Design and Implementation of a Multi-Disaster Risk Assessment Database System in South Korea

Su-Seong Chai [1] and Dongjun Suh [2],*

[1]   Korea Institute of Science and Technology Information (KISTI), Daejeon 245, Korea; sschai@kisti.re.kr
[2]   School of Convergence and Fusion System Engineering, Kyungpook National University (KNU),
      Sangju 37224, Korea
*   Correspondence: dongjunsuh@knu.ac.kr; Tel.: +82-054-530-1482

check for updates

Abstract: To cope with natural disasters and their impact, various types of risk assessment tools have been developed to assess hazards from a scientific point of view as well as the socioeconomic impacts of a hazardous event. Because most of the assessment tools have been developed using local and regional characteristics, there are limitations when applying special cases consisting of various disaster-related datasets in Korea. In this paper, we collected open data scattered among various institutions and guaranteed data interoperability through a standardization process considering various formats. In addition, this study presents a scalable design method and implementation of a database system which can be easily employed in Korean loss estimation due to natural hazards. The proposed database system will be of great help in predicting the disaster scale and degree of damage, in supporting decision-making to minimize damage, and could serve as a data platform of disaster risk reduction and prevention.

Keywords: disaster loss assessment; database system; natural disaster; risk reduction; open data

1. Introduction

In this paper, a complex disaster loss assessment database system that allows researchers to search, apply, and share database is presented. A previous study focused mainly on the development of Korean building inventory [1]; however, this paper presents an advanced system architecture design considering system scalability, various functions with specific construction information of the Korean peninsula, and a data management scheme for big disaster data processing.

Earthquake damage involving floods and fires is increasing the world over, and various types of large-scale complex disasters are occurring in succession. South Korea, which was regarded as a safe zone against earthquakes, witnessed the 2016 Gyeongju earthquake (magnitude 5.8) and the 2017 Pohang earthquake (magnitude 5.4), which were the largest in its earthquake observation history. These earthquakes confirmed that South Korea is no longer safe from earthquakes; therefore, it is necessary to develop technologies that can comprehensively predict and respond to the scale of complex disaster damage at the national level, as well as technologies for strengthening disaster resilience [2].

With the advent of the fourth industrial revolution (or the emergence of industry 4.0), the importance of big data is attracting attention. More rapid and precise damage loss assessment and disaster risk prediction and analysis are possible in the disaster response area through advancements in data collection and analysis technology, based on information and communication technology (ICT) and sufficient computing resources.

Hazards U.S. Multi-Hazard (HAZUS-MH) [3] and Mid-America Earthquake Center Seismic Loss Assessment System (ERGO) [4] are representative damage analysis tools related to earthquakes. Most analysis tools calculate earthquake-induced building damage first, assess damage losses based

on the degree of damage, and calculate the number of refugees and the number of people requiring shelters. To perform a loss assessment, data representing the information of the damaged area, such as buildings and population, are required. Relevant studies conduct research and analysis using data provided in the forms of files and databases by the country.

In South Korea, studies are being conducted using disaster damage statistical data, relevant research data, and actual measurement data. For earthquakes; however, earthquake-related data for estimating losses are insufficient. Therefore, analyses are conducted based on HAZUS-MH, which is a loss assessment tool of the Federal Emergency Management Agency (FEMA), using a large amount of accumulated actual data, earthquake-related data provided by the tool, and domestic and overseas studies. In order to utilize these specific data sets in various applications, such as HAZUS-MH or ERGO, a lot of time is required to turn the data into useful information through the initial steps of data cleaning, preparation, and preprocessing.

HAZUS-MH is developed by the Federal Emergency Management Agency (FEMA) and is used in many other regions for estimating losses after natural hazards, including earthquake, flooding, fire, and hurricane. In damage estimation studies caused by natural disasters in the Korean peninsula, Kang et al. [5] estimated loss estimation by using HAZUS-MH and ShakeMap of USGS (United States Geological Survey), and Yu and Jo et al. [6,7] investigated flood loss estimation using HAZUS-MH. Besides, Scawthron and Hancilar et al. [8,9] analyzed HAZUS-MH-based loss assessment in Canada and Oman.

As the earthquake-related data provided by HAZUS-MH (and the weight data used for loss assessments) reflect the environment and condition of the corresponding country, the data are different from those of the Korean environment. Therefore, it is necessary to provide proper datasets reflecting the Korean environment, such as the geological environment, structures, and uses of buildings, in order to evaluate the loss assessment properly. In the study by [10], only the direct economic damage of buildings caused by strong ground motion was estimated, which does not reflect the ground liquefaction and indirect damages that reflect the characteristics of the region. The studies by [11,12] presented the damage prediction and inventory construction method in a specific area and introduced the necessity of a long-term project caused by inducing the system to expand the inventory [13].

The loss assessment uses a variety of data, such as demographic figures, building-integrated information, and others, gathered and used at the national level. Most open data are managed and provided from public open data portal systems, and are easy to use for a variety of purposes. However, most of the data are generated by various institutions and for various purposes, and the compatibility and interoperability of the data are not sufficiently considered. Therefore, standard data processes should be performed with respect to various types of data formats, such as structured, semi-structured, and unstructured. This paper focused on the construction of a database system with respect to multi-disaster loss assessment analysis of earthquakes involving floods and fires, and presents a scalable design methodology of a database system considering compatibility, which can serve as a data platform of disaster risk reduction and prevention. The proposed database system plays a crucial role in providing relevant inventories and proper data sets to various applications for disaster risk reduction and prevention in Korea.

In this paper, we performed compatibility and interoperability testing of the proposed database system by applying two different systems—big data analytics framework (real-time/ distributed based analysis) and Ergo application software (batch processing/standalone based analysis). The goal of compatibility and interoperability testing is concerned with how various data sets or different types of systems perform the function properly without redundant preprocessing.

Section 2 describes the characteristics and types of public data and overseas technologies used for estimating disaster losses, Section 3 explains a scalable disaster risk analysis platform, and Section 4 explains the data design method for constructing the database system. Section 5 discusses the comparison between existing systems and the proposed system. Finally, conclusions and future research directions are drawn in Section 6.

2. Related Works

2.1. Disaster Risk Assessment System

2.1.1. HAZUS-MH

HAZUS-MH is a software package for natural disaster and risk assessment developed by the US Federal Emergency Management Agency. It helps planners and strategists to predict the consequences of disasters, and to reduce risks when hazardous events occur. Direct or indirect damage can follow when an earthquake occurs. For example, direct damage causes physical damage to various infrastructures such as residential and commercial buildings, bridges, and roads. In addition, there can be critical losses of production facilities owing to physical losses, production and restoration cost interruption, social losses owing to casualties, shelter supply issues, and indirect economic losses.

The damage and risk assessment process of HAZUS-MH for earthquakes is shown in Figure 1. Ground collapses after a major earthquake can occur for a period of time after a huge vibration owing to the initial earthquake. Direct physical damage is caused by earthquakes or ground collapses. In HAZUS-MH, physically damaged facilities can be analyzed with respect to several groups of infrastructures, such as general buildings, essential and high-risk facilities, transportation facilities, and utility facilities.
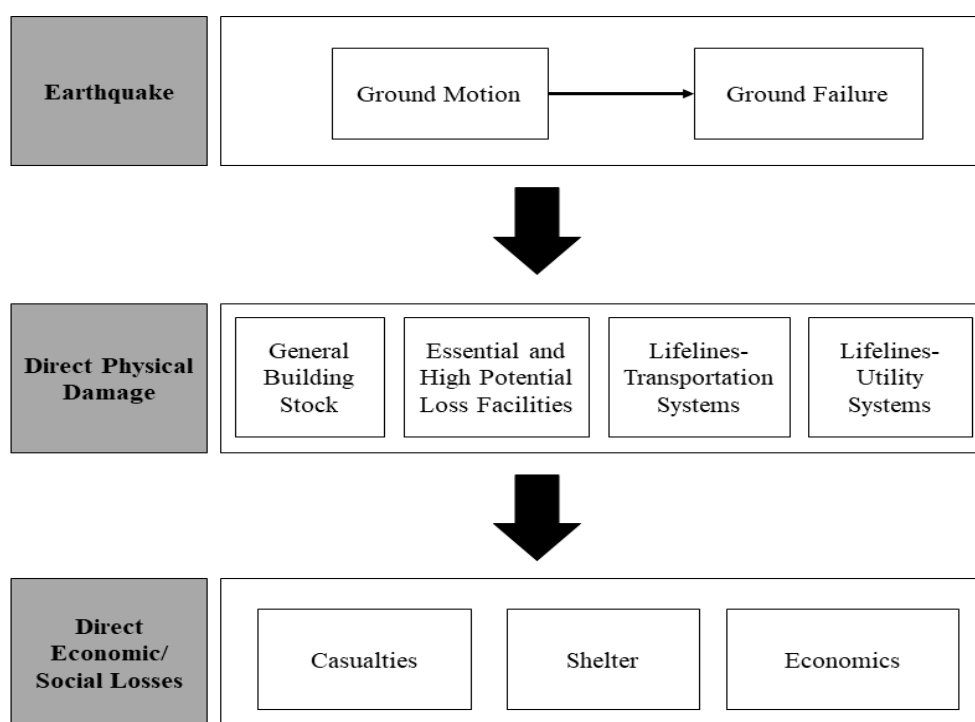
Figure 1. Flowchart of earthquake analysis of HAZUS-MH.

The general building group is divided into seven categories: residential buildings, commercial buildings, industrial buildings, agricultural buildings, religious and nonprofit buildings, government agencies, and educational institutions. More specifically, there are 33 building types according to their usage. Transportation facilities are classified into seven different specific categories: highway, bus, railway, light rail, airport, port, and ferry systems. Utility facilities are also categorized into six categories: portable drinking water, wastewater, oil, natural gas, power, and communication facilities.

Direct/indirect economic and social losses can be evaluated according to the degree of building damage caused by natural disasters or hazardous events. HAZUS calculates the casualty rates with respect to various building types and damage levels considering the number of individuals killed,

injured, or missing in buildings or areas affected by natural disasters or hazards. Focusing on the physical damage to buildings, HAZUS establishes a scale of damage from 1 to 4. The degree of building damage is divided into four grades. In addition, HAZUS computes site-specific loss estimations depending on the weight values with respect to time-based scenarios. For example, when an earthquake occurs at 02:00 a.m., the occupants of residential buildings are more likely to suffer high numbers of casualties, whereas when an earthquake occurs at 14:00, people who work in commercial buildings experience a great deal of damage. In addition, it is estimated that the number of casualties will be much higher on roads or bridges during the evening rush hour (near 17:00). In addition, in the case of shelter demand prediction, the demand capacity of a shelter is determined by the casualty estimation.

Regarding economic loss estimation, a physical damage value is converted to an economic value, and the national currency is then converted to US dollars (HAZUS-MH basis). Therefore, it is possible to calculate economic losses including the repair and replacement cost of buildings, contents, cost of inventory loss, recovery time, and rental cost from suspended operations. The following are the equations for the HAZUS earthquake model to derive the economic loss (replacement cost) of a building:

$$CS_{ds,i} = BRC_i * \sum_{i=1}^{33} PMBTSTR_{ds,i} * RCS_{ds,i}, \tag{1}$$

$$CS_i = \sum_{ds=2}^{5} CS_{ds,i}, \tag{2}$$

where $CS_{ds,i}$ is the replacement cost owing to the earthquake occurrence, BRC_i is the replacement cost of specific building use i, $PMBTSTR_{ds,i}$ is the probability of a specific building use i suffering from a specific building damage, $RCS_{ds,i}$ is the building repair and replacement cost for specific building use i that is damaged by an earthquake, and CS_i is the total sum of replacement costs owing to an earthquake occurrence.

A total of 33 building types are categorized by HAZUS with respect to building usage, and the building repair and replacement cost is calculated for each building. In addition, the degree of building damage is divided into four grades (slight, moderate, extensive, and complete), and the repair and replacement rate of the building varies depending on the use of the building and the degree of building damage.

The degree of building damage is the most basic factor in loss assessment through the risk analysis of HAZUS. Based on the degree of damage to the building, it is possible to derive the calculation value through various models such as replacement cost of the building, economic loss owing to business interruption, loss of life owing to building damage, and occurrence of casualties.

2.1.2. ERGO

ERGO, developed by the US National Supercomputing Application Center, is an open-source disaster analysis tool. Its predecessor was Mid-America EarthQuake Visualization (MAEViz), developed by the Central American Earthquake Center. HAZUS-based applications were developed to perform seismic risk assessment in the states of the central US. HAZUS provides a module to add the user's risk data, and it is possible to extend the function that combines modeling such as risk factors and vulnerabilities such as SYNER-G: Systemic Seismic Vulnerability and Risk Analysis for Buildings, Lifeline Networks and Infrastructures Safety Gain [14]. To estimate the damage, the function of estimating earthquake risks for lifelines and social and economic systems was added based on the HAZUS model and additional analysis. Risk analyses yield results such as economic losses (direct economic losses, indirect economic losses, and economic losses owing to downtime and business interruptions) and social losses (social vulnerability, casualties, and victim estimates).

ERGO is able to identify the location of a damaged building through its GIS function, evaluate the risk based on the earthquake scenario through damage calculations, and support decision making.

The following equations are the mean (μ_D) and standard deviation (σ_D) of the damage factors used in the calculation of the average expected loss of a building owing to an earthquake occurrence.

$$\mu_D = \sum_{i=1}^{4}\left[P(DS_i) * \mu_{D|DS_i}\right], \tag{3}$$

$$\sigma^2{}_D = \sum_{i=1}^{4}\left[P(DS_i) * \sigma^2{}_{D|DS_i} + \mu^2{}_{D|DS_i}\right] - \mu^2{}_D. \tag{4}$$

The mean expected loss of the building owing to the earthquake is the mean damage factors, the standard deviation, and the individual probability $P(DS_i)$ of damage to the building Table 1. DS_i is the degree of building damage owing to the earthquake occurrence. Damage states (classified as insignificant, moderate, heavy and complete) for each building is calculated from the individual probability of the damage state and the mean damage factor. Equation (3) shows how to compute expected loss ratio for each building's structural loss from discrete probabilities of damage states and mean damage factors. Using (4), the expected loss variance of building structural components is calculated using the standard deviation and mean damage factor for each damage occurrence, the individual probability of damage, and the expected loss ratio. Then, the mean expected loss for each building can be calculated, by multiplying the expected loss factor for the structural component by the value of the structural component.

Table 1. Structural damage factors [15].

ERGO Damage State	Range of Beta Distribution (%)	Mean of Damage Factor, $\mu_{D|DS_i}$ (%)	Standard Deviation of Damage Factor, $\sigma_{D|DS_i}$ (%)
Insignificant	[0, 1]	0.5	0.333
Moderate	[1, 30]	15.5	9.67
Heavy	[30, 80]	55	16.7
Complete	[80, 100]	90	6.67

From the input values (the individual probability of damage to building components, mean and standard deviation of the damage factors, and value of the building's structural components) and Equations (3) and (4), the mean and variance of the expected loss ratios for the structural components of each building and the expected loss averages of the structural values for each building are derived. Similar to HAZUS-MH, ERGO can perform various loss assessments, such as economic loss, loss of life, and shelter demand based on the loss of a building, as shown in Figure 2. Figure 3 shows a process for analyzing the damage to a building owing to an earthquake. The red box means that the necessary conditions for risk analysis have not been met. To produce a green box that meets the condition, the user should add data inventory such as buildings, fragility curves, and hazard maps.

ERGO is an open-source disaster risk analysis tool, and has the advantages of being easy to access and expandable because it is open to all users such as scholars and researchers. It is possible to share code with various researchers by directly developing and opening the risk assessment algorithm derived by the researchers. In addition, ERGO's data-generation module loads data for risk analysis, such as building, population, and geological information in the area, and it is possible to carry out risk analyses suitable to the local environment. However, there is some difficulty in collecting and loading data such as population, building, and environmental information.
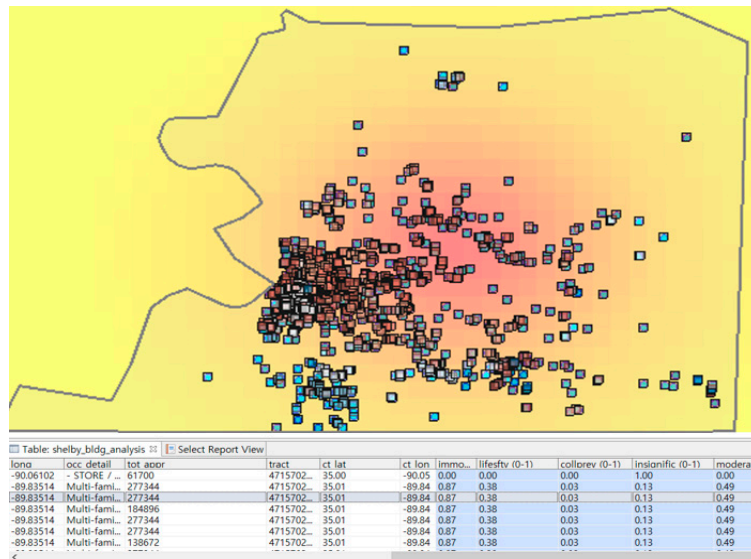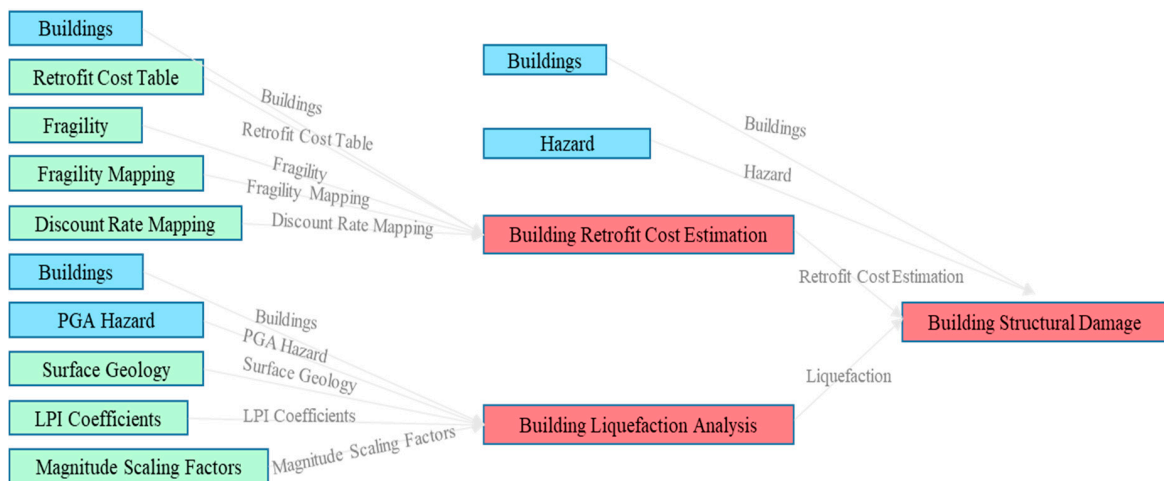
Figure 2. Building damage analysis in ERGO.

Figure 3. Process of building damage analysis in ERGO.

2.2. Public Data

As the era of big data continues, the utilization of data is increasing with the increase in data. Most people who are involved in data science focus on the economic and social value of public data, and try to makes public data open and available nationwide. Public data is freely available to everyone, and there are no rules that restrict the use of the data.

The public data portal opens open API data services with datasets of public institutions, so that it can satisfy right-to-know people and make data easier to use. Development examples using public data are increasing in various fields, such as bus information, sightseeing, weather, postal delivery, and performance display. As time passes, the number of public data provisions and utilizations are increasing exponentially through the development of mobile apps and web services using public data Figure 4.
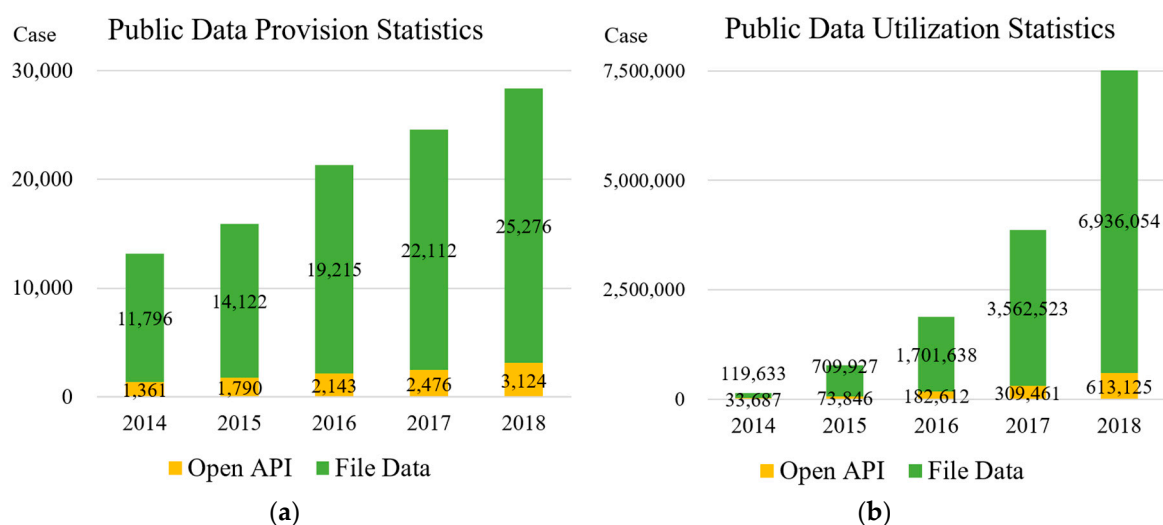
Figure 4. (**a**) Public data provision statistics; (**b**) Public data utilization statistics [16].

From institutions that specialize in earthquake research, such as the United States Geological Survey (USGS) [17] and the Korea Earthquake Research Center (KERC) [18], it is possible to obtain earthquake-related information such as GIS data, topographic maps, and geological maps. In particular, USGS data is being loaded on several modules of HAZUS-MH. HAZUS-MH uses earthquake data of shake maps provided by USGS, and the latest seismic hazard map of USGS. MAEViz, the predecessor to ERGO, was also developed by benchmarking HAZUS-MH, so it includes rich inventory data from HAZUS-MH and hazard maps of USGS. In addition to earthquakes, the USGS zoned each state in the United States into hydrologic regions, with appropriate regression curves for each zone. Based on these disaster-related data, the United States and disaster risk assessment systems around the world are building inventory, and predicting and preparing for dis-asters.

Public data portals provide several years' worth of statistical data. These public data provide a new version of the data through periodic data updates. However, for data users, it is necessary to check for data updates periodically, and if a data update is accomplished, the new data should be reflected in the existing data.

Table 2 shows the GIS complex space information of building and the GIS general space information of building using "building" as a keyword in the public data portal of Korea. A complex building is a building that can be divided into places such as apartments, coalitions, and studio apartments, and where ownership can be exercised. A general building is a single-family house that is owned by the owner of the entire building or is jointly owned, excluding a complex building. The date indicates when the dataset was updated, and it can be determined that updates are periodically performed on the same dataset. Updating a data set includes adding new data, revising previous versions, and modifying the coordinate system.

Table 2. Structural damage factors.

District	Dataset	Date	File Size	Download
Seoul	GIS complex space information of building	2017-10-20	15,792 KB	⬇ SHP
Seoul	GIS complex space information of building	2017-10-19	44,076 KB	⬇ SHP
Seoul	GIS complex space information of building	2017-03-06	23,820 KB	⬇ SHP
Seoul	GIS complex space information of building	2016-11-25	66,375 KB	⬇ SHP
Seoul	GIS complex space information of building	2016-11-25	21,481 KB	⬇ SHP

3. Scalable Disaster Risk Analysis Platform

Most of studies related to disaster risk assessment tend to deal with similar properties to influence factors, such as the characteristics of location, building, ground types, population, and socio-economic elements, to estimate loss evaluation. Therefore, this paper presents a disaster risk assessment database system, which can be easily utilized in Korean loss evaluation due to natural hazards, by using both inventories and the compatible data set.

With regard to disasters, it is difficult to assimilate the situation on the spot in real time, and it is difficult to cope with rapidly changing situations. Therefore, it is necessary to repair and prepare for a damage evaluation through a simulation beforehand. The study focuses on the design and implementation of database systems that allow researchers or stakeholders to provide appropriate data sets that take into account the characteristics of specific local environments, including geological properties and population data, through the risk analysis tools or real time analysis platforms.

The disaster risk analysis platform consists of a data preprocessing process, storage process, analysis and visualization process, and service process, as shown in Figure 5. In the preprocessing process, public data (buildings, population, and weights) scattered around various agencies are collected, and the data format is converted to store and process the data. Currently, data is collected mainly from public data, but social networking service (SNS) data collection will also be conducted to reflect disaster situations in real time.

In the storage process covered in this study, the database is constructed to store public data, which is structured data. The collected data is structured considering the characteristics between the data items. Based on this, a database suitable for data attributes is constructed. When image data such as hazard maps and disaster-related social data are added, large amounts of semi-structured and unstructured data are collected in real time. Therefore, more data storage methods will be added through schema design using the JSON (JavaScript Object Notation) [19] format and NoSQL data processing. JSON (JavaScript Object Notation) is a standard text-based format for representing structured data and objects. Besides, JSON is light weight data-interchange format which makes it easy to interchange data among web, mobile, and so on. NoSQL refers to a database that is not depending on SQL (structured query language), which is very suitable for big data processing. NoSQL databases are highly scalable, do not rely on a relational model, and provide high performance [20]. Analysis and visualization processes and services provide data that can help decision-makers to make decisions in order to predict and prepare for disaster damage based on the data they hold.
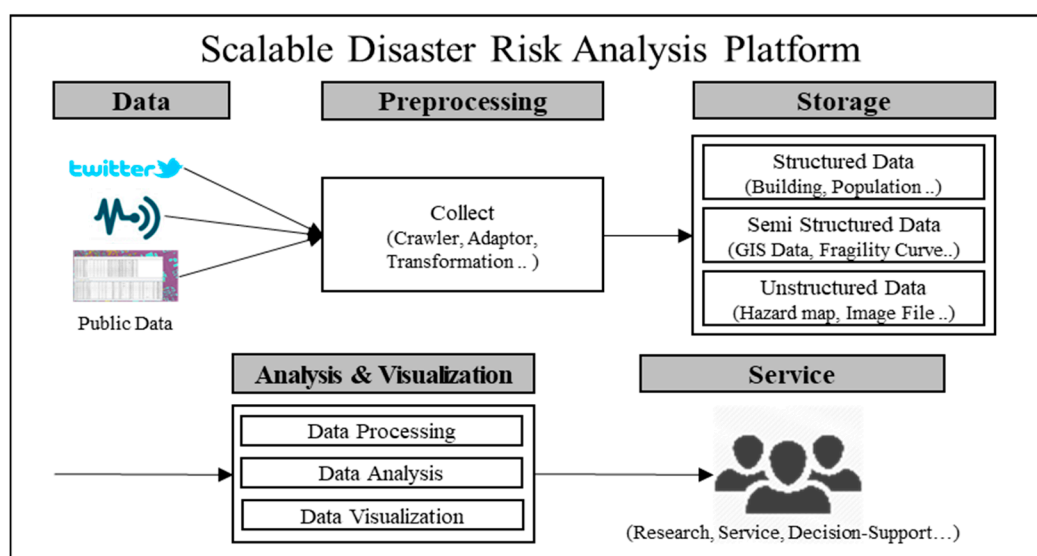

Figure 5. Dataflow of the scalable disaster risk analysis platform [21].

In this study, we conducted a compatible test with respect to the proposed database system by applying both the big data analytics framework and Ergo. The proposed system can be utilized in various types of platforms in order to assess risk due to natural hazards.

4. Multi-Disaster Loss Assessment Analysis Database System for South Korea

First, Korean environment data were collected to construct a complex disaster loss assessment system for South Korea. The database model was constructed and developed considering a standardization process after refining the data distributed to each institution. In this paper, the database was constructed by considering scalability, and the loaded data can be used as open data in various analysis systems other than disaster-related analysis systems. The system has a workflow as shown in Figure 6.
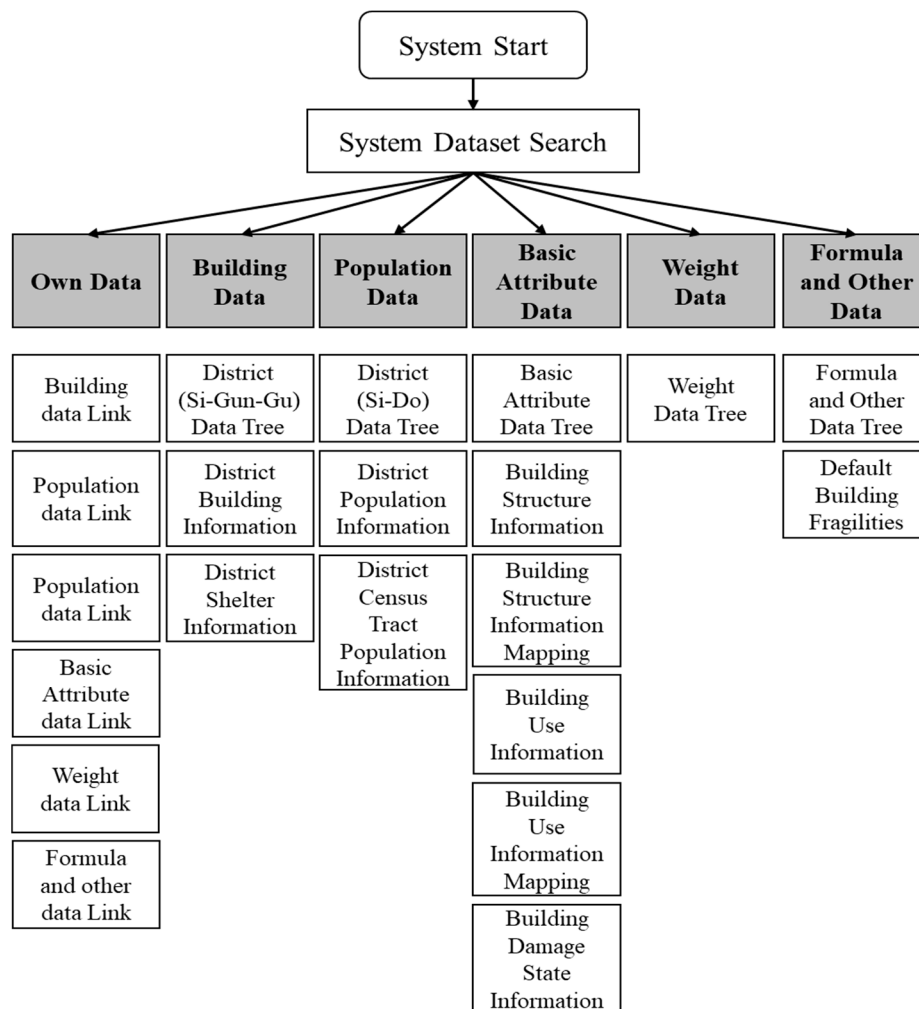

Figure 6. Database system workflow.

The database consists of 30 tables, and is mainly composed of building information data, population information data, basic attribute data used for the loss assessment, weight data, and tables related to formulas and other data. The tables in the Entity-Relationship Diagram (ERD) mainly contain building-related information, because it measures various losses that occur based on the degree of damage to the building. Figure 7 below shows an ERD of the complex disaster loss assessment and damage analysis database system designed in this paper.
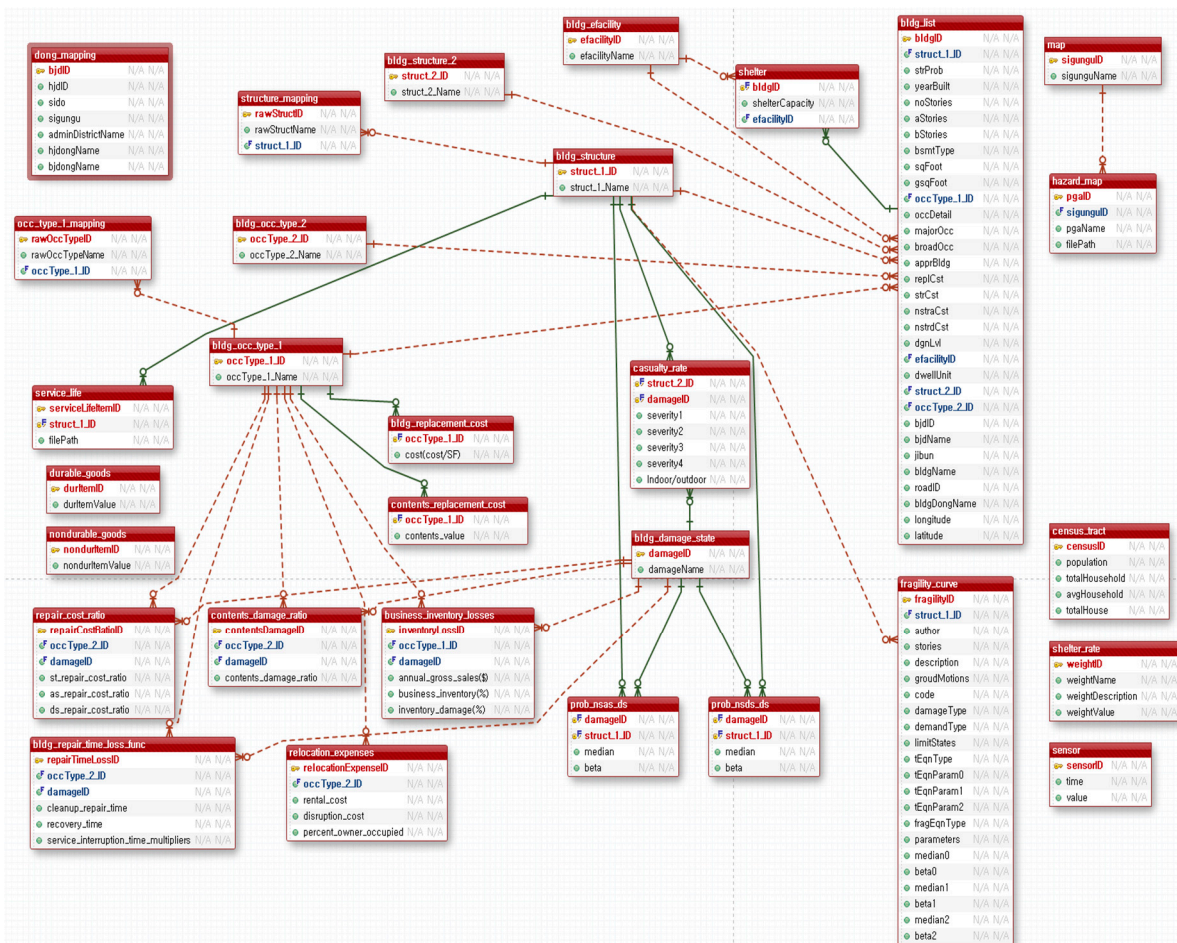
Figure 7. ERD of the loss estimation database system.

4.1. Buidling-Related Table

Relevant data to be loaded into the domestic building-related table are provided as open data in the national spatial information portal of the Ministry of Land [22]. Data can also be obtained from the road name address guidance system. One building contains various attribute data, as shown in Table 3.

The building information data are the most fundamental in predicting the risks of complex disasters occurring in a series owing to an earthquake and estimating the losses. The damage caused by disasters involves economic and human losses, and mostly occurs in buildings where people live. The magnitude of an earthquake has the largest impact on the physical damage to buildings. It is possible to predict and estimate damage such as population and economic losses by assessing the extent of damage to buildings first, through attribute information such as the structures and uses of the buildings. The economic loss can be estimated using the appraisal values of the buildings and the value attributes of the contents, and the population loss can be estimated using the buildings' resident population. The magnitude of an earthquake is variable and cannot be determined in advance because it is an unpredictable input value. As the range of the magnitude is diverse, many cases and simulations are required in the course of loss assessment and prediction. On the other hand, as the attribute information of the buildings is fixed, it is possible to construct a table. The use of the latest data also increases the reliability of the damage loss assessment and prediction results.

Table 3. Building-related table.

Table	Column	Type	Column	Type
Bldg_list	bldgID (P.K)	VARCHAR (30)	nstraCst	INT
	struct_1_ID (F.K)	VARCHAR (3)	nstrdCst	INT
	strProb	FLOAT	dgnLvl	VARCHAR(20)
	yearBuilt	SMALLINT	contVal	INT
	noStories	SMALLINT	efacilityID	VARCHAR(10)
	aStories	SMALLINT	dwellUnit	INT
	bStories	SMALLINT	struct_2_ID	VARCHAR(5)
	bsmtType	SMALLINT	occType_2_ID(F.K)	VARCHAR(10)
	sqFoot	FLOAT	bjdID	VARCHAR(10)
	gsqFoot	FLOAT	bjdName	VARCHAR(60)
	occType_1_ID(F.K)	VARCHAR(10)	Jibun	VARCHAR(10)
	occDetail	VARCHAR(30)	bldgName	VARCHAR(40)
	majorOcc	VARCHAR(20)	roadID	VARCHAR(15)
	broadOcc	VARCHAR(30)	bldgDongName	VARCHAR(40)
	apprBldg	INT	longitude	FLOAT
	replCst	INT	latitude	FLOAT
	strCst	INT		

4.2. Population-Related Table

In South Korea, population data can be obtained based on the demographic data of the statistical geographic information service of Statistics Korea [23]. A national population census is conducted by Statistics Korea every five years. It can identify population-related trends such as population by gender and age, population by educational level, population by marital status, and the number of households. The census is conducted using the output area, which is the minimum statistical zone unit constructed to provide statistical information.

The population table of Table 4 uses the output area unit and shows population information data for each output area. Based on the population table, the number of people residing in each building must be identified. Through this, in case of a disaster, it is possible to identify casualties inside the buildings by predicting and assessing the loss of the building unit. In addition, it is possible to adjust the accommodation capacities of shelters and assign additional shelters, by predicting the number of victims looking for shelters through the identification of casualties and the weight-related table. However, because the minimum unit of the population information is the output area, it is necessary to estimate the population information of buildings. Considering that the proportion of high-rise buildings is significant, and the number of apartments is higher than that of single-family houses in terms of residential buildings in South Korea, the accuracy of population estimation in a building unit is a very important factor in assessing the population loss.

Table 4. Population-related table

Table	Column	Type	Column	Type
Census_Tract	censusID (P.K)	VARCHAR (20)	avgHousehold	INT
	population	INT	totalHouse	INT
	totalHousehold	INT		

The population table consists of the output area ID, population, total number of households, average number of households, and total number of houses. However, it does not contain attribute data such as population by gender and age, single-person households, and two-person households. When there are no corresponding data among the various attribute data, the process of refining the attribute data in the blank part is additionally required because the data of the corresponding output area is missing. Later, these issues can be resolved through a data opening policy, or by linking with social data.

4.3. Basic-Attribute-Related Table

The basic-attribute-related table provides information required in the complex disaster loss assessment analysis process for South Korea, together with the building population formula and weight data, as shown in Table 5. It specifies the meaning of the attribute data that constitute the building, population, weight, formula, and other tables. Two linked (mapping) tables were designed to meet different formats and criteria.

Table 5. Basic-attribute-related table

Table	Column	Type	Column	Type
Occ_type_1_mapping	rawOccTypeID (P.K)	VARCHAR (10)	occType_1_ID (F.K)	VARCHAR (10)
	rawOccTypeName	VARCHAR (20)		
struct_mapping	rawStructID	VARCHAR (10)	structID (P.K)	VARCHAR (3)
	rawStructName	VARCHAR (26)		
dong_mapping	sido	VARCHAR (24)	bjdongName	VARCHAR (24)
	sigungu	VARCHAR (24)	bjdID	VARCHAR (11)
	adminDistrictName	VARCHAR (24)	bjdID (P.K)	VARCHAR (11)
	hjdongName	VARCHAR (24)		

The building use and structure information specified by HAZUS-MH has different criteria for South Korea, and is composed of different formats. To apply it to the complex disaster loss assessment and analysis for South Korea, it is necessary to map it to the building use and structure information table based on the actual circumstances and environment of South Korea.

HAZUS-MH places building use into seven categories, and details into 33 categories. The domestic building use information is provided by the administrative standard code management system of the Ministry of Public Administration and Security, and is classified into more than 500 items. HAZUS-MH provides a linkage table that can be mapped to HAZUS-MH's building use codes and US Industrial Standard Classification (SIC) codes.

In the previous research of this study, the method of mapping a domestic building use to the US industrial classification code and then converting it to the HAZUS-MH building use was presented, as shown in Figure 8. Through this, we designed the table, and mapped the Korean and HAZUS-MH's building use and structure data.
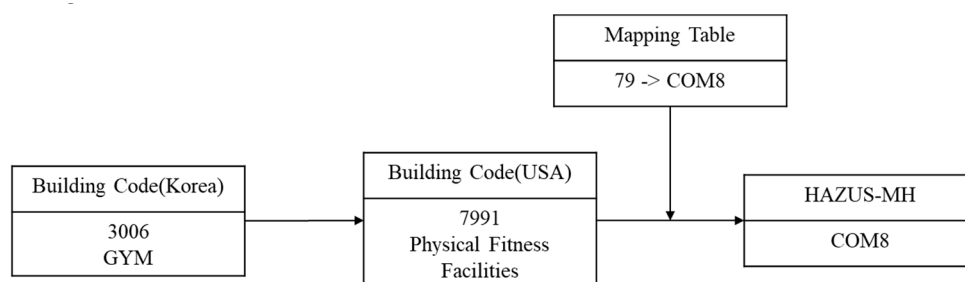
Figure 8. Building inventory mapping process.

The code for each region in South Korea differs for each agency. For bjdID of the building table, codes are assigned by the Ministry of the Interior and Safety (MOIS) on the basis of the legal districts. For the censusID (hjdID) [24] of the population table, administrative district codes are assigned by Statistics Korea based on the administrative districts. As the relationship between the administrative and legal districts varies for each region, they cannot be included in a single code. As there may be problems in calculation processes such as joining the building data using bjdID and the population data using censusID (hjdID), a clear linkage table that can be used as criteria is required. MOIS presents a linkage table between the legal district codes and the administrative district codes. Based on this, a linkage table for the calculation of the population table and building table was designed.

4.4. Weight-Related Table

The weight-related table is composed of tables representing the money information per unit area such as service life, durables and nondurables, loss of contents, building replacement cost, total production (limited to industrial buildings), and the demand on shelters according to age and income, as shown in Table 6.

Table 6. Weight-attribute-related table

Table	Column	Type	Column	Type
contents_damage_ratio	contentsDamageID	VARCHAR (3)	damage_ID	VARCHAR (2)
	occType_1_ID	VARCHAR (10)	contents_damage	FLOAT
business_inventory_losses	inventoryLossID	VARCHAR (3)	annual_gross_sales	FLOAT
	occType_1_ID	VARCHAR (10)	business_inventory	FLOAT
	damageID	VARCHAR (2)	inventory_damage	FLOAT
relocation_expenses	relocationExpenseID	VARCHAR (3)	disruption_cost	FLOAT
	occType_2_ID	VARCHAR (10)	percent_owner_occupied	FLOAT
	rental_cost	FLOAT		
repair_cost_ratio	repairCostRatioID	VARCHAR (3)	st_repair_cost_ratio	FLOAT
	occType_2_ID	VARCHAR (10)	as_repair_cost_ratio	FLOAT
	damageID	VARCHAR (2)	ds_repair_cost_ratio	FLOAT
bldg_repair_time_loss_func	repairTimeLossID	VARCHAR (3)	cleanup_repair_time	INT
	occType_2_ID	VARCHAR (10)	recovery_time	INT
	damageID	VARCHAR (2)	service_interruption_time_multipliers	FLOAT

The disaster damage in South Korea is calculated in accordance with the guidelines for natural disaster investigation and restoration plan establishment. With regard to private property loss and damage, damage to the internal contents of residential and commercial buildings is not considered, and is calculated in a way that does not depend on the building area. In addition, when the losses for commercial and industrial buildings are calculated, additional economic losses (such as business losses owing to building damage) are excluded. In case of a disaster, the money information per unit area, such as the loss of contents inside the building ($), building replacement cost ($), and total production ($) (limited to commercial and industrial buildings) based on the square footage (SF), must be provided as weights. It is possible to estimate damage to buildings, constructions, and assets (facilities, machinery, furniture, equipment, and vehicles) considering the service life and depreciation, and to calculate economic losses that are suitable for the Korean situation, such as the loss of contents inside the buildings and the total production.

To calculate or predict the population movement based on the victims of building damage owing to a disaster, weights by age and income must be considered.

The prediction model of HAZUS-MH estimates seismic risk in all regions of the United States by using past disaster history including hurricanes, floods, and earthquakes. In addition, the various risk models are incorporated with social vulnerabilities to multi-hazard losses.

In South Korea; however, there is a lack of research on social vulnerabilities to disasters, such as emergency shelter needs. Most of the parameters and weights being used are based on HAZUS-MH models.

A sociodemographic survey considering the Korean environment must be conducted first to construct data suitable for the Korean environment. It is necessary to construct income information per household and the sales amount of the building unit. Additional tables must be designed by calculating the weight information considering the officially announced local building prices to assess the economic loss of buildings more accurately.

4.5. Formulas and Other Data

Formulas and other data are tables used for analysis based on population and building data during the assessment of losses owing to a disaster, as shown in Table 7. A fragility curve, fragility curve mapping, hazard map, attenuation function, and seismic design table are constructed to draw conclusions suitable for the Korean situation using the Korean data. If an earthquake occurs, the degree of damage varies depending on the building structure even for the same magnitude. It is possible to estimate the damage degree of buildings according to the building structure using the fragility curve created based on the Korean building structures. Fragility curve mapping is a mapping table that allows for the selection of a fragility curve according to the structure of a building. In addition, different magnitudes are calculated in each region for earthquakes with the same magnitude. As each region is composed of different geological structures, a hazard map can be created through the magnitude calculation for the corresponding region using an attenuation function that reflects the geological structures of South Korea. Domestic seismic design laws were enacted for the first time in 1988. Because constructions before 1988 (and buildings of fewer than six stories) did not use seismic design, the safety of buildings without seismic design was not confirmed. The degree of loss varies depending on whether seismic design is applied. Therefore, losses owing to seismic design should be considered.

Table 7. Formulas and other data.

Table	Column	Type	Column	Type
	id	VARCHAR (6)	demandUnits	VARCHAR (3)
	fragilityIDCode	VARCHAR (10)	limitState	VARCHAR (100)
	struct_1_ID	VARCHAR (3)	Parameters	FLOAT
	author	VARCHAR (20)	median0	FLOAT
Fragility_curve	stories	INT	beta0	FLOAT
	description	VARCHAR (60)	median1	FLOAT
	groundMotion	VARCHAR (60)	beta1	FLOAT
	code	VARCHAR (10)	median2	FLOAT
	damageType	VARCHAR (15)	beta2	FLOAT
	demandType	VARCHAR (10)		

As the damping equation data and the hazard map are composed of semi-structured data (that have shapes, thus cannot be computed) and unstructured data, they cannot be applied as tables in the relational database system. Therefore, schema design using the JSON format is needed to manage semi-structured and unstructured data, and data processing through NoSQL and an additional storage plan are needed.

4.6. System User Interface

Figure 9 shows the user interface of the disaster risk database system. It is implemented based on the database designed in this study, and it is possible to search and download buildings, population, basic attributes, weight, formulas, and other data. This system will provide data to disaster researchers, and will be used in connection with the predictive assessment system for Korean complex disaster response. Through this, it is possible to anticipate the magnitude and extent of a disaster and to prepare for it in advance, and to help decision makers in the event of a disaster.

Figure 9. User interface of the proposed database system.

4.7. Compatibility Test

The risk analysis studies analyzing large-scale cities focus primarily on the huge volumes of building datasets; therefore, it is important to manage and process available data on multi-disasters efficiently. In this study, we conducted a compatible test of the proposed database system through the big data analytics framework, that we developed in a former study [21], supporting distributed and parallel processing that allows large volumes of data to be processed in a short time.

Figure 10 shows the result of the compatibility test with the dataset from the proposed database system. This test shows the performance of the proposed system, which was designed to improve compatibility and shorten the preprocessing time without additional processing step.

In case of batch processing, we evaluated performed compatibility by using the ERGO application. Figure 11 shows the test result of risk analysis of Goyang City in Korea through ERGO with building inventories, data sets provided by proposed system as well as the Korean attenuation equation [25–27]. Therefore, we confirm that the proposed system, considering system and data interoperability, enables immediate simulation without any preprocessing steps.
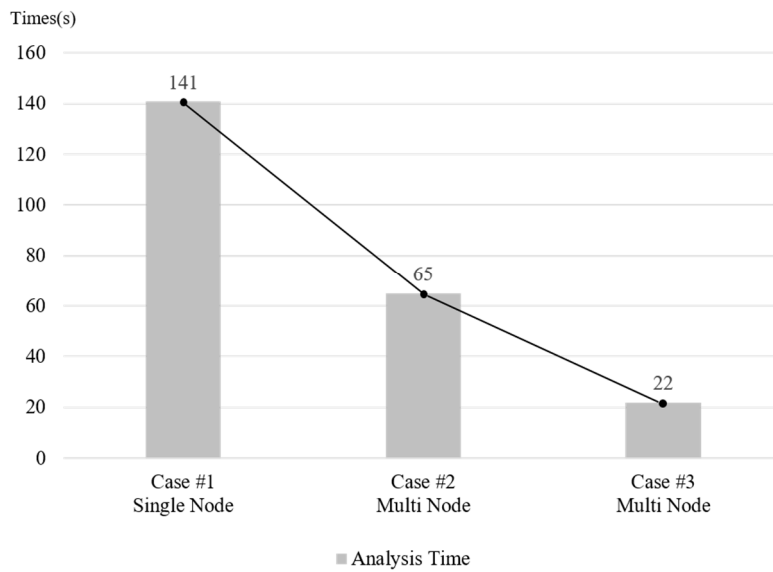
Figure 10. Compatibility test of the proposed system: Performance analysis by the distributed and parallel framework [21].
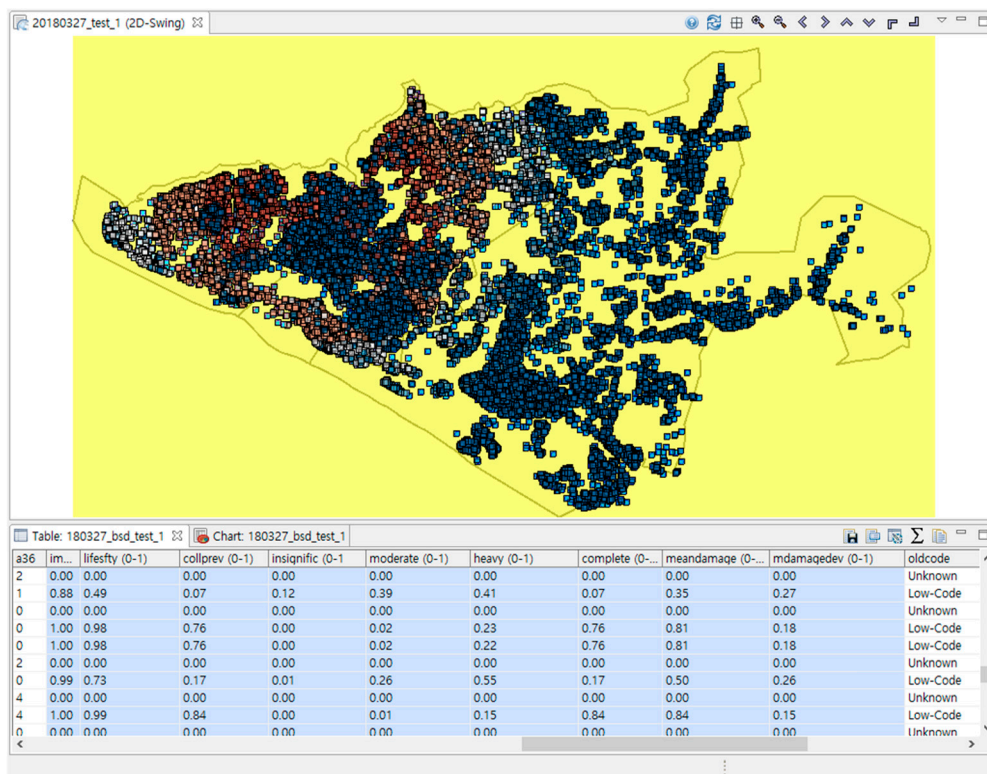

Figure 11. Compatibility test of the proposed system: Risk analysis of Goyang city, Korea.

5. Discussion

It is very crucial to collect data correctly and manage databases efficiently in disaster assessment and prevention. However, it is not easy to apply special types of raw datasets directly for the conventional database systems which can be used in assessing risk or loss of the national or local government [28]. Besides, most of applications dealing with various datasets focused on the construction and utilization of a database system to achieve the goal of researches, without considering the scalability and versatility

essential to ensuring data reuse. Therefore, these systems have limitations and cannot play a key role in a data platform expected to facilitate the various evaluations by sharing and collecting datasets [28–31].

Table 8 presents the comparison between existing systems and the proposed system. The database architecture of the proposed system allows users to easily access, manage, and modify datasets using the web, and supports useful information provided by various application tools. Users are able to upload datasets, such as building, ground information, and fragility curves, etc., that are used in their researches, to the database system; therefore, this system enables researchers to verify data as well as conduct analysis on the current research topic. We show that the proposed system not only supports data compatibility and interoperability regarding HAZUS-MH and ERGO, but also provides distributed and parallel processing which can handle large volumes of datasets.

Table 8. Comparison between existing systems and the proposed system.

	HAZUS-MH [3]	ERGO [4]	Proposed System
Data interoperability	No	No	Yes
System scalability	No	Yes	Yes
Parallel and distributed processing support	No	No	Yes
User access type	Standalone	Standalone	Standalone/Web

6. Conclusions

In this study, a database system based on the inventory of South Korea was designed to construct a loss assessment system for multi-disasters that may occur in South Korea. Data on buildings where people's main activities take place, and resident population data for each output area, are provided by public open data.

Most of public open data has various types of data formats, models, and it requires lots of time to transform into useful datasets which can be employed in various applications. Therefore, this paper proposes a scalable design method of a database system that can be easily employed in Korean loss estimation due to natural hazards. All of the datasets that were gathered and reshaped in this study consider both system compatibility and data interoperability.

In addition, through the compatible test considering both real-time and batch processing case, we showed that the proposed system is able to serve as a data platform of disaster risk reduction and prevention.

The database system construction for South Korea that was performed in this study will be provided to the loss and prediction assessment system for responding to domestic multi-disasters, and to disaster response and analysis researchers. The proposed database system will be of great help in predicting the disaster scale and degree of damage, and in supporting decision-making to minimize damage. Although we focus primarily on the case of the Korean environment in this study, we believe the design methodology introduced in this study can be generalized to establish good guideline models for other regions or countries.

References

1.   Chai, S.; Suh, D.; Shin, S. Study on the development of Korean inventory for the multi-hazard risk assessment. *J. Digit. Contents Soc.* **2017**, *18*, 1127–1134.

2. Chai, S.; Suh, D. A Review and Analysis of Earthquake Disaster Risk Assessment Tools and Applications. *J. Digit. Contents Soc.* **2018**, *19*, 899–906. [CrossRef]

3. *HAZUS-MH 2.1 Technical Manual*; Department of Homeland Security Federal Emergency Management Agency: Washington, DC, USA, 2013.

4. ERGO. Available online: http://ergo.ncsa.illinois.edu (accessed on 19 December 2019).

5. Kang, S.; Kim, K.; Kim, D. A Preliminary Study of the Global Application of HAZUS and ShakeMap for Loss Estimation from a Scenario Earthquake in the Korean Peninsula. *J. Korean Assoc. Geogr. Inf. Stud.* **2017**, *10*, 47–59.

6. Yu, S.; An, H. Study of Flood Loss Estimation using Hazus-MH 2.1. *J. Risk Manag.* **2013**, *24*, 29–58.

7. Jo, Y.; Choi, H.; Choi, S. Conceptual Design of Damage Assessment Inventory in Response to Disaster Risk for Infrastructures Close to River. *J. Korean Assoc. Geogr. Inf. Stud.* **2014**, *17*, 144–158. [CrossRef]

8. Scawthron, C.; Flores, P.; Blais, N. HAZUS-MH flood loss estimation methodology. II. Damage and loss assessment. *Nat. Hazards Rev.* **2006**, *7*, 72–81. [CrossRef]

9. Hancilar, U.; El-Hussain, I.; Sesetyan, K. Earthquake risk assessment for the building inventory of Muscat, Sultanate of Oman. *Nat. Hazards* **2018**, *93*, 1419–1434. [CrossRef]

10. Jeong, G.; Lee, H.; Kwon, O. Earthquake Direct Economic Loss Estimation of Building Structures in Gangnam-Gu District in Seoul Using HAZUS Framework. *J. Earthq. Eng. Soc. Korea* **2016**, *20*, 391–400. [CrossRef]

11. Park, M.; Yang, S.; Koo, B. A Study on Development of Village Wetlands Inventory Using GIS and Establishment of Management Methods in Asan City, Korea. *J. Korea Soc. Environ. Restor. Technol.* **2015**, *18*, 167–177. [CrossRef]

12. Kang, I.; Park, J. The Seismic Hazard Study on Chung-Nam Province using HAZUS. *J. Korean Soc. Hazard Mitig.* **2002**, *2*, 73–83.

13. Lee, C.; Kim, S.; Park, I. An Inventory Analysis for Disaster Loss Estimation Specialized in Korean Environment - Focusing on HAZUS-MH Flood Model. *J. Korean Soc. Hazard Mitig.* **2015**, *15*, 79–87. [CrossRef]

14. SYNER-G. Available online: http://www.vce.at/SYNER-G (accessed on 23 September 2019).

15. Steelman, J.; Song, J.; Hajjar, J.F. *Integrated Data Flow and Risk Aggregation for Consequence-Based Risk Management of Seismic Regional Loss*; University of Illinois: Urbana, IL, USA, 2007.

16. Open Data Portal. Available online: https://www.data.go.kr (accessed on 23 September 2019).

17. USGS. Available online: https://www.usgs.gov/media/images/2017-usgs-juneau-topographic-map-without-image-backdrop (accessed on 17 December 2019).

18. Korea Earthquake Research Center. Available online: http://quake.kigam.re.kr/html/korean/main.jsp (accessed on 23 September 2019).

19. JSON. Available online: https://www.json.org/ (accessed on 23 September 2019).

20. Moniruzzaman, A.B.M.; Hossain, S.A. NoSQL Database: New Era of Databases for Big data Analytics-Classification, Characteristics and Comparison. *Int. J. Database Theory Appl.* **2013**, *6*, 1–4.

21. Chai, S.; Jang, S.; Suh, D. Design and Implementation of Big Data Analytics Framework for Disaster Risk Assessment. *J. Digit. Contents Soc.* **2018**, *19*, 771–777.

22. National Special Data Infrastructure Portal. Available online: http://www.nsdi.go.kr/ (accessed on 23 September 2019).

23. Korean Statistical Information Service. Available online: http://kosis.kr (accessed on 23 September 2019).

24. Ministry of the Interior and Safety (Korea). Available online: http://www.mois.go.kr/frt/a01/frtMain.do (accessed on 23 September 2019).

25. Toro, G. Modification of the Toro et al. (1997) attenuation equations for large magnitudes and short distances. Available online: https://s3.amazonaws.com/academia.edu.documents/39538064/MODIFICATION_OF_THE_TORO_ET_AL._1997_A20151029-26206-13cndhr.pdf?response-content-disposition=inline%3B%20filename%3DMODIFICATION_OF_THE_TORO_ET_AL._1997_ATT.pdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIWOWYYGZ2Y53UL3A%2F20191223%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20191223T084647Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=2bf9626810491069566452743094d506f69b7a655ec3c78697b493de421c1dd9 (accessed on 23 September 2019).

26. Jo, N.; Baag, C. Estimation of Spectrum Decay *Parameter* χ and Stochastic Prediction of Strong Ground Motions in Southeastern Korea. *J. Earthq. Eng. Soc. Korea* **2003**, *7*, 59–70.

27. Noh, M.; Lee, K. Estimation of Peak Ground Motions in the Southeastern Part of the Korean peninsula (II): Development of Predictive Equations. *J. Geol. Soc. Korea* **1995**, *31*, 175–187.

28. Nam, Y.; Kim, J. A Study on the policy of disaster prevention using integrated database - Focused on The Policy for Housing welfare. *J. Korea Inst. Electron. Commun. Sci.* **2014**, *9*, 105–110. [CrossRef]

29. Park, J.; Jung, W.; Choi, H. Prevention Meteorological Database Information for the Assessment of Natural Disaster. *J. Korean Soc. Hazard Mitig.* **2017**, *7*, 41–49.

30. Lee, S.; Kim, G.; Yune, C. Development of Landslide-Risk Prediction Model thorough Database Construction. *J. Korean Geotech. Soc.* **2012**, *28*, 23–33. [CrossRef]

31. Zêzere, J.; Pereira, S.; Tavares, A.O.; Bateira, C.; Trigo, R.M.; Quaresma, I.; Santos, P.P.; Santos, M.; Verde, J. DISASTER: A GIS database on hydro-geomorphologic disasters in Portugal. *Nat. Hazards* **2014**, *72*, 503–532. [CrossRef]