Exposure Data Development for the Global Earthquake Model: Inventory Data Capture Tools

J. Bevington[1], R. Eguchi[2], C. Huyck[2]
ImageCat Inc., [1]Ashtead, UK. [2]Long Beach, CA, USA

H. Crowley
GEM Foundation, Pavia, Italy

F. Dell'Acqua, G. Iannelli
University of Pavia, Italy

C. Jordan
British Geological Survey, Keyworth, UK.

J. Morley
University of Nottingham, UK

M. Wieland, S. Parolai, M. Pittore
GFZ German Research Centre for Geosciences, Potsdam, Germany

K. Porter
SPA Risk LLC, Denver, CO, USA.

K. Saito
Cambridge Architectural Research, UK

P. Sarabandi
Stanford University, Stanford, CA, USA

A. Wright
OpenGeo, New York City, NY, USA

M. Wyss
World Agency of Planetary Monitoring & Earthquake Risk Reduction (WAPMERR), Geneva, Switzerland

SUMMARY:
The Global Earthquake Model represents a open initiative to develop software and tools for assessment of global seismic risk. This paper describes an open-source software suite and protocols for populating a Global Exposure Database of structural information being developed by GEM's Inventory Data Capture Tools (IDCT) risk global component. Several data collection methods are used: remote sensing, direct field observations, and statistically-inferred mapping schemes. Data are collected on pre-earthquake inventory, as well as post-event damage: rudimentary building information from optical satellite or aerial sensors, including the building footprints and height. Field-based observations collected on digital and paper tools by survey teams allow expert users to generate detailed structural information, including structural type, year built, and occupancy. Mapping schemes allow the generation of exposure information for user-defined areas of homogenous urban land use, using sampled field observations. These methods converge to feed GEM's Global Exposure Database and the Earthquake Consequences Database.

Keywords: GEM, inventory data capture tools, exposure development, damage assessment, remote sensing

1. INTRODUCTION

Over half a million people died in the last decade due to earthquakes, most of these in less-developed countries. However, recent earthquakes in Japan and New Zealand have reminded global populations that despite high levels of development and investment in mitigation strategies, seismic activities still account for major human and economic losses globally. A precursor to negating some of the risk to human life and economic risk, and potentially offsetting the latter through insurance, is having the ability to measure, understand and manage seismic risk. An understanding of physical vulnerability of structures in the built environment requires knowledge of the configuration and characteristics of urban building stock. These data offer valuable baseline information in responding to disasters and to estimate human and financial consequences resulting from the event. Insurance penetration in the most developed countries has led to the use of modelled probabilistic risk assessments to aid risk understanding. However, there are few risk models currently developed and operational in less-developed countries, partly due to a lack of insurance penetration in these regions. Recent years have also seen a trend towards developing open and transparent models to sit alongside the proprietary models widely used in the insurance industry. The latter models can be expensive to use and do not promote understanding of risk at the most fundamental grass-roots level of community risk assessment. Providing the tools for an open-source risk model promotes an understanding of risk in previously unmodelled or poorly-modelled regions, subsequently leading to mitigation and resilience-building activities such as an increased number of insurance products, promotion of building codes and retrofitting activities.

The public-private Global Earthquake Model (GEM) Foundation drives an international effort aimed at the development of state-of-the-art, widely accepted datasets, models and open-source tools/software. With the goal of collaboratively advancing seismic hazard and risk assessment worldwide, these will be made available through the web-based OpenQuake platform, which will be powered by a calculation engine (Silva et al., 2012 and Monelli et al., 2012). Development of a global model of seismic risk through an open, participatory process is at the core of the effort (Pinho, 2012) and international consortia are currently developing the global components of the model. Fig. 1.1 shows the inter-relationships between the risk global components, which will allow platform-users to generate new exposure data, access existing vulnerability functions and search historical damage surveys and consequences datasets. The hazard global components are aimed at developing methods, datasets and standards for assessing earthquake hazard, by combing data on historic seismic events, active faults, global instrumentation, and ground motion and strain predictions.

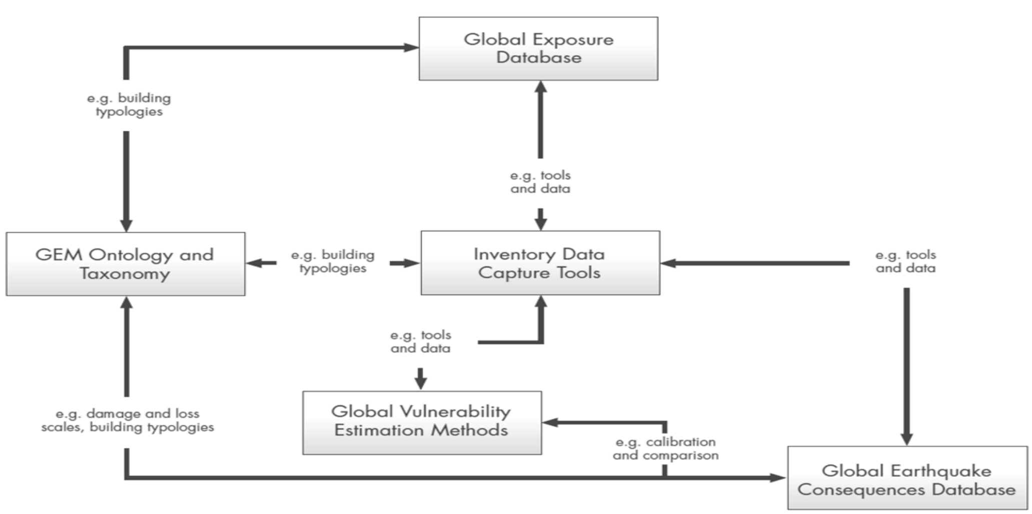

Figure 1.1. The interrelationships between the GEM Risk global components (GEM, 2011)

The GEM Risk global component is tasked with providing standards for vulnerability estimation (damage, and human and social and economic losses), and also provides a database of physical and socio-economic consequences from past events globally - the Earthquake Consequences Database (GEM). Underpinning this is the Global Exposure Database (GED); a homogenised database of global building stock that contains spatial, structural and occupancy-related information. The GED stores information on the global building stock using existing datasets and also using customised tools and protocols, whilst adhering to the standards of classification and relationships defined by the Ontology and Taxonomy consortium.

1.2. Inventory Data

A building inventory is a catalogue of the buildings and facilities in each class of a taxonomy, used in loss models to define the exposure to specific hazard, based on insurance exposure data (overview and details of insured properties for an area of interest). This underlying building characterisation that serves as input data for which the losses are calculated, and typically requires building characteristics such as type, age, height, occupancy, building value and location (Born and Martin, 2006, Erdik et al., 2010). It is often the case that catastrophe loss models are limited in their ability to provide accurate estimations due to poor estimations of exposure information. This can have a significant impact on the modelled loss or consequences (Coburn and Spence, 1992, Jaiswal et al., 2008), which is apparent by the order of magnitude difference between modelled and actual losses sustained by the insurance industry in several recent natural catastrophes (Conway et al., 2008, Global Reinsurance, 2008).

GEM aims to generate global exposure data for its open-source model through the collection of building inventory information at several geographic scales, from level 0 (country) to level 3 (per-building, Table 1.1). The GED4GEM consortium is collating global datasets for inventory characterisation at levels 0 and 1, including global housing and population data. IDCT is charged with developing the tools and user protocols for collection of inventory data at level 2 and 3 (regional to per-building precision). A similar set of tools and protocols will also be developed for the collection of earthquake damage information at granularity levels 2 and 3. These data populate the GED and ECD databases, respectively. This paper presents the rationale and workflow of GEM's Inventory Data Capture Tools alongside a description of the tools and methods under development. The three major components of 1) remote sensing, 2) direct field observations, and 3) mapping schemes are presented in context with the wider aims of the GEM Risk global components as the tools underpin much of the analytical ambitions of the model. The method of each tool is described, alongside a status report at time of writing (April 2012).

Table 1.1. Geographic scales of exposure data in the GED. IDCT tools will populate levels 2 & 3

Precision level	Source	Grid/Vector	Statistical significance
Level 0	Global (LANDSCAN, PAGER, GRUMP)	30” (~ 1 km)	Country
Level 1	Sub-country database (UN-HABITAT, national census, GEM regional programmes), Urban density, RS (aggregated, < 1 km)	30” (~ 1 km)	Region/municipality
Level 2	Regional sub-country database (Mapping Schemes)	10/15”, Vector	Region/municipality
Level 3	Per-building database (remote sensing, ground surveys)	Vector	Single building

2. THE INVENTORY DATA CAPTURE TOOLS SUITE

Developing a building exposure model requires a combination of expert judgement, established classification typologies and geo-spatial technologies and tools. With this in mind, the IDCT consortium is developing a suite of tools and user protocols that follow a three-pronged workflow (Fig. 2.1). Firstly, remotely-sensed satellite imagery provides a data source for deriving individual building footprints and larger areas of homogenous urban land use. Second, field teams characterise building attributes using a series of data collection forms for both inventory and post-earthquake damage. Once these characteristics are input into a database, the third component, the Spatial Inventory and Damage Data (SIDD) tool, provides a workbench for the iterative generation of mapping schemes using the homogeneous areas of land use (derived from remote sensing) combined with sampled field data. This workflow produces regional (level 2) and per-building (level 3) inventory datasets. Data from the field tools can also be used to directly populate the Global Exposure Database at level 3. These workflows are described in detail in Fig. 2.1.

The types of data to be collected are fundamentally defined by the end-user depending on the use case. However, the GEM tools are pre-designed to collect all parameters required to generate new vulnerability functions or calculate risk information using existing vulnerability functions. Data are generated using one or all of the following methods: automatic and manual remote sensing image interpretation (section 2.1), direct field observations (section 2.2), crowdsourced photo interpretation (section 2.2.3). IDCT offers users a flexible framework allowing data collection missions to be tailored according to the availability of engineering expertise, analysts, hardware or software. All tools described in the user protocols supplied with the GEM tools are open-source, with directions to commonly used commercial software where performance may be greater.

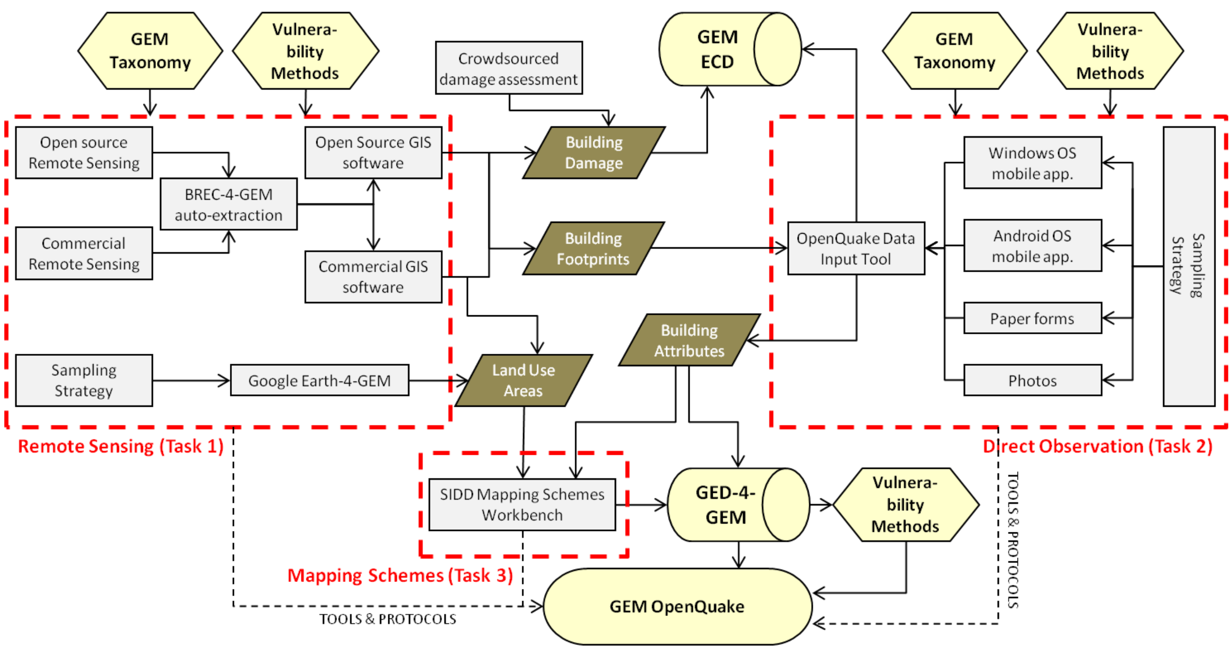

Figure 2.1. Detailed IDCT Workflow. Major tasks of remote sensing, direct observation and mapping schemes are shown in red. Main data developed are shown in green. Tools are in grey and GEM components external to IDCT are shown in yellow.

2.1. Task 1. Remote Sensing

Images captured from satellite and aerial sensors have become a staple data source for risk managers in the past decade. The availability of internet tools such as Google Earth have increased the public awareness and general uptake of image processing to extract information on the makeup of urban areas. IDCT provides a suite of software tools and user guides to enable the extraction of building outlines from imagery for level 3 exposure development, or derivation of areas of homogeneous land use for regional exposure (level 2). These GIS data captured from imagery provide vector location information by delineating in the image the outline of where the building intersects the ground (building "footprint"). Tools to derive height information based on the presence of shadows in imagery are also provided. The IDCT tools are primarily developed in the open-source GIS and Image Analysis package, Quantum GIS (QGIS). The GRASS GIS plug-in is also required, and BREC-4-GEM, a custom toolbar for automatic extraction of building footprints and height data (Gamba et al., 2009), has been developed as an extension to the software (Fig. 2.2). Two use-case scenarios provide the workflow for the user guides: 1) extraction of building footprints, and 2) delineation of urban land use. Sourcing of imagery, image pre-processing, analytical and GIS editing tools are all described in the protocols. All of these software tools and their supporting user guides will be made available as a single download from the OpenQuake site.

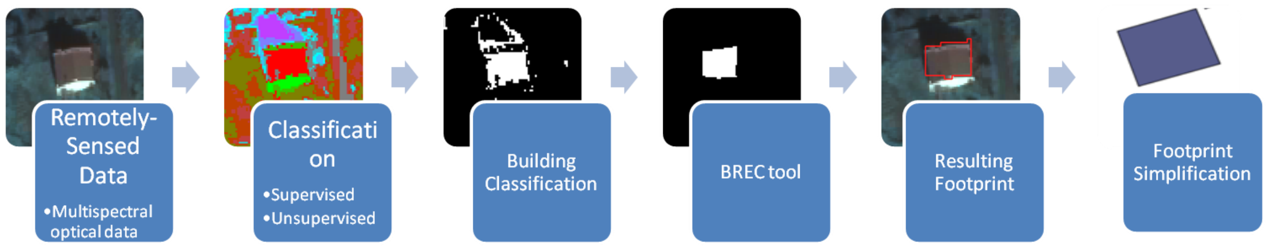

Figure 2.2. Workflow of BREC automatic footprint extraction tool

Remote sensing tools to extract seismic consequences data are also provided in the IDCT toolset. Analysis of pre- and post-event satellite or aerial imagery can be interpreted by expert users to identify buildings that have been collapsed or heavily damaged by earthquakes (e.g. Ghosh et al., 2011). Automatic change detection techniques are provided to distinguish areas of significant change between these time periods. However, the web interface that facilitates manual identification of damaged buildings provides the main RS consequences toolkit. Based on an interface developed for the Global Earth Observation Catastrophe Assessment Network (GEO-CAN), the Tomnod Disaster Mapper (Barrington et al., 2011) provides pre- and post-event satellite images for assessment by expert image analysts and engineers. Data captured from the RS toolkit can be manually uploaded to GEM's OpenQuake platform and integrated in the GED and ECD.

2.2. Task 2. Direct Observation

Remote sensing, while useful for determining location, and potentially height of buildings, has large uncertainties when manually assessing occupancy, structural type or year built. IDCT therefore fuses the image-based analytics with field observations (Task 2, Fig. 2.1). A suite of field tools is being developed by the consortium and includes both digital and paper forms for data collection, as well as protocols for collection of location information and field photographs. Digital tools are currently under development for laptop and tablet based hardware. The laptop tool is developed for Windows operating system (OS), while the tablet is being designed to run on the Android OS. The choice of which platform to support was made after extensive user consultation through a questionnaire sent to existing field reconnaissance teams.

The DO field tools are designed to directly incorporate the GEM taxonomy, allowing users to collect information using dropdown forms and pre-loaded options for all attributes to be collected. The tools will have tabbed-functionality to allow maximum ease of use, rather than prescribing a linear data collection process. Fig. 2.3 shows the preliminary wireframes of the user interface for the Windows and Android systems. The DO tools consist of several pre-loaded forms for collection of inventory and damage data. These reflect the GEM taxonomy:

1. Basic building inventory (parameters including, location, structural material type, occupancy, year of construction, lateral-load resisting system and height) - for use in GED
2. Detailed building inventory (to include additional secondary modifiers) - for use in GED
3. Post-event Damage (pre-loaded damage scales for building components) - for use in ECD
4. Interior survey (for when the survey team has access to the inside of the building. To include non-structural elements) - for use in GED and ECD.

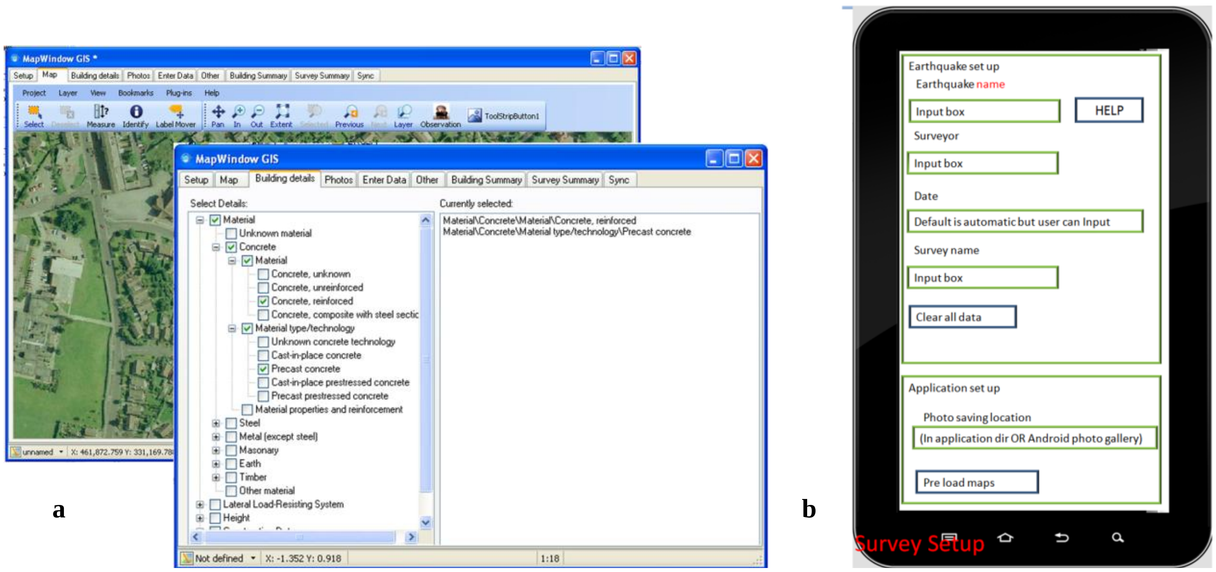

Figure 2.3. Wireframes for the Windows OS (a) and Android OS (b) digital field tools for in-field data collection

The pre-loaded forms in the digital tools also include metadata on location, date, time, surveyor details and additional project information. In addition, templates for paper forms for collecting the same information as in the digital forms are being provided so the DO capture tools do not discriminate between users based on hardware availability. The DO tools are currently under development in partnership with the UK Earthquake Engineering Field Investigation Team (EEFIT). Other global field reconnaissance teams are also being consulted through the iterative development and testing process, due for completion in late summer 2012.

2.2.1. Field Photographs

Alongside the digital and paper forms, the user will be provided with information on the types of hardware required to generate geo-tagged photographs. Photographs are an invaluable data source to be collected in field missions for many reasons, including: A photograph may be the best way to describe something (although it may need some accompanying text explaining what is shown in the photograph); if one is in a hurry and there is no time to write detailed notes, a photo can be used to record the moment and notes can be added later (see Data Input section 2.2.3). The protocols provide a 'common sense' approach, recording many parameters, including geo-location, look direction, data and time information, amongst others. It is hoped the protocol can provide best practise on the collection of field photos for both inventory and consequences data capture.

2.2.2. Sampling Strategy

Planning field deployments to collect information for statistical inferring any of the risk parameters relies on having statistically valid samples of observations. IDCT will provide users with a protocol for establishing a custom sampling strategy for each of their field deployments to identify those buildings to survey. The protocol details the following:

- A procedure that allows for variable levels of resources to carry out reconnaissance for a subset of the total population of structures
- A procedure for creating efficient, stratified samples of buildings. By “strata” is meant here consecutive ranges of some measurable building features such as building height, distance from population centres, or topographic slope.
- Procedures for extrapolating to the population.
- Procedures to elicit and employ local builder or building-department expertise, for inferring non-visible features from visible ones based on local knowledge.

GIS tools are also being made available in the remote sensing protocols to aid sampling decision making using geographic sampling techniques.

2.2.3. Data Input (PRIZM)

IDCT provides users with a user-friendly online import tool, currently under development, to manage the collation of field information once teams have returned from their mission(s). This tool is integrated within the OpenQuake online system and will allow the input of photographs, digital forms and manual input of paper forms. A unique linkage system will allow linking of GPS photos to building footprints where users can establish links to photographs for each building surveyed. The photos are then analysed by engineers and experts who can manually populate the database through visual interpretation of the field footage.

These tools will promote ultimate flexibility for field teams who have limited numbers of trained analysts or engineers, by assigning inexperienced analysts to field photograph collection while engineers can use crowdsourcing as a data input technique through visual interpretation of the field photographs in the online web interface. The system may also in the future allow the public to submit their own GPS photos or for engineers to share their findings or invite fellow scientists to validate their interpretations through the social networking functionality of OpenQuake. This potential functionality is currently in the design phase.

2.3. Task 3. Mapping Scheme Design

Mapping schemes are the IDCT tool for generation of regional exposure information (GED level 2), and serve as a way to harness the best-available data for inventory generation at a regional scale. These can often utilise inventory data extracted from individual country- or region-specific research studies (Jaiswal et al., 2008). In the IDCT suite of tools, mapping schemes are generated using the Spatial Inventory and Damage Data (SIDD) tool, a desktop-based workbench for loading, editing, testing and creating mapping schemes - statistical representations of inventory data that are applied to homogenous areas of land use or building type.

The main functions of SIDD are to: 1) receive data from the remote sensing and direct observation tasks; 2) develop mapping schemes or access a library of prebuilt mapping schemes; 3) apply mapping schemes; 4) format and export data to GEM OpenQuake. SIDD serves as a conduit between base data sets derived from both ground observations and remotely sensed sources and a GEM compliant file. Data are loaded into SIDD from the remote sensing module (Task 1, Fig. 2.1) and the field data collection tool (Task 2, Fig. 2.1) and processed into a GEM-suitable file. Data from task 1 and/or task 2 are loaded into a local data repository on the user’s hard drive. In addition, the user has the option of loading land use data, or delineations of “homogenous land use zones”. Zones that are homogenous (areas which are sufficiently similar in terms of attributes such as density and land use) will be characterized by a mapping scheme.

Once data are loaded into SIDD, mapping schemes are identified or created. Mapping schemes are tools for the statistical inference of structural parameters given data from a specific area. Data may be based on regional defaults, expert opinion, survey data, or a combination of these factors. After the mapping schemes are assembled, users apply the mapping scheme and the results output, either as exposure data or consequence data. For the development of inventories, the application of SIDD is likely to be an iterative process (Fig. 2.5 and Fig. 2.6), with users checking the results carefully and adjusting both geographic and statistical data to develop a final exposure data set that is valid according to the best available combination of field-based and remotely-sensed data. Preliminary testing of the method has revealed users may need to iteratively edit the mapping schemes and base layers before posting the final results (Bevington et al., 2012). The SIDD system is designed as a workbench of tools, rather than a linear process, to facilitate these types of adjustments. Additionally, the protocol will recommend statistical checks to validate the data set.

SIDD will primarily be a tool for processing exposure data, with consequences data will be provided through a web-based crowdsourcing interface. Lastly, the data from SIDD will be posted to the GEM databases- the GED or GCD, hosted on the GEM OpenQuake web interface. In addition, mapping schemes will be geographically-referenced, as they take into account regional variations in building parameters, such as structural material or year of construction. It is anticipated that users will submit their mapping schemes to an open library hosted by GEM, so that other users can reuse and recycle mapping schemes for other areas, especially in geographically adjacent areas, for example, a detailed survey of building inventory for one region of Indonesia may be better than using expert opinion to develop inventory in another region of Indonesia, where no surveys are planned.

3. TESTBED SITES

Development of the IDCT tools is ongoing using a number of test sites that have been identified globally. There are four main testing phases: 1) Pilot Study 1 – Tool Development, due May 2012; 2) Pilot Study 2 – Beta Testing; (inventory), due July 2012; 3) Pilot Study 3 – Beta Testing (damage) due July 2012; and 4) Pilot Study 4 – Demonstration, due November 2012.

The tools are currently (as in April 2012) under development and being demonstrated on imagery and field data collected in Pisco, Peru, Messina, Italy, Bishkek, Kyrgyzstan, Padang and Mataram, Indonesia and Pylos, Greece. Protocols have been updated during these tests, and the validity of the sampling strategy protocol is also tested. The recent earthquakes of port-au-Prince, Haiti and Christchurch, New Zealand provide valuable imagery and ground-based validation data in which to test the remote sensing and crowdsourced damage assessment techniques. It is currently envisioned that the final demonstration of the full suite of IDCT tools in late 2012 will focus on the city of Istanbul, Turkey.

4. CONCLUSIONS

One of the most poorly understood areas of disaster risk management is the building stock information that exists in a region. The IDCT component of the Global Earthquake Model is tasked with generating a set of open-source software tools and supplementary user-friendly guides to aid the understanding of seismic risk globally. These tools are aimed at being globally generalisable, with characterisation of inventory stock at multiple geographic scales; from regional to per-building capture. Remote sensing plays

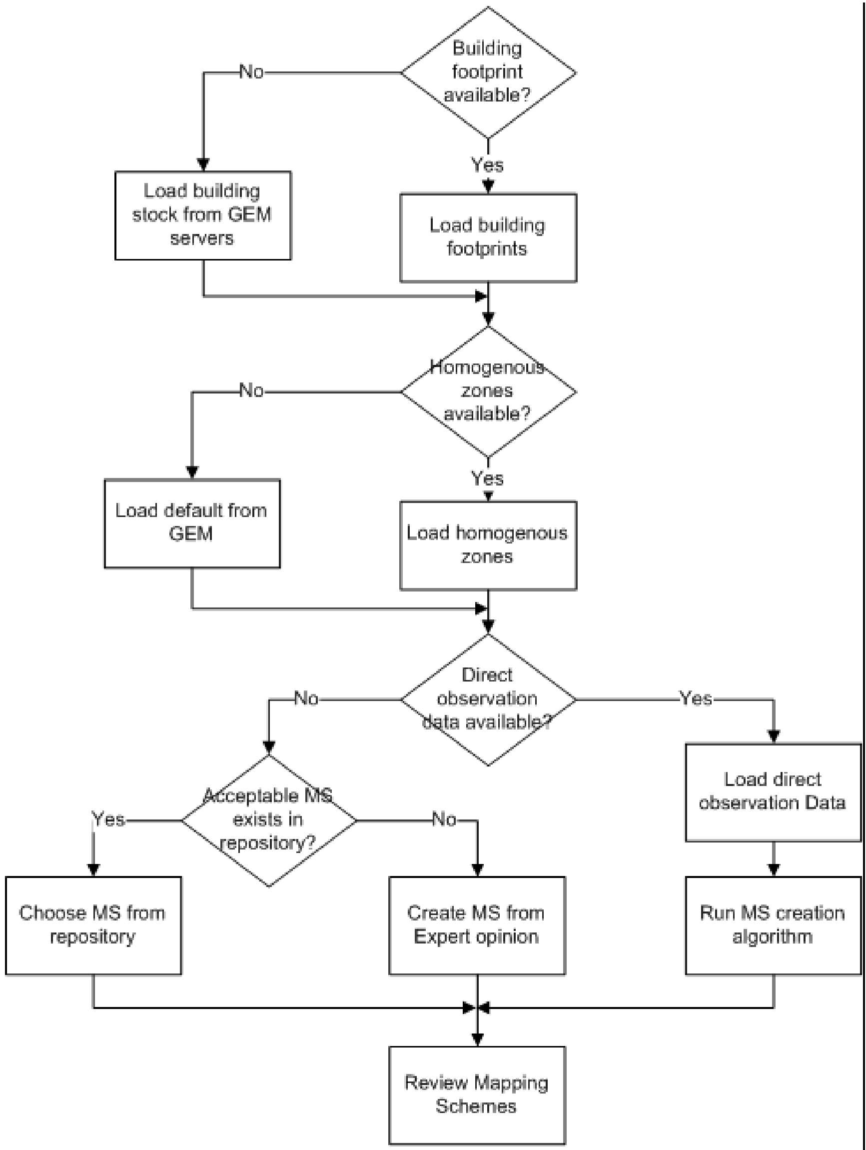

Figure 2.5. The workflow for creating a mapping scheme in the SIDD tool, the mapping scheme workbench for regional exposure development

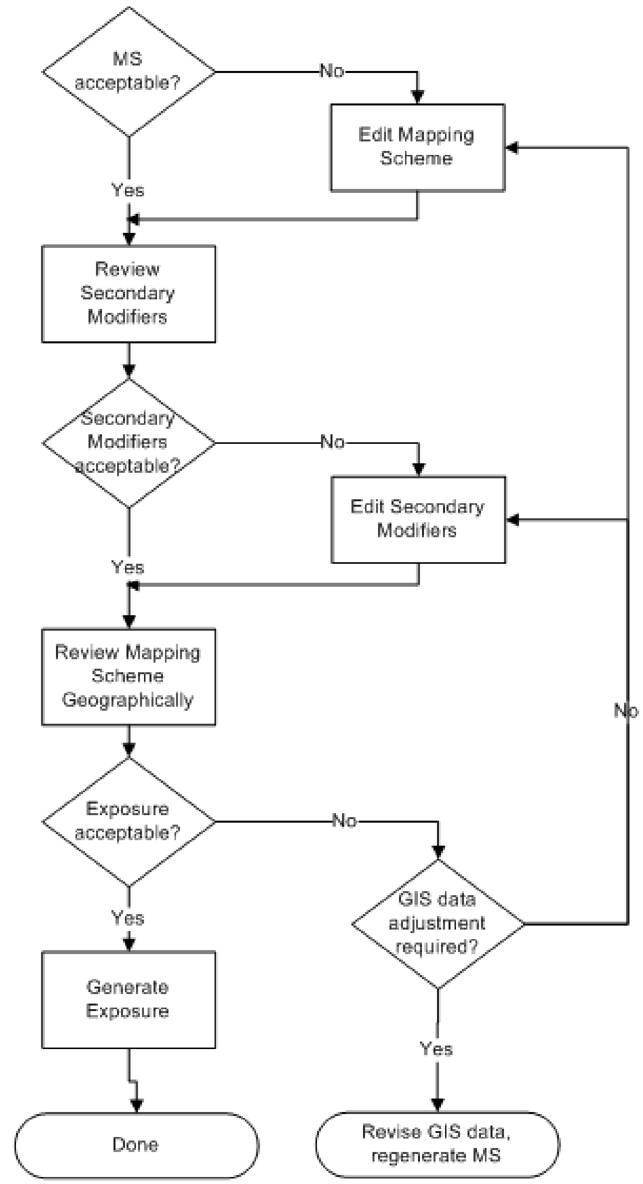

Figure 2.6. SIDD workflow, part 2

a major role in the pre-event inventory capture, but also in post-earthquake damage assessment. Emerging methods (e.g. crowdsourcing) and field capture technologies (e.g. tablet-based tools for survey teams) will be used to facilitate a range of domain experience levels. Field tools provide the majority of the inventory characterisation, but can be costly to develop for a full assessment of buildings over large areas. The development, application, and validation of mapping schemes are a primary function of the SIDD tool. These tools allow users to generate custom earthquake exposure databases and post-event consequences databases that can be made publically available through GED4GEM and GEMECD. The IDCT tools will be integrated with the GEM's OpenQuake web interface and will be released in 2013.

ACKNOWLEDGEMENT
The authors would like to thank the full IDCT team seismologists, engineers and geographers for their ongoing enthusiasm and commitment to this project. The support of the coordinators and staff at the Global Earthquake Model Foundation has been tremendous throughout. The authors would also like to recognise the support from the GEM Scientific Board and the IDCT internal advisory committee: Fumio Yamazaki, Friedemann Wenzel, Hannes Taubenboeck, Ed Parsons and Rowan Douglas.

REFERENCES

Barrington, L., Ghosh, S., Greene, M., Har-Noy, S., Berger, J., Gill, S., Lin, A. and Huyck, C. (2011) Crowdsourcing earthquake damage assessment using remote sensing imagery. *Annals of Geophysics* **54:6**, 680-687.

Bevington, J., Eguchi, R., Ghosh, S., Graf, W., Hu, Z., Amyx, P., Huyck, C., Huyck, M., Eguchi, M. and Vicini, A. (2012) *Indonesia Building Exposure Development*. Report for GEM IDCT Deliverable 13, Pilot Study 1. v.1.0.

Born, P. & Martin, W. (2006) Catastrophe Modelling in the Classroom. *Risk Management and Insurance Review* **9**, 219-229.

Coburn, A. and Spence, R. (1992) Factors determining human casualty levels in earthquakes- Mortality prediction in building collapse: *Proceedings of the Tenth World Conference on Earthquake Engineering*, Madrid, Spain, 5989-5994.

Conway, T., Zarinejad, N. and Conway, T. (2008) Raising the bar on Catastrophe data. The Ernst & Young 2008 Catastrophe Exposure Data Quality Survey. Ernst & Young LLP.

Erdik, M., Sesetyan, K., Demircioglu, M.B., Hancilar, U., Zulfikar, C., Durukal, E., Kamer, Y., Yenidogan, C., Tuzun, C., Cagnan, Z., Harmandar, E., (2010), Rapid Earthquake Hazard and Loss Assessment for Euro-Mediterranean Region, *Acta Geophysica*, **58**:**5**, 855 -892

Gamba, P. Dell'Acqua, F. Lisini, G. (2009) BREC: The Built-up area RECognition tool. 2009 *Joint Urban Remote Sensing Event*, Shanghai, China. 20-22 May 2009.

Ghosh, S., Huyck, C., Greene, M., Gill, S., Bevington, J., Svekla, W., DesRoches, R. and Eguchi, R., (2010). Crowd-Sourcing for Rapid Damage Assessment: The Global Earth Observation Catastrophe Assessment Network (GEO-CAN), *Earthquake Spectra* **27**, S179-S198.

Global Earthquake Model (2011) Global Earthquake Model 2009-2010 Report. Second edition.

Global Reinsurance (2008) Gustav/Ike losses significantly underestimated. *Global Reinsurance*. 9 October, 2008

Grossi, P. (2004) Sources, nature, and impact of uncertainties on catastrophe modeling. *13th World Conference on Earthquake Engineering*. Vancouver, B.C., Canada.

Huyck, C., Hu, Z., Bevington, J., Ghosh, S. and Eguchi, R. (2011) Inference of Structural Characteristics for Regional Building Inventories Using Remotely Sensed Data and Ground Observations. *Ninth International Workshop on Remote Sensing for Disaster Response*, Stanford, CA. 15-16 September 2011.

Jaiswal, K. and Wald, D., 2008, Creating a global building inventory for earthquake loss assessment and risk management: *U.S. Geological Survey Open-File Report 2008-1160.*

Monelli, D., Pagani, M. and Weatherill, G. A. (2012). The hazard component of OpenQuake: the calculation engine of the Global Earthquake Model. *Proceedings of the 15th World Conference on Earthquake Engineering*. Lisbon, Portugal, paper n. 4180.

Pinho, R. (2012) GEM: a participatory framework for open, state-of-the-art models and tools for earthquake risk assessment worldwide. *Proceedings of the 15th World Conference on Earthquake Engineering*, Lisbon, Portugal, paper n. 4929.

Silva, V., Crowley, H., Pagani, M., Monelli, D., Pinho, R. (2012) Development and application of OpenQuake, an open source software for seismic risk assessment. *Proceedings of the 15th World Conference on Earthquake Engineering*, Lisbon, Portugal, paper n. 4917.