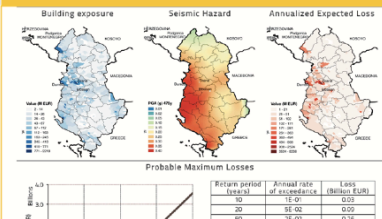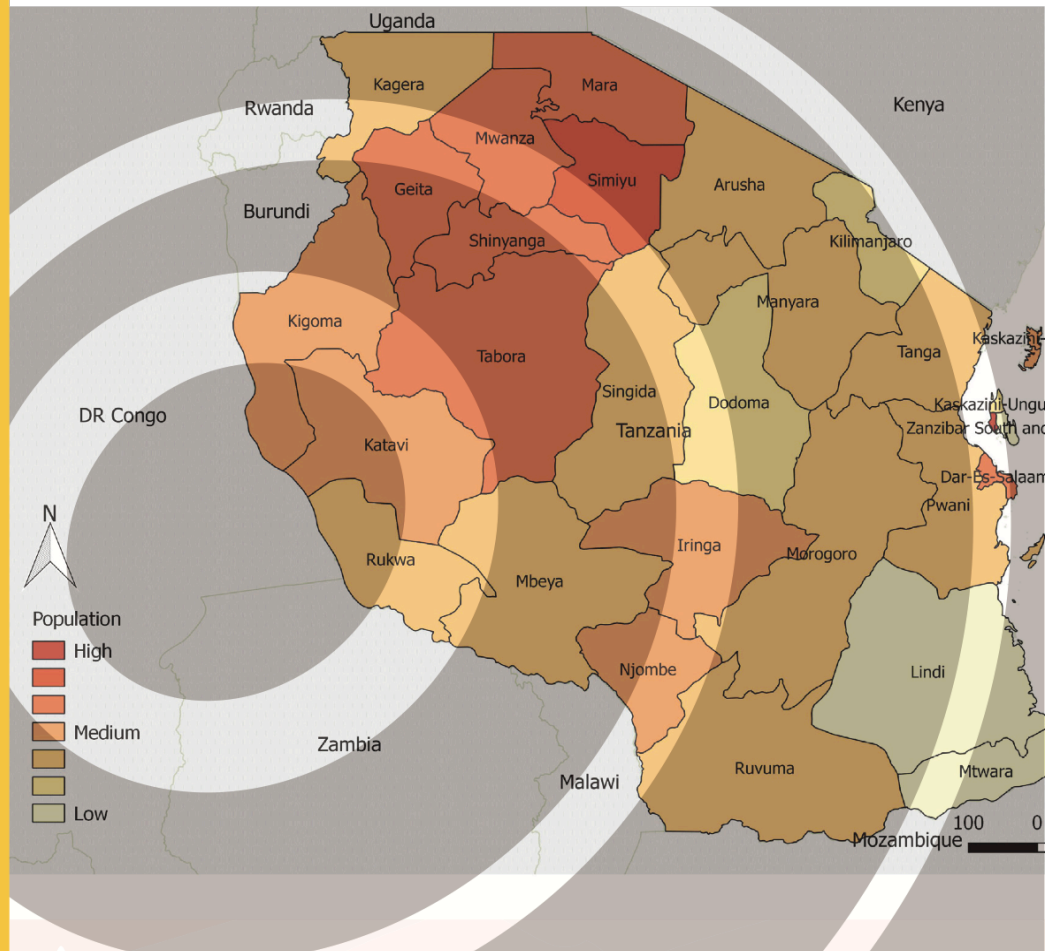# GED4ALL

Populating Global Exposure Database for All (GED4ALL) with existing databases

O'Hara M, Silva V, Henshaw P, Huyck C.

SUBMITTED BY:

The Global Earthquake
Model Foundation (GEM)
Via Ferrata 1, 27100
Italy

GED4ALL

Global Exposure Database for Multi-Hazard Risk Analysis

D3 - Populating GED4ALL with existing databases

Authors: O'Hara M, Silva V, Henshaw P, Huyck C
Date: March 2018

Abstract

The Global Facility for Disaster Reduction and Recovery (GFDRR) and the UK Department for International Development (DFID) Challenge Fund, aims to decrease disaster risk management costs and increase resilience by developing a framework that facilitates a multi-hazard view of risk. The Challenge Fund initiative is part of this effort, with the second round focusing on the creation of three open databases: hazard footprints, exposure data and vulnerability models. The Global Earthquake Model (GEM), ImageCat and the Humanitarian OpenStreetMap Team (HOT), partnered together to develop a 'Global Exposure Database for Multi-Hazard Risk Analysis'.

This project will result in an open exposure database (GED4ALL) for multi-hazard risk analysis, along with protocols to integrate existing datasets, as well as generate new exposure information. In order to achieve these goals, this consortium has proposed five main working packages, focused on the IT structure of the database, creation of a multi-hazard taxonomy, methods for the integration of existing datasets, generation of exposure information for the pilot country, Tanzania, and development of an outreach campaign. This report covers how the GED4ALL can be populated with existing databases, including datasets from GEM, GFDRR, OpenStreetMap (OSM), Landscan, GRUMP and WorldPop.

Close collaboration with the other Challenge Fund partners tasked with developing the 'Hazard Footprints' (British Geological Survey) and 'Vulnerability Models' (University College of London) databases, was employed to ensure compatibility, as well as interoperability between the three databases.

Keywords:

Exposure data, GED4GEM, GFDRR datasets, OpenStreetMap, exposure modeling, databases.

Table of Contents

List of Figures

List of Tables

Introduction

As part of the Global Facility for Disaster Reduction and Recovery (GFDRR) and the UK Department for International Development (DFID) Challenge Fund project, a Data Schema (see Deliverable 1) and Multi-Hazard Taxonomy (see Deliverable 2) to store exposure data have been developed. It is thus fundamental to understand how users can import exposure data (e.g. buildings, infrastructure, lifelines, crops, forestry, livestock, socio-economic data) into GED4ALL in a uniform and harmonized manner.

This deliverable describes a number of existing open and accessible exposure datasets that can be used to populate GED4ALL. Some of these datasets have been developed within regional programmes supported by the Global Earthquake Model Foundation or the Global Facility for Disaster Reduction and Recovery, and include information about the number of buildings, critical facilities, road network, replacement cost, main structural characteristics, and occupants and different times of the day. Other relevant sources of data include the OpenStreetMap initiative, which features building-level data, and WorldPop, which contains population data for different time periods at a 100x100 m^2 spatial resolution.

Tools that support the importing process are also described in this deliverable. These tools were either developed as part of this project, or already existed due to complementary activities led by GEM and the Humanitarian OpenStreetMap Team. Due to the open-source nature of these tools and the transparency behind the data formats, these resources can also be incorporated within other platforms to facilitate the importing of existing data into GED4ALL.

This deliverable also highlights the main steps required to scale GED4ALL at the global level.

Existing Databases

The Data Schema and Multi-Hazard Taxonomy have been developed to be compatible with the most common open databases of exposure information. This section describes some of these sources of information, while the following section explains how these data can be imported into GED4ALL.

Global Earthquake Model datasets

As part of its main goal to develop a global earthquake hazard and risk model, GEM and its partners have been developing exposure datasets for more than 110 countries. Such a model describes the location of physical assets, their replacement costs, estimated number of occupants at various times of the day and the vulnerability categorization of each asset. Exposure data that describes the type of structure, their location, height or occupancy characteristics are often unavailable. Even when some information is gathered, the fact that the built environment changes rapidly in some parts of the world (e.g. South-east Asia, Central America) poses additional challenges.

Generally, GEM covers three types of building occupancies: residential, commercial and industrial. The derivation method for the first category differs considerably (for some countries) in comparison with the steps required for the commercial and industrial building stock. The development of the global residential exposure model relies predominantly on data from the national housing census of each country. These surveys are usually performed every 10 years (sometimes with an update after 5 years – e.g. Mexico) with varying geographic resolution, often relying on statistical summaries at the smallest administrative level, or an even smaller division that are defined specifically for such purposes (e.g. United States, Canada). The quality of the data collected by each country, unfortunately, varies greatly. In the best-case scenario, the survey data comprises information concerning the number of buildings, type of structures (e.g. isolated houses, apartment buildings), main material of construction, material of the roofs, material of the floors, number of stories, year of construction and even the state of conservation of the building (e.g. Portugal, Costa Rica). For many nations (e.g. Algeria, Iraq, Zambia, Mali), the survey data is quite limited and it only provides information on type of dwelling and the main material type (e.g. wood, brick, concrete) used for the construction of walls. In these cases, it is necessary to develop a mapping scheme, which establishes a relation between the variables used by the national housing census, and the most likely vulnerability classes. For example, in Peru when the structure is classified as an isolated house with walls made of adobe, the mapping scheme can dictate that 75% of such structures will be classified as having 1 storey while the remaining will be assigned to 2 storey adobe structures. The definition of these mapping schemes is performed using peer-reviewed literature, World Housing Encyclopedia (WHE) reports, and the judgment of local experts, collected using remote expert surveys.

In cases where the housing census is not appropriate for seismic exposure and vulnerability modelling, as in most Central Asian countries, rapid screening activities are often conducted in selected locations and integrated by ancillary information derived from satellite imagery and from volunteered geoinformation (e.g. Pittore and Wieland, 2012, Wieland et al., 2015). The use of this multifaceted

taxonomy proposed by GEM, allows in these cases the development of regional exposure models and mapping schemes to accommodate the differences in building practices across countries. There are also countries that simply have no housing information available (e.g. Democratic Republic of Congo, South Sudan), or have been heavily affected by natural disasters (e.g. hurricanes, earthquakes) after the completion of the national housing census (e.g. Haiti, Nepal), with no updates. In these cases, an alternative approach that leverages population datasets, satellite imagery and OpenStreetMap data is followed.

An example of such datasets is presented in Figure 1, for ten countries located in Southeast Asia. Due to the significant differences in the extent of the administrative regions of each country (e.g. Indonesia versus Brunei), the total number of buildings was distributed in an evenly spaced grid with 0.2 decimal degrees for the sake of clarity.
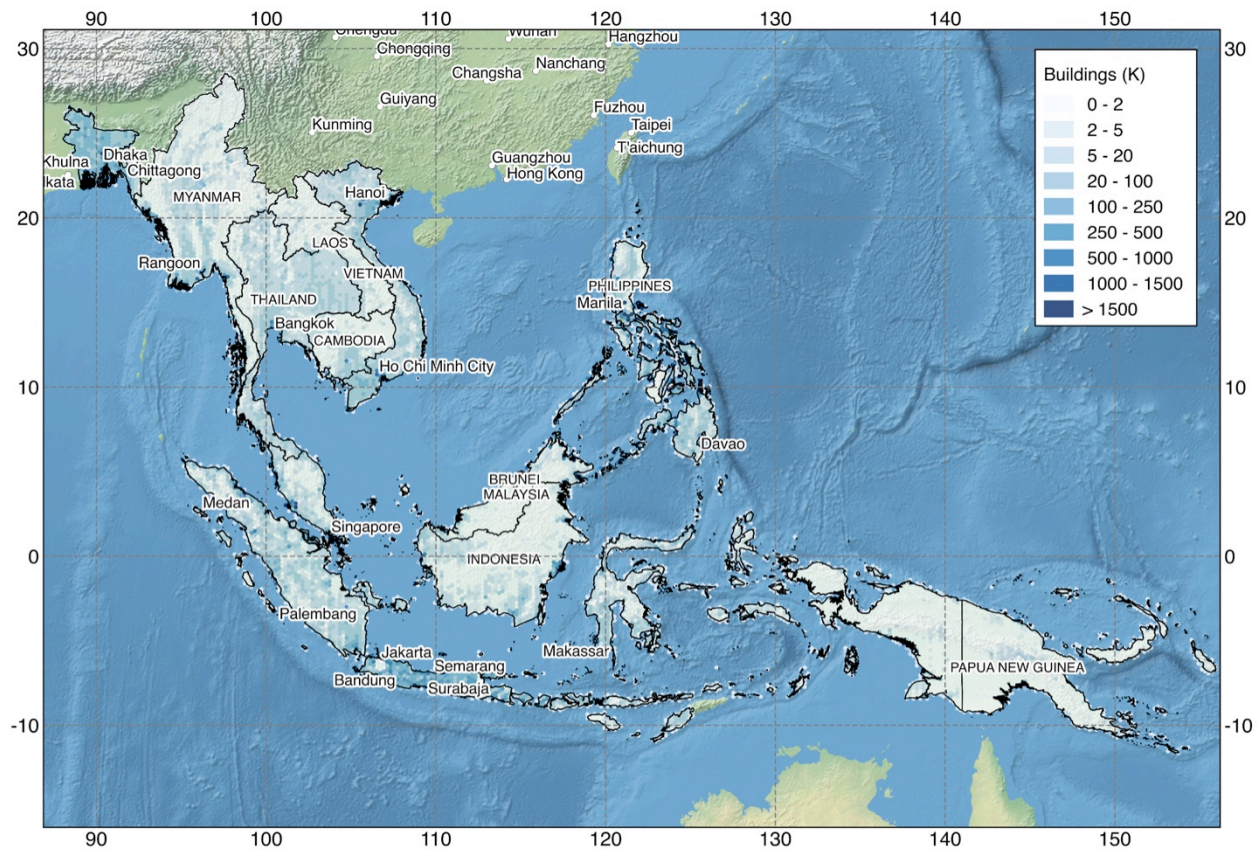
Figure 1 – Distribution of the number of residential buildings following an evenly spaced grid (0.2 decimal degrees) for ten countries in South-East Asia.

Global Facility for Disaster Reduction and Recovery datasets

As part of initiatives supported by GFDRR to evaluation disaster risk, several exposure datasets have been produced for different parts of the world (e.g. Pacific Islands, Central Asia, East Sub-Saharan Africa, Eastern Europe). These datasets usually cover the residential, commercial, industrial, healthcare and

educational building stock. Such datasets can also be used to populate GED4ALL, thus contributing significantly to the scaling of the database to the global scale. This section briefly describes the characteristics of an exposure dataset developed by ImageCAT within the scope of the Africa R5 project. Some of the datasets produced within this project have already been introduced into GED4ALL, using the methodology and tools described in the following sections.

The R5 project involved the development of a number of datasets that represent estimations of population, economic, and physical exposure for sub-Saharan African nations in 2 Phases. Developed as part of the Africa Disaster Risk Financing program of the World Bank, these data were used to model the losses from a number of natural physical phenomena (earthquake, flood, drought, landslide and volcanic activity) in Sub-Saharan Africa.

For Phase 1 countries (Ethiopia, Kenya, Niger, Senegal and Uganda), exposure data were generated for buildings and critical facilities (education, healthcare and power generation plants). A number of methods were used to generate each type of exposure data. Buildings were generated through the definition of rural and urban areas, derived from earth observation imagery. Within these areas, land classifications were used to generate mapping schemes – statistical distributions of construction practices and use (see Deliverable 4 for additional information). The final datasets were produced at a spatial resolution of 15 arc seconds. Critical facilities datasets were produced from a combination of public and open data sources, through the latitude-longitude spatial reference at the building or Admin 1 level centroids, depending on the certainty of the sites. Transportation networks for road and rail were generated from existing data sources and refined through manual interpretation of EO imagery and other data sources. Replacement cost estimations were also provided for buildings, critical facilities and transportation data based on expert judgment and a number of scientific publications. Population and GDP estimates were then developed for 2010 and 2050, using WorldPop and Shared Socioeconomic Pathway 2 projections.

The exposure data was developed in collaboration with four risk modelling companies, as well as the project leads at GFDRR. Local expert opinion was used to validate many of the assumptions made in the data generation process. A description of the datasets that are now available for each African country covered in this project is provided in Table 1.

Table 1 - Description of the datasets covered within the R5 project.

Dataset	Description
Building exposure	Building count and cost estimates at the 15" arc second level based on the "development pattern" methodology.
Building exposure breakdown	Intermediate data from which final building exposure grids were aggregated. Each record of this dataset contains construction type, count, area, and building value per a 15" arc second grid cell. This format

	is more suitable for risk modeling.
Educational building stock (with known location)	Educational facilities that have a given or known latitude and longitude at a building or campus centroid.
Educational building stock (with modelled location)	Model educational facilities locations at the 15" arc second level based on inferred human settlement locations and density. Created to fill in the number of educational facilities at the administrative boundary 2 recorded by the local census group.
Healthcare building stock (with known location)	Healthcare facilities that have a given or known latitude and longitude at a building or health site centroid.
Healthcare building stock (with modelled location)	Model healthcare facilities locations at the 15" arc second level based on inferred human settlement locations and density. Created to fill in the number of healthcare facilities at the administrative boundary 2 recorded by the local census group.

As an example, a brief summary of building count and total replacement cost is provided in Table 2 for Ethiopia.

Table 2 – Building count and total replacement cost for Ethiopia.

| Administrative Level 1	Total Building Count	Total Replacement (USD)
Addis Ababa	565,090	32,161,919,068
Afar	318,018	1,742,850,316
Amhara	3,868,695	29,518,476,967
Beneshangul Gumu	156,471	557,625,474
Dire Dawa	78,008	1,554,529,249
Gambela	76,999	212,330,056
Hareri	42,336	776,512,319
Oromia	6,321,326	56,296,434,093
SNNPR	3,387,202	21,814,497,487
Somali	1,027,765	5,206,771,689
Tigray	994,678	12,698,726,085
Grand Total	**16,836,588**	**162,540,672,803**

OpenStreetMap

OpenStreetMap (OSM) is a crowdsourced geospatial database of the world built largely by volunteers and professionals digitising aerial imagery, collecting attribute information on the ground and importing existing public sources of open data. Known as the 'Wikipedia' of maps, the data is freely accessible to all under the Open Database License (ODbL), meaning that it can be queried, used, manipulated, contributed to and redistributed in any form. OSM is the ideal database for humanitarian efforts and disaster management, as it is a great source of geographic baseline data for many cities around the globe, especially in countries with emerging economies that are not always on the map.

Geospatial data is stored in OSM as vectors, with three types of elements:
- o Nodes, which represent a point on the surface of the earth.
- o Ways, which are sets of nodes that can form lines or polygons.
- o Relations, which are sets of nodes, ways or other relations.

Each of these elements can have any number of key=value tags, that represents the attribute information for a given feature. For example, a post office may be represented by a way with the tags building=yes and amenity=post_office, to help identify the purpose of the building.

OSM can populate the exposure data in GED4ALL at the building level, which focuses on individual buildings available as polygons. OSM data can also be used to amalgamate datasets for other levels of exposure data in GED4ALL if needed, depending on the quality of other available datasets. Figure 2 illustrates OSM data for Kathmandu (Nepal) and Dar es Salaam (Tanzania).
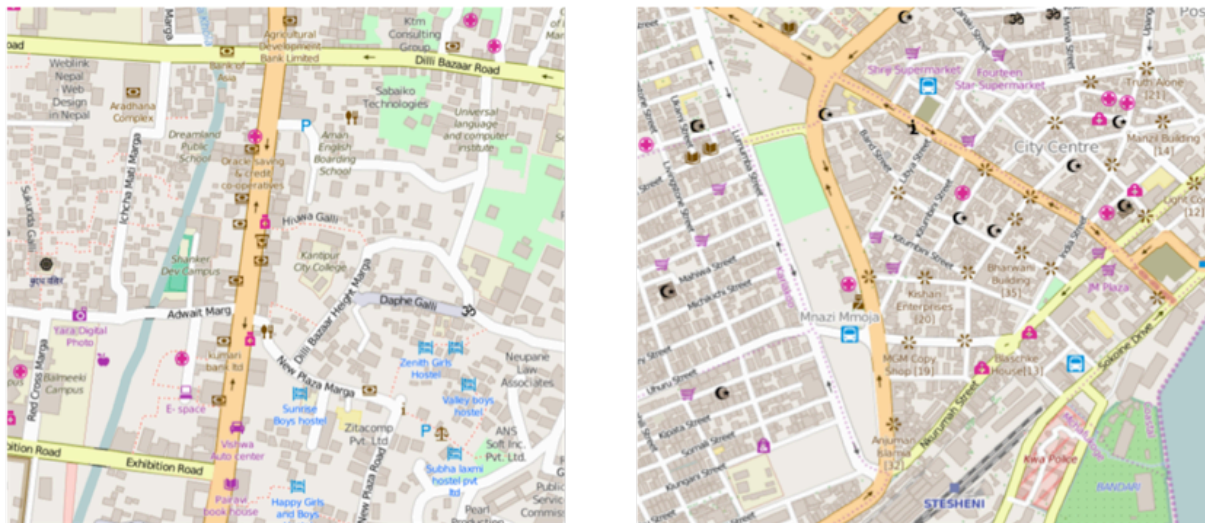
Figure 2 – Exposure data for Kathmandu (left) and Dar Es Salaam (right) from OpenStreetMap.

Aditional information is provided in the following sections about how to export data from OSM and import it into GED4ALL.

Population datasets

Currently, there are a few databases that can provide population count or density either globally (LandScan, GRUMP) or for large regions mostly composed by less developed countries (WorldPop). LandScan[1] is supported by the the Oak Ridge National Laboratory, and it has a 30 arc second spatial resolution (about one square kilometre at the equator. A great number of data sources were used to create this database such as: Digital Chart of the World, VMap1 (a map of major roads and rail networks, drainage networks, utility systems, elevation contours, coastlines, international boundaries and populated places), night-time lights, Global Land Cover Characterization (GLCC) and high-resolution aerial photography and satellite imagery. The dataset provides population estimates based on aggregate data for second order administrative units from the US Census Bureau's International Program Center. There is no specific distinction made between urban and rural areas, though urban areas can be inferred by analyzing population density (Dobson et al., 2000).

The Global Rural Urban Mapping Project (GRUMP[2]) dataset is produced by the Center for International Earth Science Information Network (CIESIN) and it also follows a 30 arc second resolution. A combination of datasets have been used to produce this database such as Census data with georeferenced human settlements data (55,000) with a population of more than 1000, Gazetteer, City Population, Digital Chart of the World, administrative boundaries datasets and United Nations (UN) national estimates. The population is assigned to the grid cells considering the classification of the areas in urban or rural. After this, checks are made on the total population in each administrative unit according to UN estimates. The urban extents are defined by a combination of datasets and not just night-time lights (recorded in a period between 1994 and 1995), that can miss small settlements in less developed countries and greatly overestimate urban extents for large settlements. The value assigned to each cell is an integer that represents the estimated population count.
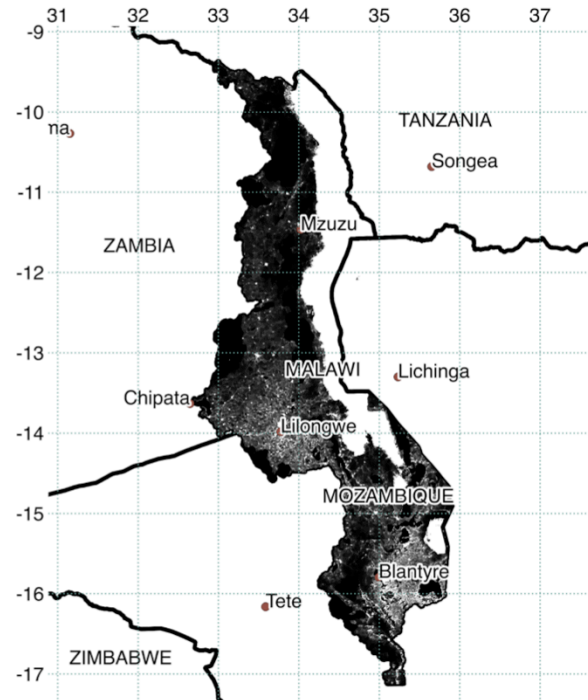
WorldPop[3] represents a recent effort to map population distribution in Africa, Asia, and Latin America at a fine resolution (100x100 m2). It uses high spatial resolution data, the methods for its generation are well documented, and the population datasets are openly accessible. Moreover, it provides estimates of population distribution throughout time, thus enabling the evaluation of trajectory of risk trends in the future. An example of the population distribution for Malawi (one of the demonstration countries covered by this project) is presented at the 100x100 m2 resolution for the year of 2010 in Figure 3.

---

[1] LandScan - https://web.ornl.gov/sci/landscan/
[2] GRUMP - http://sedac.ciesin.columbia.edu/data/collection/grump-v1
[3] WorldPop - http://www.worldpop.org.uk/

Figure 3 - Distribution of population in Malawi for 2010, at an evenly spaced grid following a 100 m resolution.

All of the aforementioned population datasets can be used to populate GED4ALL, or used to estimate the expected occupants of a particular building portfolio at different times of the day.

Populating GED4ALL

Although it is possible to populate the GED4ALL database directly using SQL INSERT statements, this is both time consuming and error prone. In most cases it is far more convenient to import data from an existing exposure model using a software tool. In this section we briefly describe some of the tools which can be used to help convert exposure data and incorporate it into GED4ALL.

Natural Hazard Risk Mark Language

As described in more detail in deliverable, *D1 - Exposure Database Schema and Complementary Tools*, the GED4ALL database is accompanied by Python programs to import and export data in the Natural Hazard Risk Mark Language (*NRML*) format. Given an exposure model in *NRML* format, this can be imported using a command of the form:

```
$ python import_exposure_nrml.py /path/to/exposure_model.xml
```

A recent development to the OpenQuake-engine and underlying libraries allows users to describe exposure models using an *xml* file for general descriptive metadata, cost types, tag name, area units along with one or more *csv* files containing the actual data values for each asset in the model. This further simplifies the import process in cases where it is possible to extract/convert existing exposure data into *csv* format using standard desktop and database tools. The import script which makes it possible to import models using this hybrid *xml+csv* approach is available on version 2.9 of the OpenQuake-engine.

This mechanism allows importing any kind of tabullar data (similar to what was previously presented for GEM,GFDRR and Populatin datasets) into GED4ALL. In addition to the import/export tools developed specifically for this project, there are a number of pre-existing software tools and services available which can be used to assist in the construction of exposure models and in converting existing models into NRML format, as described in the following section.

OpenQuake-platform Input Preparation Toolkit

As illustrated in Figure 4, the OpenQuake-platform provides an online Input Preparation Tool to assist in the construction of input models for use with the OpenQuake-engine. This includes a section for converting tabular exposure data (for example in CSV format) into NRML; this means that users with exposure data in different formats can easily construct files in *NRML* ready to be imported into GED4ALL. However, as previously explained, data can also be introduced in GED4ALL in a *csv* format.

The Input Preparation Tool is available online at the following url: https://platform.openquake.org/ipt. A locally hosted version of the IPT suitable for offline use is available as part of the OpenQuake Virtual Machine image and some installer packages, please see https://github.com/gem/oq-engine for details.

A video tutorial for the Input Preparation Toolkit is available from the following link: https://www.youtube.com/watch?v=a8NPdiv7Lf0
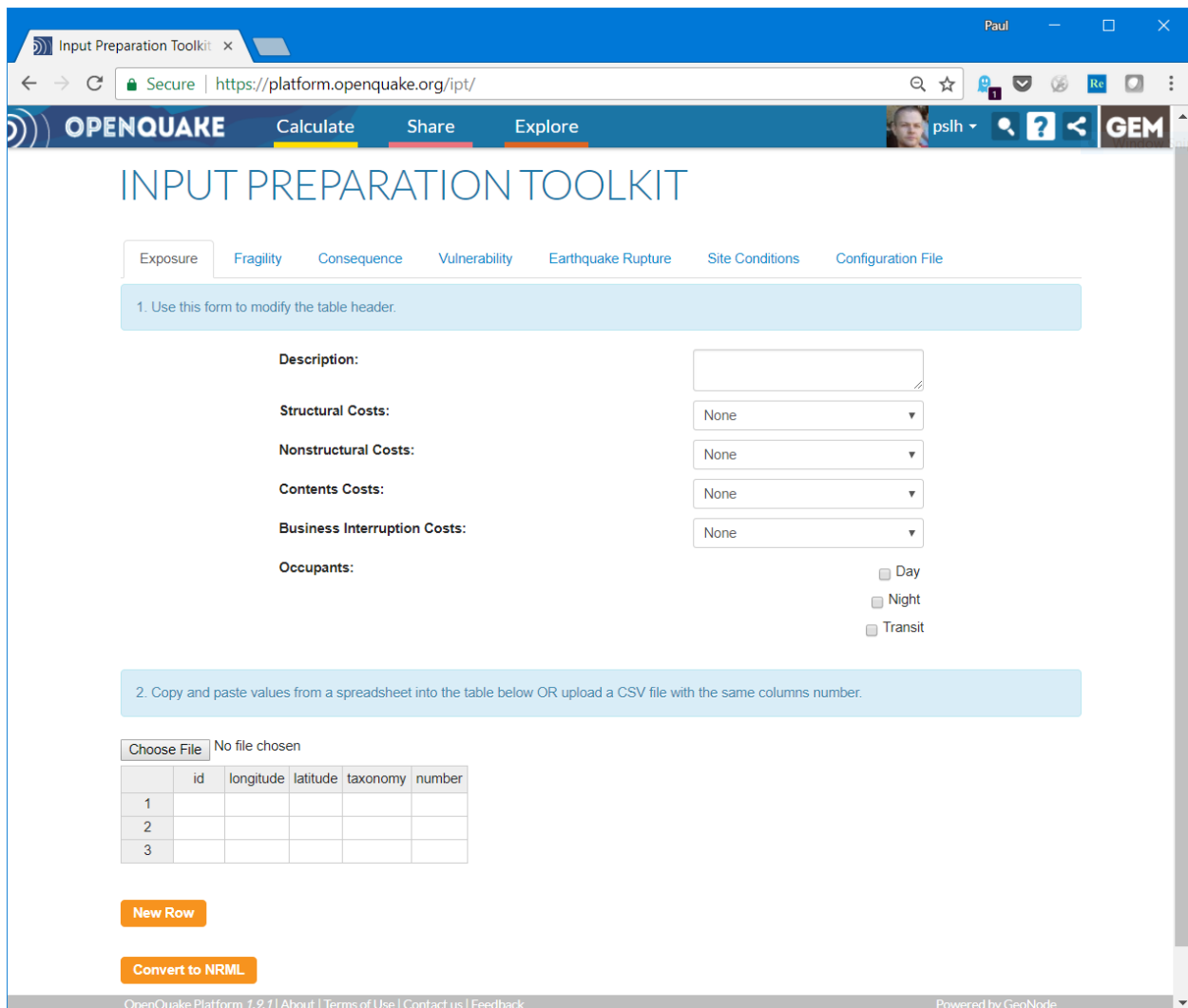

Figure 4 - Input Preparation Toolkit Exposure tool on the OpenQuake-platform.

Humanitarian OpenStreetMap Team Export Tool

GED4ALL can also support building-level data, as described within the OpenStreetMap initative. The Export Tool[4] (see Figure 5), developed by HOT, allows users to create custom OSM data extracts for anywhere in the world, simply by selecting an area of interest, the map features and finally, the file formats to convert the data into. Within minutes, up-to-date OSM data is exported, filtered and converted for use. These data can then be matched to the GED4ALL taxonomy, and imported into GED4ALL. This section describes this process.

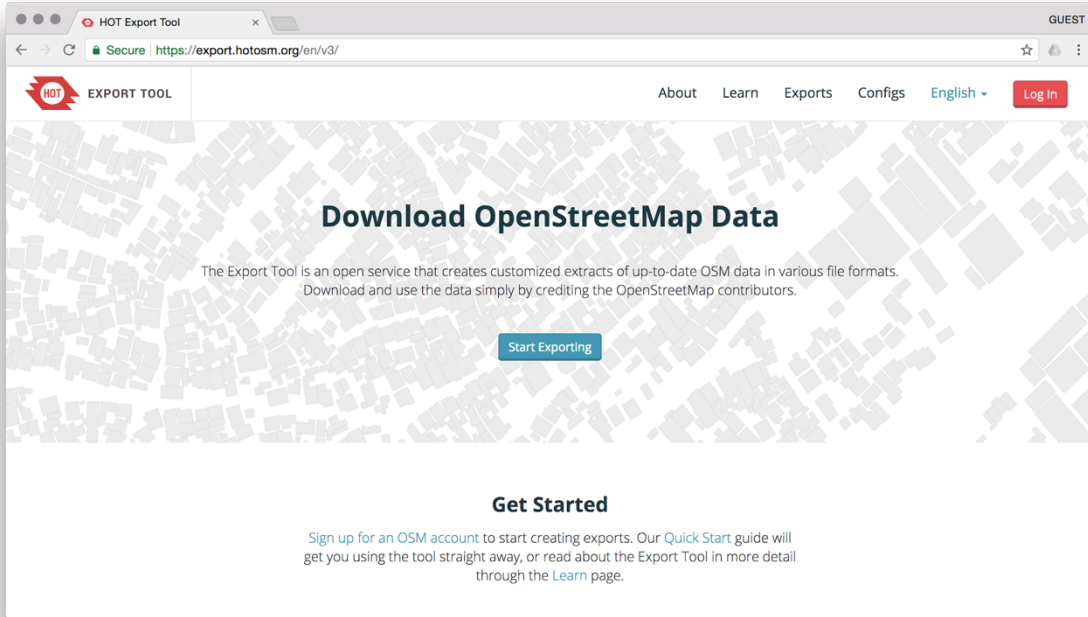[4] Export tool - http://export.hotosm.org

Figure 5 - Export Tool: landing page.

The **create page** is where export jobs are configured (see Figure 5). This section is split into four main parts, the **describe**, **formats**, **data** and **summary** tab. The describe tab is where the user will provide a name, description, project title and select an area of interest. The formats tab is where the user selects the file formats they wish to convert the exported data to. The data tab is where the user selects what map features they would like export and the summary tab provides the users with an overview of the settings and the choice to 'Publish this Export'.
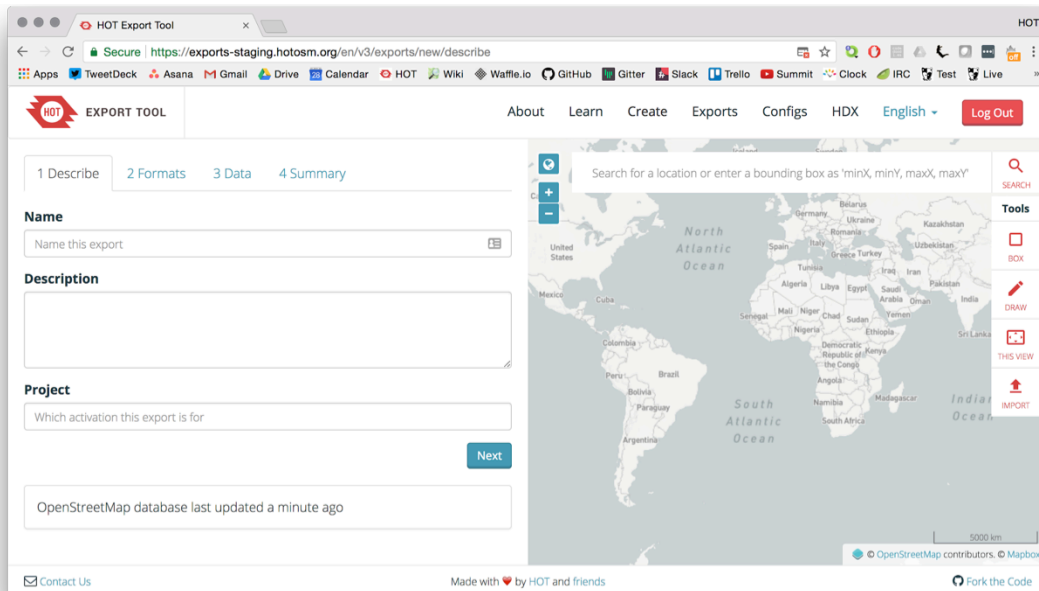

Figure 6 - Export Tool: create page.

Describe Export

The Describe tab is where you will setup your export, by describing it on the left hand side and selecting your area of interest (AOI) with the map on the right hand side. There are 6 ways to define an AOI for your export. The first and second, use the search bar, which can either find a place by typing in its name and selecting it from the provided options, or by defining the bounding box coordinates. The minimum X, minimum Y, maximum X and maximum Y (West, South, East, North) coordinates of a country can be found in a csv list[5] provided by the Humanitarian Data Exchange (HDX). The third and fourth way that an AOI can be selected in the Export Tool, is by either drawing a bounding box by selecting the 'Box' option from the column of Tools on the right of the screen. Or you can free draw a polygon directly on the map by selecting the 'Draw' tool. The last two ways in which an AOI can be selected is through the 'This View' tool on the right hand side of the map, which captures the extent of the map's current view, or by the 'Import' tool which enables you to upload a polygon of the area.
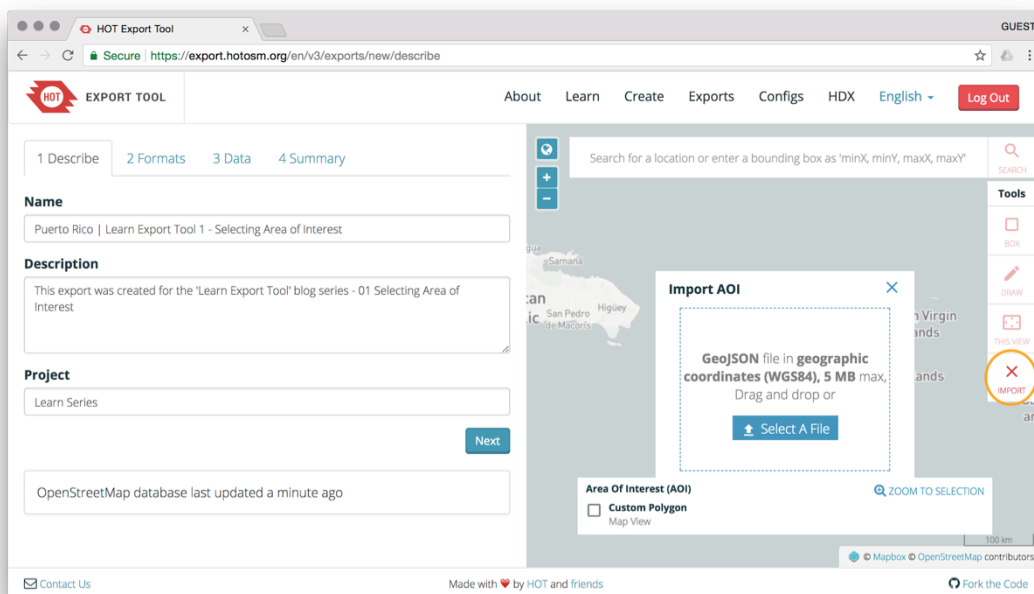

Figure 7 - Export tool: load polygon.

File Formats

The tool allows OSM data to be extracted through the Overpass API in its native Protocol Buffer Binary (PBF) file format, before filtering the data for the map features and associated tags specified by the user. Once the data has been filtered it is then converted into the file formats chosen by the user. Currently, the tool can convert OSM data into Shapefile **.shp**, GeoPackage **.gpkg**, Garmin **.img**, Google Earth **.kml**, OSM **.xml**, OSM **.pbf**, MAPS.ME **.mwm**, OsmAnd **.obf**, and MBTiles **.mbtile**s. At least one of the file formats must be selected to create the export, but there is no restriction as to how many can be chosen, which can be done by ticking multiple boxes.

---

[5] csl list - https://data.humdata.org/dataset/bounding-boxes-for-countries.
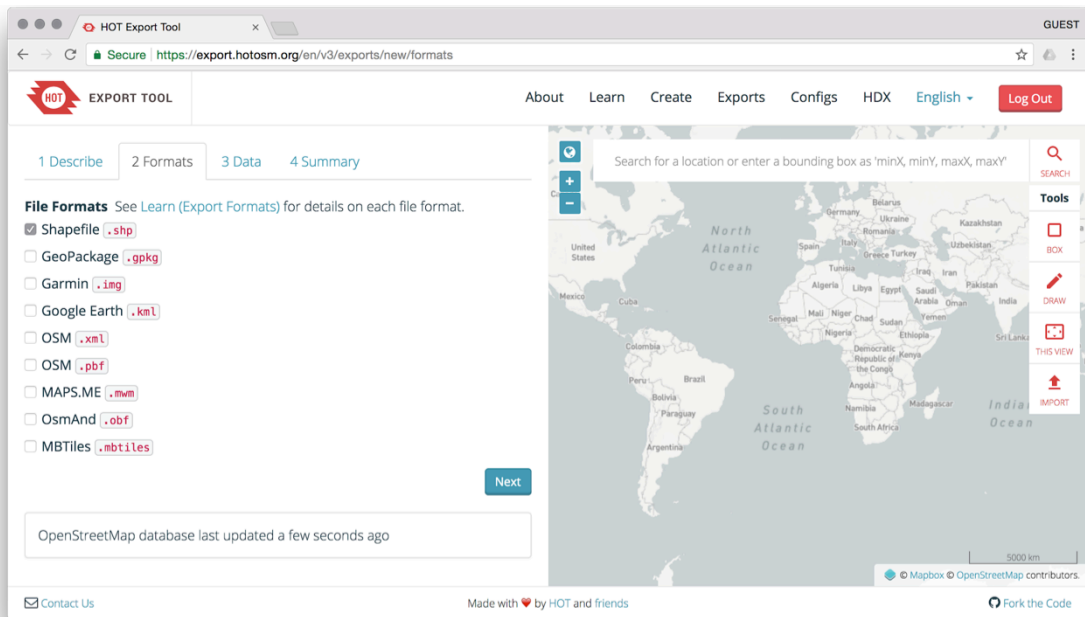
Figure 8 - Export Tool: file formats.
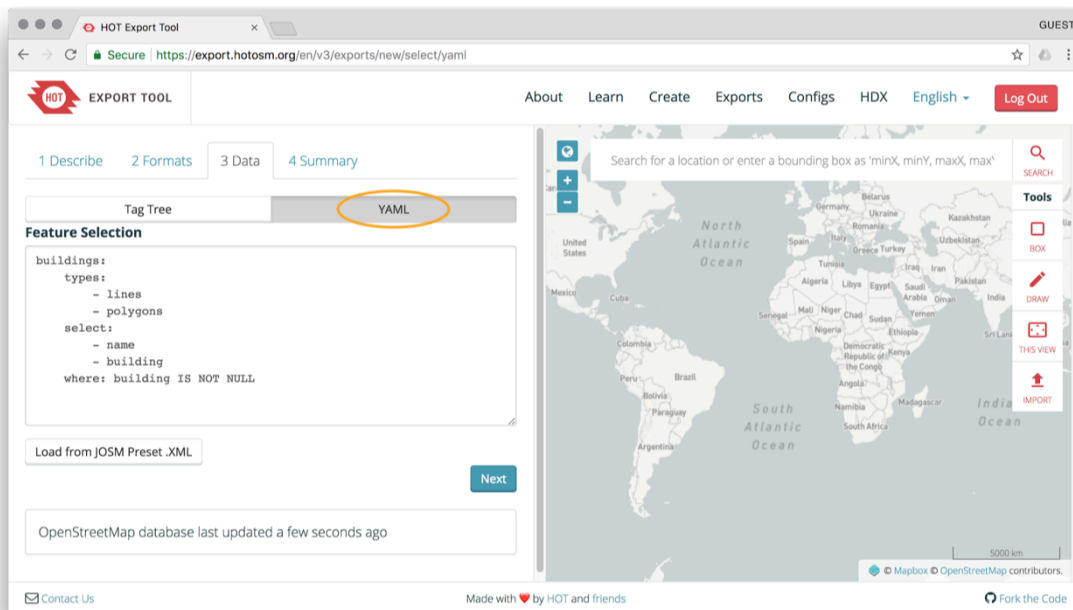
Map Features

The tool allows the user to customise the data selected within the defined area of interest. The OSM data is defined by using tag filters and key selections, with the **Tag Tree** or **YAML Form**. The Tag Tree is for common use cases, presenting a curated set of filters and selections, where the YAML format provides complete control over filters and selections, using a SQL-like filter definition.

The Tag Tree is the simplest way to get started selecting features, simply by ticking the desired parent of child checkboxes. Please note that selecting a parent checkbox will automatically select all the child checkboxes below it, as well as additional key=value tags within the theme. Each parent checkbox has a different query applied to it to filter data, so it is highly recommended that the syntax for each is explored by hovering over the checkbox which will provide an info box to the right.
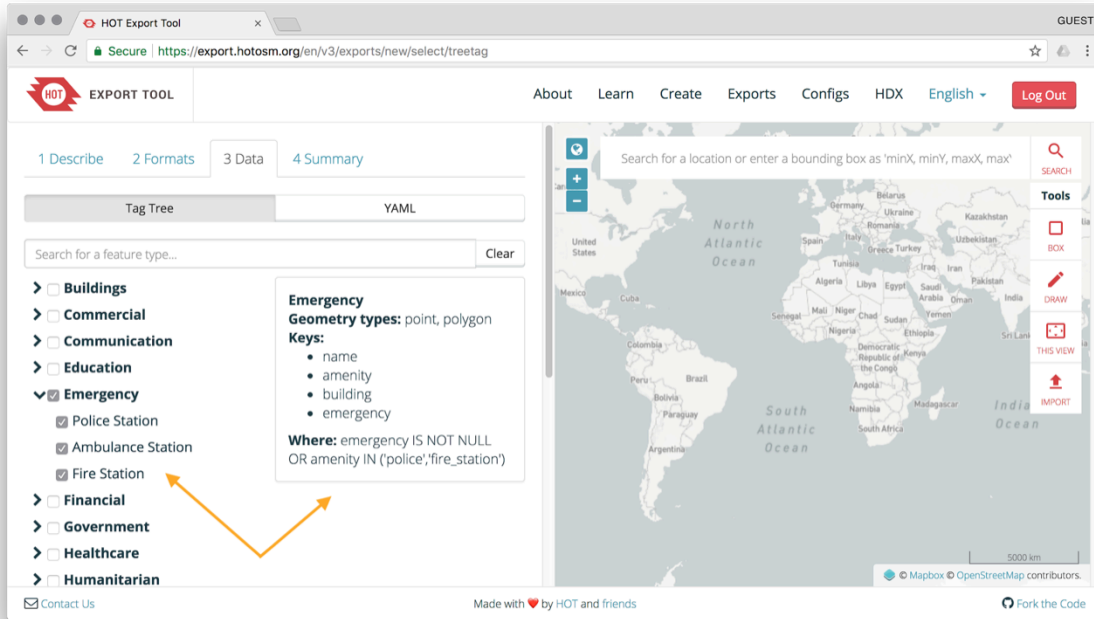
Figure 9 - Export Tool: Tag Tree Feature Selection.

Using the YAML format provides complete control over applying filters to the OSM data, by using a SQL-like filter definition to apply specific key=value tags. Please note that the Tag Tree also generates syntax on the YAML Form, so any parent and child checkboxes selected will automatically be applied to the YAML box. This acts as a starting point for the query which can be modified. The use of YAML was chosen due to its simplicity and compatibility with SQL. The YAML feature selection format is similar to style files used by programs such as osm2pgsql and imposm.

Figure 10 - Export Tool: YAML Feature Selection.

Exports

The **exports page** lists all the exports created by the user by default, with the option to show all published exports (Figure 43). Selecting an export from the list directs the user to the details of the export, where the features selected can be viewed in yaml and the option to **re-run** or **clone** the export can be executed.
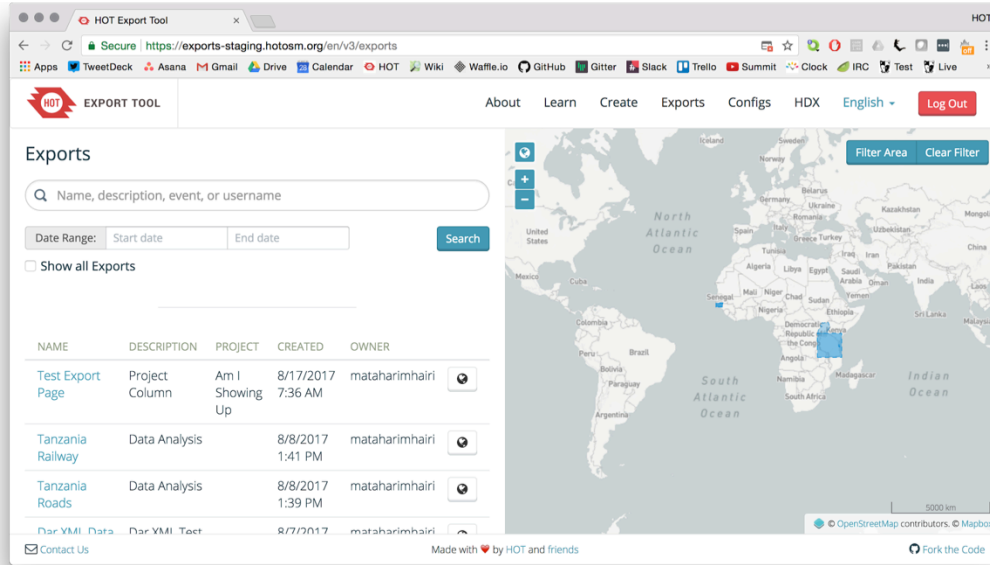

Figure 11 - Export Tool: Export Page

Configs

The **config page** lists all the configurations created by the user by default, with the option to show all published configurations. Selecting a configuration from the list shows the feature selection made in YAML, which can be applied to other exports.
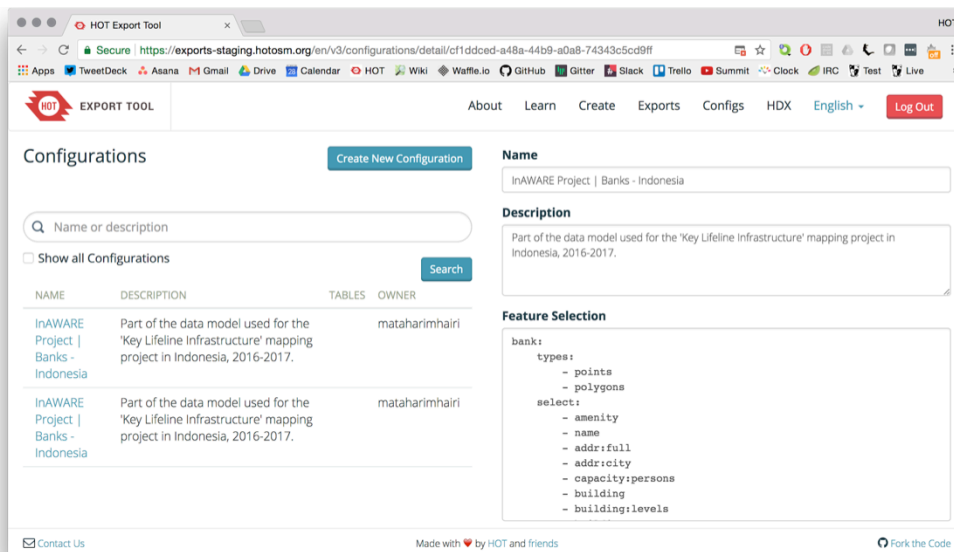

Figure 12 - Export Tool: Config Page.

Converting OSM data to GED4ALL-compatible format

Based on the taxonomy attributes identified in *Deliverable 2 - Multi-Hazard Exposure Taxonomy* report, tags were reviewed using the OpenStreetMap Wiki[6], a website that acts as the main repository of OSM information and provides steps on how data should be mapped. The second resource used to help identify suitable OSM tags. TagInfo[7], a website that provides information on OSM tags, also known as the key, and additional insight into their values, commonly used combinations, geographical dispersion, similar keys and projects they are used in.

It should be kept in mind that although OSM tags have been suggested for the use of compatibility with the GED4ALL taxonomy, based on the most suitable and widely used tags across the globe, as OSM is an open database, there is also a degree of informal tagging. This means that although the community tries to agree on certain key=value combinations for the most commonly used tags, not every individual that maps may follow these, and the tags used for the same feature can vary from country to country, and region to region. The OSM project is a great source of geospatial data for many places, but there is still some work to be done to obtain consistency of tags across borders.

The best practice to consider when populating the GED4ALL database with OSM data is to do further investigation for the appropriate tags within the area of interest. This will ensure that the data being extracted contains the correct information required for the taxonomy. The development of the GED4ALL taxonomy, however, is a great starting point to providing the OSM community with a data model that can be used to collect the right attribute information for carrying out multi-hazard risk analysis, which can also help to create consistency across the globe in the use of OSM tags for the structural integrity of buildings, roads, railways and other lifelines.

---

[6] http://wiki.openstreetmap.org/wiki/Main_Page
[7] https://taginfo.openstreetmap.org/

Global Coverage

With the completion of the data schema, classification system (taxonomy) of the elements exposed to the hazard and infrastructure to host and maintain GED4ALL, it is now important to discuss how the outcomes of this project can be scaled globally. In other words, what can be done to ensure that within a reasonable period of time the following goals are achieved:

- GED4ALL has exposure information (e.g. building stock, infrastructure, critical facilities, crops) at the national level for all countries.
- GED4ALL is the preferred data repository for national governments, international organizations and large-scale programmes.
- GED4ALL is compatible with the most common and relevant natural hazards risk assessment tools, and it is frequently used in disaster risk estimation.
- GED4ALL is frequently used in disaster risk reduction initiatives, and supports decision makers in disaster risk management.

In order to reach global coverage in the future, it is critical to leverage on the achievements from the past. The previous sections of this document have described how data from existing sources (e.g. GED4GEM, GFDRR, WorldPop, OpenStreetMap) can be used to populate GED4ALL. To maximize these valuable data, additional funding is required to port all the data into GED4ALL. Additional work is also necessary in order to ensure that the licenses of such data are suitable for an openly accessible database. An example of these existing data (GEM datasets) is presented in Figure 13 for 12 countries in the Middle East.
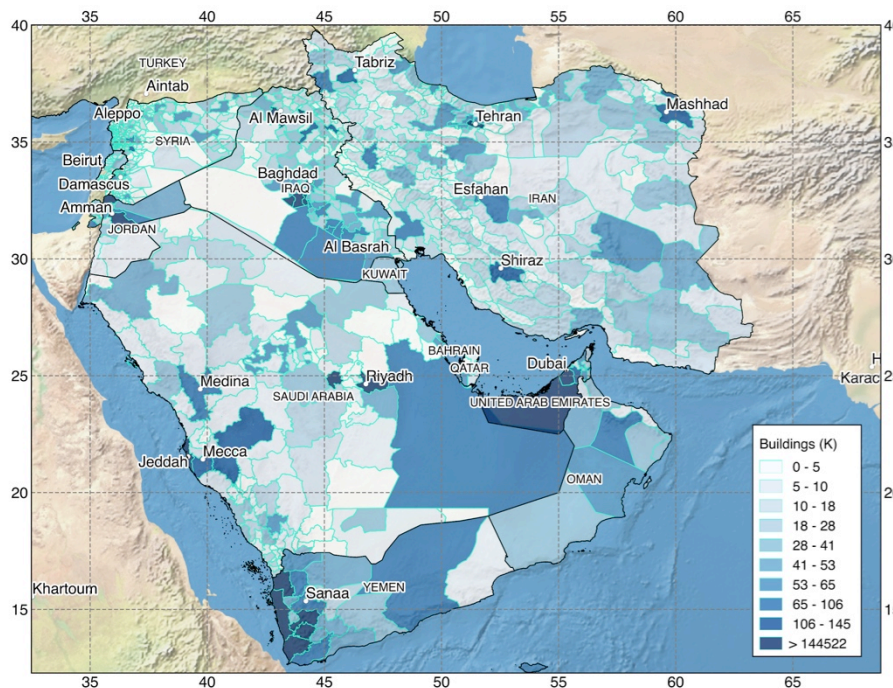
Figure 13 - Distribution of the number of residential buildings at the smallest available administrative subdivision for twelve countries in the Middle East.

Large initiatives to collect, process and develop exposure data might also represent a critical mechanism to reach global coverage. Every year, international organizations such as the European Commission, United Nations, Japan International Cooperation Agency, United States Agency for International Development and the Global Facility for Disaster Risk Recovery support projects which feature the development of comprehensive exposure datasets which can be used for multi-hazard risk analysis. Unfortunately, in the vast majority of the cases there is no obligation to store the data in a public repository that would allow other individuals or organizations to explore the data for disaster risk assessment and reduction. As a consequence, these data are often kept privately by the consultants, or made available through dedicated web portals which eventually become extinct, and therefore lost.

The aforementioned organizations could state as part of the terms of reference of future projects the release of all relevant data through GED4ALL. At a national scale, it is also worth mentioning that in less developed countries, even if there is a will to share exposure data, there might not exist a proper mechanism to do so. Once again, the adoption of GED4ALL could facilitate the dissemination of data, and consequently reduce the gap between data collectors and experts with the remit to evaluate disaster risk. These decisions should go beyond the treatment of the exposure data, and also include loss, hazard and vulnerability datasets.

Populating GED4ALL will inevitably highlight regions in the world where either data simply does not exist, or the current quality and reliability is insufficient to support the development of disaster risk reduction measures. From previous experiences from this consortium, countries such as South Sudan, Somalia, Democratic Republic of Congo, Haiti, Suriname or Turkmenistan do not have up-to-date housing census or reliable data from the national statistical offices that could be directly employed in the assessment of disaster risk. In these cases, it will be necessary to support campaigns to develop new exposure datasets at the local, regional or national level, possibly using the approaches described in Deliverable 4 of this project. Such new datasets should be stored in GED4ALL, thus ensuring that a wider audience will be able to explore them.

Finally, this consortium also believes that in order to convince national governments, smaller organizations, and even individuals to contribute with their exposure data, GED4ALL needs to demonstrate clear benefits, besides the obvious ability to make the data available to wider the community. One of these advantages could be connecting GED4ALL to powerful, but intuitive, tools for the identification of disaster risk (e.g. GeoSafe, OpenQuake, OASIS). For example, a local government might be interested in storing its dataset of healthcare facilities, in order to assess which ones might be affected by a potential flood scenario. Connecting GED4ALL to such tools would empower significantly institutions and individuals who might not have the resources (or interest) to develop their own disaster risk assessment platform.

Final Remarks

This deliverable described some of the most relevant open and available exposure datasets (GEM and GFDRR datasets, WorldPop, OpenStreetMap Data) that can be used to populate GED4ALL. These datasets have either been developed using modeling techniques (e.g. housing census data, satellite imagery) or based on field missions or mapping exercises. The former category can be used to populate GED4ALL at level 1 and 2 (i.e. administrative levels or evenly spaced grid), whilst the latter can be used at level 3 (i.e. building-level data).

A process to import these data into GED4ALL has been described, which relies on a Python program. This program processes data in the *NRML* or *csv* formats, which can be prepared using the OpenQuake-platform Input Preparation Toolkit. For the OSM data, an additional process is required to first export the data using a recently released tool developed by the Humanitarian OpenStreetMap Team. It is relevant to note that this Python Program can also be used by linked to other graphical user interfaces to facilitate the importing of additional data (for example, the one currently under development by GeoSolutions as part of a GFDRR project).

Although GED4ALL does not require the classification of the assets using the *Multi-Hazard Taxonomy* developed within this project (see Deliverable 2), it is highly recommended to convert the classification system employed in the existing datasets to this one. This will allow leveraging on other datasets (e.g. damage/loss data, vulnerability functions as part of Challenge Fund 3) that will use either precisely this taxonomy, or a similar version (e.g. GEM Building Taxonomy v2.0). The process to convert existing classification systems to the GED4ALL taxonomy required a clear understanding of the meaning of the terms previously used. The GED4ALL taxonomy and the newly proposed OSM tags should be formally introduced to the OSM community, in order to gain acceptance. There is a Formal Procedure[8] that can be followed, which is outlined on the OSM wiki. This would be the first advisable step to pursue. Once accepted, a wiki page for the taxonomy can be created so that any individual looking to map and collect detailed attribute information towards hazard assessment, can use GED4ALL as a point of reference. Ideally, the data model would not just be used by HOT for their mapping projects, but by multiple organizations across the globe. Workshops and conferences are also another great way to share the taxonomy with the OSM community, and should be employed wherever possible.

This deliverable also describes, in the opinion of this consortium, the main steps required to scale GED4ALL at the global level. To summarize, the following actions are recommended: 1) additional resources to import exposure data from past initiatives into GED4ALL, 2) a global effort to enforce future projects on disaster risk assessment to make eventual exposure models available through GED4ALL, 3) organization of programs at local, regional and national scale to model or collect exposure data in regions where the coverage is still insufficient, and 4) link GED4ALL to powerful, but yet user friendly, platform to assess disaster risk.

---

[8] Formal OSM request procedure: http://wiki.openstreetmap.org/wiki/Proposal_process