ON-LINE DATASETS

ANDRIANA GKANIATSOU

1. Introduction

Bellow we present some of the available on-line Emergency Response (ER) related datasets, schemas, vocabularies and taxonomies.

2. Available Datasets

The OFDA/CRED International Disaster Database[1] contains all the world-wide disasters, from 1900 to 2008[2].

The dataset is compiled form various sources (agencies, non-governmental organisations, insurance companies, etc). A disaster can be either natural, or technological, or complex emergencies. Then, depending its nature it is further categorised into:

- Biological
- Climatological
- Complex Disasters
- Geophysical
- Hydrological
- Meteorological

The above categories are broken down into more specific disasters, for example, *flood*, *earthquake*, *epidemic* etc. When a disaster occurs, is it up to the administrator whether he will use a category to describe it, for example storm or the specific name of the disaster (if there is one), for example *katrina* .

The dataset is organised by the disaster Id, the date(start,end), country(country name, sub-location), disaster type (category, subcategory, disaster name), cost(us dollars), persons affected (killed, injured, homeless), and is available in spreadsheets.

The National Response Centre[3] also provides an available database of recorded events, which is categorised into calendar years, and is available in spreadsheets. The schema that is used, as well as the vocabulary are also available.

US Open Data[4] makes available several ER related sets, in raw data Data is provided in a raw format. The vocabulary that is used is also available. Datasets related to disasters

[1] http://www.emdat.be/

[2] http://www.infochimps.com/datasets/disasters-worldwide-from-1900-2008

[3] http://www.nrc.uscg.mil/

[4] http://www.data.gov

and ER cases can also be found in **Australian Government's Open Data**[5]. The format of the data varies. For example, all the datasets that come from *ACT Emergency Agency* follow the xml format. Each incident is described by the following fields: incident, location, status, suburb, type, agency, incident number, updated, time of call. Datasets from other agencies are either presented in plain texts or spreadsheets .

Other online datasets can be found in :

- Natural-disaster worldwide : http://natural-disaster.findthedata.org/ (html table)
- Earthquakes : http://earthquakes.findthedata.org/ (html table)
- Tornado Fuita : http://tornado-fujita-scale.findthedata.org/ (html data)
- Natural disasters in Australia http://www.disasters.ema.gov.au/Browse/Zones.aspx

NHS also makes available some ER related data, such as the the database schema that they follow(**Ambulance Services Data Set**[6]), and the vocabulary (**NHS Data Model And Dictionary**[7]). The vocabulary is a taxonomies (classes, subclass and properties). Both the schema and the vocabulary are raw data.

Another example of taxonomies is the **Management of a Crisis Vocabulary Specification**[8]. which provides URIs for ER scenarios, by defining a set of classes and their properties. It links data from three different sources: a)traditional humanitarian agencies, b)volunteer and technical committees, c)disaster affected communities. The format of the data is in OWL/RDF. W3C has proposed the Emergency Management Ontology (2007)[9], in an attempt to offer a well defined terminology and a core set of data for exchange among agencies.

Sahana[10] is a glossary for humanitarian aid and emergency and comes from the Sahana Disaster Management System[11] that was firstly introduced as a response to the Asian Tsunami (2005). The glossary is presented in plain text. Another available glossary for emergency management purposes, is **Fema**[12] (Federal Emergency Management Agency).

3. Emergency Management (EM) Information Standards

EM languages have ben developed for system-to-system interoperability. An example is the *Common Alerting Protocol* a warning-purpose language, represented using xml schemas. Another example is the *Emergency Data Exchange Language - Distribution Element* which facilitates the routing of any xml formatted emergency message. IEEE 1512 is a family of standards for ER and EM situations.

[5]data.gov.au
[6]http://www.datadictionary.nhs.uk/data_dictionary/messages/central_return_data_sets/data_sets/ambulance_services_data
[7]http://www.datadictionary.nhs.uk/
[8]http://observedchange.com/moac/ns/
[9]http://esw.w3.org/topic/DisasterManagement
[10]http://users.on.net/ donc/sahana/sahana_draft_domain_glossary.pdf
[11]http://sahanafoundation.org/
[12]http://www.fema.gov/emergency/nims/Glossary.shtm

4. Discussion

The most common format of the available datasets is spreadsheet, the second common is raw data (no categories, no whatsoever connection between the data), and xml/rdf for some the taxonomies. Taking a closer look at the subsets, it is obvious that they do not use the same terminology. We can find different words for the same meaning, for example, *incident* and *hazard*. Another case is when the same word is used under different meaning. For example, *intelligence* is use to describe both information that leads to detection and prevention of criminal activities, as well as to describe the information that has been evaluated.