End-to-end demonstration of the Inventory Data Capture Tools

GEM TECHNICAL REPORT
2014-06 V1.0.0

DATA CAPTURE TOOLS

Foulser-Piggott, R., J. Bevington, A. Vicini

End-to-end demonstration of the Inventory Data Capture Tools

Technical Report 2014-06

Version: 1.0.0

Date: January 2014

Author(s)*: Foulser-Piggott, R., J. Bevington, A. Vicini

(*) Authors' affiliations:

Roxanne Foulser-Piggott, Cambridge Architectural Research, UK

John Bevington, ImageCat, UK

Alessandro Vicini, ImageCat, UK

Rights and permissions

Citation advice

Foulser-Piggott, R., J. Bevington, A. Vicini (2014), End-to-end demonstration of the Inventory Data Capture Tools, GEM Technical Report 2014-06 V1.0.0, 38 pp., GEM Foundation, Pavia, Italy, doi: 10.13117/GEM.DATA-CAPTURE.TR2014.06.

Photo credits

http://www.globalquakemodel.org/

ACKNOWLEDGEMENTS

The authors wish to sincerely thank Massimiliano Pittore, Marc Wieland and Sergey Tyagunov at GFZ and Ulugbek T. Begaliev, Akjol, Janysh, Yzatbek, Bektursun, Samagan, Erlan, Bulutbek and Nurlan at INTuit for their help with the field test and support for this mission.

ABSTRACT

The Global Earthquake Model (GEM) aims to provide a set of tools and models for hazard and risk analysis. The fundamental aim of the GEM Inventory Data Capture Tools (IDCT) risk global component is to provide a set of software and accompanying user protocols to enable users to collect and modify building exposure information, which can be input into the Global Exposure Database (GED) and the Global Earthquake Consequences Database.

IDCT uses two primary data collection tools to generate these pre-event inventory and post-event damage data: 1) satellite and aerial remote sensing image analysis, and 2) direct observations by individuals on the ground using a set of mobile data collection tools.

This report describes the end-to-end testing of the suite of IDCT tools for 2 specific workflows:

1. Generation of GED Level 3 data: per-building inventory data
2. Generation of GED Level 2 data: statistically-aggregated inventory data

Keywords

Inventory Data Capture Tools; remote sensing tools; mobile tools; spatial inventory data development; field-survey; sampling strategy; mapping schemes; land-use maps; homogeneous zones

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iii

ABSTRACT iv

TABLE OF CONTENTS v

LIST OF FIGURES vii

LIST OF TABLES vii

1 Introduction 1

1.1 Aims 2

2 Study Location 3

3 Inventory Data Capture Tools 4

3.1 Remote sensing 4

3.2 Direct Observation 4

3.3 Mapping Schemes 4

4 Method 6

4.1 Remote Sensing 6

4.1.1 Data used 6

4.1.2 Method 6

4.2 Sampling 8

4.3 Direct Observation 9

4.3.1 Survey study area characteristics and description 10

4.3.2 Mission requirements 11

4.3.3 Data management 11

4.4 Outputs 13

5 Results and Validation 14

5.1 Using SIDD to Process Level 2 and Level 3 Exposure Data 15

5.2 Validation of Remote Sensing IDCT 18

5.2.1 Land-use maps 18

5.3 Comparison and Validation of Level 2 and Level 3 Exposure Data 19

5.3.1 Level 3 Field data validation 19

5.3.2 Building Distributions 19

5.3.3 Modal building types 22

5.4 Utility of Collecting Level 3 Exposure Data 23

6 Conclusions..24
6.1 Recommendations..24
REFERENCES..26

LIST OF FIGURES

Page

Figure 1.1 Detailed workflow of the suite of Inventory Data Capture Tools 2

Figure 2.1. Overview of study areas. OpenStreetMap data overlaid with high resolution WorldView-2 satellite image (© European Space Imaging/DigitalGlobe) 3

Figure 4.1. Overview of the building clusters locations and examples of delineated footprints 8

Figure 4.2. Locations of the surveyed buildings 10

Figure 5.1. SIDD output containing aggregated exposure data and building-by-building survey locations 16

Figure 5.2. Figure 5.1 magnified to Level 3 study area, showing overlay of Level 2 data output to RS IDCT defined homogeneous zones of land-use, SIDD Level 3 data output to GED4GEM grid cells and Level 3 survey points. 17

Figure 5.3 Land use maps derived from remote sensing (GFZ - left and IDCT - right) 18

Figure 5.4 Occupancy categories distribution by GFZ (left) and IDCT (right) 19

Figure 5.5. Mapping scheme creation in SIDD 20

Figure 5.6. SIDD building type distributions in (from top left, clockwise): RS IDCT Land-use zone C2; RS IDCT Land-use zone PDI; RS IDCT Land-use zone MRD; PAGER residential and non-residential building distributions; RS IDCT Land-use zone HR 21

Figure 5.7. Comparison between Level 2 and Level 3 building type distributions for MRD land-use class 22

Figure 5.8. Level 3 exposure data information 23

LIST OF TABLES

Table 4.1 Main occupancies identified from remote sensing in the Area Of Interest 7

Table 4.3. Most commonly collected building attributes 12

Table 5.1. csv output of SIDD for Level 3 data 16

Table 5.2. Modal building types from different sources of exposure data information. 22

1 Introduction

The Global Earthquake Model (GEM) Foundation is an international public-private initiative aimed at development of state-of-the-art global 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 a web-based platform, OpenQuake, which will be powered by a model calculation engine (Silva *et al.*, 2012, Monelli *et al.*, 2012). Combined, the GEM Risk Global Components 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.

The GEM Risk Global Components are 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 (GEMECD). 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 GEM Building Taxonomy v2.0 (Brzev *et al.*, 2013).

The GEM global exposure database requires the collection of building inventory information at several geographic scales, from Level 0 (country) to Level 3 (per-building). The GED4GEM consortium is collating global datasets for inventory characterisation at Levels 0 and 1, including global housing and population data. These are stored in the Global Exposure Database.

The Inventory Data Capture Tools Global Component was charged with developing a suite of open-source tools and user protocols for collection of inventory data at Level 2 and 3 (regional to per-building level precision). These data populate the GED and Earthquake Consequences Database (ECD), respectively. IDCT uses two primary data collection tools to generate this pre-event inventory and post-event damage data: 1) satellite and aerial remote sensing image analysis, and 2) direct observations by individuals on the ground using a set of mobile data collection tools.

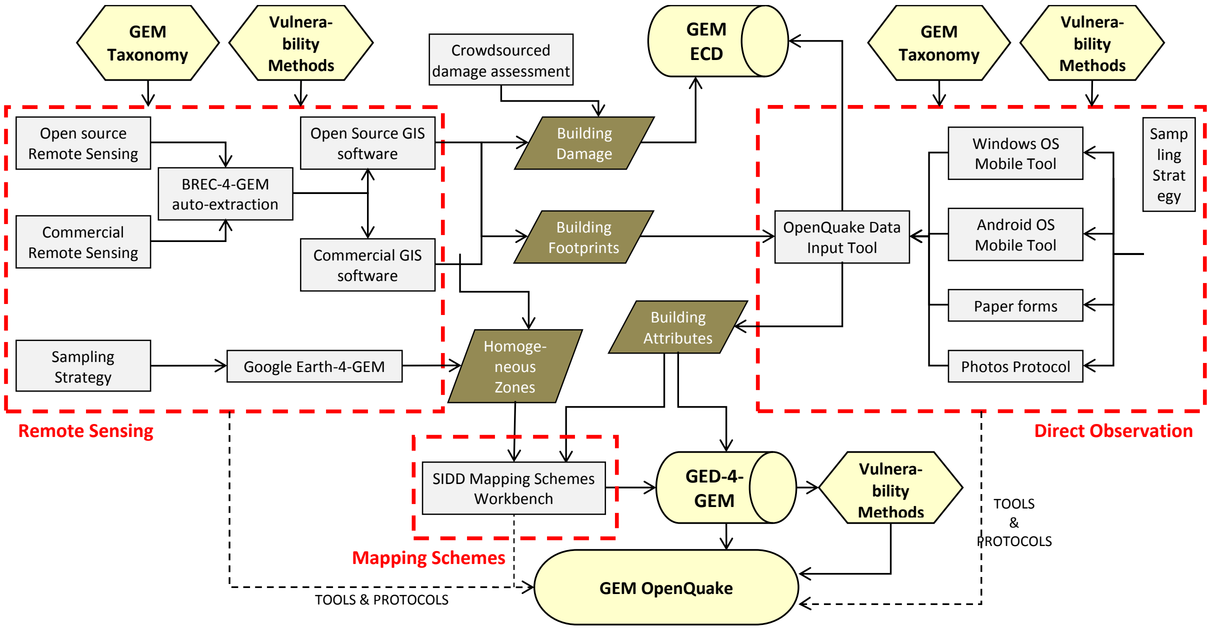


Figure 1.1 Detailed workflow of the suite of Inventory Data Capture Tools

1.1 Aims

This report describes the end-to-end testing of the suite of IDCT tools for 2 specific workflows:

- Generation of GED Level 3 data: per-building inventory data
- Generation of GED Level 2 data: statistically- aggregated inventory data

By providing an end-to-end test, the IDCT consortium is demonstrating the full workflow for the specific use case of generating inventory data for populating the Global Exposure Database. The main aims of this test are:

- To capture both Level 2 and Level 3 exposure data.
- To test remote sensing protocols to pre-process satellite imagery and extract information to be used in the sampling strategy delineation and the field data collection.
- To test functionality of the Android and Windows operating system (OS) Mobile Tools and paper tool for in-field inventory data collection.
- To test functionality of the Mobile Tools for Windows tool in terms of basic data management requirements.
- To outline necessary pre-mission technical and non-technical requirements and suggestions for an effective and realistic sampling strategy.
- To demonstrate the utility of the Mobile Tools for the local engineers so they can use them in future projects and contribute to GEM.
- To generate an input for the Global Exposure Database

2 Study Location

The study site selected for the complete workflow testing for the IDCT suite was the city of Bishkek, capital of Kyrgyzstan. The city covers an approximate area of 127 Km^2 with a population of 874,400, and presents a mixture of Soviet-era public buildings and apartment blocks, surrounded by smaller, privately built houses.

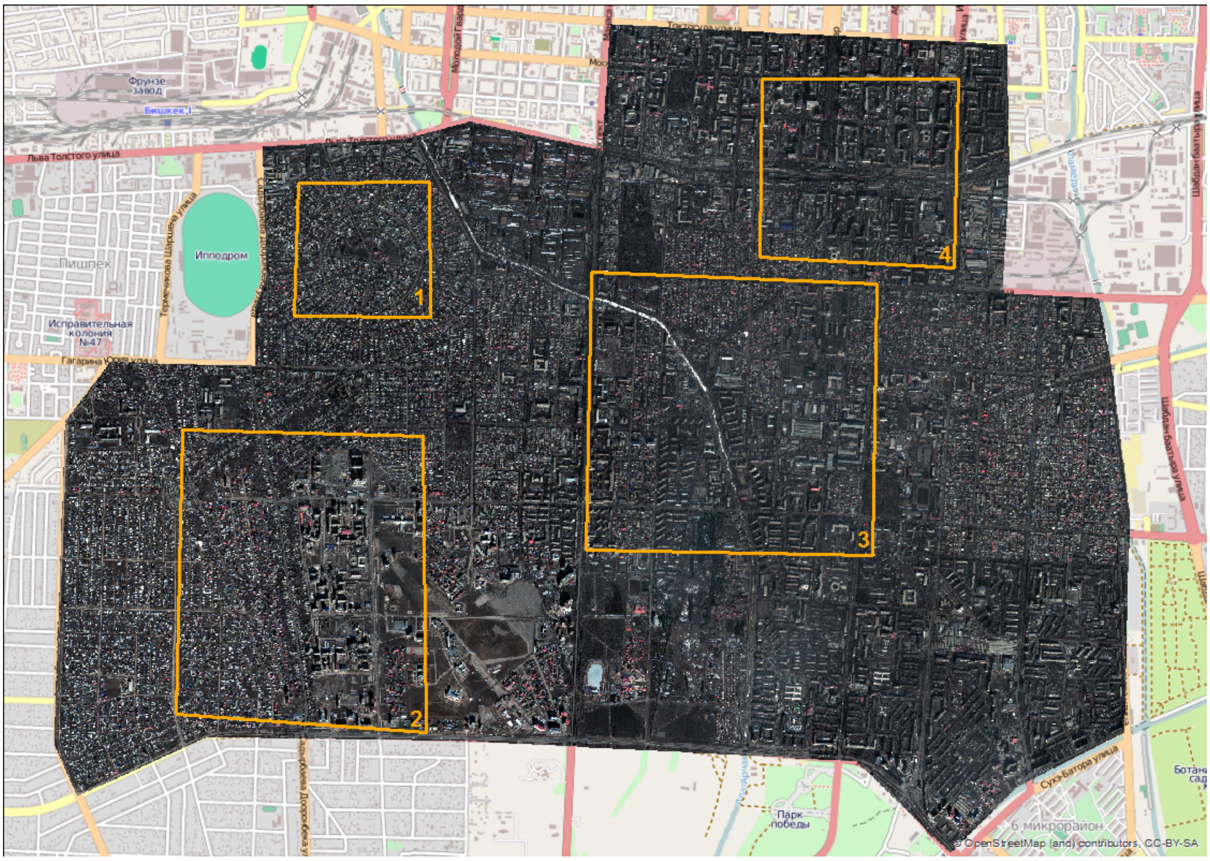


Figure 2.1. Overview of study areas. OpenStreetMap data overlaid with high resolution WorldView-2 satellite image (© European Space Imaging/DigitalGlobe)

A high spatial resolution WorldView-2 satellite imagery acquired on the 19th February 2013, covering an area of 28 km^2 of the central districts of the city was used as a base for the extraction of remotely sensed information and to define the areas for the survey. Four study areas were visited as shown in Figure 2.1. In areas 1 – 3 a sample of the buildings was surveyed. In area 4 a building-by-building survey was conducted. The areas were defined considering the range of building types observable, availability of validation data and ease of access.

The field team from Cambridge Architectural Research, ImageCat and GFZ was assisted during the survey by engineering collaborators the Earthquake Model for Central Asia (EMCA) and local counterparts from the International University for Innovation Technologies, Bishkek. The team was made up of trained structural engineers and engineering and GIS students.

3 Inventory Data Capture Tools

Tools developed by the IDCT Global Component are designed to offer a flexible data collection solution. Options for usage are tailored based on user experience, availability of analysts/engineers, and availability of hardware. All tools are open-source and available for free usage from GEM NEXUS. The tools are described in Figure 1.1 and briefly in this section according to the three major thrust areas.

3.1 Remote sensing

IDCT has produced a detailed set of protocols for the extraction of structural information from remotely-sensed imagery (Vicini *et al.*, 2014); both aerial and satellite photographs. These protocols describe the collection, pre-processing and feature extraction steps required to develop data from raw imagery. All outputs from the remote sensing tools take the form of GIS-compatible geo-information layers and structural attributes.

Two procedures were tested respectively for Level 3 and Level 2 data generation: extraction of individual building footprints (L3) and generation of maps of homogenous zones - areas of predominant land use used for development of mapping schemes.

3.2 Direct Observation

The DO component of IDCT has developed two digital data collection tools - the Mobile Tools - that run on Windows (Jordan *et al.*, 2014) and Android (Rosser *et al.*, 2014) operating systems(OS). Each of these tools is released pre-populated with the GEM Building Taxonomy, allowing a standardised and systematic collection of building attributes anywhere in the world. Paper forms provide the third data collection medium, also with a representation of the Taxonomy to allow rapid collection using a relatively low-tech collection method. Alongside the digital and paper tools sits procedures for selecting sample sizes for Level 2 building sampling as well as user guides for capturing and inputting photographic and other media that informs the inventory generation.

Testing of the two digital tools - for Windows and Android OS - as well as the paper forms is described in this report. A more in-depth assessment of each of the features and functions of the tools is described in the sister report produced by the IDCT consortium (Foulser-Piggott *et al.*, 2013).

3.3 Mapping Schemes

The final stage of Level 2 data development is the development of Mapping Schemes for the generation of statistically-aggregated distributions of exposure data. The Spatial Inventory Data Developer (SIDD) is a workbench to generate this types of data by use of the best-available input data in the form of attributes from a sample set of buildings, homogenous zones (from maps or remote sensing) or population statistics from census. SIDD facilitates the development, testing and iteration of mapping schemes, either from scratch or from previously published Mapping Scheme distributions (from GEM, USGS PAGER or other sources). Data from the Remote Sensing and Direct observation thrust areas act as input data for SIDD, with outputs being GEM-compliant exposure data for populating the GED and undertaking risk assessments.

SIDD was tested using sampled data collected in Bishkek to generate Level 2 data. The procedures and validation of this approach are described in the following sections.

4 Method

This section describes the remote sensing, sampling and direct observation methods used to obtain the inventory data for Bishkek.

4.1 Remote Sensing

This section describes the pre-processing and analysis steps necessary to extract the base map and reference data used in the Mobile Tools, the sampling strategy and for the mapping schemes creation from remotely sensed imagery. A detailed description of all of these procedures can be found in the remote sensing user guide (Vicini *et al.*, 2014). Archive satellite image was pre-processed using open source software and analysed to define (1) homogenous zones, (2) counts of buildings, and (3) a set of individual building footprints to be used in the generation of mapping schemes in SIDD.

4.1.1 Data used

The image used in the test was a high spatial resolution WorldView-2 satellite image, from 19th February 2013. The image extent covered 28 km^2 in the central area of Bishkek. The image comprised of a panchromatic band image with 50cm resolution and four bands (Red, Green, Blue, Near infrared) with 2m resolution.

Part of the pre-processing stage required the use of a base map for the geo-referencing, based on satellite images provided by Google Maps.

4.1.2 Method

The raw satellite data was processed according to the workflow identified in the Remote Sensing protocol developed by the IDCT consortium (Vicini *et al.*, 2014). The workflow covers all the steps necessary to derive from the raw imagery a set of geospatial data to be used as a base for field data collection, sampling strategy and mapping schemes creation. The processing was carried out using the combination of open source GIS and remote sensing software supported by the protocols, QuantumGIS with the GRASS toolbar plugin.

The raw satellite image was enhanced via two separate processes: pan-sharpening, to augment the spatial resolution of the image, and geo-referencing, which results in a more accurate image position relative to the Earth surface. Pan-sharpening is a data fusion process to derive an image characterised by high spatial and spectral resolution, from the combination of two separate images of the same area. In this case, the very high resolution (50 cm) WorldView-2 panchromatic band was combined with the corresponding multi-spectral (red, green, blue and near infrared bands) image covering 4 bands of the visible and near-infrared regions of the electromagnetic spectrum, with a spatial resolution of 2 m. The resulting image was ideal to represent in high spatial and spectral detail the study area.

All the parameters necessary to run the pan-sharpening function are explained in detail with examples, resulting in a fairly straightforward use of the algorithm. During the geo-referencing process, couplets of Ground Control Points in the satellite image and in the base map were used to correct the geographic position of the satellite image. The IDCT Remote Sensing protocols explain how to use the geo-referencing tool in QuantumGIS, where the corresponding points in both images are selected, defining the transformations needed to perform the geo-referencing. The GCPs selected to geo-reference the Bishkek image were mainly corners of buildings at ground level, to minimise the influence of perspective in the transformations.

The creation of Level 2 inventory data through mapping schemes required a set of building footprints and a set of polygons representing homogenous zones in the area of interest (AOI). Following the IDCT Remote Sensing protocols, five main occupancies were identified for the AOI:

Table 4.1 Main occupancies identified from remote sensing in the Area Of Interest

Medium Residential Density (MRD) A township with equal amount of land used for open space as is used for housing	
High Residential Density (HRD) Predominately reserves for high residency density in a cities and metropolitan areas with 85% and up impervious surfaces.	
Commercial (C2) Large structures, shopping centres, grocery markets, business offices, hotels, resorts, amusement parks, and commercial sports arenas.	
Public Domain / Institution (PDI) All school/ universities, government buildings, plazas squares, and public recreational areas(and depending on the circumstances cemeteries & landfills may be part of PDI or LRD or OS)	

Open Space Includes agricultural land, forest, nation reserves, inland bodies of water and other natural landscapes	

A series of neighbourhood-sized polygons were delineated in the AOI and classified according to the predominant occupancy identifiable from the satellite imagery. Indications and examples on how to identify different occupancies from visual assessment of the images are provided in the IDCT RS protocols, alongside a list of useful websites to aid with this type of identification.

A set of building footprints was manually delineated following the instructions in the protocols. The buildings selected for this step have been identified from visual assessment of occupancy and height to comply with the sample characteristics suggested in the sampling strategy document.

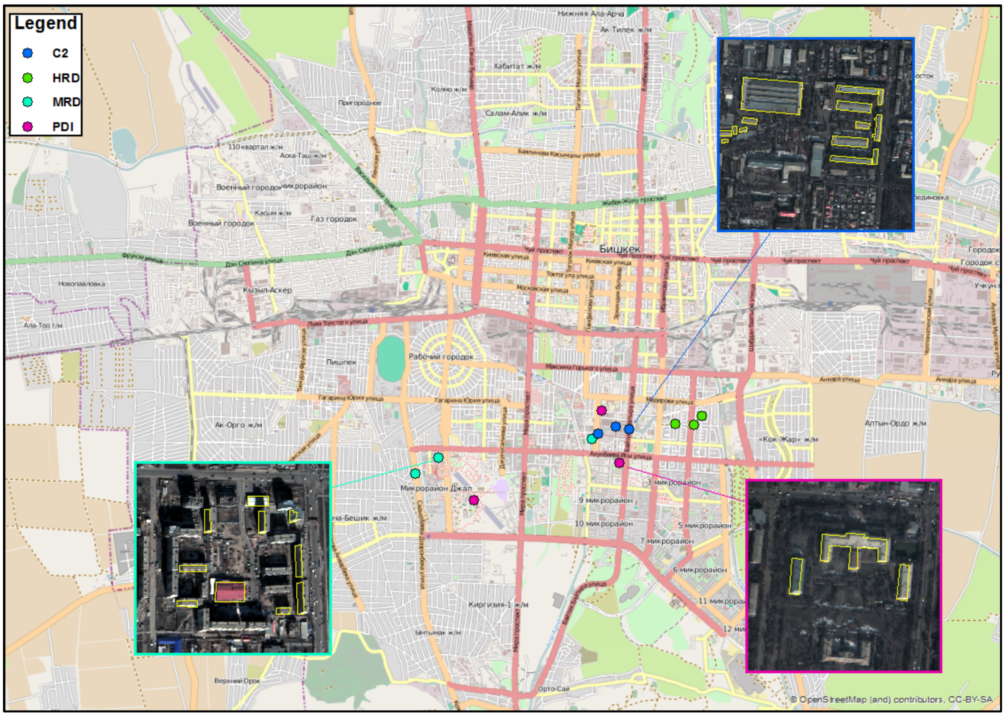


Figure 4.1. Overview of the building clusters locations and examples of delineated footprints

4.2 Sampling

The sampling strategy document developed by IDCT (Porter *et al.*, 2014) explains how to define a significant sample of buildings to be surveyed to contribute to the creation of mapping schemes for the Area Of Interest. Given the heterogeneity of the buildings stock in terms of height, the suitable approach to follow was deemed to be a stratified sampling without prior expert judgement.

The homogenous zones polygons defined in the remote sensing pre-process were used as a first stratification to identify the sample of buildings. Each homogenous zone was visually inspected on the satellite image and three buildings height thresholds were identified, based on the approximate percentage of buildings lower than each value. For the Medium Residential Density, for example, the identified thresholds were: less than 2 stories, 2 to 5 stories and more than 5 stories.

Buildings in each homogenous zone were grouped in clusters with average height equal to the defined thresholds, formed by a constant number of buildings. However, different numbers of buildings across different occupancies have been selected to reflect the varied distribution of buildings in each occupancy class. From each one of the homogenous zones, three (one per each average height threshold) were selected among all the delineated clusters, forming the strata for the buildings sample. Each building was attributed with values to indicate Object number, Cluster number, Occupancy class and the relative footprints have been digitised to allow the calculations of square footage in SIDD.

The remote sensing and sampling phases produced the complete set of data necessary to perform the direct observation survey:

- Pan-sharpened high resolution satellite image for the AOI in Bishkek, to be used as a base map for the DO tools.
- Homogenous zones polygons, to be used in the definition of the sample of buildings to be surveyed, including a count of all the buildings in each zone to be used in the creation of mapping schemes.
- A set of building footprint polygons, used in the DO tools to identify the buildings to be surveyed and to serve as an indication of building square footage for the creation of mapping schemes.
- A total number of buildings in the AOI and in each of the homogenous zones.

To be used effectively in the field, the Mobile Tools need to be loaded with the resulting GIS layers from the remote sensing analysis as well as building locations defined by the sampling strategy. The Windows Mobile Tool is built on a customised GIS interface that allows for any standard raster and vector format to be imported and used seamlessly to the user. The Android Mobile Tool is based on a render engine derived from web applications, and requires data to be in compatible formats.

The satellite imagery was saved as a 3-bands natural colour GeoTIFF file to be used in the Windows tool and converted to PNG tiles for the Android tool using MapTiler, a freeware software able to translate many image formats to Google Maps and Google Earth compatible tiles. The OpenStreetMap base map for the AOI was compiled using a freeware software called Mobile Atlas Creator (MOBAC), that allows the user to extract a subset of the OSM database and create standard format imagery and tiles at different levels of zoom. More information on these procedures can be found in Rosser *et al.* (2014) . All the vector GIS data were converted to KML format to be displayed on the Android tool and imported in the Windows tool in the standard ESRI shapefile format.

4.3 Direct Observation

This section describes the methodological aspects of the field test that are relevant to the overall performance of the complete set of IDCT and the results obtained. This document does not provide details on the functionality of tools, for a detailed description of the field test, the reader should consult Foulser-Piggott et al. (2013).

4.3.1 Survey study area characteristics and description

Four study areas were visited as shown in Figure 2.1. In areas 1 – 3 a sample of the buildings was surveyed. In area 4 a building-by-building survey was conducted.

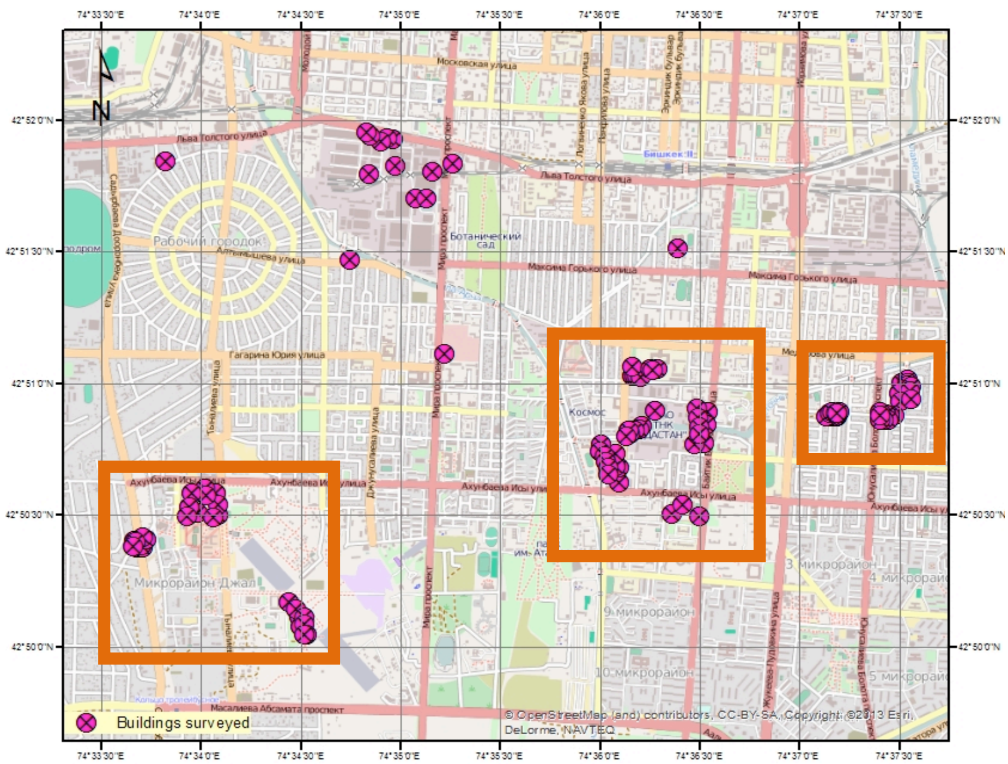
For the Level 2 exposure data, the workflow for the collection of data is as follows:

- Identification of homogeneous land use zones within the 2 areas of interest.
- Application of a stratified sampling technique to identify buildings to be surveyed.
- Survey of sampled locations using the Direct Observation (DO) tools.

The surveyed buildings can be divided into three main study areas: West, centre and East, shown in Figure 4.2 (delineated using orange boxes). The buildings shown clustered in the North of the map were added to the main study areas in order to test the sampling strategy. Two teams collected data in these areas:

- One team recorded data for all buildings in the three study areas (84 buildings) on the Android and paper tools
- The second team visited a section of the western study area during the tool familiarisation session and collected data on Windows, Android and paper tools.

Figure 4.2. Locations of the surveyed buildings



For the Level 3 exposure data, a building-by-building survey was conducted in area 4 (Figure 2.1). The workflow for the collection of data is as follows:

- The BREC footprint extractor was used to generate building footprints and attributes at a per-building level.
- All buildings (100%) in the study area were surveyed.
- The Windows, Android and paper Direct Observation (DO) tools were used.

4.3.2 *Mission requirements*

An important deliverable for this project is a clear statement of non-technical and technical pre-mission and during mission requirements. Additionally, a successful survey depends to a great extent on a well thought out sampling strategy. These requirements are outlined in this section.

Non-technical guidelines

Pre-mission planning is well described as part of the US Federal Emergency Management Agency's FEMA-154 (FEMA, 2002a) and 155 (FEMA, 2002b) procedures and it is recommended that these documents should be referred to before undertaking field data collection. Particularly, the recommendation to gather as much information as possible pre-mission on the attributes of buildings in the location to be surveyed. The following additional points were noted both in the Athens and Bishkek studies (Foulser-Piggott *et al.*, 2013):

- It is useful to have official document from your institution stating the project you are working on and the purpose of the data being collected.
- *On average,* allow 15 minutes to do a full external survey of a building.
- The knowledge of local engineers is invaluable in both the pre-planning and in-field phases.
- Local engineers generally have good knowledge of normal local construction practices and are able to make logical and relatively accurate inferences when necessary.

Technical guidelines

- It is recommended that the imagery and vector data is loaded into the tool and checked in a location without internet access before starting a survey to simulate actual in-field conditions.
- Glare on screen is an ongoing issue for the tools.
- Where possible, multiple devices should be brought to the field and paper forms and maps carried as a backup.
- An in-depth knowledge of the GEM taxonomy is essential.

Sampling strategy

- It is important to perform a thorough visual assessment of the AOI to be able to accurately evaluate the stratification thresholds.
- The selected clusters of buildings should strike a balance between the need for diversity in the sample (include different areas of the AOI) and the capacity of the field team to reach the designated areas during the time allocated for the survey.
- It is necessary to carefully consider access issues to certain areas when selecting buildings to survey, or prepare a contingency plan with guidelines on in-field election of alternative structures to include in the sample.
- Applying a travelling salesman algorithm to the suggested sampling strategy is a quick way of calculating time and resources necessary to conduct a survey.

4.3.3 *Data management*

After the survey has been completed, data management of the survey points collected using the Android DO tool was conducted using the Windows tool.

During the survey, the collection of different levels of data was further investigated in order to understand how the number of fields in the database may be reduced for rapid assessment applications. The most commonly used fields during the survey are shown in the table below.

Table 4.2. Most commonly collected building attributes

COMMENTS			building has many different sections with mixed material types. Main building appears to be RC frame and is roughly rectangular. some shear walls present
PLAN_SHAPE	Rectangular, solid	Square, solid	Irregular shape
POSITION	Detached building	Detached building	Detached building
NONSTRCEXW	Masonry	Concrete	Unknown material
ROOFSYSMAT	Unknown roof material	Wood	Metal
ROOFCOVMAT	Unknown roof covering	Metal or asbestos sheets	Metal or asbestos sheets
ROOF_SHAPE	Unknown roof shape	Pitched and hipped	Pitched and hipped
ROOFSYSTYP		Wood, unknown	Metal beams or trusses supporting light roofing
MAT_TECH_L	Fired clay solid bricks	Concrete blocks, hollow	Concrete, unknown
MAT_TYPE_L	Masonry, unreinforced	Masonry, unreinforced	Concrete, reinforced
LLRS_T	Wall	Wall	Moment frame
LLRS_DCT_T	Non-ductile	Non-ductile	Ductile
STR_HZIR_P	No irregularity	No irregularity	Other horizontal irregularity
STR_HZIR_S	No irregularity	No irregularity	Torsion eccentricity
STR_VEIR_P	No irregularity	No irregularity	Other vertical irregularity
STR_VEIR_S	No irregularity	No irregularity	
STR_IRREG	Regular structure	Regular structure	Irregular structure
FLOOR_MAT		Unknown floor material	Concrete
FLOOR_TYPE			Concrete, unknown
STORY_AG_Q	Exactly	Exactly	Approximately
STORY_AG_1	5	1	2
STORY_AG_2	0	0	0
STORY_BG_Q	Approximately	Exactly	Approximately
STORY_BG_1	0	0	1
STORY_BG_2	0	0	0
HT_GR_GF_Q	Approximately	Approximately	Approximately
HT_GR_GF_1	0	0	0
HT_GR_GF_2	0	0	0
YR_BUILT_Q	Pre	Approximately	Approximately
YR_BUILT_1	0	0	0
YR_BUILT_2	2000	2010	2010
YR_RETRO	0	0	0
OCCUPCY	Residential	Residential	Commercial and public
OCCUPCY_DT	Residential	Residential	Commercial and public

The data collected by multiple survey teams on buildings in the study areas were merged successfully and exported to both comma-separated values (csv) and ESRI shapefile (shp). Additionally, the data collected on different tools were merged successfully and converted to csv and shp. It is interesting to consider how this information might be merged in future developments. If the points are not added at exactly the same coordinates, but are on the same building, a merge to building would be very useful. This issue may also arise if a person visits a region pre-event and collects building data and a second person visits after an event and collects consequences and damage data.

4.4 Outputs

The Level 2 information gathered from the Bishkek test was used to review the sampling strategy selection procedures and to review the integration of the mobile tools and remote sensing results into the Spatial Inventory and Damage Data tool for the generation of mapping schemes (see section 5).

The Level 3 has been compiled in the GEM database format and has been integrated directly into the Global Exposure Database (GED4GEM) component.

The field tests have also:

- Provided an opportunity to review the training documentation and media.
- Provided indicators on the performance of Remote Sensing IDCT in generating the Level 2 and 3 exposure data
- Aided the development of suggested methods for validating field data.

The following section describes both the use of SIDD to produce a GED4GEM-compatible set of results from the Bishkek mobile tool testing and the validation of the field testing results.

5 Results and Validation

This section describes the way in which both Level 2 and Level 3 exposure data captured as previously described in this document can be validated. The captured exposure data is compared with available building footprint and building attribute information to indicate the performance of the Remote Sensing and Direct Observation (DO) tools and protocols in generating accurate Level 2 and 3 exposure data.

The following data were produced or collected in the full test of the remote sensing and mobile DO IDCTs in Bishkek:

- Level 2 exposure data, including: homogeneous land use areas; data collected in the field using the mobile tools.
- Level 3 exposure data: building footprints digitised manually from imagery; data collected in the field using the mobile tools.

The exposure data produced and captured in this demonstration were processed using SIDD to obtain building distributions and main building types in each land-use class based on Level 2 and Level 3 data. The exposure created for Levels 2 and 3 were then compared.

As well as comparisons between the Level 2 and Level 3 datasets, the following datasets are used for an independent comparison with the IDCT results. However, as neither of these datasets are ground-truth data, this comparison will be of limited value.

- Land-use (and predominant building-type map) created by GFZ (Wieland et al., 2012)
- PAGER building distributions.

A true validation of IDCT captured exposure data would require ground-truth data for the study area. However, it is likely that in the majority of cases where IDCT is used to collect exposure data, ground-truth data will not be available. Therefore, this study explores basic methods of examining the accuracy of captured exposure data, including how captured exposure data can be compared with and validated using different exposure data sources.

For the Level 3 exposure data, the study area was treated as follows:

- All buildings (100%) in the study area were surveyed using Direct Observation (DO) tools
- Footprints of surveyed buildings were manually delineated.
- SIDD used to build exposure based on the Level 3 data, GED4GEM compatible grid, building footprints and mapping scheme.
- SIDD built exposure output as csv and shapefile as shown in Table 5.1 and Figure 5.2 respectively.

The following datasets were available for validation:

- The majority of building footprints in this area were previously digitised by GFZ and the remainder were digitised manually.
- PAGER building distributions
- Land use map (GFZ)

For the Level 2 exposure data, the 3 study areas were treated as follows:

- Homogeneous land use zones (based on IDCT RS protocol) produced. The methodology for this is described in section 5.1
- Stratified sampling technique to identify buildings to be surveyed applied, as described in Section 4.2.
- Footprints of buildings to be surveyed were delineated.
- Sample of buildings surveyed in the field using DO tools.
- SIDD used to build exposure based on the Level 2 data, homogeneous land-use zones, building footprints and mapping scheme.
- SIDD built exposure output as csv and ESRI shapefile formats, examples of shapefile output are as shown in Figure 5.1 and csv in Table 5.1.

The following datasets are available for validation:

- Land-use map (GFZ)
- Level 3 exposure data collected using mobile tools

5.1 Using SIDD to Process Level 2 and Level 3 Exposure Data

SIDD was used to aggregate Level 2 exposure data and output the exposure as both gridded values (using the GED4GEM compatible grid) and homogeneous zone values corresponding to land-use map zones produced using remote sensing IDCT. SIDD was also used to aggregate Level 3 data and output exposure as gridded values (using the GED4GEM compatible grid). Both the Level 2 and Level 3 SIDD outputs were passed to GED4GEM to form part of the global exposure dataset. The results presented in this section have been produced using the version 1.0.1 of SIDD, and therefore might contain minor differences from the taxonomy strings generated from v.2.0 of the Building Taxonomy.

Figure 5.1 shows:

- The GED4GEM grid cells class (blue polygons), within which the Level 3 building exposure data is aggregated in SIDD and defined as a single GEM taxonomy class.
- The homogeneous zones created using remote sensing IDCT (pink patterns), within which the Level 2 building exposure data is aggregated and defined as a single GEM taxonomy.

The attributes of the building exposure information within these zones and grid cells are discussed in more detail in later sections of this report.

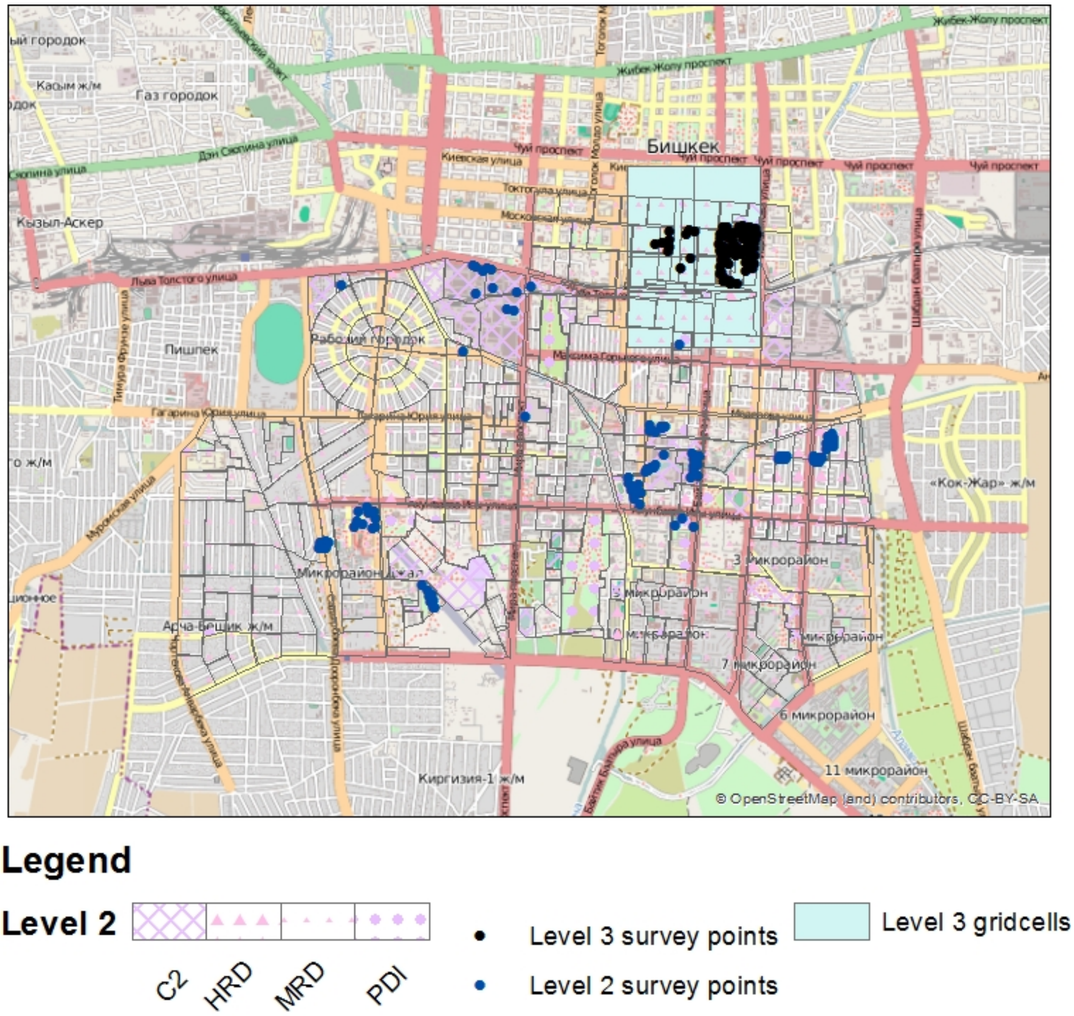


Figure 5.1. SIDD output containing aggregated exposure data and building-by-building survey locations

SIDD is intuitive to use and allows the user to explore characteristics of the exposure data collected. The data can be exported to both shapefile (the basis of Figure 5.1 and Figure 5.2) and csv formats (Table 5.1) and is therefore easy to view and analyse. SIDD can also be used to simplify the building exposure results, by using a mapping scheme based on a subset of the GEM taxonomy attributes, i.e. allowing the user to explore the distribution of building types on the basis of material type and lateral load resisting system. However a feature that appears to be missing and would be useful is the ability to export the SIDD calculated building distribution for different land-use classes.

Table 5.1. csv output of SIDD for Level 3 data

GID	LON	LAT	TAXONOMY	ZONE	NUM_BLDGS	AREA	REP_COST
586748952	42.87083	74.6125	MR+CLBRS+MOC/RC+RC4/FC+FC4/HEX:5/YBET:1960,1970/IRRE/RES+RES2F		1	0	0
586748951	42.8625	74.6125	CR+CIP/LFM+DUC/RC+RC1/FN/HEX:1/Y99/IRIR+REC/COM+COM5		1	0	0
586748952	42.87083	74.6125	ME+MEO/R99/FO/HEX:1/YBET:2000,2005/IRRE/COM+COM1		1	0	0
586748952	42.87083	74.6125	C99+CT99/LWAL+D99/R99/F99/HEX:1/YPRE:2005/IRRE/COM+COM5		1	0	0
586748952	42.87083	74.6125	M99+CLBRS+MR99+MOC/LWAL+DNO/R99/FC+FC1/HEX:3/YAPP:2005/IRIR+TOR+POP+CHV/COM+COM5		1	0	0

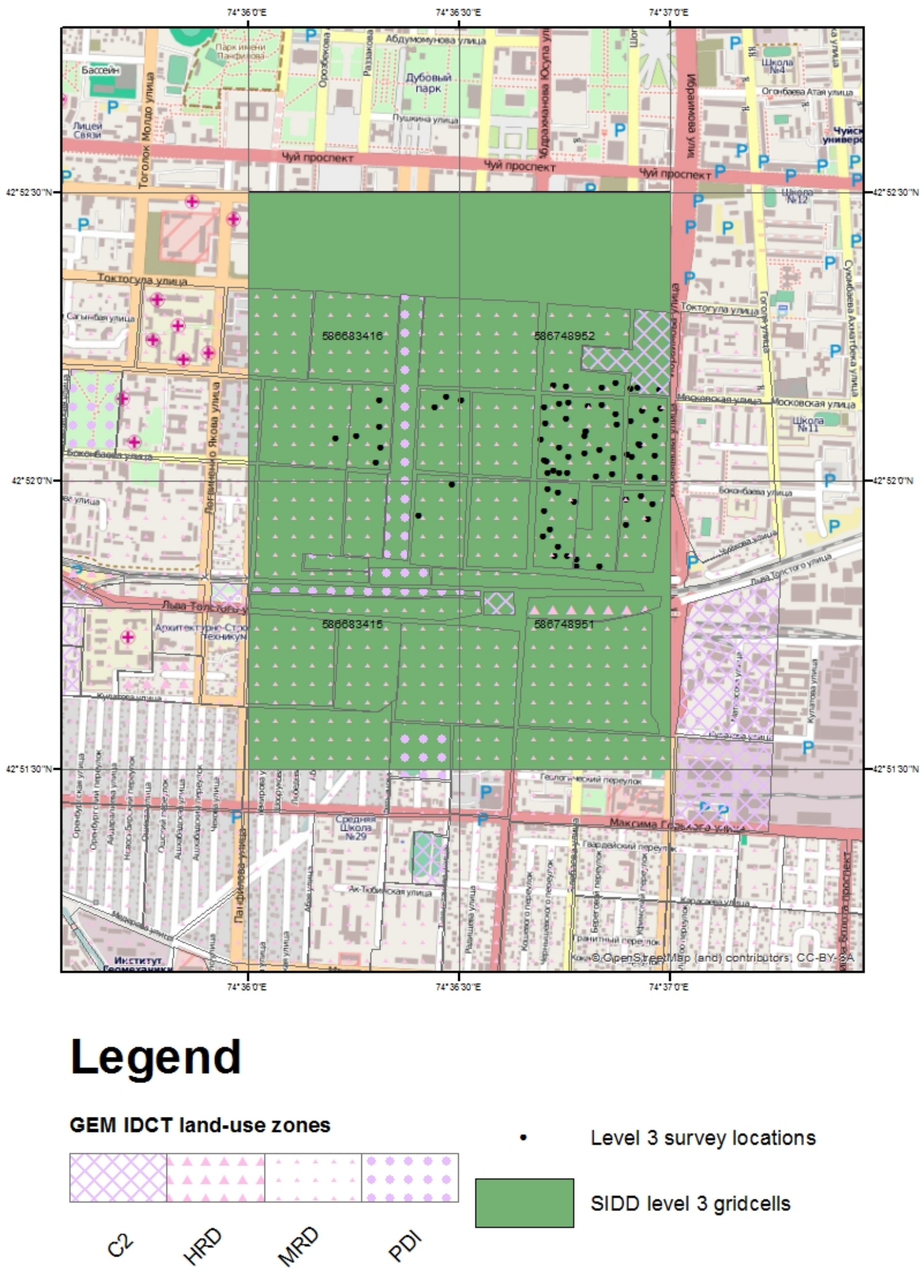


Figure 5.2. Figure 5.1 magnified to Level 3 study area, showing overlay of Level 2 data output to RS IDCT defined homogeneous zones of land-use, SIDD Level 3 data output to GED4GEM grid cells and Level 3 survey points.

In using SIDD to manage exposure data, some shortcomings of the GEM taxonomy were also noted. The first is that the GEM taxonomy does not lend itself to comparisons of building types found in different areas due to the level of complexity of the building descriptor string. For example, it is difficult to compare building types on a simplified basis, e.g. 50% of buildings in Area A are masonry compared with 30% in Area B. This is well demonstrated by the x-axes in Figure 5.6; the GEM taxonomy output string is long, complex and without decoding does not provide the user with useful information on building types. This is in contrast to PAGER (also Figure 5.6), where material type can be identified by the first one or two letters.

5.2 Validation of Remote Sensing IDCT

5.2.1 Land-use maps

As part of the validation process, the Homogenous Zones map derived from satellite imagery was compared with the land use map produced by GFZ. The two maps presented differences in coverage, classes analysed and in the classification technique, with the IDCT map derived from visual inspection of a high resolution satellite image and the GFZ one based on automatic classification from ground surveys. Furthermore, the GFZ map covered a considerably larger area and was therefore clipped to the extent of the satellite image used as a base for the IDCT analysis.

The GFZ analysis considered four classes, RES1 (single family dwelling), RES2 (multi-family dwelling), COM/IND (commercial/industrial) and a mixed use class where was not possible to identify a predominant one. IDCT classified in four slightly different categories, MRD (medium residential density), HRD (high residential density), PDI (public domain / institutions), C2 (commercial). To make the maps comparable, the different residential classes (RES1, RES2; MRD, HRD) have been aggregated and colour coded to represent a generic residential class (blue shades in the map below). The red areas in both maps represented the commercial/industrial class, and the pink ones the PDI.

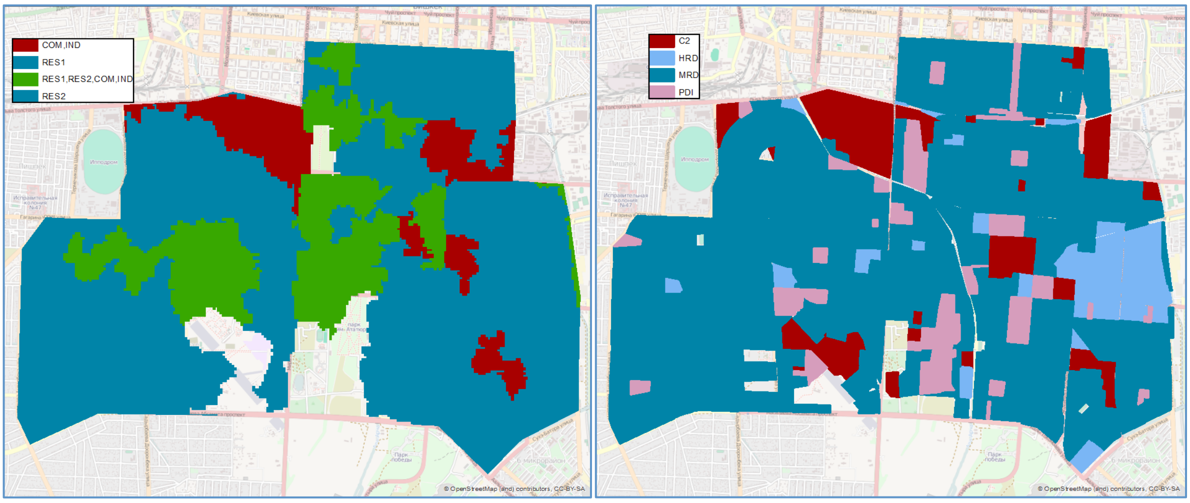


Figure 5.3 Land use maps derived from remote sensing (GFZ - left and IDCT - right)

The percentage of total area of each class was used as a comparison measure, resulting in 1% difference for commercial, and 18% in residential. The mixed use class of the GFZ land use map includes both the residential classes, and, if considered in the residential areas tally, the difference between the two classifications is reduced to 4%. Therefore the two classifications presented comparable results notwithstanding the difference in the approach taken to compute the statistics. The variations between the results can be reduced further with the adoption of a standardised (or comparable) set of occupancy/land use classes.

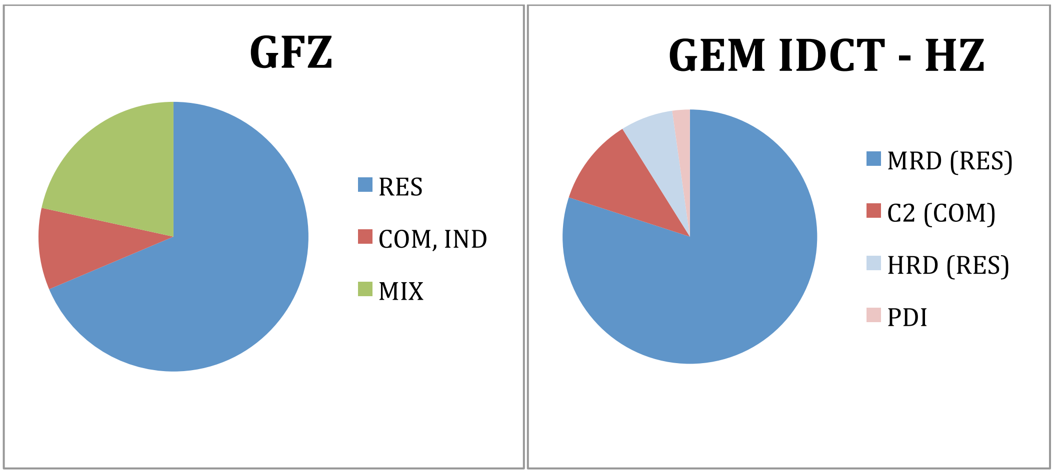


Figure 5.4 Occupancy categories distribution by GFZ (left) and IDCT (right)

5.3 Comparison and Validation of Level 2 and Level 3 Exposure Data

5.3.1 Level 3 Field data validation

- In this study, as well as in future studies, the Level 3 exposure data can be used as an important comparison for the Level 2 exposure data, and as a validation for some outputs of the remote sensing IDCTs. In order to ensure that the use of the Level 3 data as 'ground-truth data' is valid, the quality of the Level 3 data should be confirmed. The first step in the validation of Level 3 data is to use quality control within the teams in the field collecting data. The following observations on the quality of data captured were made following the full test of the tools in Bishkek:
    - **Expert judgment.** In order to collect accurate data on building types, local expertise should be sought to improve the accuracy of assignments. As building type assignment involves expert judgment, multiple surveys of the same building by different team members would help understanding of the uncertainty in the building type assignment.
    - **Unknown and uncertain elements of the taxonomy.** Certain elements of the taxonomy can be interpreted in different ways. This is particularly problematic when the taxonomy is used by non-native English speakers. As the use of an unfamiliar taxonomy can cause errors, a solution is to record the building types using both a local or locally familiar taxonomy or code (e.g. PAGER) and the GEM taxonomy. A comparison of the two recordings of building types would demonstrate if any terms in the taxonomy have been misinterpreted and would allow results to be corrected.

5.3.2 Building Distributions

Using SIDD, two distributions of building types were generated, one based on Level 2 data and the second on Level 3 data. The Level 2 output is building distributions for particular land-use classes (MRD, HRD, PDI and C2: as previously defined in Section 5.1 and the Level 3 output is a building distribution for the MRD land-use class. Additionally, the PAGER distribution of buildings in the urban-residential class and urban non-residential class (N.B. for Kyrgyzstan, these distributions are the same) was used as an independent comparison.

The Level 2 and 3 output distributions of building types in each land-use zone were obtained using the Level 2 exposure data, homogeneous land-use zones, building footprints and a mapping scheme. In order to provide a simplified distribution of the building types in each land-use zone, the mapping scheme was defined using four attributes of the taxonomy, namely: age, material, lateral-load resisting system and occupancy. The method of selecting the mapping scheme is demonstrated in the screen shot in Figure 5.5.

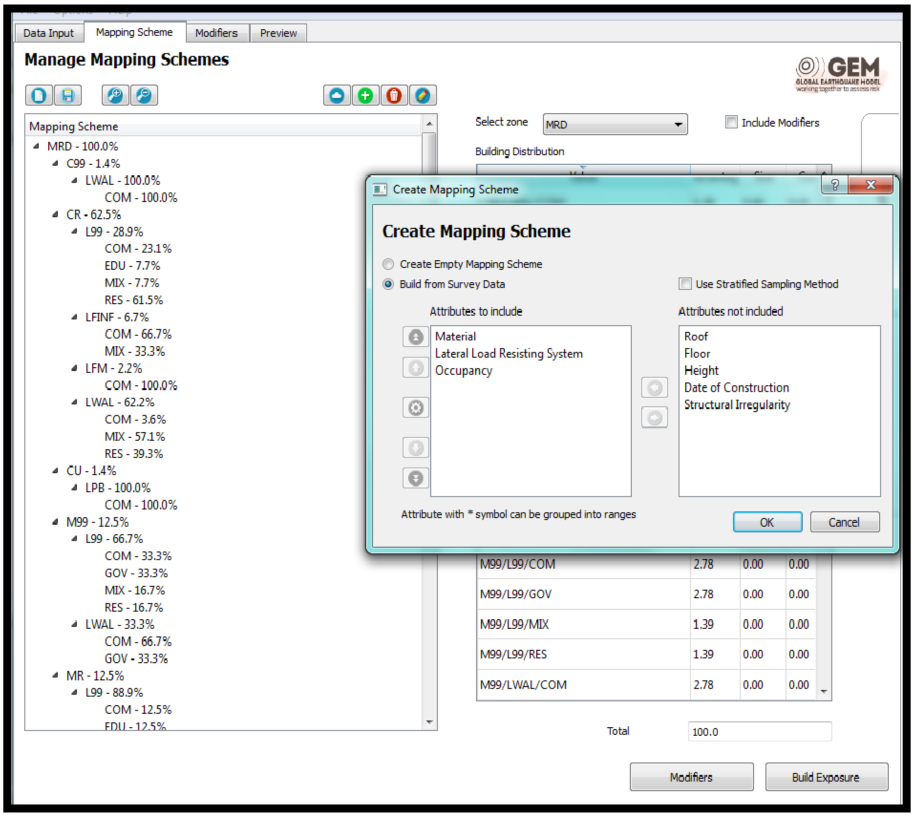


Figure 5.5. Mapping scheme creation in SIDD

The Level 2 SIDD output building distributions (Figure 5.5) were used to plot the graphs shown in Figure 5.6. The PAGER distribution of building types (using the PAGER classes mapped to simplified classes from the GEM Building Taxonomy (Brzev *et al.*, 2013)) is also provided for comparison. The figures show the distributions of buildings in each land-use zone and demonstrate the following:

- The SIDD output suggests that whilst there is variation in the distribution of building types in each land-use zone, similar building types are present in all zones, e.g. MUR – unreinforced masonry of various ages.
- Differentiating between MRD and HRD appears worthwhile as there is a difference in the range of building types in these zones.
- The distribution of building types in PAGER is very different in constituent building types and proportions to the distributions produced by SIDD based on Level 2 data. This is discussed further in the next section.
- There are shortcomings (mentioned previously) of using the GEM taxonomy as a way of defining buildings, particularly when only comparisons of basic building characteristics are required.

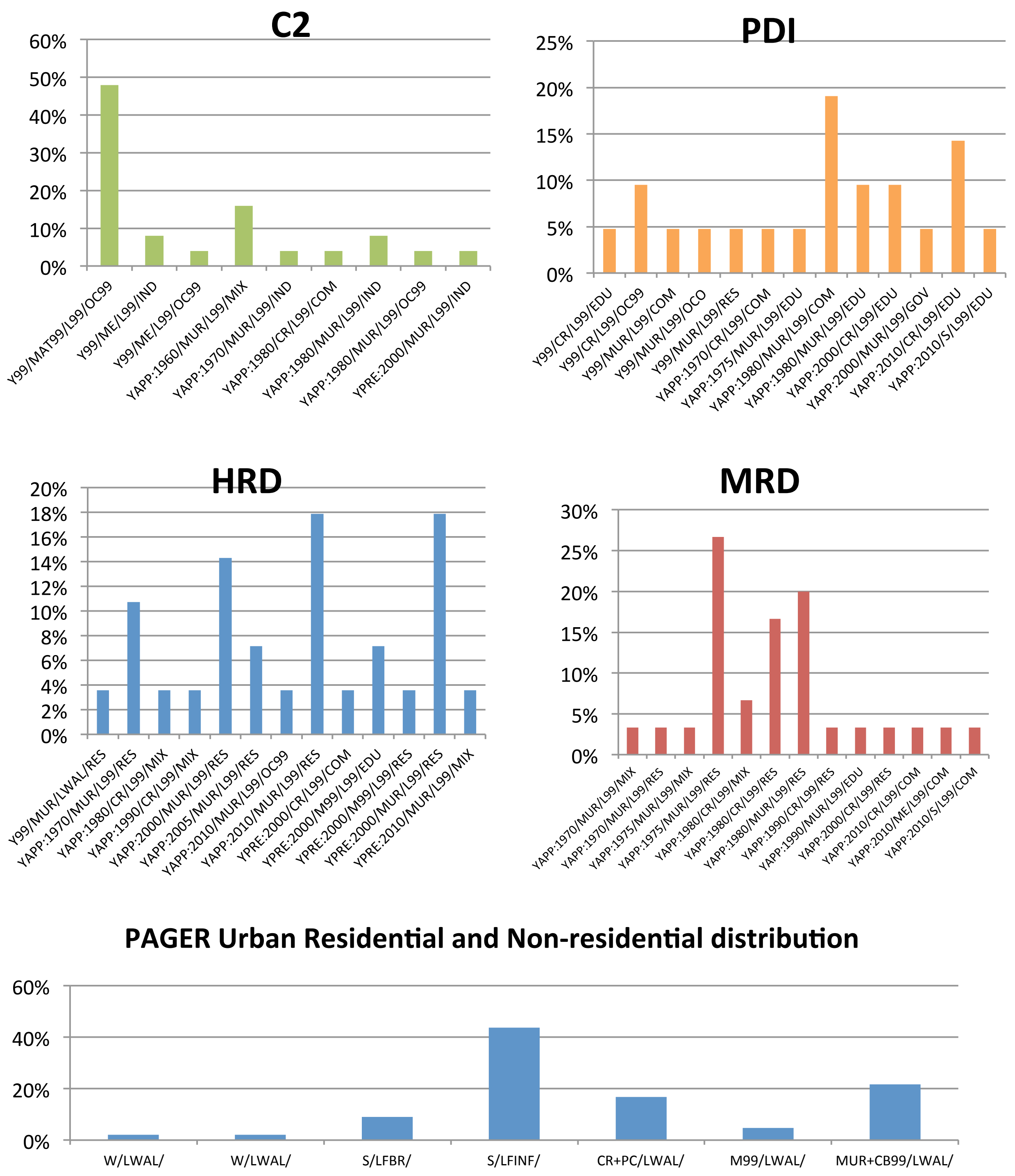
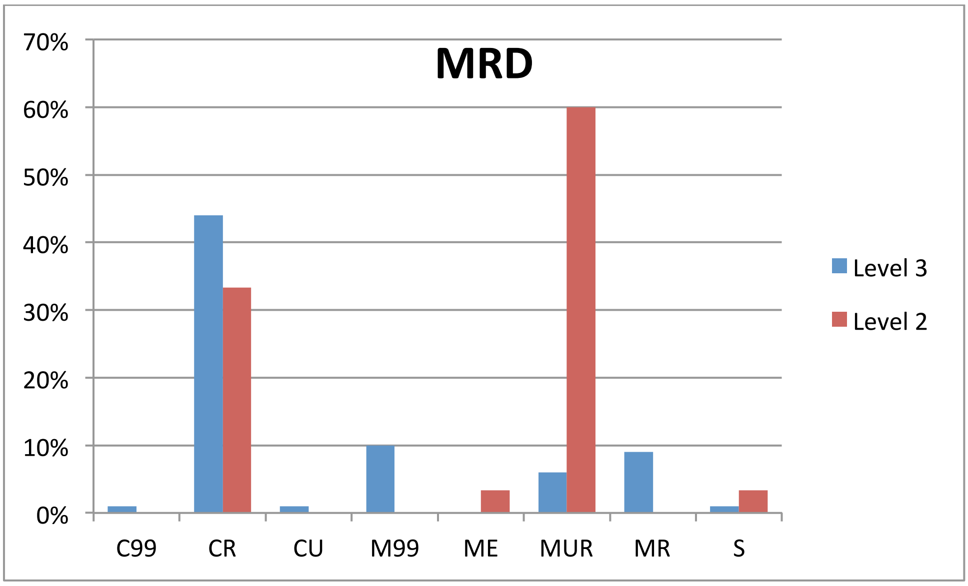


Figure 5.6. SIDD building type distributions in (from top left, clockwise): RS IDCT Land-use zone C2; RS IDCT Land-use zone PDI; RS IDCT Land-use zone MRD; PAGER residential and non-residential building distributions; RS IDCT Land-use zone HR

- Figure 5.7 compares the distributions of buildings obtained from Level 2 SIDD outputs with the distributions defined using Level 3 SIDD outputs, both for zone MRD. It is clear that there is a large discrepancy between Level 2 and Level 3 results. This is discussed further in the next section.

•



•

Figure 5.7. Comparison between Level 2 and Level 3 building type distributions for MRD land-use class

5.3.3 Modal building types

Referring to the building distributions shown in Figure 5.5, Table 5.2 lists the modal (most common) building types identified from each source for residential zones or GED4GEM grid cells (for Level 2 and Level 3 this corresponds to MRD land-use zones or grid cells).

Table 5.2. Modal building types from different sources of exposure data information.

Source	**Class**	**Description**
PAGER	S5	Steel frame with URM infill
Level 3 (MRD)	CR/LWAL	Reinforced concrete wall
Level 2 (MRD)	MUR	Unreinforced masonry
Level 2 (HRD)	MUR	Unreinforced masonry
Level 2 (PDI)	MUR	Unreinforced masonry
GFZ Landuse	RES 1	1-2 storey masonry, brick
GFZ Landuse	RES 2	Masonry, brick, concrete, panel buildings

The predominant building types identified in each land-use area were then compared with the predominant building types identified in PAGER. In a similar way to the distributions shown in Figure 5.6, the modal building type in PAGER does not correspond to the modal building type identified by other sources. This is to be expected, as the PAGER building type distribution for Kyrgyzstan is obtained “by Neighbour”, i.e. it is set as the same as the Russian distribution. Therefore, it is not based on ground-truth data and is likely significantly less accurate than the Level 2 and 3 data collected.

As a second check of the results of the IDCT full test against independently collected exposure data, the modal building types identified in each land-use area from Level 2 data were compared with the predominant building types identified in the GFZ land-use map. As shown in the table, these were found to be in agreement which suggests that the IDCT process produces results which are representative of the modal building types in Bishkek.

As may be anticipated from the difference in the distributions of building types resulting from the Level 2 and Level 3 SIDD outputs, the Level 2 and Level 3 main building types are different. Based on field-experience, a possible explanation is that the areas in which the Level 3 surveys were conducted were misclassified as MRD. The Level 3 survey areas were located close to the city centre and as such, more likely to be mixed medium to high-density residential and commercial. Additionally, the area was more modern than other parts of the city with buildings generally built or retrofitted post-2000. Therefore, different types of buildings were found in this area compared with other, correctly classified MRD areas. This demonstrates the importance of an accurate land-use classification map and ground-truth data to validate results.

The land-use classification map used for the SIDD homogenous zones in this test was derived from imagery without the input from local engineers or geographers. Having local understanding included when defining homogenous zones in a study area will increase the accuracy of the derived maps. Collecting Level 3 reference data from multiple areas around a city will also aid the validation of Level 2 data. Due to its time-consuming nature, this could not be achieved in this test.

5.4 Utility of Collecting Level 3 Exposure Data

The collection of Level 3 exposure data is time consuming and it is therefore worth exploring whether the time and effort taken to collect more detailed exposure information (Level 3) has substantial benefits in terms of the utility of the data collected.

It has already been discussed that Level 3 exposure data potentially identified a land-use misclassification and the data was therefore very useful. Figure 5.8 shows the building distribution in the Level 3 grid cells in terms of building type (LHS) and occupancy type (RHS). These provide useful information, as firstly, the distribution of occupancy types in an area identified by the land-use mapping as predominantly residential are shown. This suggests that the validity and accuracy of land-use mapping and extrapolating Level 2 sample results to land-use zones in this part of Bishkek should be explored in more detail.

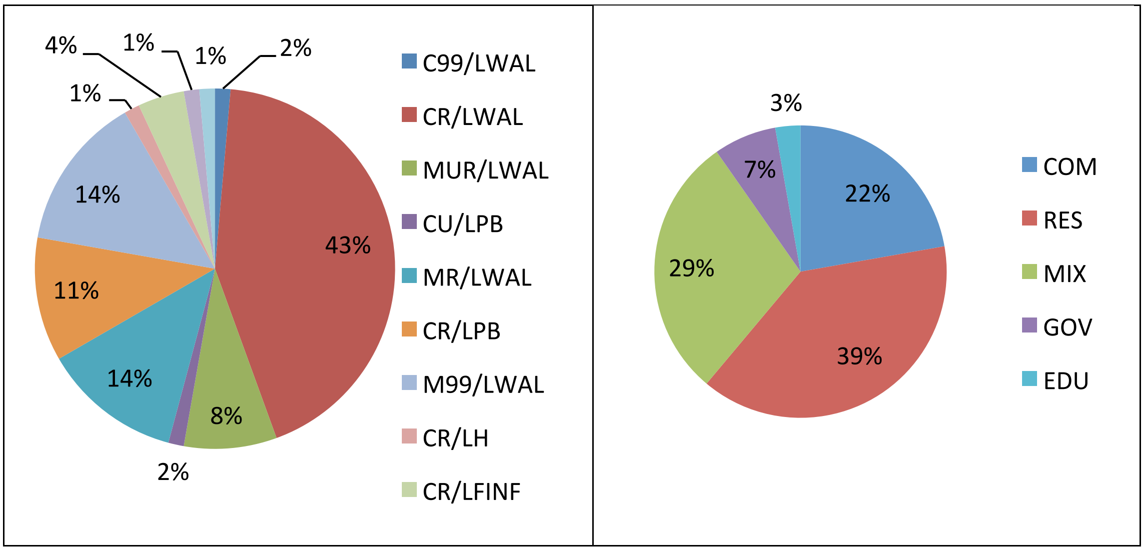


Figure 5.8. Level 3 exposure data information

6 Conclusions

This report has described a demonstration of the end-to-end workflow of the Inventory Data Capture Tools Global Component. The work described to generate inventory datasets in Bishkek has successfully met the objectives outlined in the introduction to this document:

- To capture both Level 2 and Level 3 exposure data.
- To test functionality of the Android and Windows Mobile Tools and paper tool for in-field inventory data collection.
- To test the remote sensing protocols to pre-process satellite imagery and extract information to be used in the sampling strategy delineation and the field data collection.
- To test functionality of the Windows Mobile Tool in terms of basic data management requirements.
- To outline necessary pre-mission technical and non-technical requirements and suggestions for an effective and realistic sampling strategy.
- To demonstrate the utility of the Mobile Tools for the local engineers so they can use them in future projects and contribute to GEM.
- To generate an input for the Global Exposure Database

6.1 Recommendations

There are a number of recommendations that can be made following this end-to-end test:

The Level 2 and Level 3 datasets obtained are different in their composition, accuracy and utility:

- The Level 2 exposure data capture procedure provides a relatively rapid way of collecting data which can be used to create a building inventory for a city or region. The dataset obtained should be validated or compared with existing datasets as described in Section 5.
- It is time consuming to collect Level 3 data and it is of little use in obtaining a city-level or regional distribution of buildings. However, Level 3 data is useful for several applications:
  - in depth analysis of the buildings in a particular area (e.g., critical facilities and campus locations (e.g., hospitals, factories or airports). This is particularly suitable for single site seismic risk assessment activities for key facilities.
  - calibration and validation of Level 2 exposure data
  - understanding the validity of land-use mapping
- Both datasets are dependent on the quality of the data collected at each stage in the IDCT workflow. Where possible, an indication of the data quality and accuracy at each stage should be provided.

Use of local expertise:

- Local engineers or other experts participating in the study are invaluable in providing local expertise to improve the quality of data collected at each point in the workflow. Local knowledge should be included throughout, particularly in understanding land use and urban configuration of the study area or providing insight into local construction practices. Collaboration with local engineers and students should be encouraged, where possible, at each stage in the inventory development process.

- It is recommended that the data collection team implement a method to validate the results of the field test. In particular, it is important to check that local engineers or less experienced engineers are correctly interpreting the GEM Building Taxonomy. A short workshop should be held before major inventory collection efforts to acquaint the data collection teams with the information contained in the taxonomy. The Mobile Tools contain the GEM Glossary of all terms contained in the Building Taxonomy and this should be updated before each mission. The data collected in the field will be limited by surveyors' understanding of taxonomy and interpretation of key terms arising from language barriers and assumptions.

Remote sensing:

- The image pre-processing steps need to be performed and validated thoroughly, and the resulting data needs to be used as the only base for all the subsequent analyses to prevent registration errors.
- It is recommended that the vector data produced, such as the homogenous zones, are validated against pre-existent datasets or local knowledge. Alternatively, cross-checks by different users are encouraged.

Sampling:

- For Level 2 data development, an appropriate sampling size should be encouraged and protocols have been developed to guide the user in developing a strategy (Porter *et al.*, 2014).
- Users should be aware of the time and resources required to implement a survey with a large sample size.

Mapping schemes:

SIDD provides a flexible platform for generating mapping scheme distributions of structural attributes: This can be done from scratch or can begin with an existing mapping scheme (e.g., a country-level scheme from PAGER) and iteratively improve the distributions by adding data generated by the GEM IDCTs.

REFERENCES

Document References

Brzev S., C. Scawthorn, A.W. Charleson, L. Allen, M. Greene, K. Jaiswal, V. Silva (2013), GEM Building Taxonomy Version 2.0, GEM Technical Report 2013-02 V1.0.0, GEM Foundation, Pavia, Italy.

Foulser-Piggott, R. Vicini, A., Verrucci, E., Bevington, J. and Shelley, W. (2013) IDCT Mobile Tools - Field Test Reports, *GEM Inventory Data Capture Tools Project Report,* Version 1.0, June 2013, available from http://www.nexus.globalquakemodel.org/gem-idct/posts

Jordan, C. J., K. Adlam, K. Lawrie, W. Shelley, J. Bevington (2014), User guide: Windows mobile tool for field data collection, GEM Technical Report 2014-04 V1.0.0, GEM Foundation, Pavia, Italy.

Federal Emergency Management Agency (2002a). *Rapid Visual Screening of Buildings for Potential Seismic Hazards: A Handbook*, FEMA 154, Edition 2, Applied Technology Council Redwood City California.

Federal Emergency Management Agency (2002b). *Rapid Visual Screening of Buildings for Potential Seismic Hazards: Supporting Documentation*, 2nd ed., FEMA report 155, 2nd ed.

Monelli, D., M. Pagani, and G.A. Weatherill (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.

Porter, K., Z. Hu, C. Huyck, and J. Bevington (2014), User guide: Field sampling strategies for estimating building inventories, GEM Technical Report 2014-02 V1.0.0, GEM Foundation, Pavia, Italy.

Rosser, J., Morley, J.G. and Vicini, A. (2014), User guide: Android mobile tool for field data collection, GEM Technical Report 014-03 V1.0.0, GEM Foundation, Pavia, Italy.

Silva, V., H. Crowley, M. Pagani, D. Monelli, R. Pinho (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.

Vicini, A., J. Bevington, G. Esquivias, G-C. Iannelli. and M. Wieland (2014), User guide: Building footprint extraction and definition of homogenous urban areas, GEM Technical Report 2014-01 V1.0.0, GEM Foundation, Pavia, Italy.

Wieland, M., M. Pittore, S. Parolai and J. Zschau (2012), Exposure Estimation from Multi-Resolution Optical Satellite Imagery for Seismic Risk Assessment. ISPRS Int. J. Geo-Inf., 1, 69-88, doi:10.3390/ijgi1010069.

Budetta, G. and D. Carbone (1998), Temporal variations in gravity at Mt. Etna (Italy) associated with the 1989 and 1991 eruptions, Bull. Volcanol., 59, 311–326.

THE GLOBAL EARTHQUAKE MODEL

The mission of the Global Earthquake Model (GEM) collaborative effort is to increase earthquake resilience worldwide.

To deliver on its mission and increase public understanding and awareness of seismic risk, the GEM Foundation, a non-profit public-private partnership, drives the GEM effort by involving and engaging with a very diverse community to:

- Share data, models, and knowledge through the OpenQuake platform
- Apply GEM tools and software to inform decision-making for risk mitigation and management
- Expand the science and understanding of earthquakes

GEM Foundation

Via Ferrata 1
27100 Pavia, Italy
Phone: +39 0382 5169865
Fax: +39 0382 529131
info@globalquakemodel.org
www.globalquakemodel.org

JANUARY 2014

GLOBAL EARTHQUAKE MODEL
working together to assess risk