our universe of discourse is presented as consisting of event-centric processes, rather than of static analysis and classification of products out of the human and spatiotemporal context. Thus the core model describes fundamental notions of the core model that answer basic questions of type "who did what when, where with what", and expressed, for example, in form of queries, such as "give me all the observations/ measurements made by actor X at time Y, give me all the tools/software used for the outcome of activity X, give me all the documents or maps that refer to this event, etc.

We distinguish classes and properties in the sense of the knowledge representation languages proposed for the Semantic Web such as RDFS. Any particular individual item in the scope of the model is regarded instance of a class Entity. Any instance of Entity may be further classified by suitable terms - instances of Type - via the property "has type". Additionally any instance of entity can be identified by an appellation or an identifier and can be described by a short note-description or comment.

For clarity and readability we have followed the following naming scheme throughout the definition and description of the geoscientific observation model:

- Classes that come from already existing standard models are named as *Exx* and the class name

- Classes that are declared specifically by the geoscientific observation model are named as *Sxxx* and the class name
- Properties that come from already existing standard models are named as *Pxxx* and the property name
- Properties that are declared specifically by the geoscientific observation model are named as *Oxxx* and the property name

More detailed descriptions of the class and properties of the whole model can be found at the Annex accompanying this deliverable.

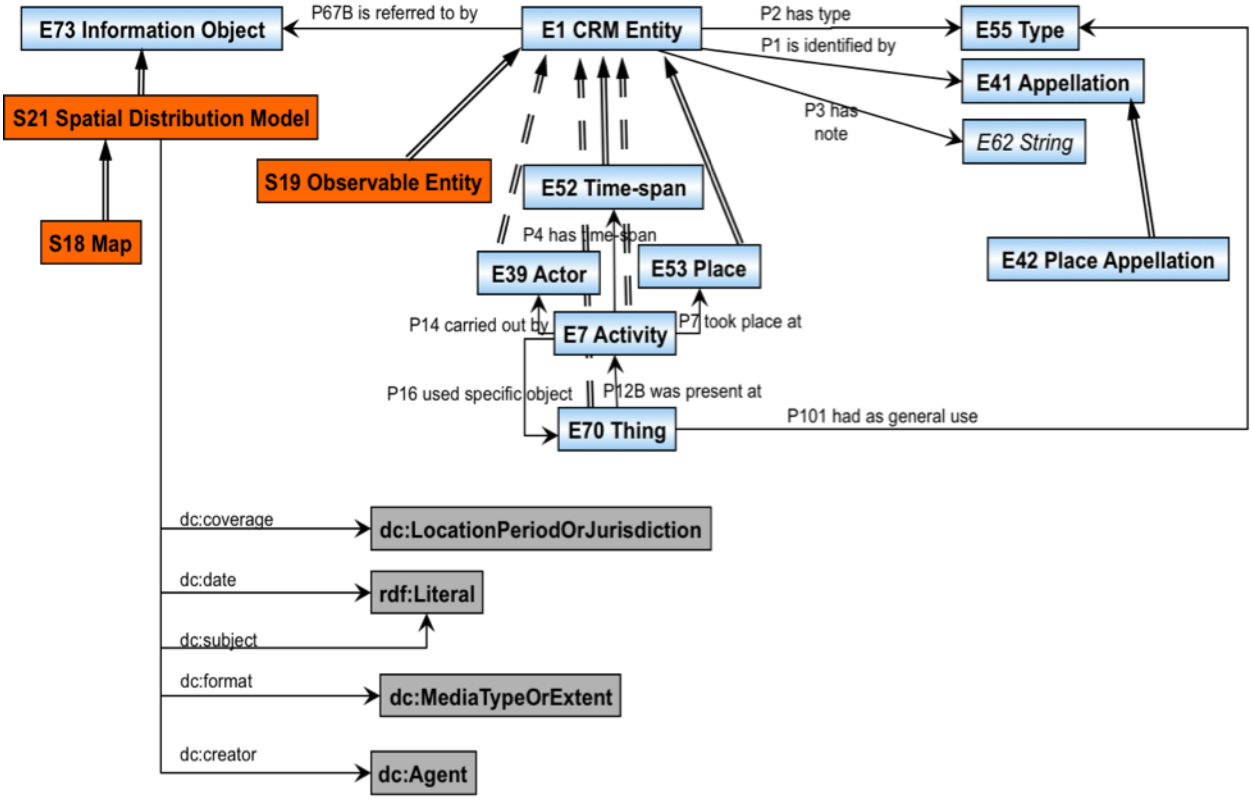

Figure 1: Basic Concepts [ING12]

The next level can be described as a simple structure for the description of activities of people and things taking place in time and space.

More formally, we distinguish three entities:

- *Actors*, i.e. people with the capacity to actively or passively *participate in* activities. Machines are not Actors.
- *Activities* are real processes carried out by people of any scale. Activities lay in the *immediate* or *remote past*.
- *Things*. Things may be material or immaterial. The project focuses on **material things and features**, material substantials with a relative stability of any form or structure, or containing a fixed amount of matter, solid and fluid bodies. So

geological features, water constellations, acquifer concepts are all classified under S10 Material Substantial (material things, Figure 2).

All these things can be present at Activities, at their creation, during use, as derivatives or results, etc. Things also may have types of uses, which imply the methods and techniques of use to be documented.

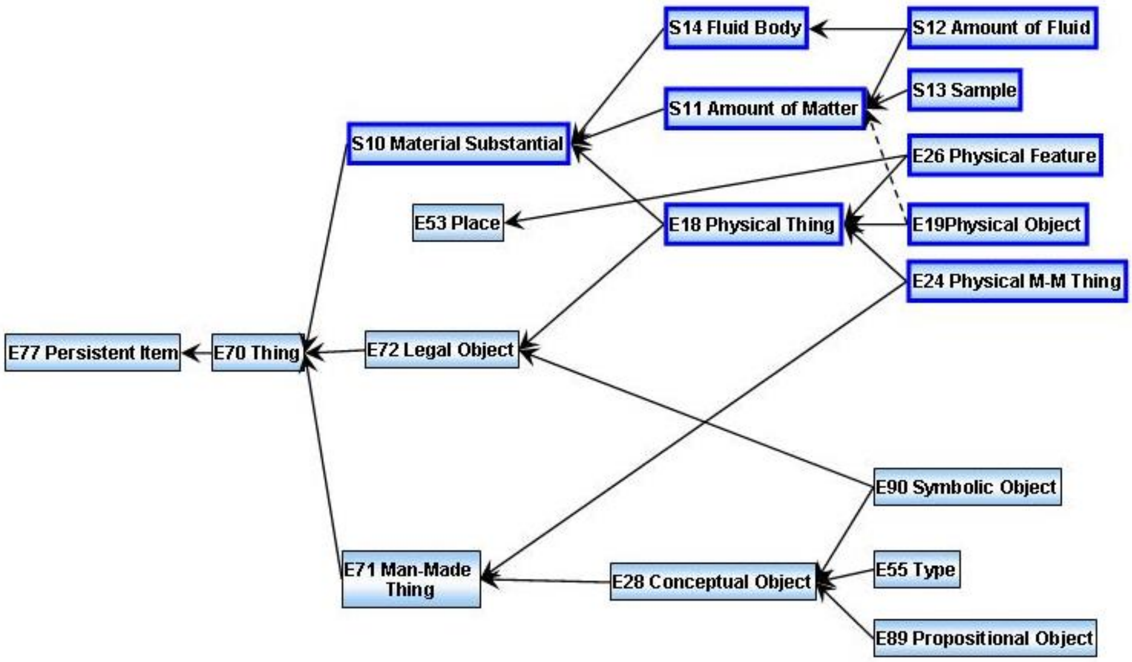

Figure 2: Material and immaterial things [ING12]

Visual representations and references of things, places, activities and other phenomena in forms of spatial distribution models and maps, can be documented by connecting these to a Spatial Distribution Model, represented by class S21 Spatial Distribution Model, which is a general concept that has many expressions such as geological models and paper maps. This class comprises propositions about the spatial distribution of phenomena typical on, above or below the surface of the earth.

Our project deals with datasets, which can be grouped into three main thematic categories.

- Floods, their monitoring and their recording (which is an extension to the original model since such data category did not exist)
- Earthquakes (seismic data), their monitoring and their recording (which is adopted from the original model)
- Landslides, their monitoring and their recording (which is adopted and extended from the original model)

3.3 THEME 1: FLOODS CORE MODEL

Describing floods is a rather complex effort and includes observations by different specialists and measurements by different sensing devices (sensors) being either freely carried around or situated in static observation stations. Compared to earthquakes and landslides, floods are a more local phenomenon since they happen around areas where some water sources already exist (e.g. rivers, lakes). So the core observation model is used here more to capture the different observation types as well as the different actors that might be influenced by such an event.

Thus as we can see in Figure 3 we define the concepts of different stations (e.g. S102 Meteorological Station, S30 Geophysical Station that carry different unique identifiers (so we can distinguish between measurement from different stations) and can record measurements (P100 Record Measurement). For that purpose we define different type of measurements (e.g. S103 Hydrological Measurement and S104 Meteorologic Measurement), all of which are derived from the E16 Measurement class, which is part of the core model. Each measurement carries properties that allow us to define its place (E54 Dimension), its time span (E52 Time Span) and the actual measurement in terms of values (through the P20F has value data property) and units (E56 Measurement Unit).

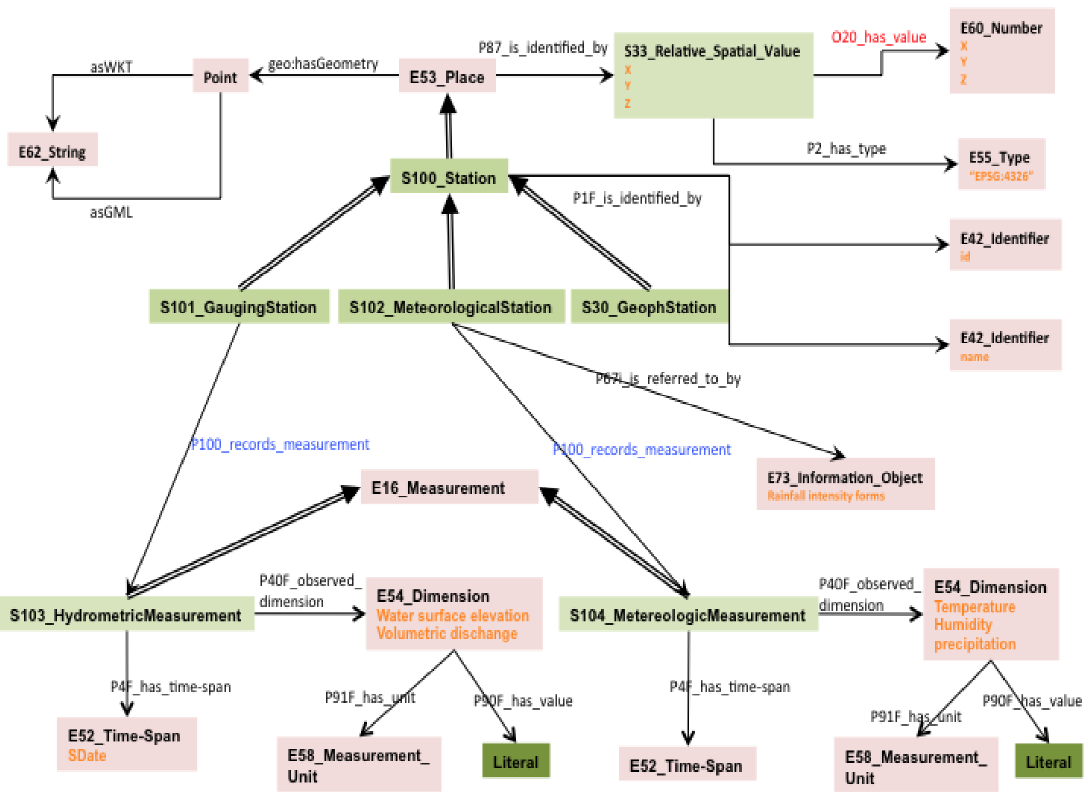

Figure 3: Flood measurements and related measurement stations

In Figure 4 the reader can observe the different hydraulic structures (S105 Hydraulic Structure class and its subclasses) that can be involved while describing a flood event. Being a subclass of the E26 Physical Feature class all the hydraulic structures carry the properties that are already described for the physical feature. These properties are detailed for completeness in Figure 5, there we can see that we can register the place where the structure is located (E54 Dimension and S33 Relative Spatial Value), what is its type (E55 Type), its plan or inlet/outlet design (E29 Design or Procedure) and any measurements (P90F has value property) or other information associated with it (E73 Information Object).

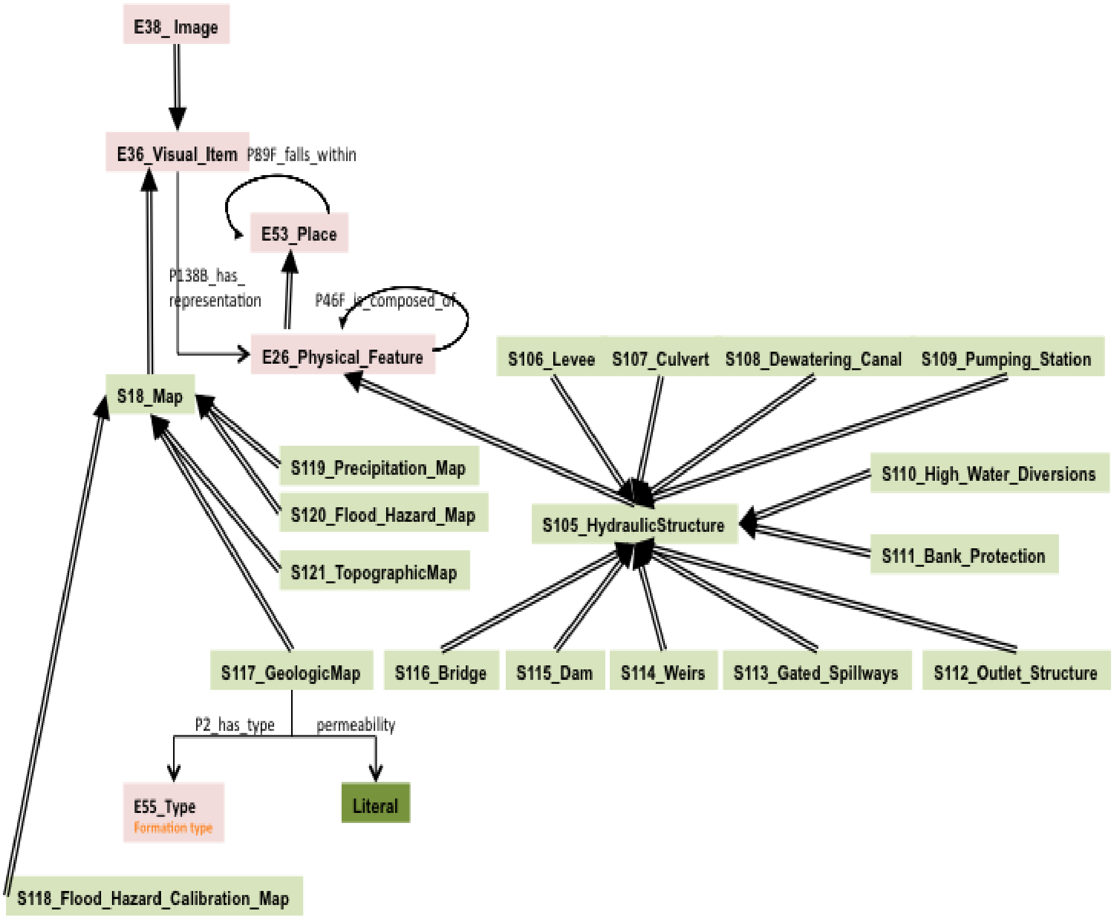

Figure 4: Hydraulic structures and corresponding maps

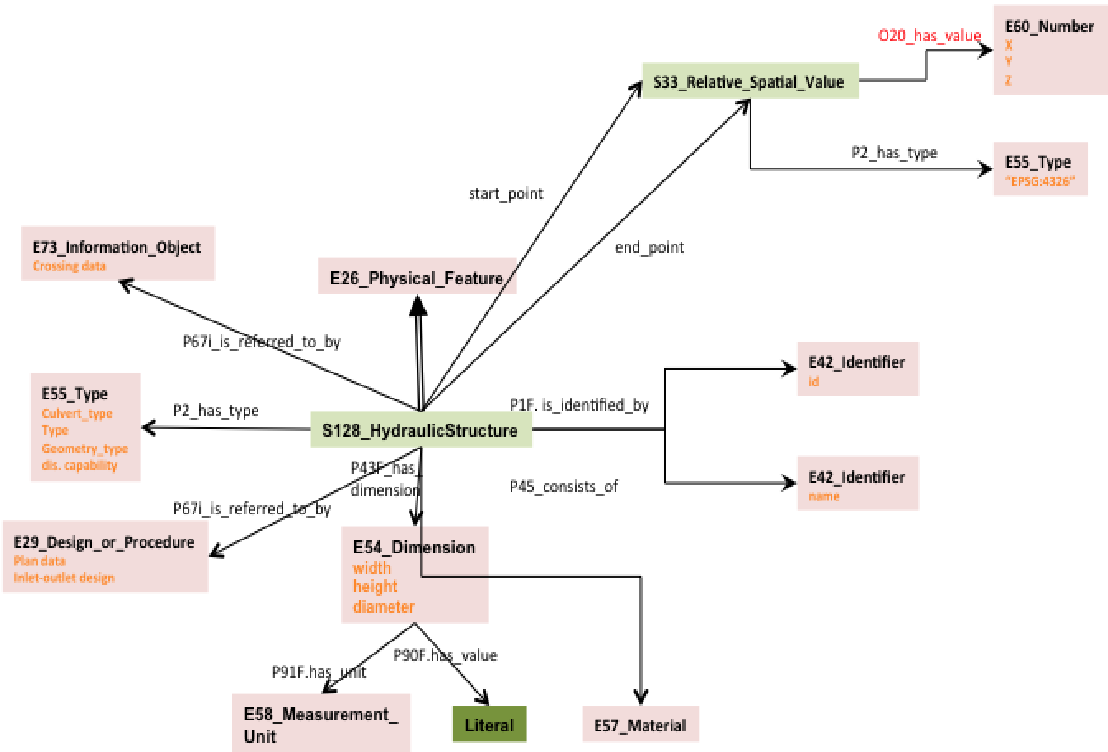

Figure 5: Hydraulic structure's properties

Usually we also use a series of maps (class S16 Map) to picture information related to floods (Figure 4). Here we additionally introduce specific types of maps like the S119 Precipitation Map that shows the rain that fell in the area, the S120 Flood Hazard Map that depicts the affected areas by the flood, the S117 Geologic Map, the S128 Flood Hazard Calibration Map and the S121 Topographic Map that are usually used by the flood prediction models and as background information for the other corresponding maps. We record mainly metadata (since they are of type (E36 Visual Item)) about these maps like information about the map creation, information carried, etc. and the map itself is stored in one of the available formats not necessarily (but it could be the case nevertheless) in the knowledge base.

As we can also see in Figure 6 we also described as part of our data and metadata model a series of concepts related to the topographic features that need to be captured and associated with floods when we want to use them in flood modeling activities. Thus we define affected areas (like S130 Urban Area or S132 Airport) and affected networks like the S136 Road Network and S137 Railway Network. This allows the researcher to capture information not only about the flood itself but also about the entities affected by the flood and thus record both observations but also results of the modeling and observation procedures as well with potential issues that affect thode neighborhood entities.

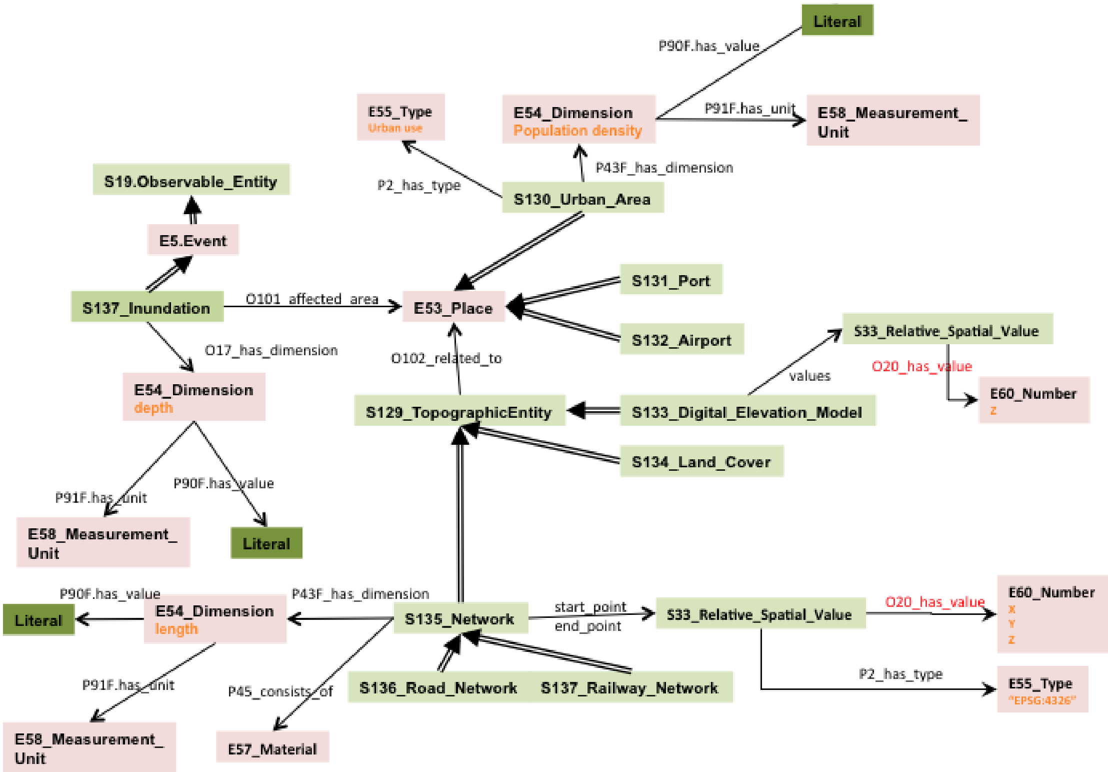

Figure 6: Entities and Networks involved in floods management

Finally the model also defines additional elements of floods management, which refers to hydraulic networks, since floods have to do with the ad hoc creation of those due to external extenuation circumstances. These networks as a substance deal with fluid matter and this is described in Figure 7. There we define notions at the S123 Water Body and the S124 Stream. Additionally we define the important concept of the S127 Catchment Area, which defines the place where the water will be gathered while a flood will be occurring.

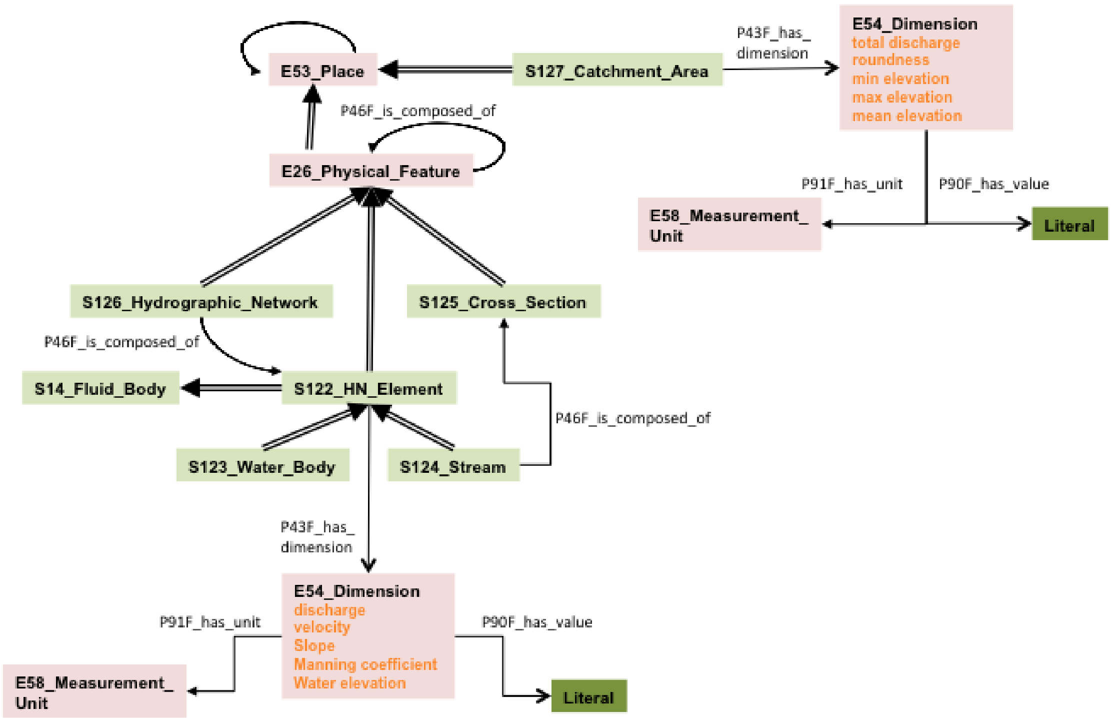

Figure 7: Definition of floods' elements

Given all the above-introduced notions we cover the data captured and are available in the project by the project partners. This core flood model is an extensible model that allows anyone interested or needing additional concepts to be introduced them, to propose new concepts as long as they are in line with the existing ones and do not break the validity and the conceptual clarity of the model. The definition of all the concepts around flood management has been provided entirely by the SciNet NatHaz project and will be contributed to the community using the model.

3.4 THEME 2: SEISMIC CORE MODEL

The general core observation model can also be used for describing fuzzy large scale power natural phenomena and destructive events such as seismic phenomena and earthquakes that show features of observable seismic action and energy-related behavior (Figure 8). We use two types of scientific assignments–approaches on these phenomena: a measurement event (E16 Measurement) which is a kind of observation, in order to measure the magnitude of an impact, and an S6 Data Evaluation event (a kind of Inference Making) for evaluations and quantitative – qualitative observations on ground displacements and calculations used on evaluating the measured data. Simulations - Predictions and Hypothesis building are other types of making propositions about behavior and properties of real phenomena, which are both based on theories and inference rules (but having different usage, e.g. for predictions).
Geodetic observation to tectonic motion, evaluation of dynamic properties of rock and seismic hazard analysis involves repeated measurements.

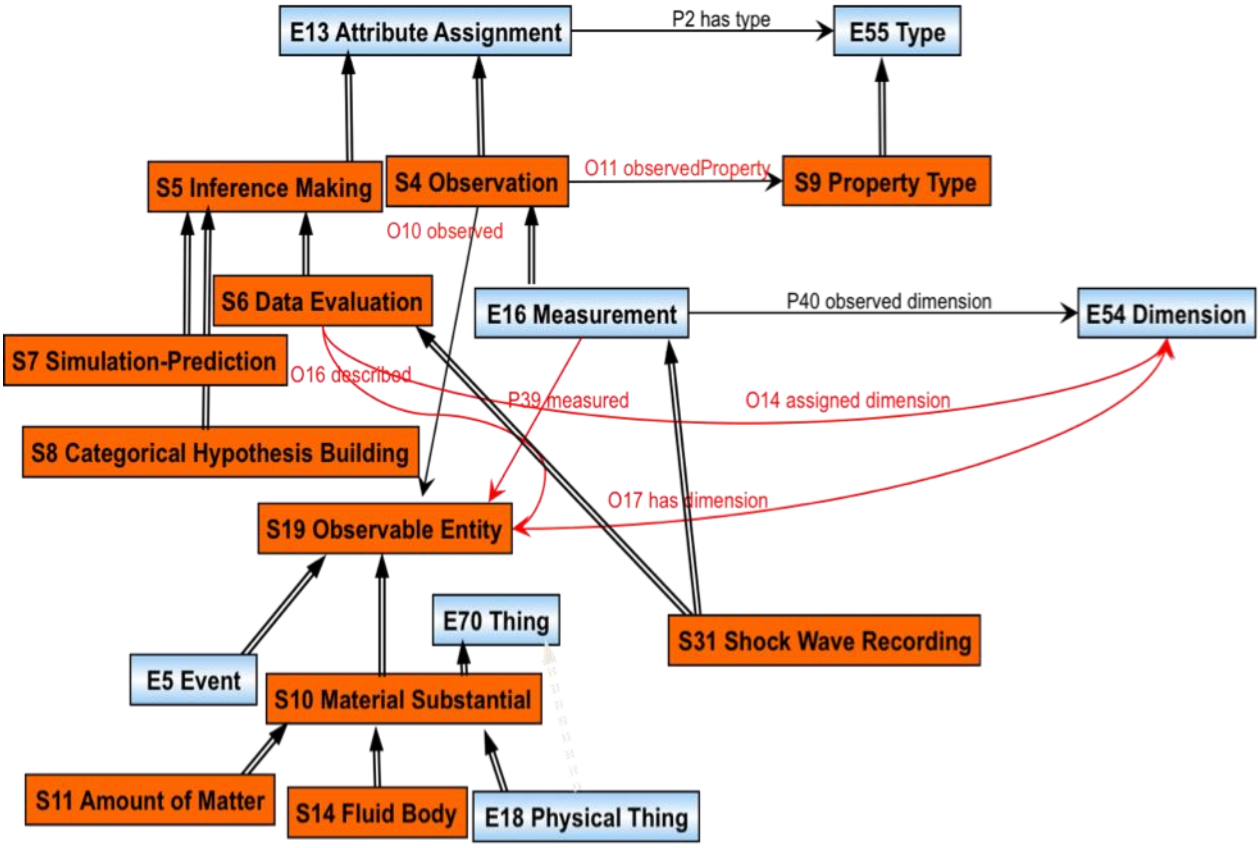

Figure 8: Observation on earthquakes [ING12]

In our model an (S23 Earthquake) is a special case of a seismic phenomenon (S22 Seismic Phenomenon) which is an event. A seismic phenomenon is a more general notion for generating seismic waves resulting as a consequence of different other phenomena, such as a volcano eruption, etc. In that sense, we have a classification of earthquakes (see Figure 9), such as Cluster, Shock, Mainshock or Shockwave, based on their form and identity, if the description implies one or two earthquakes, if it is specified as a group or a sequence, a main shock or a pro-shock, and other conditions that are used to distinguish an earthquake from another. This classification is under discussion and it is based on specific criteria. These types are also related by a part-whole relation between them (so there is a part-whole relationship implied in their definition, too).

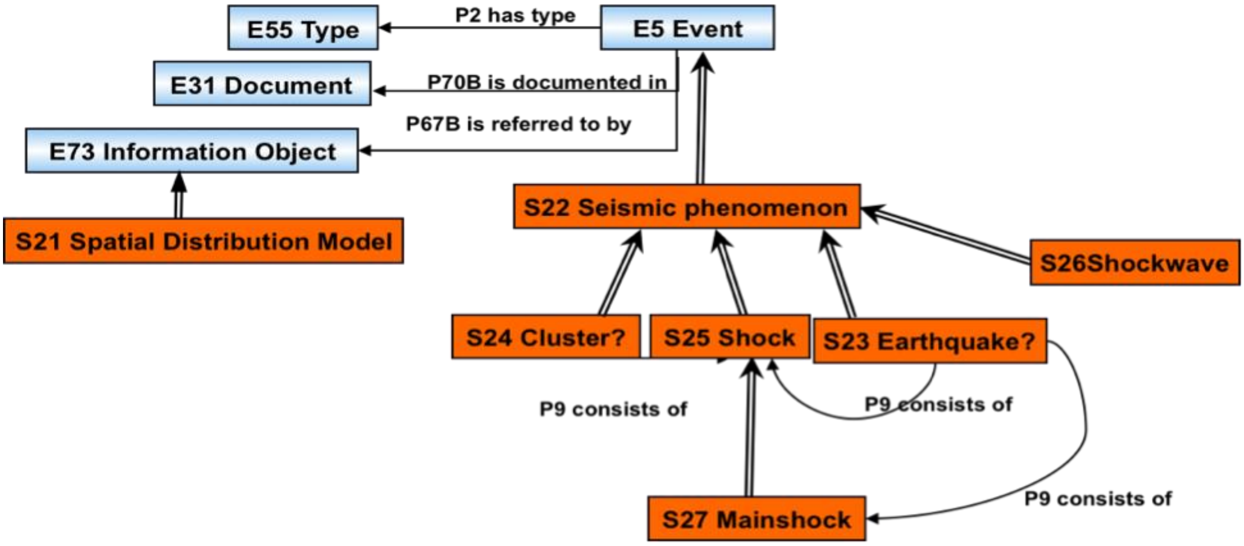

Figure 9: Types of seismic phenomena [ING12]

For every earthquake event we keep information about its spatial and temporal distribution (see Figure 10) by the size of the phenomenon (the measure type of E54 Dimension of an earthquake is magnitude). So we register the time-span of the earthquake (inner and outer temporal boundaries of the occurrence of the phenomenon) and the location in which the earthquake took place at (E48 Place Name that identifies it, coordinates representation (E47) and a short description of that location relative to landmarks). The documentation of the description and the classification of an earthquake is provided by an authority document (E31 Document) with metadata information on the source of the document and the date it was written (for preserving, integrating and updating provenance information).

We also model the composition of a network of instruments required for the earthquake recordings. This network for the data transmission (S28 GeophSensorNetwork) consists of accelerometers and other sensors (components – types of accelerometers), which are located in stations. A Station (S30 GeophStation) is defined as the place of the recording equipment, and in that sense it is represented by spatial coordinates, place name and by an identifier. A station also has physical parameters (geological, geophysical, geotechnical) related to the specific location such as velocity dimension (see Figure 11).

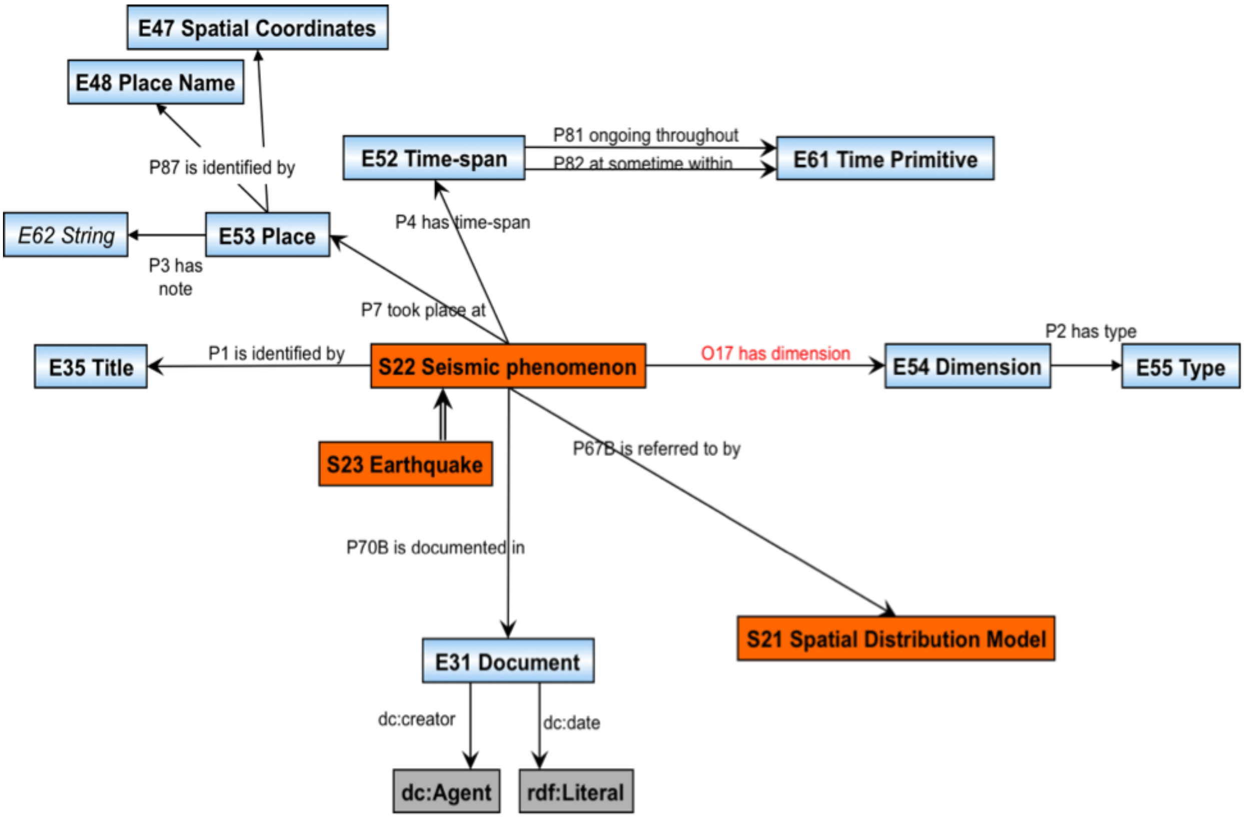

Figure 10: Distribution of an earthquake [ING12]

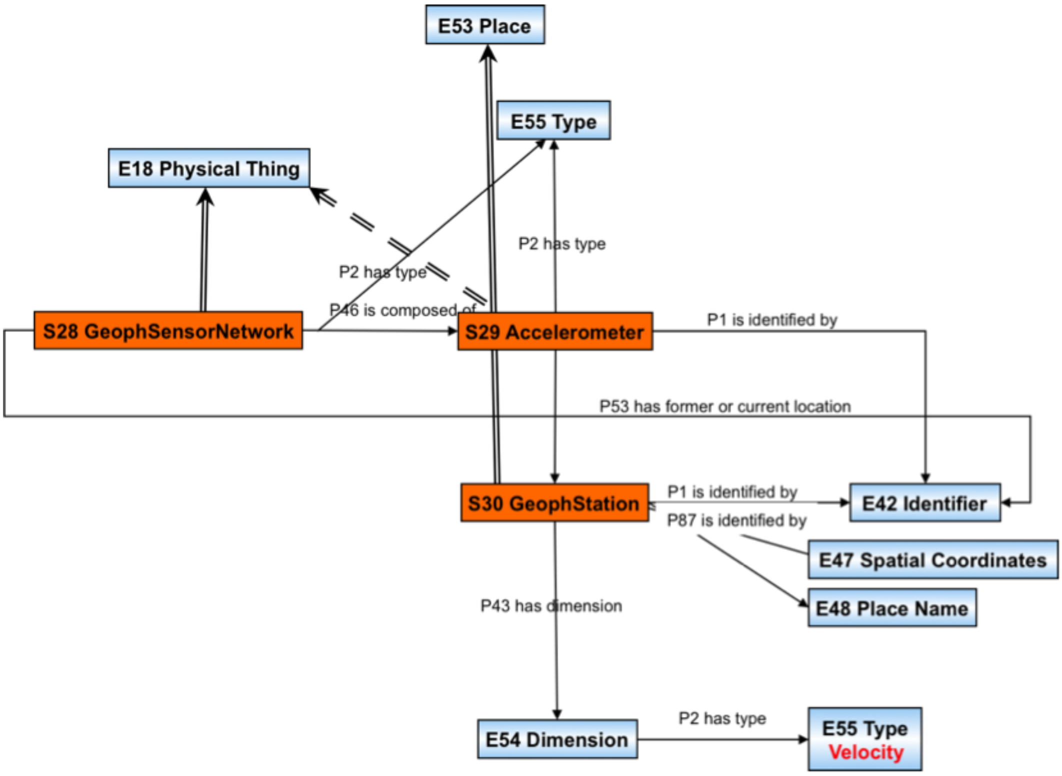

Figure 11: Stations and accelerometers [ING12]

Accelerometers are the special equipment used for the recording of a shock wave (S31 Shock Wave Recording). This class is multi-classified; as a measurement of e.g. an intensity dimension of the earthquake, and at the same time as data evaluation event, a process filtering the raw data of measurements, using a specific software, which produces derivatives. Seismic waves produced by an earthquake are also part of an earthquake. A Shockwave Recording starts with the phenomenon of a shockwave; the recording is coherent with the part of the shockwave reaching in the instrument (see Figure 12).

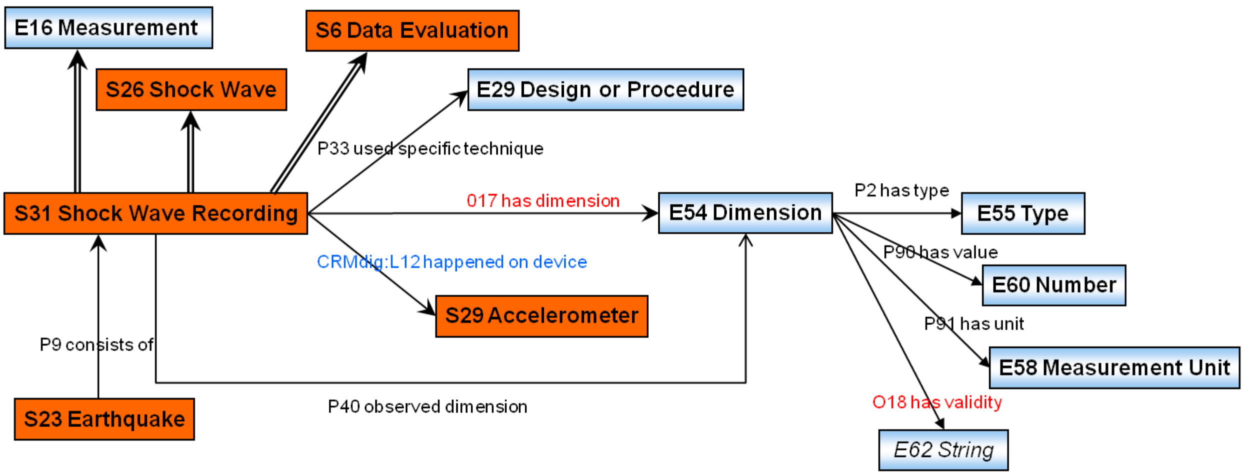

Figure 12: Shock wave recording [ING12]

3.5 THEME 3: LANDSLIDES CORE MODEL

The general core observation model can also be used for describing events that cover a large area and also might have specific repercussions to the neighborhood areas. The model allows for capturing the diverse nature of landslides by specifying their generation, by declaring the S32_Landslide class as a subtype of the S35_FeatureGenesis class. This means that from the beginning the landslide is defined as an event (subclass of the E5_Event) class and thus carries all the corresponding properties of this class, i.e. it can consist of other (sub)events, it can carry the time and the place it happened and of course we can observe various properties associated with this event (subclass of the S19_ObservableEntity class), etc.

The landslides core model specifies also (mainly physical) features[6] that can be associated with a landslide which include the matter involved in the event (e.g. the amount of earth moved during the landslide) and of course registers the spatial coordinates related to any of the features of the landslide event. All these are depicted in Figure 13.

[6] The notion of a feature here is used in accordance with its definition in various existing standard models like the ones produced by OGC (www.opengeospatial.org), INSPIRE (http://inspire.ec.europa.eu/) and the series of ISO standards for specifying spatial metadata and data models (e.g. ISO19115, ISO19119 for more look at www.iso.org). The reader is directed to the specific web sites for a more detailed description of the feature concept since describing it in more detail is beyond the scope of this document.

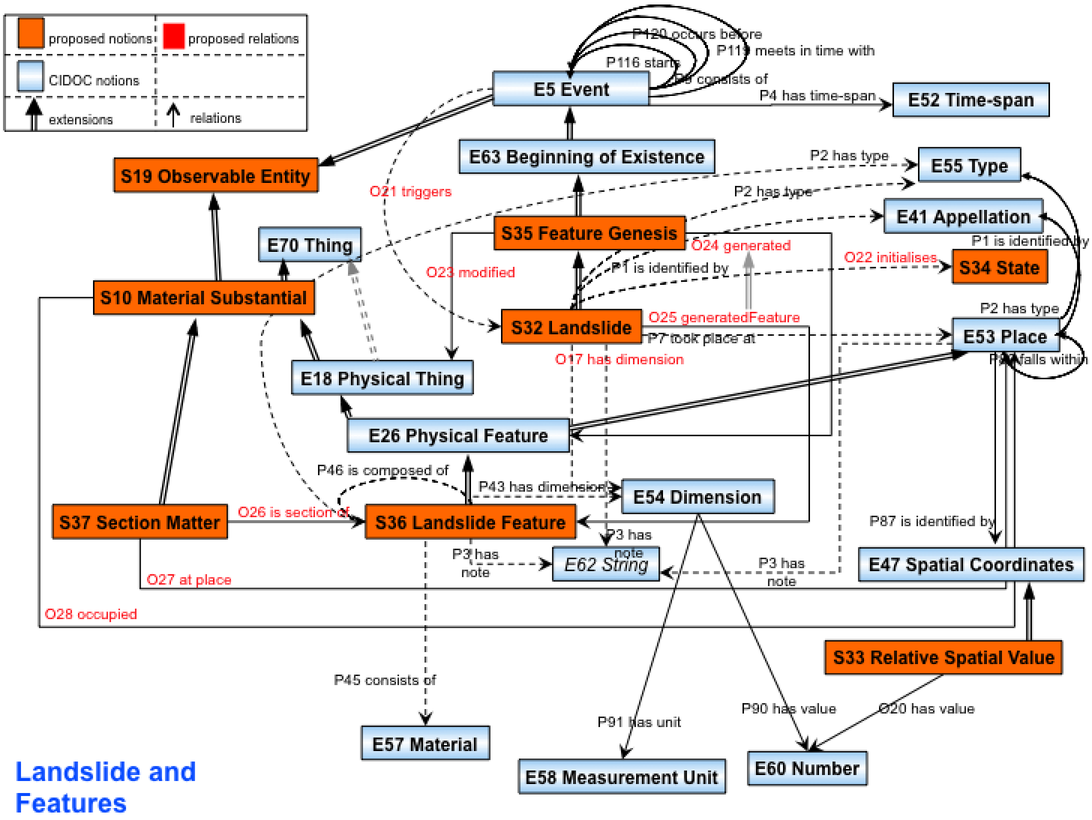

Figure 13: Landslides core model: features

Each landslide can be observed by any number of human actors (either through the use of corresponding sensing devices or visually with local inspection) and these observations can be recorded using the model. Any number of observations can be associated with a specific landslide and information can be recorded about the type of the observation, the time and the place it was made and the actor who made it. Additionally with this observation we can register the physical features involved like the materials changed (ground, waters, etc.) and we also have the ability to record the evaluation of the data (using the S6_DataEvaluation class) that can be done in post processing mode in the laboratory after the observation has been recorded in the field. This has been deemed necessary since on the one hand observations could contain errors and on the other hand we could have different evaluations for the data by different actors in the landslide evaluation process, i.e. data evaluation could be done by a geologist but also an environmentalist. As shown in Figure 14 each landslide observation can be associated by any number of data evaluations, which in turn can carry a series of physical features associated with them.

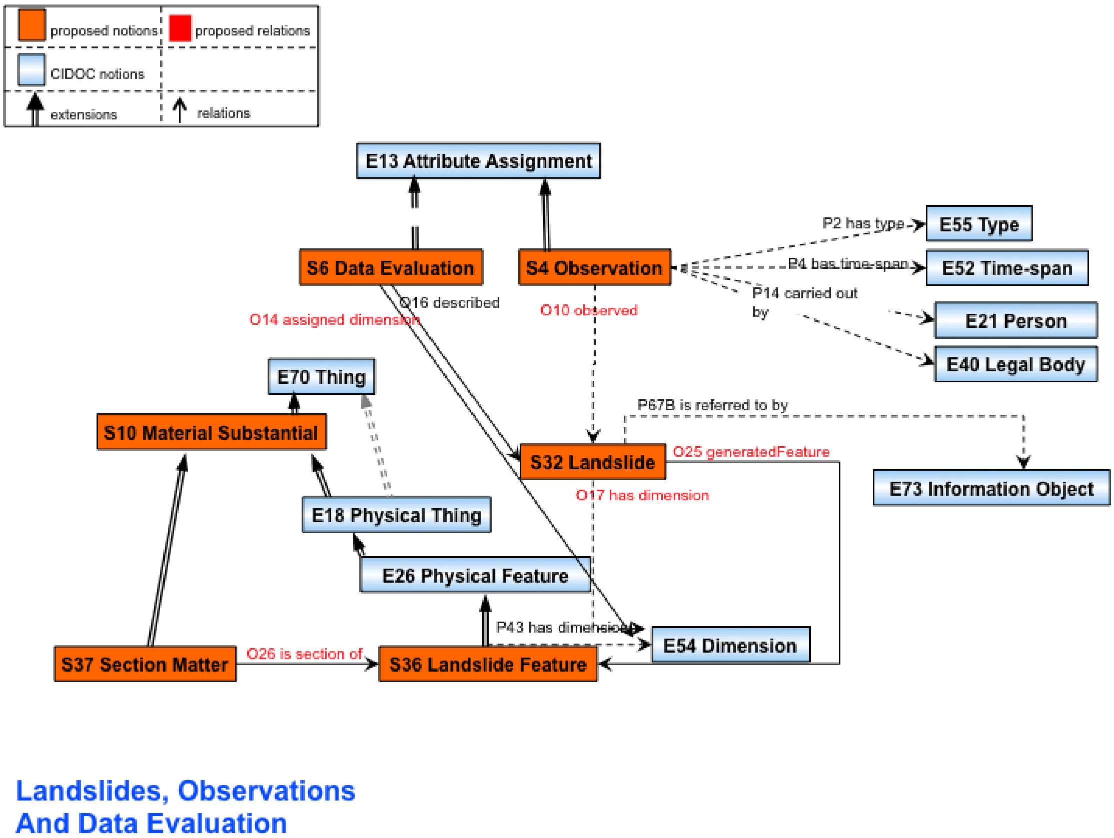

Figure 14: Landslides core model: Observations and Data Evaluation

The Landslides core model was briefly introduced in the first version of the GSOM (GeoScientific Observation Model) and was extended in the second version; here it was further extended and some notions were re-described to better fit the concepts as defined by the SciNet NatHaz project partners.

4 EXAMPLE: DETAILED ANALYSIS OF THE SEISMIC DATA THEME

In this section we distinguish the different thematic categories, which refer on earthquake events and measurements over them as described by the project's partners. This is given as an example to the reader to show that further analysis of the concepts and their relationships can be done. This is provided as an example to facilitate the reader's comprehension of the model and the interworking of the different concepts (either classes or properties). Doing this kind of analysis for the whole model and each and every class, which is based on their detailed descriptions in the Annex, is straightforward and would only add to the size of this deliverable without adding anything to the reader's comprehension. For further details about the model the reader is referred to the Annex.

Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε. shows the different themes we captured from the datasets and next, we will show how these themes correspond to specific usage examples of our General Core Model. In the following pictures, we mapped the notions, which are used in the datasets on the corresponding notions of the core model and created the sub - models for each category.

Table 1. Categories of the different concepts in the Seismic Model

Categories
Stations and Sensors Network
Earthquake Event
Seismic Recording

4.1 STATIONS AND SENSORS NETWORK

Seismic events are recorded by geophysical stations. We used the entity *S30.GeophStation* to describe the stations and for each station we keep its location (Latitude, Longitude), its name (notion Site) and we measure the shear-wave velocity between 0 and 30 meters depth (notion VS30) (see Figure 15). Each station consists of sensors, which are the recording instruments for earthquakes. Sensors are described using entity *S29.Accelerometer* and we used the relation *P53F.has_former_or_current_location* to show that a sensor belongs in a station. Each sensor has information about its communication type (analog or digital) using notion commtype and a description of the instrument type (notion insttype). The available sensors from a specific station constitute a network, which is described using the notion *S28.GeophSensorNetwork*. Every network is identified by a network identifier string (notion netid) and the agency that maintains each station (notion source).

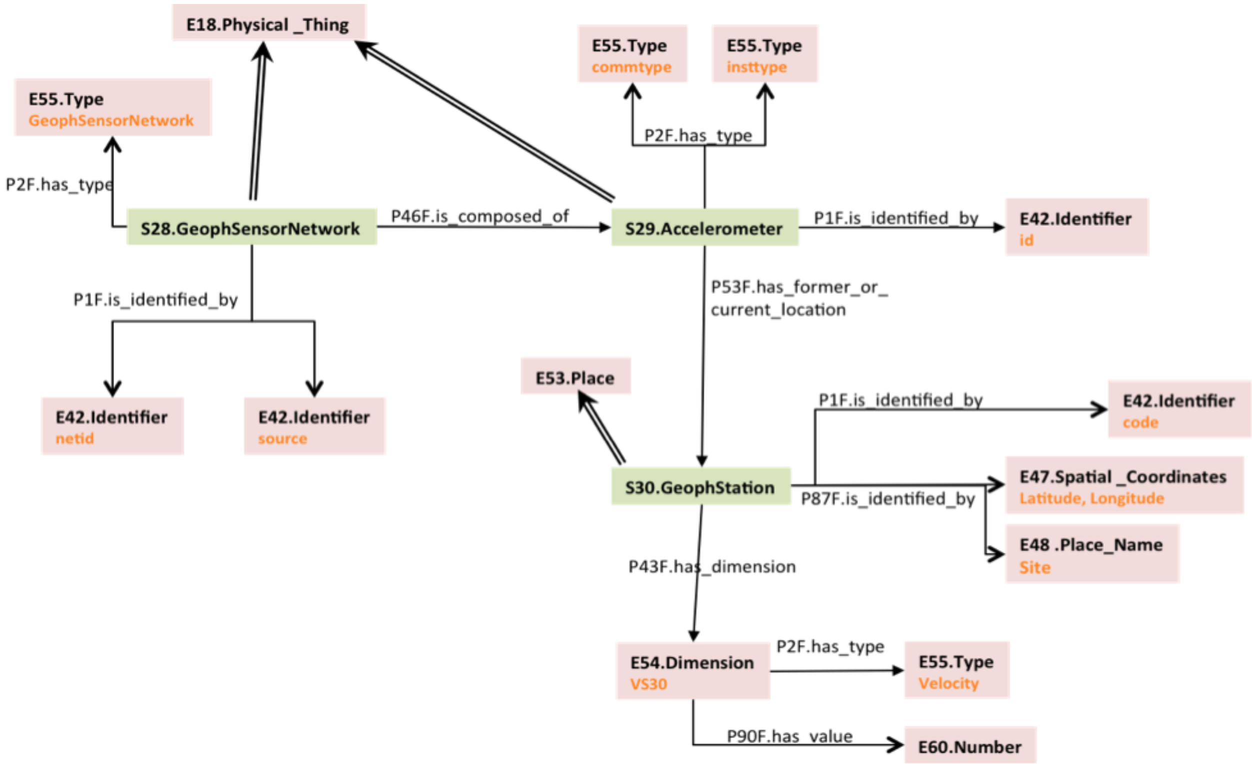

Figure 15: Stations and Sensors Network

4.2 EARTHQUAKE EVENT

In our Core Model every earthquake event is described using the entity *S23.Earthquake* and is considered to be a special case of a seismic phenomenon. As we can see in Figure 16, a seismic phenomenon is described using the entity *S22.Seismic_Phenomenon* and it refers on a general event that generates seismic waves (e.g., an earthquake, a volcanic eruption etc.). For every earthquake event we keep information about some origin time parameters (notion Origin time) and this parameter is described using the relation *P81F.ongoing_throughout*. Moreover, we keep information about the location the earthquake took place at (latitude and longitude), a string representation of the location (notion locstring) along with the depth within the earth where the hypocenter is located. Another important feature of each earthquake is its magnitude (notion mag) which is described using the *E54.Dimension* connected with the relation *O17.has_dimension*. The project's partners maintain documents which contain metadata about seismic phenomena and we capture this functionality using the entity *E31.Document* along with some features i.e., the creator of the event's record and the date of the registration. Finally, we describe each seismic phenomenon using the entity *S21.Spatial_Distribution_Model*. This holds because such phenomena can be spatially distributed (e.g., the earthquake locations on a map) and corresponds to real world situations i.e., the seismic events.

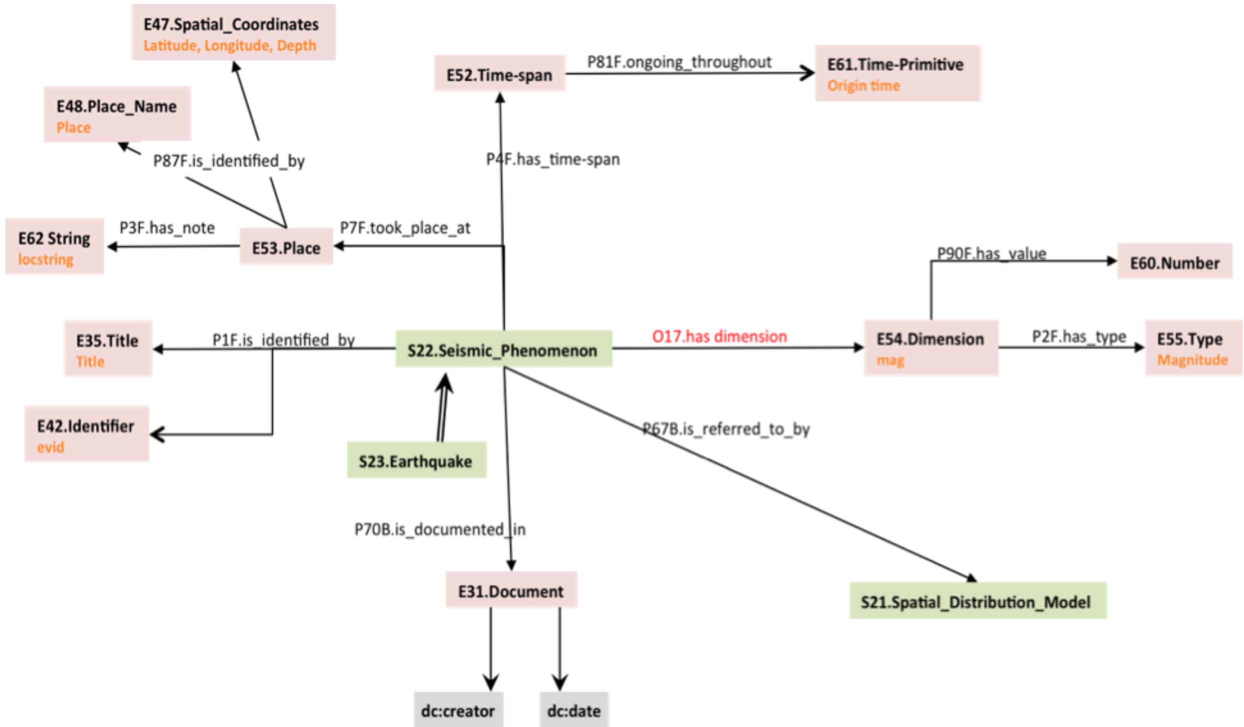

Figure 16: Earthquake Event

4.3 SEISMIC RECORDING

The emergence of an earthquake event triggers its recording by the sensors which are located in the geophysical stations. As we can see in Figure 17, for every such event we relate its recording with the sensor (*S29.Accelerometer*) which measured it using the relation *CRMdig:L12 happened on device*. Every recording can be multiply classified as the original shockwave recording (*S26.Shockwave*) and as the result of the evaluation over the recordings which came from the sensors (*S6.Data_Evaluation*) which can be the result of filtering or other type of processing. This multiple classification leads to different related properties. Considering a recording as an original shockwave recording, we use relation *O17.has_dimension* to describe the intensity value of the observation (notion Intensity) and the regression distance (notion Distance). On the other hand if we consider a recording as a derivative of the original recording we measure different features of the recording. Every recording comes from three different channels namely, HHE, HHN, HHZ. For each one of these channels we are interested in the peak velocity (notion pgv), the peak acceleration (notion pga) and the spectral acceleration at 0.3, 1.0, 3.0 seconds (notions psa03, psa10, psa30).

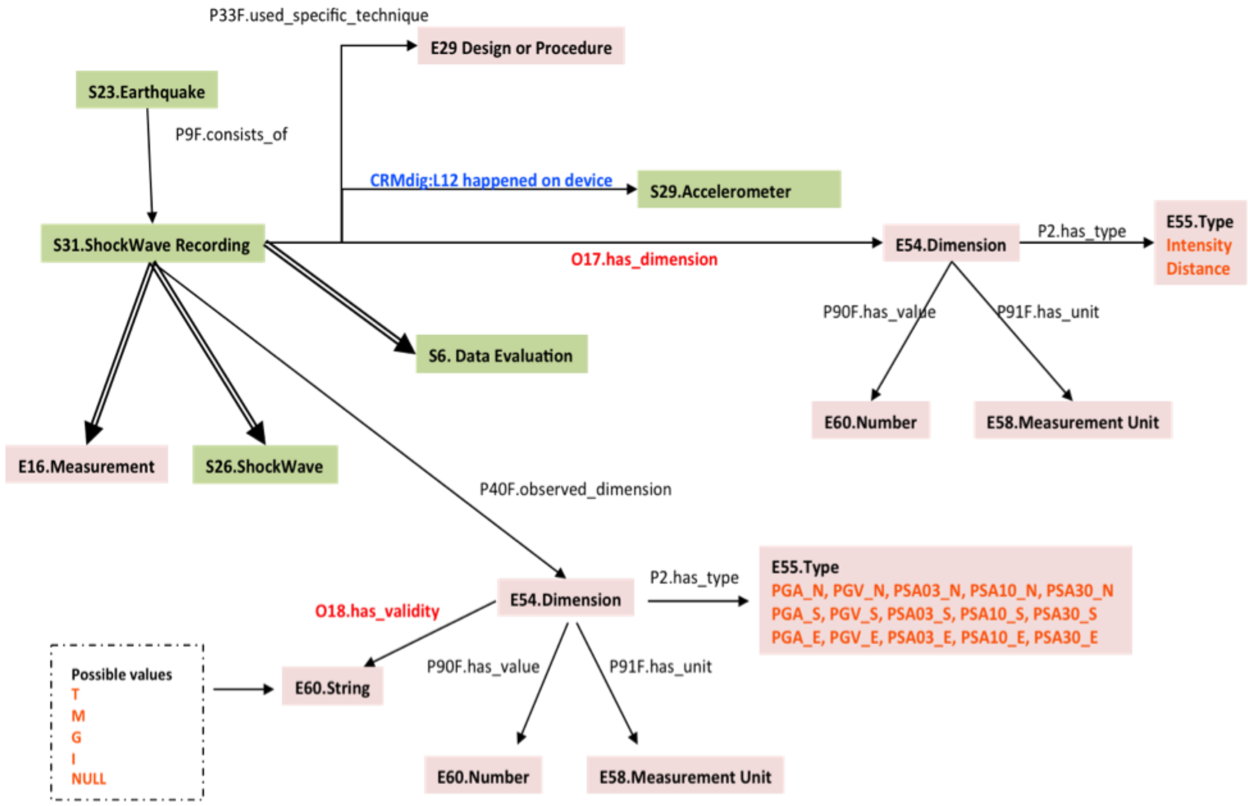

Figure 17: Seismic Recording

5 CONCLUSIONS

The data model described in this deliverable and its annex covers fully the available datasets for the SCInetNatHaz project and allows for useful extensions in the future, if and when this will be required. The proposed model is easily extensible since it is built using semantic web principles and with data integration in mind.

The model covers conceptually three broad areas that all deal with Natural Hazards and their management issues and of course the data that are collected for those purposes. The areas are: Floods, Earthquakes and landslides. We offer an extensive and extensible coverage of the area allowing researchers to better understand the information collected so far and to collect better information in the future. The model is an extension of a previously existing model (so it reuses a tested model in information description, management and integration) and is based on well established standard models that are been published by international standards organizations (like ISO, OGC, W3C, etc.).

6 REFERENCES

[**HFM06**] Halevy A. Y., Franklin M. J., and Maier D.. Principles of dataspace systems. In PODS, pages 1–9, 2006.

[**PCF+12**] Paton N., Christodoulou K., Fernandes A. Parsia B., Hedeler C.: Pay-as-you-go Data Integration for Linked Data: opportunities, challenges and architectures. In SWIM 2012, May 20, 2012, Scottsdale, Arizona, USA.

[**MCD+07**] Madhavan, J., Cohen, S., Dong, X.L., Halevy, A.Y., Jeffery, S.R., Ko, D., Yu, C.: Web-scale data integration: You can afford to pay as you go. In: CIDR'07. (2007) 342 – 350

[**JFH08**] Jeffery, S.R., Franklin, M.J., Halevy, A.Y.: Pay-as-you-go user feedback for dataspace systems. In: SIGMOD'08, ACM (2008) 847 - 860

[**DS2006**] Dittrich, J.P., Salles, M.A.V.: idm: A unified and versatile data model for personal dataspace management. In: VLDB'06, ACM (2006) 367 - 378

[**FHM05**] Franklin, M., Halevy, A., Maier, D.: From databases to dataspaces: a new abstraction for information management. SIGMOD Record 34(4) (2005) 27 - 33

[**HFM06**] Halevy, A., Franklin, M., Maier, D.: Principles of dataspace systems. In: PODS'06, ACM (2006) 1 - 9

[**SDH08**] Das Sarma, A., Dong, X., Halevy, A.: Bootstrapping pay-as-you-go data integration systems. In: SIGMOD'08, ACM (2008) 861 – 874

[**BFM05**] Berners-Lee T., Fielding R., Masinter L.: Uniform Resource Identifiers (URI): generic syntax. January 2005. Internet RFC 3986. URL: http://www.ietf.org/rfc/rfc3986.txt

[**DS06**] Duerst M., Suignard M.: Internationalized Resource Identifiers (IRI). January 2005. Internet RFC 3987. URL: http://www.ietf.org/rfc/rfc3987.txt

[**TSS11**] Tschirner S., Scherp A., Staab S.: Semantic access to INSPIRE, How to publish and query advanced GML data. The 10th Internation Semnatic Web Conference, 23 – 27 October 2011, Bonn, Germany (2011).

[**ING12**] InGeoCloudS Consortium: Interface of web services and models of data, D2.2, Version 1.0, Avaiable at: http://www.ingeoclouds.eu/sites/default/files/D2.2-INGC-IntfWSandDatamodels-V1-Approved_0.pdf

[**CID11**] ICOM/CIDOC Documentation Standards Group; & CIDOC CRM Special Interest Group.: Official Version of the CIDOC CRM, the version 5.0.4 of the reference document. Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead, Matthew Stiff (editors): Definition of the CIDOC Conceptual Reference Model, December 2011. Available online at

http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.doc

7 ANNEX I – DETAILED DESCRIPTION OF THE (GEOSCIENTIFIC OBSERVATION) CONCEPTUAL MODEL