GED4ALL

Exposure Database Schema and Complementary Tools

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

SUBMITTED BY:

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

	A	B	C	D	E	F	G	H	I
1					RESIDENTIAL			COMMERCIAL	
2		COUNTRY	Year (RES, COM, IND)	Population	Dwellings	Buildings	Repl. cost (USD Million)	Buildings	Repl. cost (USD Million)
3	BLZ	Belize	2010	322,459	49,107	42,607	2,462	7,456	1,084
4	BRB	Barbados	2010	277,819	78,934	74,697	4,941	5,640	1,921
5	CRI	Costa Rica	2011	4,301,716	1,286,211	1,186,046	75,871	37,829	25,325
6	CUB	Cuba	2012	11,167,328	3,644,001	3,160,517	112,033	8,779	42,940
7	DOM	Dominican Republic	2010	9,445,367	2,652,794	2,239,180	122,658	16,642	46,060
8	GTM	Guatemala	2002	11,237,106	2,574,874	2,417,206	67,362	226,352	45,639
9	HND	Honduras	2013	8,303,765	1,837,855	1,646,054	33,372	134,659	17,169
10	HTI	Haiti	2015	10,733,022	2,316,754	2,149,306	17,471	0	0
11	JAM	Jamaica	2011	2,697,053	861,021	815,031	28,152	44,361	15,202
12	NIC	Nicaragua	2005	5,142,161	985,970	927,484	21,581	142,982	13,364
13	PAN	Panama	2010	3,405,830	895,763	751,089	49,677	57,157	34,896
14	SLV	El Salvador	2007	5,744,092	1,372,831	1,266,063	41,178	140,872	21,646
15	TTO	Trinidad and Tobago	2000	1,114,777	313,032	288,537	12,620	8,413	2,654
16	MEX	Mexico	2015	119,491,628	31,914,272	26,035,512	2,596,187	3,741,675	915,047
17	USA	United States of America	2010	308,758,105	131,704,730	103,610,750	27,304,678	6,069,528	5,278,661
18	ARG	Argentina	2010	40,134,430	13,810,262	10,512,754	564,351	613,455	403,648
19	BOL	Bolivia	2012	10,058,586	2,803,654	2,458,560	20,038	163,928	25,349
20	BRA	Brazil		163,820,945		16,923,573	446,382		
21	CHL	Chile	2002	15,112,940	3,898,549	3,119,138	228,594	306,750	182,991
22	COL	Colombia	2005	41,123,763	9,731,154	7,591,208	294,361	117,171	257,940
23	ECU	Ecuador	2010	14,474,386	3,746,602	3,018,688	62,563	22,738	52,840
24	GUF	French Guiana		162,125		30,120	851		
25	GUY	Guyana		662,928		117,613	1,604		
26	PER	Peru	2007	27,402,300	6,397,862	5,694,792	100,244	413,668	124,339
27	PRY	Paraguay		5,471,323		764,696	23,146		
28	SUR	Suriname		434,276		43,494	896		
29	URY	Uruguay		3,235,899		452,325	38,785		
30	VEN	Venezuela	2011	27,222,751	6,928,701	4,962,678	161,942	70,416	211,200
31	AUS	Australia		21,869,924	9,004,942	7,535,573	3,541,434	215,311	1,814,509
32	DZA	Algeria		41,077,590	6,767,694	6,043,285	282,274	92,294	25,486
33	AGO	Angola		13,384,135	1,800,056	1,800,056	114,326	71,911	19,857
34	BEN	Benin		7,163,807	822,565	822,565	18,583	27,125	7,490
35	BWA	Botswana		2,034,425	478,268	373,693	13,295	7,215	1,992
36	BFA	Burkina Faso		11,043,356	1,008,951	1,008,951	9,428	38,323	4,020
37	BDI	Burundi		8,053,574	1,685,553	1,417,453	13,884	4,285	1,183
38	CMR	Cameroon		19,406,100	3,391,817	2,393,249	67,449	72,513	20,014
39	CPV	Cabo Verde		491,683	114,469	94,933	2,543	3,120	862
40	CAF	Central African Republic		3,685,441	493,882	493,882	8,433	10,181	2,811
41	TCD	Chad		7,599,548	1,015,516	1,015,516	19,144	29,502	8,147
42	COM	Comoros		495,073	53,318	53,318	3,690	2,116	584
43	COG	Congo		3,264,359	275,850	275,850	9,546	9,842	3,718
44	CIV	Côte d'Ivoire		15,974,797	1,356,063	1,356,063	8,876	63,109	6,620
45	DJI	Djibouti		818,159	86,977	82,288	6,613	2,120	585
46	EGY	Egypt		92,280,504	18,634,136	14,097,552	419,016	236,680	65,358
47	GNQ	Equatorial Guinea		444,344	50,019	50,019	615	3,172	333
48	ERI	Eritrea		5,617,046	893,782	708,818	31,327	15,303	4,226
49	ETH	Ethiopia		74,074,340	15,116,855	13,162,865	207,833	146,134	40,353
50	GAB	Gabon		684,992	73,895	73,895	3,048	4,762	1,315
51	GMB	Gambia		926,042	62,612	62,612	1,022	4,942	518

Exposure Summary | Building Fractions | Risk Summary | GDP 2015 | Action Points | Country Checklist | Social

GED4ALL

Global Exposure Database for Multi-Hazard Risk Analysis

D1 - Exposure Database Schema and Complementary Tools

Authors: P. Henshaw, V. Silva, M. O'Hara

Last update: March 2018

Citation: Henshaw P, Silva V, O'Hara M, (2018). GED4ALL D1 - Exposure Database Schema and Complementary Tools 2018-1, X pp., GEM Foundation, Pavia, Italy, DOI: to be included.

Abstract

A consortium comprised by 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, crops, socio-economic data), while considering relevant attributes for six natural hazards: earthquakes, floods, volcanoes, strong winds, tsunamis and drought. The outcomes of this project will be applied to the country of Tanzania to develop an exposure model at the national scale, and to the city level of Dar es Salaam to demonstrate how an exposure dataset at a building-by-building resolution can be created. Moreover, this process will be demonstrated for five countries around Tanzania (Ethiopia, Uganda, Kenya, Malawi and Mozambique).

This report provides a detailed description of the structure of the tables comprising the GED4ALL exposure database schema and the relationships between them. We will also describe the software components required for installation and basic operation of the database and the associated tools for converting, importing and exporting data.

Keywords:

Exposure, databases, schema, software, building stock, disaster risk reduction.

Table of Contents

List of Figures

List of Tables

Introduction

The main objective of Challenge Fund 2 is the development of an open exposure database, for the purposes of performing multi-hazard risk analysis. This goal will be achieved through five components:

1. IT structure of the database;
2. Multi-hazard taxonomy;
3. Protocols to populate the database;
4. Development of exposure information for pilot countries;
5. Dissemination of the outcomes through workshops.

This report provides a detailed description of the structure of the tables comprising the GED4ALL exposure database schema and the relationships between them. We will also describe the software components required for installation and basic operation of the database and the associated tools for converting, importing and exporting data.

Scientific Requirements

As discussed in the inception report, GED4ALL is required to support the following features:

Multiple Spatial resolutions: the database will allow three types of resolution: gridded distribution (e.g. GED4GEM, WorldPop, LandScan, GRUMP, GDP data); aggregated numbers based on administrative regions (e.g. GADM, socio-economic data, building census surveys, CRESTA); and site-specific assets (e.g. OpenStreetMap, insurance portfolios). The first two categories usually refer to modelled exposure data (top-to-bottom approach), most likely developed using government data (e.g. Census) or satellite imagery. The last category includes the digitization of building footprints from aerial imagery, field missions to collect building-by-building attribute information, geocoded building data, or the collection and compilation of existing building specific databases with coordinates provided- such as critical facilities, vulnerable buildings, schools, or government buildings.

GED4ALL addresses this requirement by removing the GED4GEM level0/1 limitation forcing all assets to be located on a regular 30 arc-second grid and instead allowing each asset to be located at any point on the globe. In addition, the database allows for an optional second geometry to be specified, providing for example a building footprint, a road/rail network or an administrative boundary.

Multiple types of assets: although this initial version of the database will not be populated with all the types of assets for a given country, a decision was made to make it compatible with the following types of assets: buildings (e.g. residential, industrial, commercial, healthcare, educational or governmental); infrastructure (lifelines, energy generation facilities, bridges); crops, livestock and forestry; and socio-economic data (e.g. gross domestic product, population, education indices).

We address this requirement by allowing each exposure model to specify a taxonomy, and each asset to provide a taxonomy string valid for the specified system. In this way, buildings can be described using the GED4ALL taxonomy, while crops and forestry can use the FAO taxonomy, or any other classification system. The multi-hazard exposure taxonomy developed as part of this project is described in detail in Deliverable D2.

Multiple versions: the database will support multiple exposure datasets for the same regions. These might represent the region at different times, contain different levels of detail, or belong to different institutions. For example, both GEM and ImageCAT have developed exposure datasets for the residential, commercial and industrial building stock in Ethiopia, Uganda and Kenya. Moreover, some cities might be covered with modelled and field-collected data.

Time component: each exposure dataset is linked to a particular date, representing the time that should be considered for the impact analysis (e.g. time that the building stock represents or season characterizing the crops). Moreover, agriculture, population and lifelines (transportation system) will have a seasonal component (variation throughout the year).

Each exposure model present in the GED4ALL database provides basic metadata describing the date and source of the data. The database does not limit the number of exposure models that can specified for any given region.

Interoperability: given the existence of several exposure databases and loss estimation tools, the database will be capable of exporting data in a well-documented and useful format for these other resources (i.e. containing all of the most common fields).

GED4ALL provides a view (virtual table) to facilitate exposure export in the popular *csv* format; this enables exposure data to be loaded into desktop spreadsheet applications, statistics packages, GIS tools and other software.

The Python export tool developed as part of this project also allows data to be exported in the Natural Risks Markup Language (*NRML*) format used by the OpenQuake-engine. This is an open *xml* format, which can be used freely either via the Python libraries provided with the OpenQuake-engine, or using any other standards compliant with a *xml* parser toolset.

Using *NRML* format for exposure data export guarantees that the information stored in the database is compatible with at least one loss estimation tool, the OpenQuake-engine. Please see the OpenQuake-engine manual[1] for further details on the exposure model and *NRML* format.

Visualization: the technology and structure of the database will allow the connection to visualization tools. Although this project does not cover the development of such tools it is clear that users will need some form of visualization support in order to import, explore and export exposure data.

Data exported in *csv* format can be loaded into Excel, QGIS or other common desktop tools for visualization. It is also possible to visualize exposure data as map layers using the QGIS Database Manager to connect to the PostGIS database directly. We also plan to experiment with open-source web-GIS server solutions such as QGIS server, GeoServer and/or GeoNode to provide access to GED4ALL via open OGC standards; in this way common clients such as QGIS and OpenLayers are able to provide basic visualization support.

[1] OpenQuake-engine manual: https://storage.globalquakemodel.org/openquake/support/documentation/engine/

GED4ALL Exposure Database and Tools

The basic infrastructure for the GED4ALL exposure database is composed of a relational, geospatial database and software tools for importing and exporting exposure data developed specifically in the context of this project. We also describe how some existing external tools can be used to assist users in tasks such as converting exposure data into the *NRML* format for use with the import tool.

Software and Hardware requirements and Installation

The GED4ALL development database is currently hosted at the GEM headquarters in Pavia on a Linux workstation running Ubuntu Linux (16.04) and using Postgres 9.5 and PostGIS 2.2 for basic geospatial relational database storage. It is also possible to install GED4ALL on other modern Linux operating systems such as CentOS and RedHat Enterprise Linux. The exact installation and configuration procedure will depend on the operating system and local network configuration. Please consult the PostgreSQL and PostGIS documentation for further details.

GED4ALL Database Schema

The GED4ALL database schema contains the following tables:

- exposure_model
- contribution
- model_cost_type
- asset
- cost
- occupancy
- tags

The tables and their relationships are depicted in Figure 1 below:

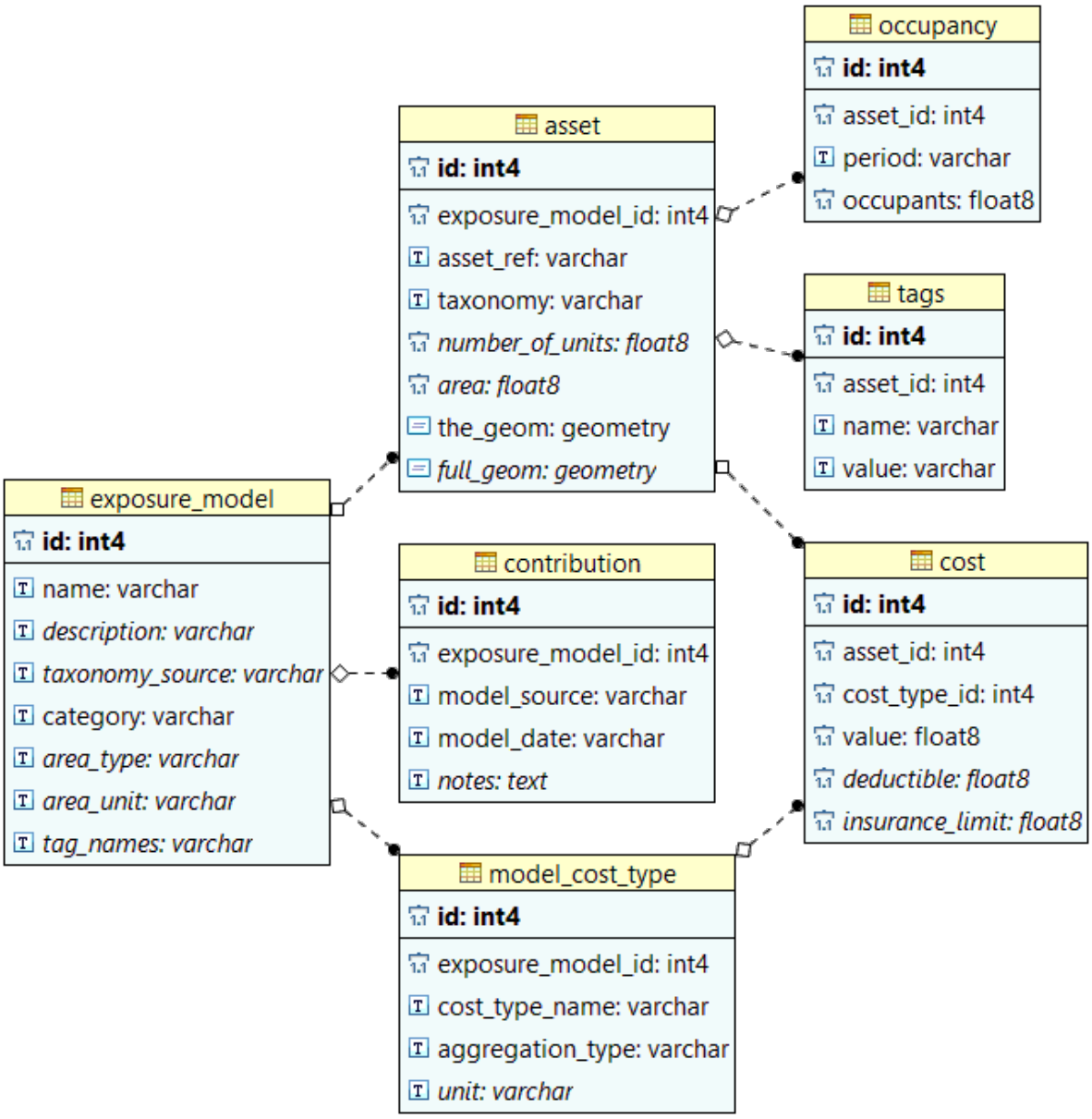

Figure 1 - GED4ALL Level 2 Schema Entity Relationship Diagram

Fields shown in *italic text* are optional, fields in normal type face are mandatory (NOT NULL). Lines between tables indicate a relationship. A hollow diamond shape indicates a "many" cardinality, a solid dot indicates a "one" cardinality. For example, each asset is part of exactly one exposure model while an exposure model has many assets.

In the following sections we describe each table present in the data schema.

Table "exposure_model"

Each row in this table represents a collection of assets with a taxonomy_source (GEM Taxonomy, Syner-G, FAO, etc) and category (buildings, road network, etc). A model can optionally contain one or more cost types (see model_cost_type), e.g. structural, contents and can also optionally contain information about the surface area occupied by the assets. Finally, a model can optionally specify any number of "tags", which can be applied to each asset (see Table 1).

Table 1 - Attributes of Table "exposure_model".

Field Name	Type	Notes
id	Integer	Unique identifier, generated automatically
name	String	User supplied name of the exposure model
description	String	Optional description
taxonomy_source	String	Taxonomy system used for assets in this model, e.g. GED4ALL, GEM, Syner-G, FAO, etc.
category	String	Category of assets in this model, e.g. "buildings", "agriculture", "livestock", "lifelines", "socioeconomic", "population"
area_type	String	Optional area type - aggregated / by asset
area_unit	String	Optional unit used for area e.g. "sqm"
tag_names	String	Optional space separated list of tags present in the exposure model (see tags table)

Table "contribution"

This table contains metadata which is not strictly speaking part of an exposure model; it is not used for risk assessment and not present in the *csv* export. This information is intended to help users choose the most appropriate model.

Table 2 - Attributes of Table "contribution".

Field Name	Type	Notes
id	Integer	Unique identifier, generated automatically
exposure_model_id	Integer	Foreign key: exposure_model.id. Identifies the exposure_model this contribution belongs to
model_source	String	Where this model came from
model_date	String	Model date
notes	Text	Optional notes (arbitrary length text)

Table "model_cost_type"

Each exposure model can optionally have one or more types of cost associated with loss or damage to assets. For example, the cost of the building structure by square meter or the cost of the contents of a single building. Each cost type entry describes a category of cost, an aggregation type and a unit.

Table 3 - Attributes of Table "model_cost_type".

Field Name	Type	Notes
id	Integer	Unique identifier, generated automatically
exposure_model_id	Integer	Foreign key: exposure_model.id. Identifies the exposure_model this cost_type belongs to
cost_type	String	The type or category of cost, e.g. "structural"
aggregation_type	String	How the costs are aggregated, e.g. "per_asset" or "aggregated"
unit	Text	The unit used for costs, usually a currency code e.g. "USD"

Table "asset"

Each row in the asset table represents a single 'exposable good' such as a building, aggregated collection of buildings, an item in an infrastructure network or an area of farmland. The taxonomy value must be

set to a valid taxonomy string using the taxonomy system specified in the exposure_model table, and identifies the typology of this asset. The number_of_units field is used to specify the number of assets of this type at this location. There are two geometry fields providing geospatial information:

- the_geom is a mandatory field point geometry used to identify the location of the good;
- Full_geom is an optional geometry of any type supported by PostGIS (polygon, multiline, multipolygon...) which can be used to store information such as building footprints, regional boundaries, or lifeline networks.

Table 4 - Attributes of Table "asset".

Field Name	Type	Notes
id	Integer	Unique identifier, generated automatically
exposure_model_id	Integer	Foreign key: exposure_model.id. Identifies the exposure_model this asset belongs to
asset_ref	String	User supplied identifier
taxonomy	String	Taxonomy string using the system identified by exposure_model.taxonomy_source
number_of_units	Double precision floating point number	Number of assets represented
area	Double precision floating point number	Optional area of the asset in the units specified in exposure_model.area_unit
the_geom	Point geometry	Point location of the asset. Please see Geometry Fields notes below.
full_geom	Arbitrary geometry	Option geometry of any kind - used for vector building footprints, regional boundaries etc. Please see Geometry Fields notes below.

Table "geometry Fields"

PostGIS geometry fields are implementations of the "Simple Features" specifications as defined by the OpenGIS Consortium (OGC[2]) which specifies common storage and access model of geometries (point,

[2] OpenGIS Consortium: http://www.opengeospatial.org/standards/sfa

line, polygon, multi-point, multi-line, etc.) used by geographic information systems. Each geometry field provides a reference to a spatial reference system (map projection) and the description of the specific shape (point, polygon, line, etc). While PostGIS supports a variety of spatial reference systems, in GED4ALL we adopt the 1984 revision of the World Geodetic System (the WGS84/EPSG:4326)

The "the_geom" field specifies a single point location and can be considered as a composite representation of latitude, longitude and a reference to WGS84. Only *Point* geometries may be stored in this field.

The "full_geom" field is an optional representation of any geometry type (point, line, polygon, multi-point, multi-line, etc.) supported by the standard and can be used to describe the footprint of a building (including complex multi-polygon geometries such as a castle with a tower in a courtyard), or the shape of a field, or a multi-line segment of a road network.

Table "occupancy"

The occupancy table is used to store information about the occupants of an asset in a given period - for example, the number of people in a given building during the day and during the night. For some cases the period might refer to a season, for example when considering livestock or agricultural assets. In some communities the term "occupancy" is used to refer to building usage, for example "industrial" or "residential"; this concept can be modelled in the GED4ALL schema using tags and/or by using a taxonomy system that supports building usage specification, such as the GEM Building Taxonomy.

Table 5 - Attributes of Table "occupancy".

Field Name	Type	Notes
id	Integer	Unique identifier, generated automatically
asset_id	Integer	Foreign key: asset.id. Identifies the asset this occupancy refers to
period	String	The period considered - e.g. "day", "night", "transit".
occupants	Double precision floating point number	Number of occupants in the asset during the given period

Table "cost"

Each row in the cost table represents a cost of a given cost type for a given asset. The asset_id and cost_type_id fields are foreign keys used to identify an asset and model_cost_type. The value field provides the actual cost value in the unit specified in the corresponding model_cost_type entry.

Table 6 - Attributes of Table "cost".

Field Name	Type	Notes
id	Integer	Unique identifier, generated automatically
asset_id	Integer	Foreign key: asset.id. Identifies the asset this cost refers to
cost_type_id	Integer	Foreign key: model_cost_type.id. Identifies the model cost-type
value	Double precision floating point number	The cost of the specified asset for the given model cost-type

Table "tags"

Each row in the tags table represents a name=value pair for a given asset. The asset_id is a foreign key used to identify an asset. The name field represents the name of the tag while the value field contains the associated value . Tags may be used to store any named scalar information with an asset, but in particular are useful for storing socio-economic indicators and their associated values, and also fro grouping assets into for example Cresta-zones or census tracts.

Table 7 - Attributes of Table "tags".

Field Name	Type	Notes
id	Integer	Unique identifier, generated automatically
asset_id	Integer	Foreign key: asset.id. Identifies the asset this cost refers to
name	String	The name of the tag. Examples might include "census_tract_id", "cresta_zone_id" or "poverty_index"
value	String	The value associated with the tag for this asset.

Table “all_exposure” VIEW

In addition we have created a database view named all_exposure which allows us to access information regarding all assets along with their cost and occupancies as a single “virtual table”.

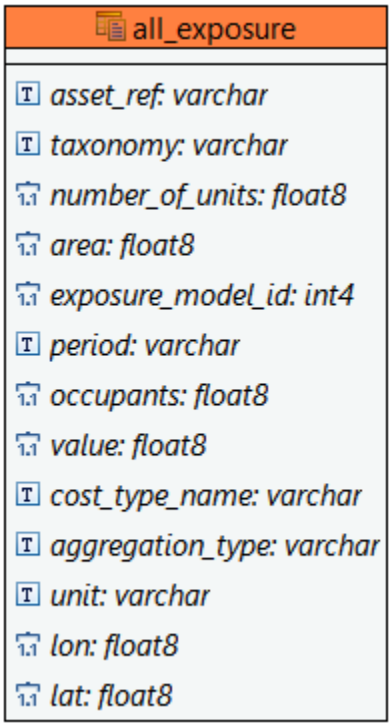

Figure 2 - GED4ALL “all_exposure” view for simple CSV export

This view can be used to export exposure information in *csv* format with a SQL command such as:

```
COPY (
 SELECT * FROM level2.all_exposure WHERE exposure_model_id=65
) TO STDOUT WITH CSV HEADER
```

Note that the output produced includes the asset position as lat and lon coordinated but not full geometries such as building footprints.

GED4ALL Import/Export Tools

Alongside the database infrastructure, two Python programs have been developed to import exposure models in the *NRML* format into the database and to export data from the database *NRML* format. As already described earlier, it is also possible to extract data in the *csv* format using a simple *SQL* query. Using the *NRML* import tool in combination with pre-existing software tools and web services that facilitate exposure model format conversion (described briefly in the next section) allows us to import existing exposure data in a variety of formats and from different sources.

Import/Export tool Installation

The import/export tools make use of GEM's OpenQuake engine software in order to parse and generate exposure files in the NRML format. It is possible to examine the Python source code by opening the following URL in a web browser: https://github.com/gem/ged4all/blob/master/python

The GED4ALL schema and tools can be obtained from GitHub using the following command:

```
$ git clone https://github.com/gem/ged4all/
```

In order to install these tools, a functional Python 2.7 environment with VirtualEnv support is required. On Ubuntu 16.04, the following commands can be used to setup a Python virtual environment:

```
$ sudo apt install python-virtualenv python-dev
$ virtualenv gedenv
$ source gedenv/bin/activate
```

The import/export tools provide a setuptools compliant requirements.txt file which allows the software stack to be installed easily using pip:

```
$ pip -U install pip
$ pip install -r requirements.txt
```

Before the import and export tools can be used for the first time, it is necessary to supply the connection details and credentials to use to connect to the database. Execute the following command to copy the db_settings.py.template:

```
$ cp db_settings.py.template db_settings.py
```

Then edit the db_settings.py file with your favourite text editor and replace the text between angle brackets with username, password and hostname/IP Address to use. There are two database connection sections: 'geddb' and 'gedcontrib'. The user in the 'geddb' section must have read access rights to the database, and is used by the exporter. The user provided in the 'gedcontrib' section must

have both read and write access to the GED4ALL database and is used when importing data into the database. While it is possible to use a single read/write account for both sections, for security reasons we strongly recommend using a write-enabled account only when strictly necessary.

Once the db_settings.py file has been updated with valid credentials, the import tool can be used as follows:

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

If the import procedure is completed without error, the automatically generated unique identifier of the newly imported model will be displayed.

In order to export an exposure model in NRML format, the export tool can be used as follows to display data in NRML format to the standard output (console):

```
$ python export_exposure_nrml.py model-id
```

Use the following command to save the data in a file:

```
$ python export_exposure_nrml.py model-id > /path/to/file.xml
```

External Tools

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.

OpenQuake Platform Input Preparation Tool

As illustrated in the figure below, 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.

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

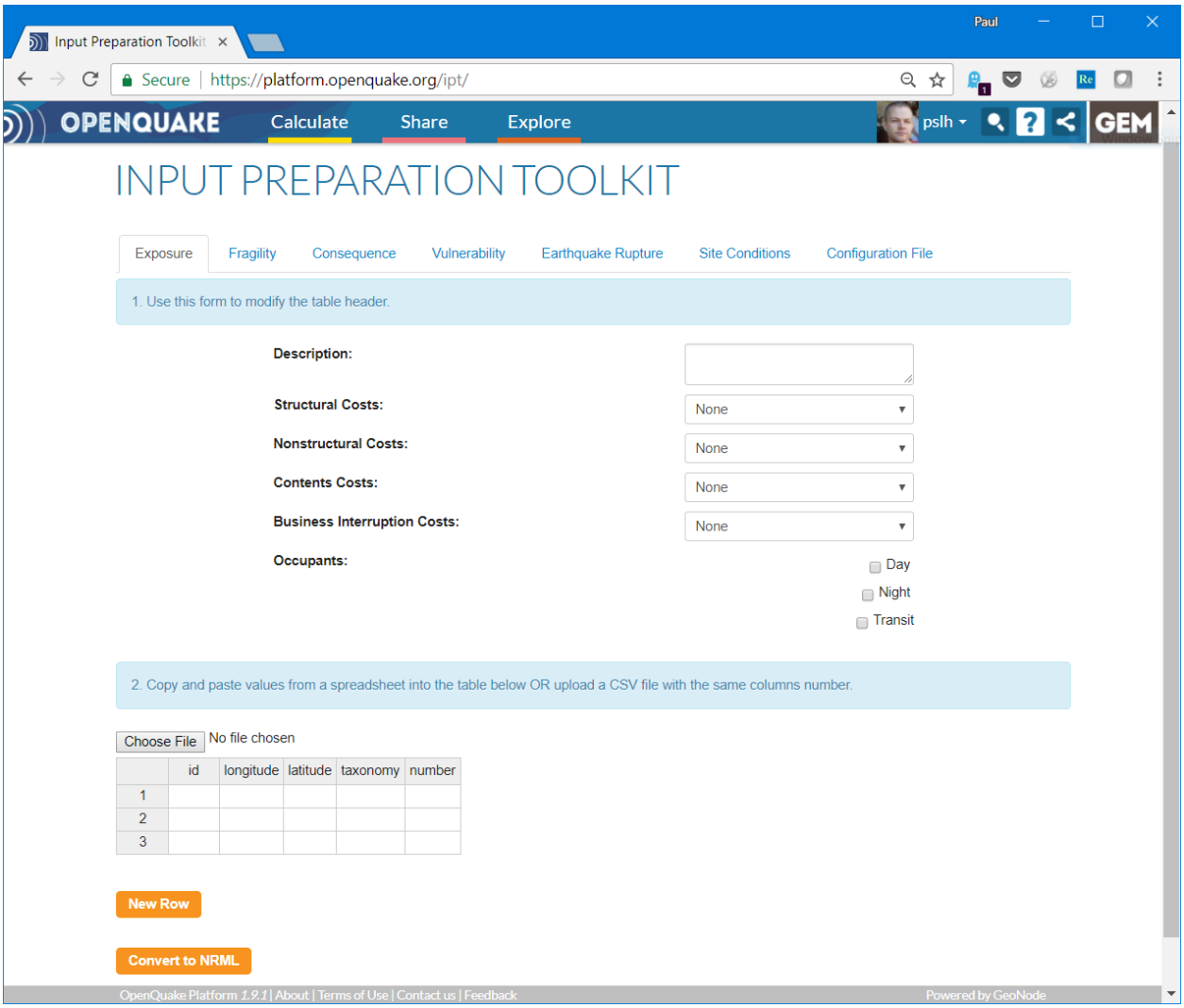

Figure 3 - Input Preparation Toolkit Exposure tool on the OpenQuake-platform.

Taxtweb + Taxonomy glossary

The OpenQuake-platform also provides a web based tool (TaxTWeb[3]) for constructing valid GEM Building Taxonomy strings by selecting options from drop-down menus. The same tool can be also used in the reverse direction, automatically populating the menus for a user provided taxonomy string. A searchable glossary[4] of taxonomy terms, with explanations and illustrations is also available from

A more detailed discussion of the GEM Building Taxonomy and other supported taxonomy systems is provided in Deliverable D2.

[3] TaxTWeb: https://platform.openquake.org/taxtweb/
[4] GEM Glossary: https://taxonomy.openquake.org

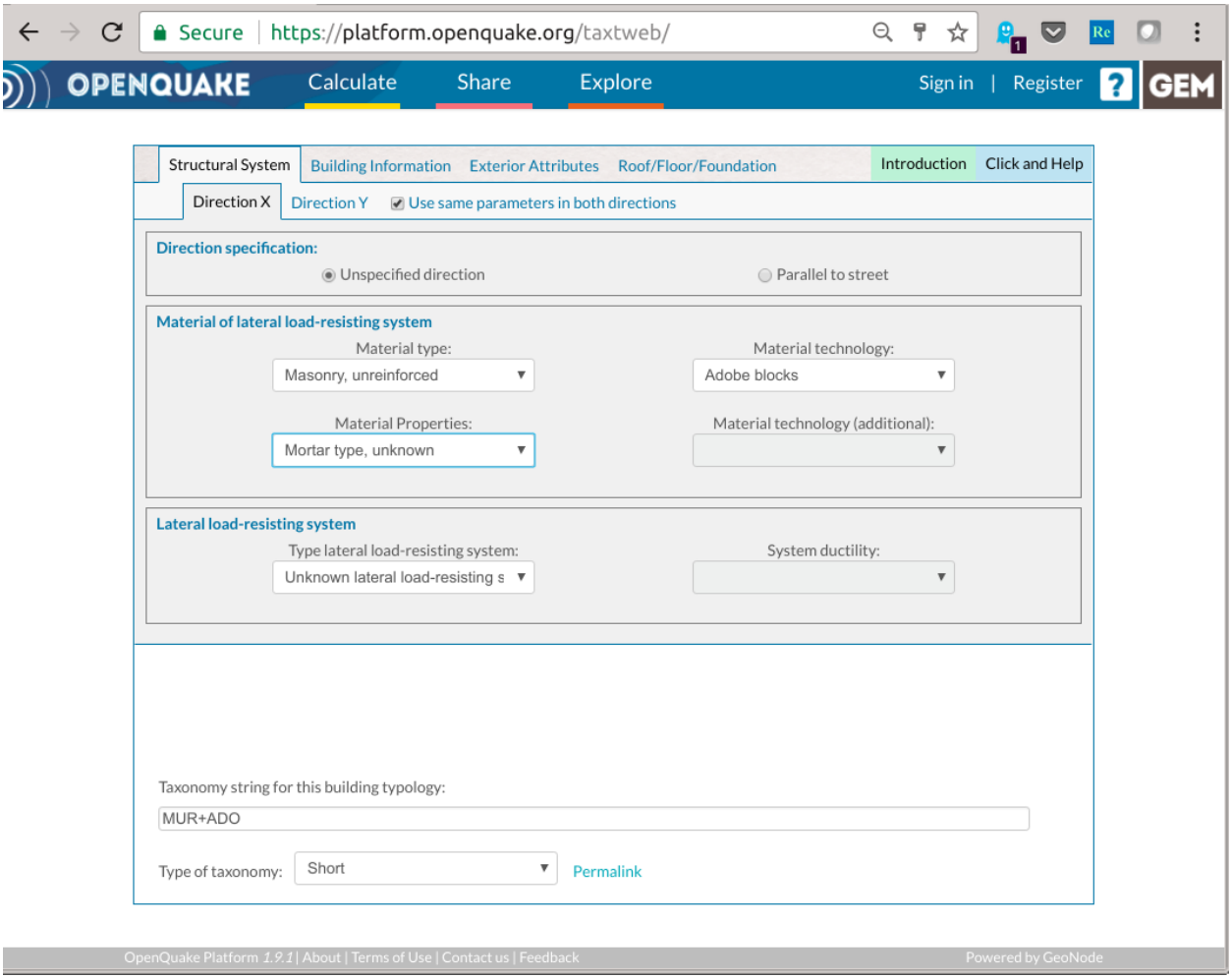

Figure 4 - GEM Taxonomy web tool on the OpenQuake Platform

Humanitarian OpenStreetMap Team Export Tool

The HOT team recently released version 3 of their Export Tool[5], a powerful web based system allowing users to search for OpenStreetMap data for a specified region, filter according to a user supplied criteria andexport the data in a wide range of file formats, including, but not limited to ESRI Geopackage and Shapefile formats.

A more detailed description of this tool and its use in the context of GED4ALL is provided in Deliverable D3, Populating GED4ALL with existing Databases .

[5] HOT Export Tool: https://export.hotosm.org/en/v3/

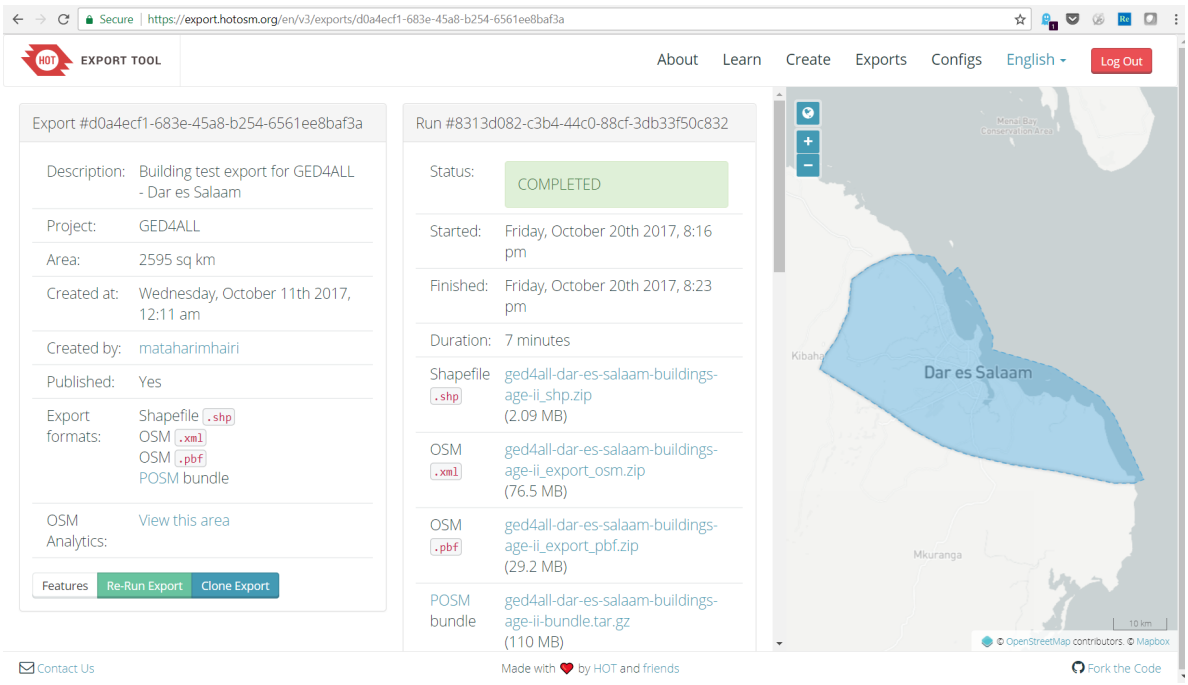

Figure 5 - HOT Export Tool.

Test Data

In order to demonstrate that the schema is capable of storing data at all three different levels of data identified in the

Scientific Requirements section, we have imported into the database the following test data sets:

- Gridded exposure data for Tanzania
- Transportation network for Tanzania (from OpenStreetMap)
- Buildings aggregated to a subnational regional level - Peru
- Buildings in Zanzibar City, with Individual building footprints - data from OpenStreetMap, preliminary mapping from OSM attributes to GEM taxonomy string.
- Socio-economic indicators for Tanzania and the surrounding countries

It is important to note that these data have been used only to demonstrate the suitability of the database schema as part of development and testing activities and do not constitute part of the final database contents. Deliverable D4 will describe the process and datasets used to populate the GED4ALL database.

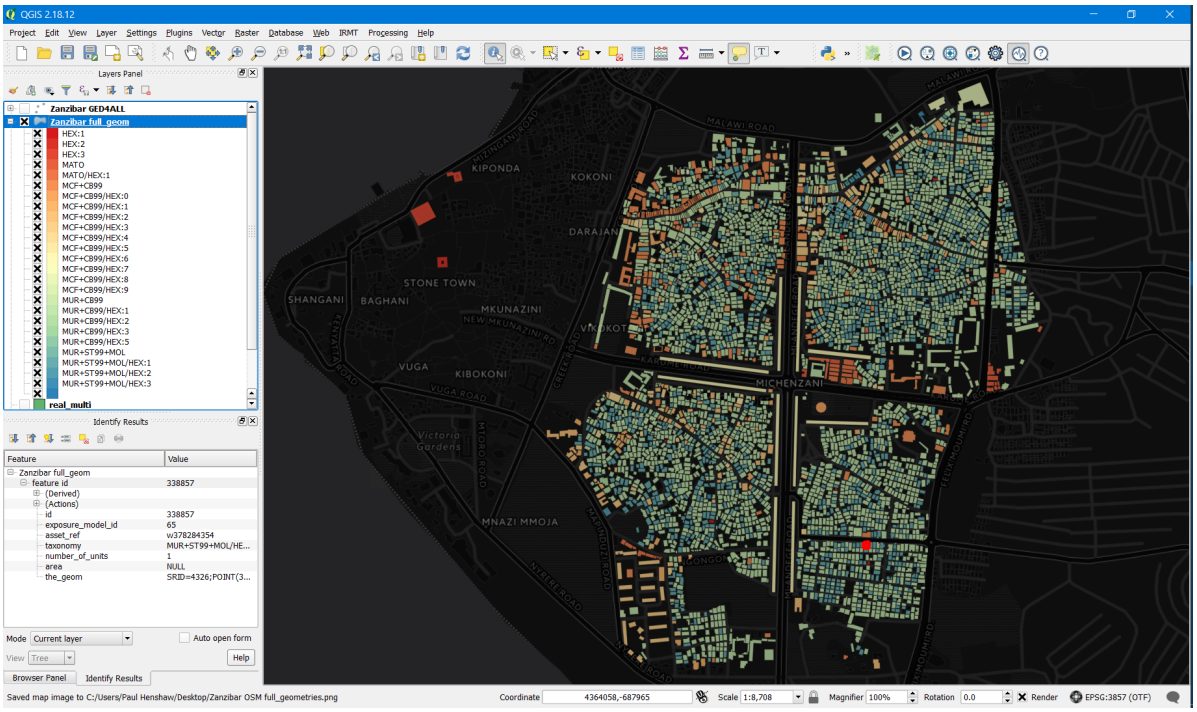

Figure 6 - QGIS visualizing building footprint geometries in Zanzibar City, imported from OpenStreetMap.

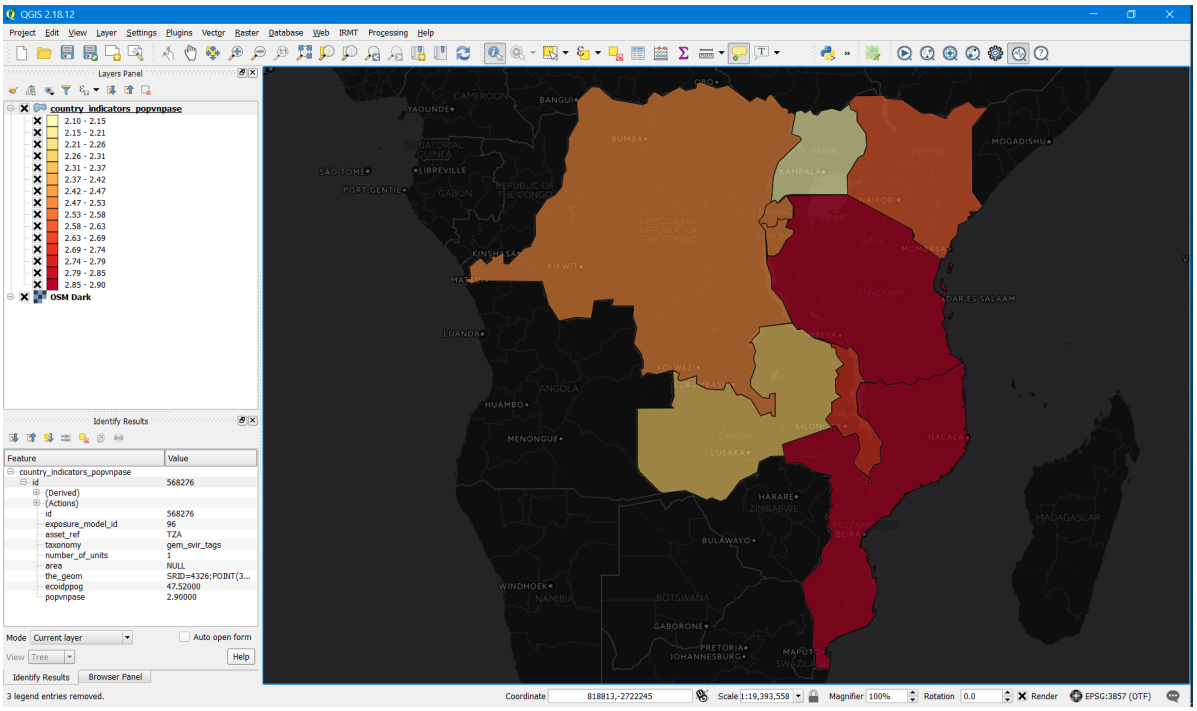

Figure 7 - QGIS visualizing building simplified country boundaries for Tanzania, styled by a socioeconomic index, data imported from the OpenQuake-platform.

In order to extract the Zanzibar City OSM data used to test the building-by-building support for GED4GEM, we made use of the HOT Export Tool, along with the open-source GIS package QGIS and the QuickOSM plugin to load the OSM data as a GIS layer.

We then used the Taxtweb tool to produce GEM building taxonomy strings for the common building typologies identified in the OSM data, and the IPT to convert the resulting tabular data into NRML format. While there were a number of manual steps in this process and the resulting exposure model is intended for testing purposes only, this does demonstrate that the pre-existing tools can be used to prepare exposure models suitable for ingestion into GED4ALL.

A more detailed discussion of how to construct exposure models will be provided in deliverable D4, Development of Exposure Datasets.

Final Remarks

The database schema presented in this report 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 described earlier in this report and in particular of handling the 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. 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 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.

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. 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. In contrast, GED4ALL's flexible approach to asset locations means that there is no easy way to share geographic information and so each asset must provide a geographic location. 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.

Another limitation of the current toolset is that the import tools reads the entire XML tree into memory, while this has been sufficient for the datasets encountered so far, it is possible that attempting very large exposure models would cause the import to fail due to an Out Of Memory error. There are a number of possible solutions to this problem including the use of a streaming XML parser, and support for importing data directly in a less verbose format such as *csv*.

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.