GED4ALL

Global Exposure Database for Multi-Hazard Risk Analysis

D5 - FINAL REPORT

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

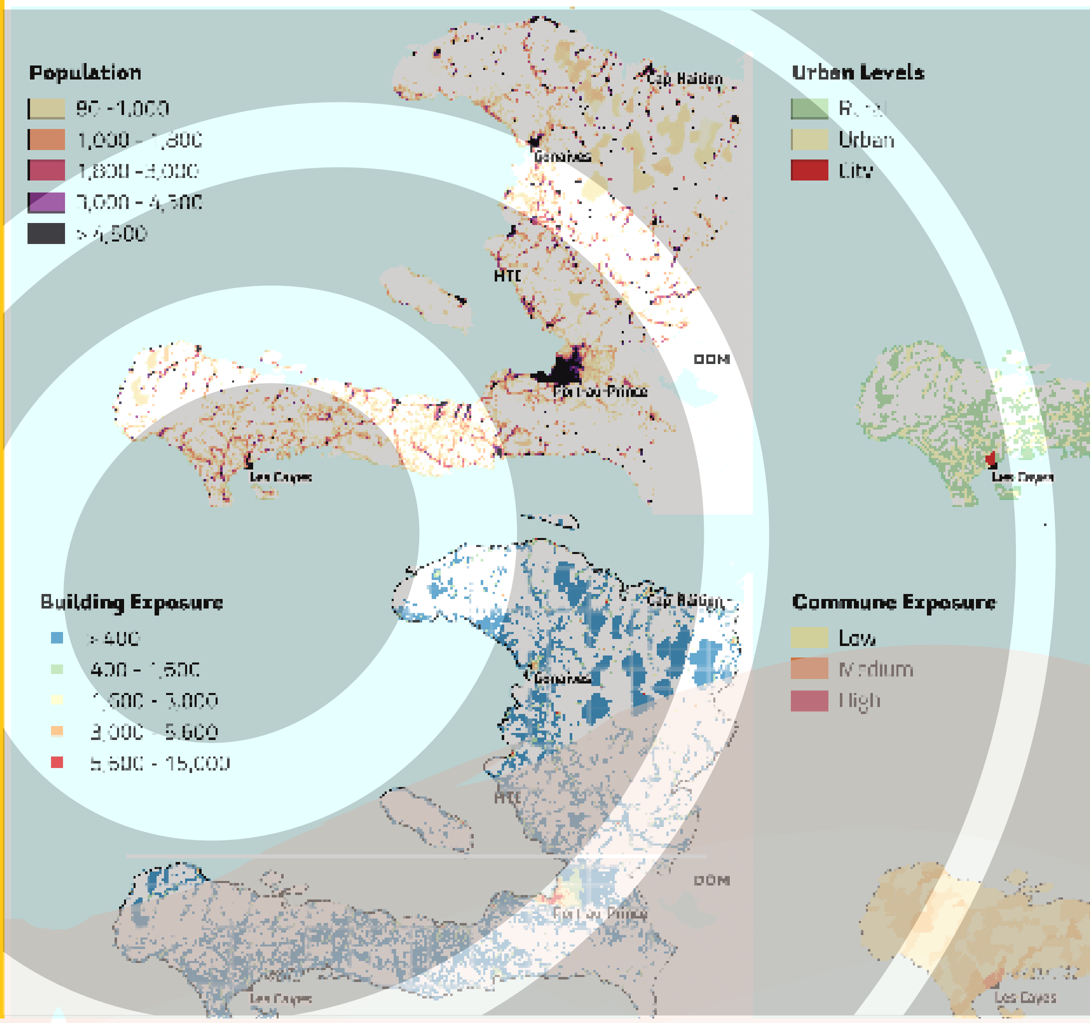
GEM GLOBAL EARTHQUAKE MODEL working together to assess risk

ImageCat

HOT Humanitarian OpenStreetMap Team

SUBMITTED BY:

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



GED4ALL

Global Exposure Database for Multi-Hazard Risk Analysis

D5 - Final Report

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

Citation: Silva V, Henshaw P, Huyck C, O’Hara M (2018). D5 - Final Report. GEM Technical Report 2018-05, GEM Foundation, Pavia, Italy.

Abstract

A consortium comprised of the Global Earthquake Model Foundation, ImageCat Inc. and the Humanitarian OpenStreetMap Team has been chosen by the Global Facility for Disaster Risk Reduction and Recovery (GFDRR) to develop an open exposure database for multi-hazard risk assessment, as part of the Challenge Funds supported by the Department for International Development of the United Kingdom. This database is capable of storing different assets (building stock, lifelines, infrastructure, crops, forestry, livestock and socio-economic data), while considering relevant attributes for six natural hazards: earthquakes, floods, volcanoes, strong winds, tsunamis and drought. The outcomes of this project were applied to the country of Tanzania to develop an exposure model at the national scale, and to the local level for Zanzibar. This process was also demonstrated for five countries around Tanzania (Ethiopia, Uganda, Kenya, Malawi and Mozambique). This final report summarizes the main outcomes of the four components that comprised this project.

Keywords:

Exposure data, buildings, infrastructure, crops, taxonomy, natural hazards.

Table of Contents

Introduction

The Global Facility for Disaster Reduction and Recovery in partnership with the Department for International dEvelopment of the United Kingdom supported a number of Challenge Fund projects to develop open schemas to store hazard, exposure and vulnerability data. The Global Earthquake Model Foundation, ImageCat Inc. and the Humanitarian OpenStreetMap Team were selected to developed the open data schema to store information about the built environment as part of Challenge Fund 2. This data schema (GED4ALL) is capable of storing information concerning buildings, critical facilities, lifelines, crops, forestry, livestock and socio-economic data in a uniform and consistent manner, and following a classification system (taxonomy) which captures a number of attributes relevant for multi-hazard risk analysis. This classification system considered six natural hazards (earthquakes, floods, volcanoes, strong winds, tsunamis and drought) for the identification of the attributes necessary to classify assets according to their vulnerability. In addition to the development of the data schema and the multi-hazard taxonomy, this project also developed tools to import and export data into GED4ALL and identified existing sources of information that can be used to increase the current coverage.

This projected was composed by four main components, whose outcomes have been documented in the following deliverables:

- **D1 - Exposure Database Schema and Complementary Tools**: Documentation describing the exposure database schema, possible complementary tools to access, extend and visualize data, and strategy to maintain and host the database in the long-term.
- **D2 - Multi-hazard Exposure Taxonomy**: Description of the comprehensive multi-hazard exposure taxonomy, and simplified version of the classification system with limited attributes. This deliverable also comprises a number of examples in which the different versions of the taxonomy are applied.
- **D3 - Populating GED4ALL with existing Datasets**: This deliverable describes a list of existing exposure datasets that can be used to populate GED4ALL, as well as tools that can be explored to export data from OpenStreetMap, and import to GED4ALL.
- **D4 - Development of Exposure Datasets**: This deliverable provides a comprehensive description of how exposure datasets can be derived using satellite data, National Census information, and campaigns to collect field data, amongst other approaches. This report also provides a description of the exposure models for the African countries that were imported into GED4ALL.

Also as part of this project, the data schema has been populated with data for 11 African nations (Mozambique, Tanzania, Malawi, Kenya, Uganda, Ethiopia, Mali, Rwanda, Senegal, Cabo Verde and Senegal), and several other existing datasets that could be used to improve the covered of GED4ALL have been identified. Figure 1 illustrates the distribution of the number of buildings in a 15x15 arcsecond grid for these African countries.

Figure 1 - Distribution of the number of buildings from the African nations currently featured in GED4ALL, following an evenly spaced grid with a 15x15 arcsec resolution.

This final report summarizes the main outcomes of each of this components, and includes a discussion regardings future improvements and way forward to scale the current coverage of the database to the global level.

Component 1 - Exposure Database Structure

The first component covered the development of the database schema of GED4ALL, and addressed issues related with the IT development, installation, basic operation of the database and the associated tools for converting, importing and exporting data, possible connection to visualization tools, interoperability with other exposure databases (e.g. OpenStreetMap - OSM). All of the outcomes of this component are described in Deliverable 1 - Exposure Database Schema and Complementary Tools[1].

The database schema developed within this component has proved sufficient to store exposure information of different types, resolutions and sources. In particular we are confident that the schema is capable of satisfying the scientific requirements imposed by the various exposure experts and in particular of handling three different levels of spatial resolution identified as targets at the inception phase of this project.

We have also demonstrated that the OpenQuake *NRML* format is sufficiently expressive for use as a *lingua franca* for exposure data interchange, and that there is a pre-existing set of tools available to facilitate the process of data discovery, extraction, and format conversion. Using these tools in conjunction with the import and export tools developed as part of this project we have been able to import a wide range of different types of exposure datasets into GED4ALL for the purposes of testing.

The GED4ALL schema is based on the level 2 schema present in GED4GEM with extensions and modifications to improve support for multiple perils. In comparison with the GED4GEM level 0/1 schema, the GED4ALL schema is considerably simpler, containing a smaller number of tables linked by both fewer and less complex relationships, as illustrated in Figure 2.

The GED4ALL schema is also much more flexible, supporting arbitrary locations and full geometries whereas GED4GEM supports a fixed spaced 30 arc-second grid of points and is focused primarily on residential buildings. In GED4GEM, all studies share the geometry information stored in the grid, so multiple studies for a large geographic area occupy a comparatively small amount of storage space. If a large number of very high-resolution exposure models are stored in the GED4ALL schema, it is likely that the resulting total disk space occupation will exceed that of GED4GEM. Given that the GED4ALL allows exposure models to be added incrementally and that the storage space requirements increase in proportion with the size of the stored models, we feel that this is a reasonable trade-off in line with the objectives of the project.

[1] Henshaw P, Silva V, O'Hara M, (2018). GED4ALL D1 - Exposure Database Schema and Complementary Tools 2018-1, X pp., GEM Foundation, Pavia, Italy.

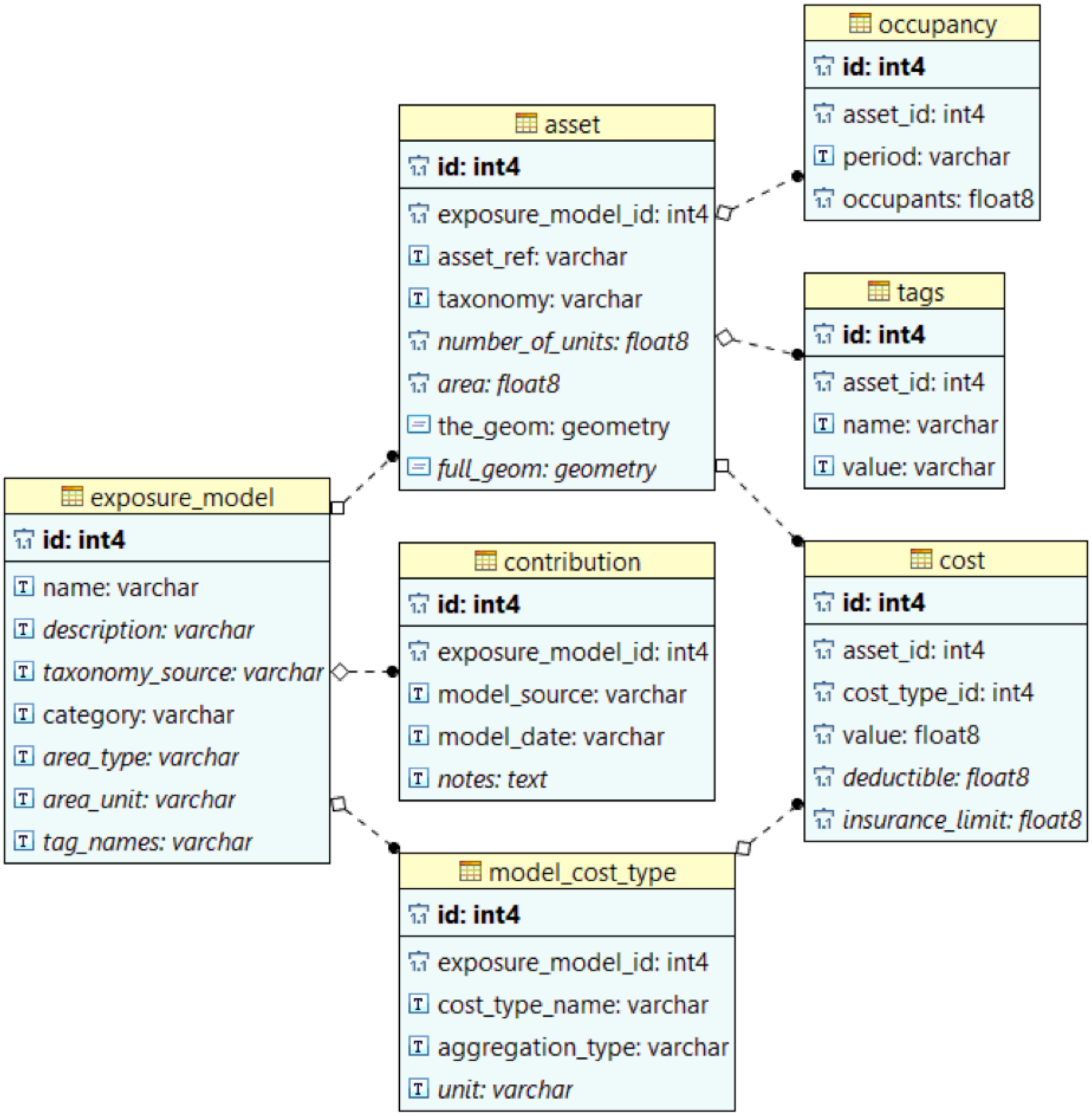


Figure 2 - GED4ALL Level 2 Schema Entity Relationship Diagram.

While visualization and exploration support are the subject of a separate project, we have made use the QGIS Desktop tool to show that the data can indeed be visualized using standard tools and that the schema provides sufficient information for styling and presentation. This schema is also currently being used within another GFDRR-supported project to explore and download exposure datasets.

Component 2 - Multi-hazard Exposure Taxonomy

A uniform classification system to categorize the elements exposed to natural hazards was developed within this project. This taxonomy is able to capture attributes relevant for various perils (e.g. the ones featured by the ThinkHazard![1] platform), cover a wide range of assets, and is sufficiently flexible and comprehensive in order to be applicable at a global scale. All of the outcomes of this component are described in Deliverable 2 - Multi-hazard Exposure Taxonomy[2].

This taxonomy also assumes special importance for the integration of the different products of the Challenge Fund, since it creates the link between the exposure datasets and the vulnerability functions, or to any existing damage or loss databases. Four main categories of assets were considered for the development of the taxonomy:

- Buildings
- Lifelines and infrastructure
- Crops, livestock and forestry
- Socio-economic data

For the development of the uniform taxonomy for structures (i.e. residential, commercial, industrial, educational and healthcare), the GEM Taxonomy v2.0 was considered as a starting point (Brzev *et al*.[3] 2013). This classification system was created considering a wide spectrum of existing taxonomies (e.g. PAGER-STR, WHE), and despite the fact that it was originally developed with a focus on earthquakes, it featured several parameters relevant for other perils such as floods and hurricanes (e.g. existence of floors below ground, type of roof). Following the feedback collected at the Inception Workshop in Pavia, Stakeholders meeting in London, and from a comprehensive literature review, other attributes and options were added to this existing classification system, thus leading to the GED4ALL taxonomy. The structure of this new taxonomy is depicted in Figure 3.

The Syner-G and HAZUS classification systems for infrastructure (e.g. roads, railways, bridges, storage tanks, electric grid, water supply network and gas network) were used to develop the taxonomy for these elements. This taxonomy is based on a number of structural attributes, similar to what was described for buildings. Each attribute has a number of options that can be combined to generate a taxonomy string or category.

[2] V. Silva, C. Yepes-Estrada, J. Dabbeek, L. Martins, S. Brzev (2018). D2 Multi-Hazard Exposure Taxonomy. GEM Technical Report 2018-01, GEM Foundation, Pavia, Italy.
[3] GEM Building Taxonomy version 2.0:
https://www.globalquakemodel.org/resources/publications/technical-reports/gem-building-taxonomy-report/

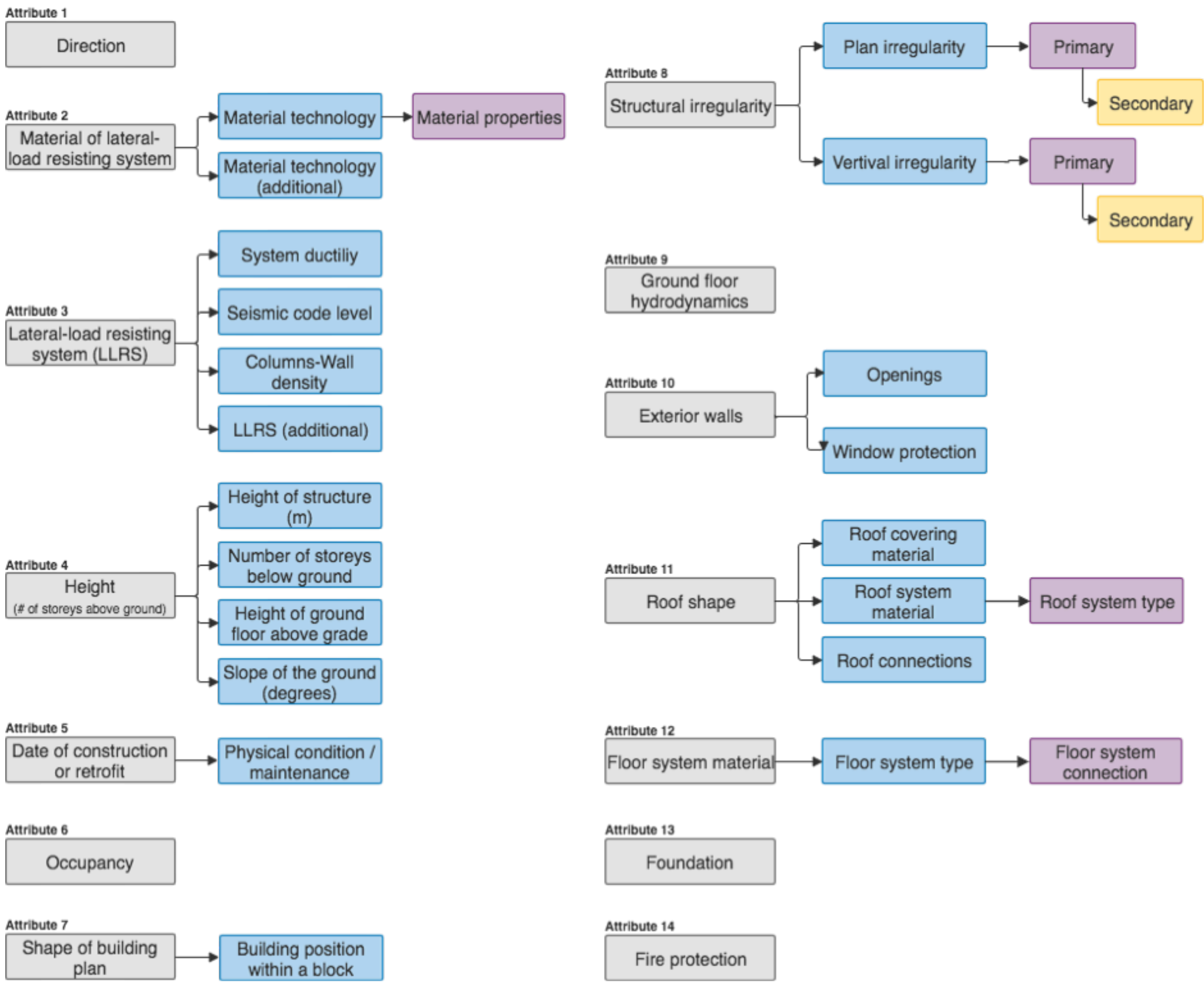


Figure 3 - GEM Building Taxonomy v2.0 structure.

The definition of the taxonomy for crops, livestock and forestry was performed considering the classification system proposed by the Food and Agriculture Organization (FAO) for the first two types of elements, and the categories of forestry defined by UNESCO. The classification of livestock and forestry was included in this project as part of a request from the stakeholders, received during the London meeting.

For what concerns the taxonomy for the socio-economic data, this consortium leveraged on the work performed by the Karlsruhe Institute of Technology and GEM, who have created sub-national databases of socio-economic data for South East Asia, East Sub-Saharan Africa and Central/South America for multi-hazard disaster risk assessment.

Despite the recognized need for a uniform taxonomy to classify exposure data for the purposes of disaster risk assessment, it is also fundamental to understand that such systems can be overwhelming (and even intimidating) for less experienced users. To overcome such issue, within this project a simplified simplified version of the classification system was also developed, as illustrated in Figure 4.

The selection of the lower number of attributes was performed based on the feedback of dozens of experts regarding the minimum attributes necessary to consider in order to characterize a building stock for multi-hazard risk analysis.

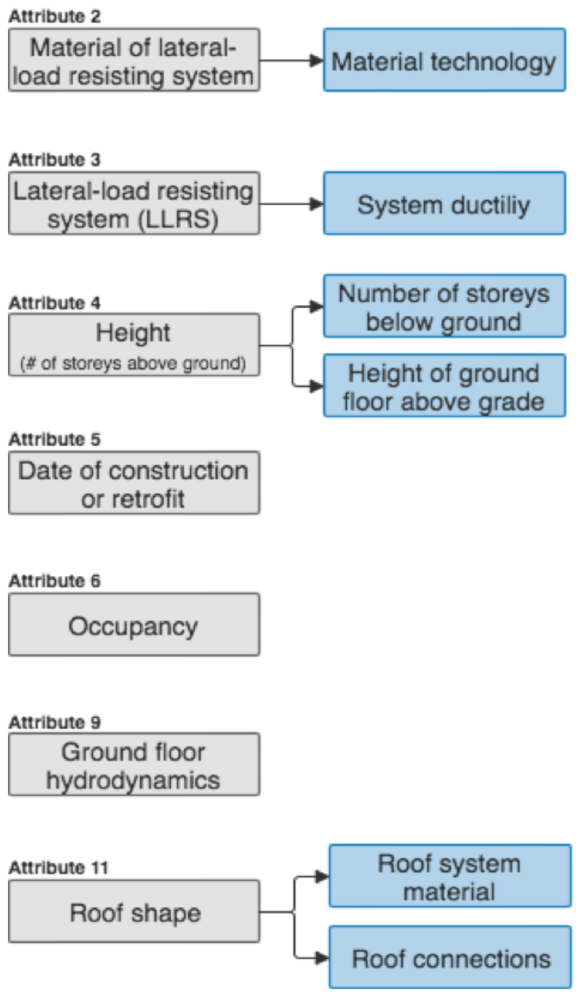


Figure 4 - Simplified multi-hazard GED4ALL building taxonomy.

Moreover, we have also described a number of supporting tools that have been used by risk modellers and mappers around the world in risk assessment exercises. We propose that similar tools should be developed or adapted for this taxonomy, in order to facilitate its use and purpose.

Component 3 - Populating the exposure database

This component of the project described 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 WorldPop, which contains population data for different time periods at a 100x100 m^2 spatial resolution, and OpenStreetMap initiative, which features building-level data, as illustrated in Figure 5 for the cities of Kathmandu and Dar es Salaam. All of the outcomes of this component are described in Deliverable 3 - Populating GED4ALL with existing databases[4].

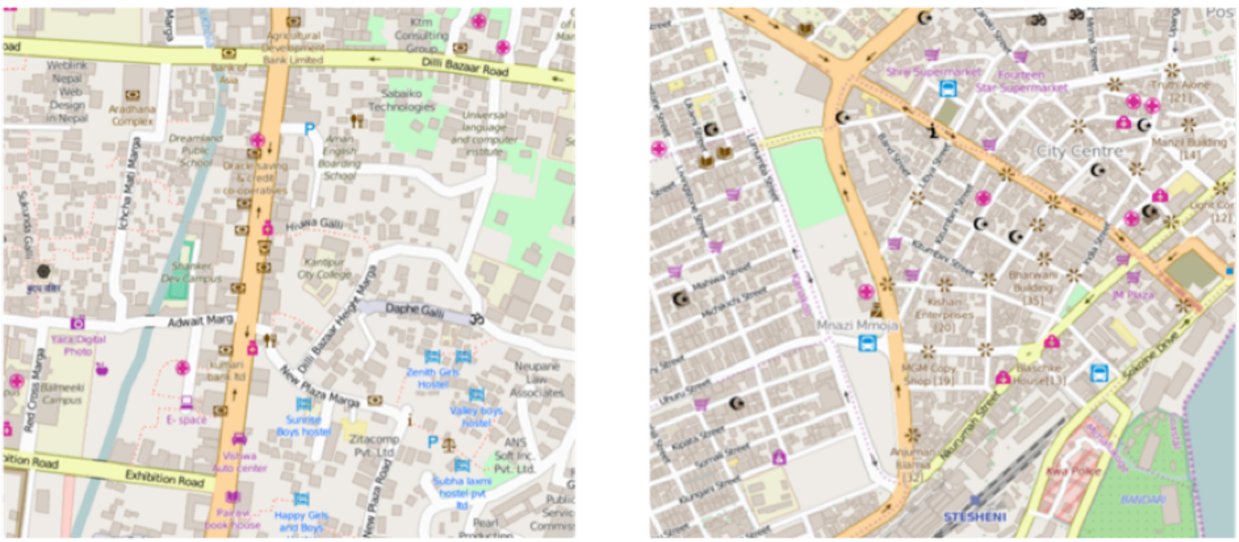


Figure 5 - Exposure data for Kathmandu (left) and Dar Es Salaam (right) from OpenStreetMap.

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. In particular, we have shown that it is possible to import OSM data into GED4ALL, taking advantage of detailed building material information, when available, in order to produce a classification sufficient for calculations with a loss estimation engine. It should be noted however that due to the crowd source driven approach adopted by OSM, there is significant variability in the building classification information provided by different contributors, which suggests that in many cases it will be necessary to supplement the information from OSM with other data sources in order to classify assets and perform loss estimates. Even in cases where the OSM

[4] O'Hara M, Silva V, Henshaw P, Huyck C (2018). D3 Populating GED4ALL with existing databases GEM Technical Report 2018-03, GEM Foundation, Pavia, Italy.

building classification is rich and detailed (as in the case of the Zanzibar city example), manual processing is required in order to produce a suitable taxonomy string.

This deliverable also highlights several tasks that can contribute to the scaling of GED4ALL to the global level. One of these efforts is related with existing databases of exposure data. To maximize these valuable data, additional support 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 6 for 10 countries in the Middle East.

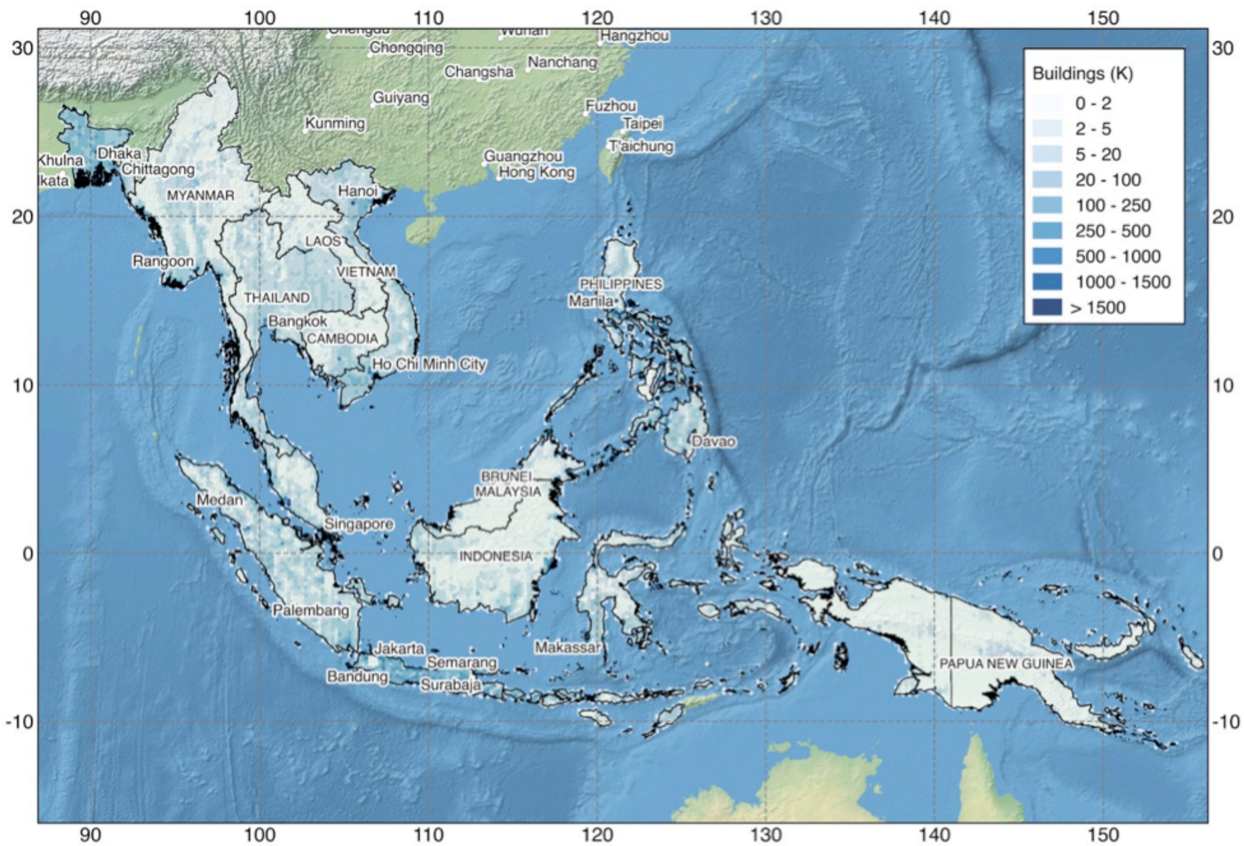


Figure 6 - Distribution of the number of residential buildings following an evenly spaced grid (0.2 decimal degrees) for ten countries in South-East Asia.

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 (see next section). 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.

Component 4 - Development of exposure datasets

Whilst there are several international efforts to characterize the built environment for the purposes of disaster risk assessment and reduction, there are still large portions of the world that remain without a proper exposure model that can be used in multi-hazard risk analysis. This component of the project presented several approaches that can be followed to develop such datasets, from satellite imagery to field missions. All of the outcomes of this component are described in Deliverable 4 - Development of Exposure Datasets[5]

With the exception of a few countries around the world (e.g. Australia, Canada, Italy, Portugal, United States), exposure models at the national scale are not usually publicly available, or fit-for-purpose for disaster risk assessment. For this reason, in the last decade several national and international organizations have supported the development of such models. Despite these encouraging developments, there are still parts of the world not covered by a robust and up-to-date exposure model that can allow the evaluation of disaster risk, and support decision makers in the development of risk mitigation measures. For these regions, it is urgent to follow one of the many approaches to collect data regarding the built environment, and develop an exposure dataset.

Amongst the many methodologies for exposure modeling, two main categories can be identified: 1) “modeled” exposure datasets and 2) “collected” exposure data. In the former group, an exposure model is developed using one or multiple auxiliary datasets (e.g. satellite imagery, census data, cadastral data), which are utilized to estimate the number, economic value and vulnerability classes of the building stock. This approach is usually adopted for large-scale risk analysis (e.g. national or regional scale). In the latter group, usually field missions are organized to collect asset-by-asset data, using mobile apps or similar types of applications. This methodology has naturally a much higher reliability and accuracy, but it is inevitably time-consuming and unpractical at the national scale.

This component of the project also provided a description of several data sources, exposure modelling methodologies, data collection techniques and tools that can be used in the development of exposure datasets. Several examples have been presented using the country of Tanzania (see Figure 7), one of the nations that were covered in this project.

[5] C.K. Huyck, M. O’Hara, P. Henshaw, V. Silva, Z. Hu, M.T. Eguchi, G.R. Esquivias, M.M. Huyck. D4 - Development of Exposure Datasets. GEM Technical Report 2018-04, GEM Foundation, Pavia, Italy.

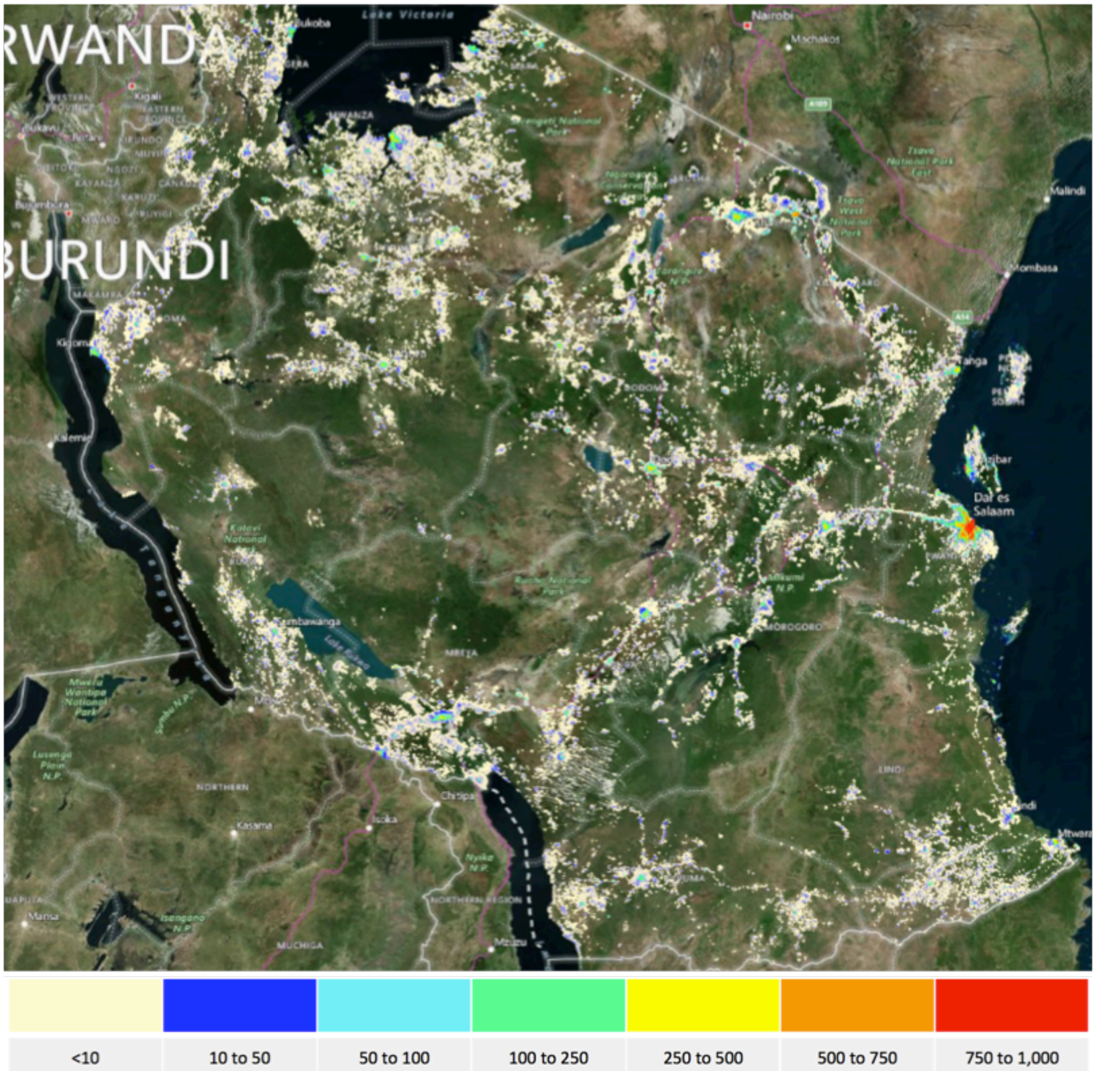


Figure 7 - Exposure model for Tanzania in terms of replacement cost following an evenly spaced grid with 15x15 arcsec resolution (~500 Meters by 500 Meters).

The selection of the most appropriate approach to develop an exposure dataset must be done considering a number of criteria, including the available resources, the geographical extent of the study, and the final use for the datasets. For example, the development of a disaster risk profile at the national scale will require large geographical coverage, and not necessarily the precise location of all the assets. Such assessment would fall into the main exposure modelling category. On the other hand, the development of an emergency plan at the municipal level will require the definition of structural characteristics and location of critical facilities, roads and other lifelines with precision, which would most likely require the field missions to collect the necessary data. A combination of both approaches is also possible, in which modelling techniques are calibrated with data collected on the field.

Final Remarks

This final deliverable described the main outcomes of each project component. It is the opinion of the consortium that the GED4ALL database and associated deliverables represent a valid and robust response to the original challenge: providing an open exposure database for multi-hazard risk assessment. This consortium recognizes the various challenges (technical, scientific and political) that have to be addressed in order to scale the three Challenge fund project globally, and to ensure that these schema become a standard used worldwide. An exploratory plan has been already provided within this deliverable, which has now to be discussed by the wider community, and followed by an implementation strategy.

Feedback and Possible Future directions

During the joint Challenge Fund final workshop held in Dar es Salaam, March 20 to March 22 2018, there were a number of discussions and break-out sessions during which local stakeholders and consortium partners exchanged views on possible use cases and future extensions of the results of the three challenge fund projects. Many break-out groups identified improved planning and policy making as the priority use case, with local stakeholders expressing appreciation for the fact that the schema associated entries with geographic location and time: it was felt that it was important to be able to compare the data with the actual local situation and where necessary and appropriate provide an updated view.

During the workshop, participants also indicated that they would like to contribute datasets to the databases, which is a very encouraging signal indicating that the at least some members of the community see value in the work done so far and are willing to contribute to its ongoing success. To continue seeking the interest of the community, a side event will be held at the 2018 Understanding Risk Conference in Mexico City. The outcomes from this event will provide additional insight regarding the plan to reach global coverage, and to ensure that the current data schemas become the standard in disaster risk assessment and reduction.