European Public Sector Information Platform

Topic Report No. 2014/01

Open Data in Natural Hazards Management

Authors: Vassilios Vescoukis and Charalampos Bratsas

Published: January 2014

Table of Contents

Keywords:

Natural hazards, natural disasters, environmental data, open spatial data.

Abstract/ Executive Summary:

The increasing frequency and severity of natural hazards has greatly raised awareness of local authorities and governments all around the world, towards more efficient planning and management of natural disasters. Environmental incidents in the past couple of decades have resulted in a loss of thousands of human lives, as well as severe damage to infrastructures and economies alike. Traditional methods for managing natural disasters usually focus on the assignment of resources in real-time operations, as well as on the support of the recovery processes. Information technologies such as Geographical Information Systems have been of great help in such approaches.

The latest developments in all areas of Information and Communication Technologies offer a huge and largely unexploited potential for supporting the whole lifecycle of natural hazards management. Computational tools such as environmental simulation, operational logistics and knowledge representation are now available to support scenario-based incident management, as well as real-time decision support. Wireless sensor networks, satellite imaging and communications, as well as terrestrial wireless communication technologies, today allow close-to-real-time monitoring of environmental parameters all around the globe; furthermore, use of such technologies in more than military applications is now possible, which enables a previously unimaginable range of services related to the protection of the environment. Web-GIS and modern mobile personal computing technologies complement the image by enabling virtually unlimited access to up-to-date land images and spatial data of the highest possible resolution, available to everybody without the need of old-time GIS.

There are countless imaginable applications that integrate any or all of the above technologies into Information Systems for the Management of Environmental Hazards. Such systems are by nature distributed and heterogeneous, and require a huge amount of spatial data, ranging from land models, images and road network vector data, to meteorological data flows, real-time measurements, land usage, infrastructures, vegetation coverage, and even real-time spatial distribution of population. This paper discusses the challenges of using Open Data in environmental applications, especially focusing on the management of natural hazards.

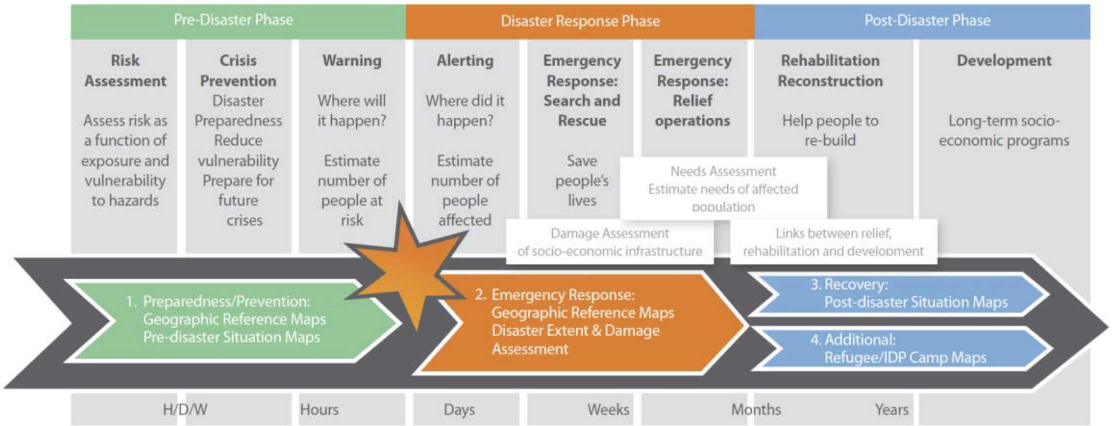
1 The context

1.1 Lifecycle of natural hazards management

To properly understand the requirements of open data for supporting Natural Hazards Management, we first need to establish a terminology for the life cycle of what is referred to by "Natural Hazards Management". Surprising as it may sound, there is no global consensus about what the exact phases of this lifecycle are. Whether this is due to the nature of the problem itself, to conceptual differences or even to political factors is out of the scope of this work; having said that, we shall briefly discuss the terms used by major organizations around the world, in order to define a concise semantic terminology framework that will be used in this paper.

The **United Nations** recognize 7 phases related to disasters and their management, namely prevention, mitigation, preparedness, response, rehabilitation, reconstruction and recovery (United Nations, 2013).

There are numerous actions and initiatives in the **European Union** which constitute a less that totally comprehensive policy area, which is probably due to the cultural, political and administrative diversity across Europe. A generic discrimination of the different phases of emergencies follows the three-phase pattern: preparedness/prevention (pre-disaster), emergency response (disaster) and recovery (post-disaster) (GMES Emergency Response Service, 2010). These three phases are further analyzed into 8 sub-phases: risk assessment, crisis prevention, warning (pre-disaster), alerting, and response: search-and-rescue, response: relief (disaster), rehabilitation-reconstruction and development (post-disaster). Other sources make word about prevention, preparedness and response (European Commission, 2013), which probably indicates a differently focused point of view.



Source: (GMES Emergency Response Service, 2010)

On the other side of the Atlantic, the **US Federal Emergency Management Agency** (FEMA, 2013) recognizes four phases of emergency management: mitigation, preparedness, response and recovery. Mitigation is about prevention of emergencies and reduction of their damaging effects; preparedness deals with planning of actions that will improve the effectiveness of response to an emergency; recovery includes actions needed to return to a normal or safer than before the disaster, state.

The lack of global common terminology about disaster management can be attributed to the different viewpoints from different organizations and administrative entities. Everybody recognizes the pre-disaster, response, and post-disaster phases, which is the simplest approach one can think of; however, depending on each entity of organization, its role, and the degree of operational authority it has, more detailed sub-phases can be recognized. From our point of view, as the discrimination of phases becomes more detailed, new requirements for open data arise in order to support operations, monitoring and decisions across the lifecycle of Natural Hazards Management, which justifies this discussion. Table 1 shows how the three viewpoints presented so far relate to each other.

<table>
<tr><th></th><th colspan="3">Pre-disaster</th><th colspan="3">Disaster response</th><th colspan="3">Post-disaster</th></tr>
<tr><td>UN</td><td>Prevention</td><td>Mitigation</td><td>Preparedness</td><td colspan="3">Response</td><td>Rehabilitation</td><td>Construction</td><td>Recovery</td></tr>
<tr><td>EU (GMES)</td><td colspan="2">Assessment, Prevention</td><td>Warning</td><td>Alert</td><td>Rescue</td><td>Relief</td><td colspan="2">Rehabilitation, Reconstruction</td><td>Development</td></tr>
<tr><td>FEMA</td><td colspan="2">Mitigation</td><td>Preparedness</td><td colspan="3">Response</td><td colspan="3">Recovery</td></tr>
</table>

Table 1. Terminologies and phases recognized in disaster management

In the sequel we follow the most generic discrimination of the pre-disaster, response and post-disaster phases and list the responsibilities and actions that fall into these phases, with regard to the established terminology presented so far.

During the **pre-disaster** phase the following activities are recognized:

- Risk assessment: identify vulnerabilities in population and infrastructures
- Mitigation: reduce damaging effects of unavoidable natural hazards
- Preparedness: develop operational plans for operational teams; develop awareness plans and advice to the general public

During the **disaster response** phase, the following activities are recognized:

- Detection and alert: timely detect hazards that could not be predicted (such as earthquakes).
- Operations planning: real-time situational assessment, field and resource management
- Operations execution: deploy available resources to minimize casualties and damages.

During the **post-disaster** phase, the following activities are recognized:

- Damage assessment and recovery planning: measure the effects of the disaster; create plans for immediate and long-term measures and actions
- Run recovery activities: deploy and monitor all kinds of recovery activities
- Improve preparedness and prevention: mitigate effects of future hazards

The above outline will be used in the next sections to establish functional and non-functional requirements that need to be satisfied by Information Systems used anywhere in the lifecycle of Natural Hazards Management.

1.2 Options and functional requirements

Information Systems can be of indispensable help in supporting all actions in the lifecycle of natural hazards management. Computational tools and methods can support all phases from preparedness, to operational scenarios, to real-time decision support, to recovery monitoring planning and management. Wireless sensor networks and modern communication technologies make possible the close-to-real-time detection and monitoring of all natural disasters, which in turn enables a previously unimaginable range of services. Web-GIS and similar technologies enable virtually unlimited access to all kinds of up-to-date data needed.

However, no silver bullet exists, and as is the case in any other computer application, requirements specification is a key in understanding what Information Systems can do; formal specification of requirements specification from information systems for natural hazards management is far beyond the scope of this paper. Nevertheless, a qualitative specification of all kinds of services that can be offered in every phase of natural hazards management can be of great value in understanding what data is needed to provide this functionality. In this regard, the following qualitative requirements can be recognized:

Digital mapping services. This is the most essential information system service, which will offer the mapping presentation layer to be used almost in every other service of the natural hazards management lifecycle. Current options include traditional GIS software, as well as global mapping services offered by Apple, Bing, Google, Yahoo, etc. Traditional desktop GIS, although will do the job, is rather outdated; modern web-based digital mapping solutions are a much better option since the costly update of spatial data needs to occur once and data is made available to all possible users and applications. Furthermore, advanced presentation services such as 3D or ground views only need to be implemented once and are usable by all. Needless to say, such web-based systems are the ecosystem where open data feels at home and therefore can be of more impact. Even restricted to specific geographical areas of authority, mapping services have several critical non-functional requirements that need to be met in order to serve their role in Natural Hazards Management, as will be discussed in the next section.

Spatial data services. Digital mapping provides only the essential layer of functionality needed to support Natural Hazards Management. On top of this layer we need to place all human traces on Earth, including but not limited to all kinds of infrastructures, transportation networks, economic activities, resources and any other geo-tagged info useful in emergency management, and even real-time population distribution. This service needs to update its data much more frequently than the essential digital mapping service discussed above, which is the first reason for which we regard it as a different service. The second and probably more important reason is that the spatial data that needs to be maintained by this service come from many different sources, ranging from administrative entities that are responsible for specific fields, to local volunteer initiatives, to crowdsourcing. For example, the legal administrative entity responsible for ground transportation networks may not be the same as the one responsible for city structures and population distribution; in some cases the situation might be complicated enough to create a need for complicated workflows required to keep this data up-to-date, reliable and dependable, as is required for its mission.

Risk assessment services. The identification of vulnerabilities and the assessment of risks related to natural hazards requires processing of all available data, using various computational methods, depending on the type of hazard, the local characteristics and even the local culture and behavioral attributes. The assessment of risks, as well as the classification and ranking of vulnerabilities are among the main functions of this service.

Environmental modeling and simulation services. Clearly, not all natural hazards can be modeled using mathematical tools; however, using such tools where available and within their limitations, can greatly improve planning and risk assessment long before any real natural disaster occurs. This service offers simulation of natural hazards such as landslides, floods, tsunamis and wildfires; doing so, it requires more spatial data as will be discussed in more detail in section 2. When dealing with a real disaster, environmental modeling and simulation is a family of services useful in supporting authorities' decisions about the optimal field deployment of the usually limited resources; when used for planning in the pre-disaster phase, simulation can be used in the formulation of scenarios to increase preparedness of both the general public and operational teams.

Operational logistics services. Operational logistics is a relatively new term that refers to the optimization of field operations, including natural disaster response; operational logistics implement adjusted variations of algorithms commonly used to solve classical logistics problems. A typical example of such a service is the routing of emergency response vehicles, from any known location (station, point of observation, current position) to critical positions for the management of emergencies, such as fire fronts or rescue missions. Another useful service in this category is the evacuation risk assessment, useful in estimating the load of the road network during evacuation of an area under panic, which may occur in emergency situations. The service points out which roads will be most loaded in a panic evacuation, which helps the authorities to take proper action for keeping these roads open and avoid congestions that can lead to human casualties. Either of the above services is useful both in the pre-disaster and the response phase. Optimal operational resources placement is another service in this category, useful in estimating the number of vehicles (resources in general) optimally required to cover an area of interest by achieving a required emergency response performance. Operational logistics services need to be integrated to each other as well as to the environmental simulation services. For example, vehicle routing needs to take into account the spatial evolution of a natural disaster so as not to consider using blocked road network segments.

Scenario management services. Scenarios are widely used as strategic decision support and preparedness improvement tools during the pre-disaster phase. Running any combination of the services discussed above during in a pre-disaster phase, produces a set of coherent results under a certain environmental disaster scenario, useful in improving preparedness, response, and operational planning; scenarios can also be used as public awareness material, or even as educational material by authorities responsible for preparedness and response planning.

Field monitoring and data flow services. Disaster response is about saving people in a real world situation whose critical details can be neither predicted nor estimated by using mathematical methods. Real-time field data flow is today possible by several technologies, including satellite observations, wireless sensor networks, vehicle location tracking, ad-hoc networks, mobile network operators, and even people themselves. Integration of field data flows with operational logistics during the emergency response phase is needed to support the assessment of the operational needs using real situational data instead of making assumptions and estimations.

Recovery monitoring and management services. Last but not least, all actions taking place in the post-disaster phase need to be supported so as to improve the efficiency of actions taken and track the implementation of decisions aiming to lead to a better-than-before-the-disaster situation. A wide range of services ranging from landscape recovery, to reconstruction damages, to monitoring and analyzing of the economic activities, can be considered as members of this family of services.

Table 2 below, correlates functional requirements to disaster management phases, as discussed in section 1.1.

Service / phase	Pre-disaster			Disaster response			Post-disaster		
	Risk assessment	Mitigation	Prepared-ness	Alert	Operations planning	Execution	Assessment & Planning	Recovery actions	Prepared-ness & prevention
Mapping	•	•	•	•	•	•	•	•	•
Spatial data	•	•	•	•	•	•	•	•	•
Risk assessment	•				•				
Environmental modeling	•		•		•				
Operational logistics			•		•	•			
Scenario management		•	•		•	•			
Field monitoring				•	•	•			
Recovery monitoring			•				•	•	•

Table 2. Functional requirements from Natural Hazards Management Information Systems applicable to different phases of the Natural Hazards Management lifecycle.

As expected, mapping and spatial data services are essential in all phases of Natural Hazards Management, as they offer the spatial reference layer for any other service. Risk assessment services are mainly used in the risk assessment and operational planning phases, focusing on the assessment of vulnerabilities of areas in-focus, rather than the dynamics of natural phenomena; the latter is where environmental modeling focuses on, which is why this family of services is useful in the preparedness phase, as well.

Operational logistics are also useful in the response phase itself, since such services are about managing field operations. Scenario management is about putting together several hypotheses and collect results about risk, disaster strength and response, so as to have a full image about the dynamics of a possible disaster before it actually happens. Field monitoring is focused on operations, and last but not least, recovery monitoring could be also useful in the preparedness phase as well. Having said the above, let us clarify that different perspectives on the semantics and relevance of services are also possible in specific contexts.

1.3 Non-functional requirements

Although, as already stated, a complete specification of information systems for natural hazard management is out of the scope of this paper, it is important to extend the discussion so far to non-functional requirements, in order to better understand how Open Data can be used in Natural Hazards Management. In this regard, and focusing on the non-functional requirements most relevant to data, there are four major families of such requirements: availability, standardisation, security, and reliability, which will be briefly discussed in the sequel.

Availability is about data been accessible at all times. Depending on the application architecture, availability may be or may not be a critical issue; architectures that use imported data in controlled database environments have total control on the availability of this data; on the other hand, if due to technical or other limitations data is provided online by any third party, then it is important to take measures and policies to deal with the availability of services offered by others. Depending on the kind of data this may or may not be a failure point for mission critical applications such as Natural Hazards Management.

Standardization is a known issue in open data, which is very important to achieve the interoperability some services require. Although standards appear one after another, de-facto proprietary formats, legacy data, transition costs, the huge diversity of data sources, or any other limitation, make standardized data not as widely available as one would expect. If on-the-fly transformations can be applied, then the standardization requirement can be easily met. Otherwise, different approaches need to be taken, that may require possibly scarce resources such as time and big budgets, which may lead to different design and even strategic decisions.

Security is a big issue in all distributed applications and becomes an even bigger one in the case of mission critical applications that deal with civil safety. Security has several aspects, the most critical being the safe transmission of sensitive data, and the certification of the identities of parties involved in any data transaction. Cryptography technologies can handle the former and, considering the relevance of Natural Hazards Management to public safety it is acceptable to assume that there will be some solution available. Note that although cryptography is important, it is not a major concern for Open Data as such. It may be such a concern, however, if due to the nature of the application there is need to protect not the data itself, but the fact that it is communicated between to parties at a time critical for public safety, then cryptography comes into play even if all data is open data. The latter (identity verification) requires infrastructures PKIs and services such as Certification Authority chains; any restrictions on the availability or applicability of security technologies may lead to more centralized system architectures, which may be totally acceptable for mission-critical systems. Even then, all data involved may be Open Data.

Reliability is about the quality of the data, which is extremely critical for any civil safety application. There is not much to discuss here since there is no simple way to tell if a dataset is as dependable as an application related to civil protection might require, which stands even for proprietary, closed datasets. In this regard, the use of data that comes from sources whose reliability is not granted should be very cautious in mission-critical civil safety applications as is the case of Natural Hazards Management.

2 Data required in natural hazards management

Having defined the boundaries of what kinds of services can be offered by an Information System for Natural Hazards Management, we can now establish what the needs of such a system in terms of data are. As is the case with any kind of software application that offers geo-spatial services, Natural Hazards Management requires a great big deal of geo-spatial data of the highest possible quality; on top of that, if possible this data needs to be literally up-to-date. As will be discussed in the conclusions section, these requirements are not always met by open data.

There are 11 kinds of datasets useful in Natural Hazards Management that will be briefly discussed in the sequel:

- Satellite imagery
- Elevation and surface models
- Meteorological data
- Transportation networks data
- Demographics and population density data
- Country and urban borderline data
- Use of land and buildings
- Utility networks
- Critical infrastructures
- Hospitals, schools, and other vulnerable locations and last but not least
- PoI data, which is data practically about any point of interest (PoI) directly or indirectly related to natural hazards management.

Each one of the above has different specific attributes such as spatial resolution, sources, update frequency, limitations, terms, etc, which makes integration into one system quite a challenge. Some are of strategic importance and barely available as open data, and some others are updated by crowdsourcing which is good because it puts pressure on commercial vendors and governments alike join the Open Data practice.

Satellite imagery. Initially available only for military use, satellite images today have many civil applications in areas such as regional planning, geology, meteorology, agriculture and more, including Natural Hazards. Satellite images are useful in the representation of the earth surface in mapping systems, as well as in the interpretation of the surface. The two most important properties of satellite images are spatial or geometric resolution and color resolution. Geometric resolution refers to smallest earth surface unit that maps to a single pixel; low-res images have a pixel resolution of few kilometers to few hundred meters, while spatial resolution of high-res images varies from 30x30m to 0.5x0.5m for civil applications. Color resolution varies usually from 8 to 16 bits, the latter offering a significantly higher degree of detail about the covering of the surface. Satellite imagery can be supplemented by aerial photos, which offer even better resolution but on the other side are more expensive and practically impossible to keep-up with the update pace of satellite images, which may be in the range of a few hours for a single satellite. Natural Hazards Management requires as recent satellite images as possible. As money is always an issue, compromises need to be made which under certain assumptions can be quite acceptable, as will be discussed next in section 3.

Elevation and surface models. A digital terrain or elevation model is a representation of the bare ground surface of the earth without any objects on it. Digital surface models in addition offer surface covering data such as forests and vegetation, which complements satellite imagery by offering data useful in the interpretation and 3D representation of the earth surface images. They are extremely useful in Natural Hazards Management as they support modeling of water flows (as in floods) or mass movements (as in landslides). However, resolution of such models lags behind the sub-meter resolution capacity of current satellite imagery, usually being in the range of 30 meters for freely available data (NASA, 2013); a model with 12-meter resolution is expected to be available in 2014 (DLR, 2013; European Environmental Agency, 2013). Satellite images along with elevation and surface models are the essential data sets required by digital mapping services.

Meteorological data. There is no need to argue on how important meteorological data for Natural Hazards Management is, in all phases of its lifecycle and most of the required services discussed so far. Ranging from prediction, to risk assessment, to environmental simulation, to real-time operations support, to recovery management, meteorological data plays an important role in Natural Hazards Management. Historical meteo data is available from national or international organizations, as well as from private entities and initiatives of all types either free or at a fee. Real-time meteo data is also critical in disaster response situations and can be available through many sources ranging from official national authorities to crowdsourcing using private-owned weather stations. Meteorological data along with surface models is essential for environmental modeling and simulation services.

Transportation networks data. Surface transportation networks include road networks, railway networks and possibly sea lines; they are usually represented as layered graphs of nodes and vectors that connect nodes; vehicle types, vehicle and road capacities, as well as other attributes such as slope, signaling, direction, valid turns, and other data fields that relate to the kind of transportation means. Transportation networks can be damaged or otherwise altered by Natural Disasters, which may change the boundary conditions or other properties of operational logistics problems that need to be solved over such networks during the planning of emergency response operations. In this regard, data about transportation networks is essential for risk assessment and operational logistics services as discussed above.

Demographics and population density. Natural Hazards Management is above anything about protecting and saving lives, which makes the requirement for demographics and spatial population density data self-explanatory. Such data is usually available by national statistical authorities and its use in operational logistics and emergency response planning should take into account the fact that it represents a static image of the population distribution. Dynamic or even real-time population distribution can be monitored using cellular communications networks, which could drastically change the capabilities of emergency response in Natural Hazards; however this remains a relatively new area of research and limited applications are available as of today.

Country and urban borderlines. Vector data that represent perimeters of building blocks as well as of any type of urban or country settlements is another kind of data useful in many services in Natural Hazards Management, such as risk assessment, evacuation planning, damage assessment, and recovery planning and monitoring. Such data is usually available by national authorities and can be regularly updated as well. The spatial distribution of population can be either referenced to this kind of data or to the building-level, as discussed next.

Land use and buildings. This is a more detailed level of geo-referenced data about borderlines up to the single building; for non-urban or populated areas, this data is about the characterization of land uses, such as agricultural, mines, forests, industrial, wetlands, etc. Both are useful layers of information about risk assessment, emergency response planning and recovery monitoring. However, availability of detailed data, especially to the building level is still limited.

Utility networks. Utility networks are very important in several aspects of emergency management. Although not as directly related to emergency response operations as is the case with transportation networks, mapping data about utility networks are relevant to Natural Hazards Management both ways: on the one side, they can support life-saving operations; on the other side, damages to utility networks during natural disasters may trigger events that create new natural disasters. Typical examples of the latter are the eruption of wildfires due to the collapse of power lines, or the starting of urban fires due to damages in natural gas pipes. Availability of data about utility networks is, therefore, important to the risk assessment of natural hazards.

Critical infrastructures. The same goes for critical infrastructures, which need to be defined as a different category due to their importance in the functioning of societies and economies. Examples of such infrastructures are power generation plants, water supply reserves, airports, seaports, and more. The responsibility, authority and most likely all data about critical infrastructures are usually in the hand of central governments. This data is likely not to be openly available, however it may be a critical factor in decision making during emergencies, which is why it is mentioned here.

Vulnerable locations. Another kind of data important in several aspects of Natural Hazard Management is about most vulnerable locations; such locations are hospitals, schools, stadiums, as well any building or site where people are gathered. Especially for hospitals, schools, elderly people nursing homes and the like, special attention needs to be paid in planning of emergency management. They all need to be given higher levels of protection and attention during unavoidable evacuation. Data about hospitals is also critical in planning life-saving operations, even if hospitals themselves are not in danger in a disaster situation.

Points of interest. Last but not least, any geographically tagged information about Points of Interest (PoIs) of any type not discussed above, is welcome, and might be relevant as a decision-making factor in emergency management. The good news here is that there is plenty of such data available by crowdsourcing activities.

Table 3 shows the correlation of the 11 kinds of data discussed above, to the Information System services related to Natural Hazards Management, as mentioned in section 2.

Data/Service	Mapping	Spatial data	Risk assessment	Environmental modeling	Operational logistics	Scenario management	Field monitoring	Recovery monitoring
Satellite imagery	•				•			•
Elevation and surface models	•			•				
Meteo		•	•	•	•	•	•	
Transportation networks		•	•		•	•	•	•
Demographics and population		•	•		•	•		•
Borderlines		•	•	•	•			
Land use		•	•	•	•	•		
Utility		•	•		•	•		•
Critical infrastructures		•	•		•	•		•
Vulnerable sites		•	•		•	•		•
PoIs		•				•		

Table 3. Data vs Information System services in Natural Hazards Management

As shown in Table 3, satellite imagery along with elevation and surface models are the two families of data essential to mapping services. Spatial data services can use any kind of data with spatial reference, to offer simple yet quite essential presentation and mapping useful in all aspects of Natural Hazards Management. The same stands, more or less, for risk assessment services; the main difference of these services from spatial data services is that risk assessment requires more detailed and quantitative data than what is needed for presentation only. Environmental modelling is a data-demanding family if services. Depending on the environmental phenomenon, these services may require a combination of detailed surface and elevation models, along with historical and even real-time meteorological data; this kind of services is the most computing-intensive as well. Operational logistics requires any vector data available, as well as any population distribution data, including real-time population distribution in case of emergency response. Practically all sorts of data are relevant to scenario management, including results produced by environmental simulation and operational logistics services. Meteorological data flows along with transportation network monitoring are the data families most relevant to field monitoring, whereas recovery monitoring may require the data families indicated in Table 3, although this may vary depending on each specific case.

3 Open data for Natural Hazards Management

Open Data is a great idea that can tremendously change the way we think about applications and value chains in today's distributed computing. Open Data can shift the focus of many activities on the Internet, from offering "information" to adding value by dynamically linking unlimited amounts of all sorts of data to create new value, useful in new applications in all aspects of human activities.

As noted in the Open Data Handbook (Open Knowledge Foundation, 2012), *Open data is data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike.* According to the same source, the most important attributes of Open Data are the following:

- **Availability and Access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet, and in a convenient and modifiable form.
- **Reuse and Redistribution:** the data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets.
- **Universal Participation:** everyone must be able to use, reuse and redistribute - there should be no discrimination against fields of endeavor or against persons or groups.

How do the above apply when Open Data is about Natural Hazards Management as discussed above? Is there any limitation, compromise or assumption in using Open Data in Information Systems that support Natural Hazard Management? Well, there are several limitations and compromises that need to be made. This does by no means imply that Open Data is unfit for the purpose; it does mean, however, that until Open Data becomes the dominant, if not the only, way of distributing data, we have to accept certain limitations and focus on what *can* be done with Open Data in Natural Hazards Management, instead of discussing what *cannot* be done, at least as we would ideally like to do it.

Limitations are not uniform for all kinds of data required in Natural Hazards Management. High-resolution and **up-to-date satellite imagery** as of today is not free and nobody knows if it will ever be, due to the high cost it takes to build and operate satellites, as well as to the strategic significance of satellite information. Only limited datasets for selected areas of the globe in low resolutions (ranging from kilometres to hundreds of meters per pixel), barely suitable for presentation purposes only in Natural Hazards Management applications, are freely available from sources such as ESA (ESA, 2013) and NASA (NASA, 2013).

A notable development has been announced by the Copernicus Earth observation program in the European Union: Starting this year (2014), 5 new missions with launch 5 families of "Sentinel" satellites loaded with state-of-the-art instruments that will provide high-resolution spatial data and measurements (ESA, 2013-2). The first satellite, Sentinel-1, is scheduled for launch in the first quarter of 2014 and will provide imagery for user services. Less than 12 months later, Sentinel-2 is expected to be launched to offer high-resolution data quite relevant to natural disasters and emergency response, including up-to-date land coverage, floods, landslides, etc. What makes Copernicus quite a notable development is a recent announcement that Sentinel data will be free and open, which will undoubtedly be a breakthrough in the use of satellite data.

Global mapping services such as Google Maps are current alternatives with quite a few but under certain conditions acceptable limitations and compromises. However, as of today, global mapping services do not offer any sort of access to Open Data and in this regard they are out of scope of this paper.

There is also good news, however: All other kinds of data required in Natural Hazards Management are not as restricted as satellite imagery. There are several limitations and drawbacks here as well, but compromises that need to be made are more realistic compared to those it takes to use open satellite imagery in Natural Hazards Management. It is by no means an ideal situation, either. Free data comes at a huge variety of formats, availability and use restrictions, time and space resolution; some of this data is indeed Open Data as defined above but most carry restrictions especially in the "Reuse and Redistribution" and "Universal Participation" areas.

Elevation and surface models are available as Open Data with not as high spatial resolution as high-res satellite images, but still perfectly usable within the scope of Natural Hazards Management, especially after some post-processing. Surface models usually include **land use**

data at fair spatial resolutions, as is the case with CORINE (European Environmental Agency, 2006).

Meteorological data is still not widely available as Open Data. Although there are many sites offering historical meteorological data for specific areas of the world, there is no generic trend towards opening such data. Some services already offer real-time meteorological data along with weather forecast APIs for free; a typical example of such a service is Open Weather Map (OpenWeatherMap, 2013), which, however, does not comply with the third point of the definition of Open Data. Usage of meteorological data in other strategic applications such as optimal placement of green energy plants (wind, solar) is undoubtedly an obstacle to the general opening of meteorological data. There are, however, developing efforts for opening meteorological data, such as the Open Meteo Foundation (Open Meteo Foundation, 2013) that are very promising and worth supporting.

There has been a long-lasting dispute, especially in Europe, about opening of **transportation networks data**. Such data, extremely critical in Natural Hazards Management, is still offered at a fee by private companies and with significant limitations on the permitted uses. However, in April 2013 European Governments have agreed to open up public data; this process is still in progress and may take a couple of years to be completed, but the idea is quite straightforward: public data, collected or produced by public bodies using public resources, needs to be offered as Open Data (EU, 2013). Such data include geographical information, **weather and road network** data, **demographics and population** data, **utility networks**, **borderlines**, and possibly other layers of data involved in Natural Hazards Management.

There are several other options for obtaining **road network data** as Open Data; one of the most important being Open Street Map (OpenStreetMap.org, 2013), which is totally based on crowdsourcing; Open Street Map offers lots of spatial data sets at a variety of standard and usable formats, including **road networks**, **PoIs**, public transport, and more. Waze (Waze, 2013) is another crowdsourcing-based service, focusing on real-time road network status data, and PoIs management; Waze has recently been acquired by Google; it does not offer datasets as Open Data, but is a very good example of what crowdsourcing can do even in real-time applications.

Critical infrastructure data is by nature restricted from being made available as Open Data.

Authorities involved in Natural Hazards Management usually have access to critical infrastructure data under several administrative terms. Although opening of critical infrastructure data is largely a political issue, it is good to know that those in charge of Natural Hazards Management have access to critical infrastructure data, even if not as Open Data.

4 Conclusions and discussion

Natural Hazards Management is a critical responsibility of administrations from the local to the central government level. It is about civil protection, as actions taken during a war or a terrorist attack are; as such, they are by nature sensitive in all aspects. There have to be response plans for all types of emergencies, however the borderline is in some case blurred; for example evacuation of an area because of a forest fire is not different to the evacuation of the same area due to some terrorist attack. There are differences and discriminating factors, of course; however it makes little, if any, sense for administrations to maintain at least two discrete sets of response plans or, even worse, to pay for double the operational resources.

All the above are fair reasons to expect limitations in the general availability of sensitive information and data related to Natural Hazards Management. One of the most important families of data for this purpose is high-resolution and up-to-date satellite imagery; if there was only one dataset that could de made available to some entity responsible to develop Information Systems for Natural Hazards Management that would be satellite imagery of the highest possible resolution. Clearly, however, this kind of data is not available as Open Data without practical limitations that are not acceptable for this purpose. So, here comes the first and probably most important compromise for any authority involved in Natural Hazards Management: either accept this fact and pay the cost to purchase the satellite imagery needed, or compromise with the limitations of a global mapping service. The choice that will be made here, leads to possibly different architectures of the supporting Information Systems, as well. This might change as the Copernicus Sentinel missions are expected to offer open access to high-resolution and up-to-date satellite imagery, useful in emergency management.

The availability of the rest of the families of data shown in Table 3 as Open Data is a matter of time and persistence, especially for Europe. Policies are already being introduced to force the opening of datasets that have been created using public money. Datasets about transportation, demographics, borderlines, land use etc., will sooner or later become available as Open Data from their official owners; it may take time, but it will eventually happen, as it already has in the United States and elsewhere. This will signal a mentality shift of the computing services industry from selling bare data, to offering added value.

One might argue that in Europe, opening all kinds of public data required in Natural Hazards Management might take years. Possibly, but life has its own way of doing things and lately there have been some very encouraging developments related to the opening of all but satellite imagery, data: we refer to crowdsourcing, which has shown its strength to change things that regulators, for whatever reason, cannot.

Crowdsourcing is the most revolutionary development in the opening of data that has long been considered as intellectual property of private entities that have also been able to influence policy makers to legislate in their favour. The idea is simple: everybody, literally everybody, may contribute a tiny little piece of data in a dataset. Sooner or later this dataset will be complete and publicly available for free. Life has shown that this will be sooner rather than later, and the case of OpenStreetMap is a brilliant example that has encouraged other communities to be organised towards collectively gathering and openly offering data sets.

The most relevant such community to Natural Hazards Management is the meteorological data community. Life will either leave behind those who insist on selling such data, or force them to open their datasets and focus on asking money for value services. All other families of data shown in Table 3, with the exception of critical infrastructures, are already even partly available as Open Data. There is still much work to be done, but knowing that it has already started, is quite encouraging, indeed.

There is a big dispute, however, about data collected by crowdsourcing: *accuracy*, *completeness* and, when it comes to mission-critical use such as Natural Hazards Management, *dependability*. Well, this is a really critical concern that definitely needs to be taken into account. But before stating that "free is free, so accepting some limitations is expected", we shall emphasize on something else: are there any strict liability terms of commercial datasets that relate to Natural Hazards Management? The answer is unfortunately "no"; which means that if some emergency vehicle routing fails due to inaccurate digital road map data, the commercial provider of this data bears no responsibility. The conclusion is that neither Open Data collected by crowdsourcing is of questionable quality just because of the way it has been collected, nor commercially available data is dependable and thus liability can be claimed, just because it is commercial.

Natural Hazards Management is a critical and demanding task that goes far beyond data. Undoubtedly, Information Services are valuable assets in planning for, responding to and recovering from natural disasters; however, there is a whole chain of value that has to work in order Natural Hazards Management to be as effective as possible and to exploit current advances in all areas of computing. Open data may be part of this value chain, but due to the nature of this mission and its relation to homeland security and civil protection, there are certain limitations that do not stem from the classical dispute "Open vs non-Open" data, but instead they are based on the strategic nature of some critical data elements.

Keeping the above in mind, one can only be optimistic about any non-strategically-restricted dataset related to Natural Hazards Management that will, sooner than later, become available as Open Data._

5 References

DLR. (2013). *TanDEM-X.* Retrieved 12 10, 2013, from DLR Microwaves and Radar Institute: http://www.dlr.de/hr/en/desktopdefault.aspx/tabid-2317/3669_read-5488/

ESA. (2013). *EO data distributed by ESA.* Retrieved 12 10, 2013, from Earthnet Online: https://earth.esa.int/web/guest/data-access/how-to-access-eo-data/earth-observation-data-distributed-by-esa

ESA. (2013-2). *Copernicus - Observing the Earth Overview*, Retrieved 12 10, 2013, from ESA Portal: http://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Overview4

EU. (2013). *EU Open Data Portal.* Retrieved 12 10, 2013, from EU Open Data Portal: https://open-data.europa.eu/en/data

EU. (2013). *Open Data.* Retrieved 12 10, 2013, from Digital Agenda for Europe: https://ec.europa.eu/digital-agenda/en/public-sector-information-raw-data-new-services-and-products

European Commission. (2013). *The Common Emergency Communication and Information System (CECIS)*. Retrieved 12 10, 2013, from Humanitarian aid and Civil Protection: http://ec.europa.eu/echo/policies/disaster_response/cecis_en.htm

European Environmental Agency. (2006). *CORINE Land Cover.* Retrieved 12 10, 2013, from European Environmental Agency: http://www.eea.europa.eu/publications/COR0-landcover

European Environmental Agency. (2013). *Datasets.* Retrieved 12 10, 2013, from European Environmental Agency: http://www.eea.europa.eu/data-and-maps/#tab-datasets

GMES Emergency Response Service. (2010). *General Service Desctiption.* Retrieved 12 10, 2013, from http://www.emergencyresponse.eu/gmes/docs_wsw/RUB_119/ssp_gmes_datasheet_00_general_vi.pdf

NASA. (2013). *ASTER.* Retrieved 12 10, 2013, from NASA Jet Propulsion Lab: http://asterweb.jpl.nasa.gov

NASA. (2013). *Discovering Data.* Retrieved 12 10, 2013, from EOSDIS - Earth Observing System Data and Information System: https://earthdata.nasa.gov/data/discovering-data

Open Knowledge Foundation. (2012). *What is Open Data?* Retrieved 12 10, 2013, from Open Data Handbook: http://opendatahandbook.org/en/what-is-open-data/

Open Meteo Foundation. (2013). *Open Meteo Data.* Retrieved 12 10, 2013, from Open Meteo Foundation: http://openmeteofoundation.org

OpenStreetMap.org. (2013). *Open Street Map.* Retrieved 12 10, 2013

OpenWeatherMap. (2013). *Free weather data and forecast API.* Retrieved 12 10, 2013, from Open Weather Map: http://openweathermap.org

United Nations. (2013). *Disaster Management Cycle*. Retrieved 12 10, 2013, from United Nation Platform for Space-based Information for Disaster Management and Emergency Response: http://www.un-spider.org/glossary/term/6016

US NOAA. (2013). *Climate Data Online: Dataset Discovery.* Retrieved 12 10, 2013, from National Climatic Data Center: http://www.ncdc.noaa.gov/cdo-web/datasets

About the Authors

Vassilios Vescoukis - Assistant Professor of geo-spatial Information Systems at the National Technical University of Athens. He has been head researcher in national projects on the specification, design and development of Natural Hazards Management Systems for wildfires in Greece. His research interests include distributed and heterogeneous software systems, integrated simulation and combinatorial optimization in operational problems focused on environmental crisis management, web science, as well as multi-dimensional web analytics.

Charalampos Bratsas – Dr Charalampos Bratsas is the founder of the Greek Chapter of the Open Knowledge Foundation and the President of its Administration Board. He participates as a lecturer in MSc in Web Science at AUTH. His main areas of interest are Open Data, Semantic Web, Social Semantic Web, Linked Data (Coordinator of Greek Dbpedia), Data Mining and algorithms, Big Data, Applications of Artificial Intelligence. He also is a post-doctoral research associate in the Complex Systems Analysis Laboratory at Mathematics Department of AUTH.

Copyright information