



**A QUERY-DRIVEN SPATIAL DATA WAREHOUSE  
CONCEPTUAL SCHEMA FOR DISASTER MANAGEMENT**

**SAFIZA SUHANA BINTI KAMAL BAHARIN**

**DOCTOR OF PHILOSOPHY**

**2016**



**Faculty of Information and Communication Technology**

**A QUERY-DRIVEN SPATIAL DATA WAREHOUSE CONCEPTUAL  
SCHEMA FOR DISASTER MANAGEMENT**

**Safiza Suhana Binti Kamal Baharin**

**Doctor of Philosophy**

**2016**

**A QUERY-DRIVEN SPATIAL DATA WAREHOUSE CONCEPTUAL SCHEMA  
FOR DISASTER MANAGEMENT**

**SAFIZA SUHANA BINTI KAMAL BAHARIN**

**A thesis submitted  
in fulfillment of the requirements for the degree of Doctor of Philosophy**

**Faculty of Information and Communication Technology**

**UNIVERSITI TEKNIKAL MALAYSIA MELAKA**

**2016**

## DECLARATION

I declare that this thesis entitle “A Query-Driven Spatial Data Warehouse Conceptual Schema for Disaster Management” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : .....

Name : Safiza Suhana Binti Kamal Baharin

Date : .....

## APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Doctor of Philosophy.

Signature : .....

Supervisor Name : Prof. Dr. Nanna Suryana Herman

Date : .....

## DEDICATION

In the name of Allah, The Most Beneficent, The Most Merciful. May the best prayers, and the most abundant peace, be upon our exemplar The Prophet Muhammad ﷺ and upon his family and companions.

This thesis is dedicated:

To Allah ﷻ, for the guidance, remove all obstacles from and lighted my path, inspired me, and always there in every moment. Thank You Allah!

To both my parents: *Mak & Abah*, for their endless love, support and encouragement.

To my dearest husband, a partner in life: *Ahmad Kamal* who leads me through the valley of darkness with light of hope and support,

To my beloved kids whom I can't force myself to stop loving: *Muhammad AmmarFaris, Adriana Camellia and Aryssa Iris.*

And to all my family and friends.

## ABSTRACT

Malaysia has experienced various types of disasters. Such events cause billions of USD and posing great challenges to a nation's government to provide better disaster management. Indeed, disaster management is an important global problem. The National Security Council's (NSC) Directive No. 20 outlines Malaysia's policy on disaster and relief management demonstrates government efforts and initiatives to efficiently respond to disasters. In this regard, decision making is a key factor for organizational success. Positive outcomes are dependent on available data that can be manipulated to provide information to the decision maker, who faces the difficult and complex task of anticipating upcoming events and analyzing multiple parameters. Disaster management involves multiple sources for data collection at various levels as well as a wide array of stakeholders. Hence, accessibility to heterogenous spatial data is challenging. It is crucial to address this problem in terms of data distribution, query operation, and the analyzation task because each resource, level, and stakeholder involved has personal preferences with regard to its format, structure, syntax, and schema. The main purpose of this research is to *support the complex decision-making process during disaster management by enriching the body of knowledge on spatial data warehousing, particularly for conceptual schema design*. A major research problem identified are the heterogeneity of a spatial resource data model, the most appropriate approach to schema design, and the level to which the schema is dependent on the given tools. These problems must be addressed as they are main roadblocks to the process of accessing and retrieving information. The existence of heterogeneous data sources and restricted accessibility to relevant information during a disaster causes several issues with spatial data warehouse design. It can be classified into three considerations namely, the need for guidelines and formalism, schema generation model and a schema design framework and finally, a generalized schema. Four strategies have been designed to address the aforementioned problems: identifying relevant requirements, creating a conceptual design framework, deriving an appropriate schema, and refining the proposed method. User queries are prioritized in the conceptual design framework. Outputs from the formalization process are used with a schema algorithm to effectively derive a generalized schema. The conceptual model framework is taken to be representative of a potential application/ system that has been developed to design a conceptual schema using the problematic heterogeneous data and a restricted approach concerning any corresponding query formalisms. In the schema derivation phase, the conceptual schema that was produced by implementing the proposed framework is presented along with the final conceptual schema. This design is then incorporated into a tool to run an experiment demonstrating that queries from a heterogeneous context are capable of performing context-appropriate conceptual schema design in generic way. Such results outshine the capabilities of a restricted design approach and could potentially answer any relevant queries in less time.

## ABSTRAK

*Malaysia telah mengalami pelbagai bencana. Peristiwa tersebut menuntut berbilion USD dan memberikan cabaran kepada kerajaan untuk menyediakan pengurusan bencana yang lebih baik. Pengurusan Bencana adalah masalah global yang penting. Majlis Keselamatan Negara (MKN), Arahan No. 20 menggariskan dasar mengenai Bencana dan Pengurusan Bantuan sebagai satu usaha dan inisiatif kerajaan untuk menguruskan bencana dengan cara yang berkesan. Dengan itu, membuat keputusan adalah faktor utama bagi kejayaan sesebuah organisasi. Kejayaan organisasi bergantung kepada ketersediaan data untuk menghasilkan maklumat berguna kepada pembuat keputusan yang menghadapi tugas sukar dan kompleks untuk menjangka peristiwa akan datang dan menganalisis dengan pelbagai parameter. Pengurusan bencana melibatkan pelbagai sumber data di pelbagai peringkat serta pihak berkepentingan. Dengan itu, proses capaian kepada kepelbagaian sumber data spatial menjadi suatu yang kompleks dan mencabar. Ia penting untuk ditangani dari segi pengedaran data, operasi pertanyaan dan tugas menganalisis kerana setiap sumber, peringkat dan pihak berkepentingan yang terlibat mempunyai pilihan mereka sendiri termasuk format, struktur, sintaksis dan skema. Tujuan utama kajian ini adalah untuk menyokong proses membuat keputusan yang kompleks semasa bencana melalui pengayaan ilmu pengetahuan berkaitan gudang data spatial terutamanya dalam reka bentuk skema konseptual. Masalah utama penyelidikan yang dikenalpasti adalah kepelbagaian model sumber data spatial, pendekatan reka bentuk skema yang paling sesuai, dan alat-bantuan skema. Masalah tersebut perlu ditangani segera kerana mereka adalah punca utama kesukaran dan sekatan dalam mengakses dan mendapatkan maklumat. Kewujudan sumber data yang perbagai dan akses yang terhad dalam proses membuat keputusan semasa bencana disebabkan oleh beberapa isu yang berkaitan dengan reka bentuk gudang data spatial yang diklasifikasikan kepada tiga aspek keperluan iaitu, keperluan untuk mempunyai garis panduan dan maklumat formalisasi keperluan, model penjanaan skema dan Rangka Kerja Rekabentuk Skema serta aspek terakhir adalah keperluan Skema Teritlak (umum). Empat strategi telah direka untuk masalah yang dikenalpasti iaitu keperluan, rangka kerja reka bentuk konsep, menerbitkan skema yang sesuai dan penghalusan skema. Dalam rangkakerja rekabentuk konsep, pertanyaan pengguna diberi keutamaan paling atas. Rangka kerja model konseptual di ambil untuk dijadikan sebagai wakil aplikasi/sistem bagi merekabentuk skema konseptual dengan menggunakan masalah kepelbagaian data dan pendekatan yang terhalang kedalam proses formalisasi pertanyaan. Dalam terbitan skema, skema konseptual yang dihasilkan oleh rangka kerja yang dicadangkan telah dilaksanakan. Rekabentuk ini kemudian bekerjasama dengan satu alat untuk melarikan eksperimen yang mana kemudiannya menunjukkan pertanyaan yang mempunyai kepelbagaian konteks boleh mempersembahkan rekabentuk skema konseptual berkonteks dan teritlak berbanding pendekatan terhalang, dan juga boleh menjawab pertanyaan dalam masa yang pantas.*



## ACKNOWLEDGEMENT

First and foremost, all praises to Allah ﷻ the Lord of the worlds for His help and bless in the whole process of the preparation, compiling and writing of this thesis. I am totally sure that this work has never came truth without His guidance, blessing, strength and help.

Special appreciation goes to my supervisor, Prof. Dr. Nanna Suryana Herman for his supervision and support. Not forgotten, I would like to express my gratitude to my co-supervisor, Dr. Abdul Samad Shibghatullah for his help.

I owe a deep dept of gratitude to the Malaysia Government on the *Skim Latihan Pensyarah IPTA (SLAI)* under Ministry of Higher Education (MOHE) and Universiti Teknikal Malaysia Melaka (UTeM) to the generous financial support and allowed me to pursue the PhD endeavor.

I am grateful to the numbers of people, En. Daeng from JPAM, En. Zul from JPBD, Melaka and to the admin and technical of FTMK staff for all the instances in which their assistance helped me along the way.

Last but not least, my deepest gratitude goes to my family and family in laws for their support they provided me through my entire life and to all my colleagues, friends and people who took part in making this thesis real directly or indirectly; without whom this thesis might not be completely written.

## TABLE OF CONTENTS

	<b>PAGE</b>
<b>DECLARATION</b>	
<b>APPROVAL</b>	
<b>DEDICATION</b>	
<b>ABSTRACT</b>	<b>i</b>
<b>ABSTRAK</b>	<b>ii</b>
<b>ACKNOWLEDGEMENT</b>	<b>iii</b>
<b>TABLE OF CONTENTS</b>	<b>iv</b>
<b>LIST OF TABLES</b>	<b>viii</b>
<b>LIST OF FIGURES</b>	<b>x</b>
<b>LIST OF APPENDICES</b>	<b>xiii</b>
<b>LIST OF ABBREVIATIONS</b>	<b>xiv</b>
<b>LIST OF PUBLICATIONS</b>	<b>xvi</b>
<b>CHAPTER</b>	
<b>1. INTRODUCTION</b>	<b>1</b>
1.1 The importance of Disaster Management: An Overview	1
1.2 Research Background	8
1.3 Research Motivation	10
1.4 Research Problems	13
1.4.1 Lack of Guidelines and Formalism	16
1.4.2 The Absence of Schema Generation Model	17
1.4.3 The Absence of Generic Schema Derivation And Evaluation	18
1.5 Research Questions	20
1.6 Research Objectives	21
1.7 Research Scope	22
1.8 Research Contributions	22
1.8.1 Conceptual Contributions	23
1.8.2 Technical Contributions	24
1.9 Organization of the Thesis	26
1.10 Chapter Summary	28
<b>2. LITERATURE REVIEW</b>	<b>29</b>
2.1 Introduction	29
2.2 The Disaster Management System in Malaysia	30
2.2.1 Levels of Disaster Management in Malaysia	33
2.2.2 Disaster Management and the Decision Making Process	35
2.3 General Concept of a Spatial Data Warehouse	37
2.3.1 The Architecture of a Spatial Data Warehouse	41
2.4 Spatial Data Warehousing Design Issues	57
2.5 Disaster Management SDW Design Properties	66
2.5.1 Multidimensionality	77
2.5.2 Spatial Aggregation	82
2.5.3 Spatial Measures	83
2.5.4 Spatial Hierarchy and Pre-aggregation	84

2.5.5	Spatial Aggregation According to the Hierarchy	85
2.6	Current Approaches to Conceptual Schema Design	86
2.6.1	Top-down Design	86
2.6.2	Bottom-up Design	91
2.7	Roles of Queries	99
2.7.1	Context-Driven Schema Design	99
2.7.2	Query-Driven Design Approach	101
2.7.3	Query Types	102
2.7.4	Analytical Comparison with Contextual Query-Oriented Schema Design	104
2.8	Evaluation Methods	110
2.9	Conclusion	114
<b>3.</b>	<b>RESEARCH METHODOLOGY</b>	<b>115</b>
3.1	Introduction	115
3.2	Research Approach	116
3.3	Research Framework and Development	118
3.3.1	Theoretical Background and Conceptual Framework at the Empirical Study Stage	119
3.3.2	Analysis and Design Stage	124
3.3.3	Implementation and Evaluation	128
3.4	Research Implementation and Evaluation	131
3.4.1	Requirements Analysis Implementation	131
3.4.2	Conceptual Schema Design Framework Implementation	135
3.4.3	Conceptual Schema Generation and Derivation Implementation	136
3.4.4	Conceptual Schema Refinement Implementation and Evaluation	137
3.5	Summary	139
<b>4.</b>	<b>DEVELOPMENT OF CONTEXT-BASED-QUERY-DRIVEN CONCEPTUAL SCHEMA DESIGN FRAMEWORK</b>	<b>140</b>
4.1	Introduction	140
4.2	Requirements/ Organizational and Formalism	141
4.2.1	Defining Organizational Performance Requirements	142
4.2.2	Categorizing the Organizational Structure Process	149
4.2.3	Basic Foundation Management	155
4.3	Query Management	164
4.3.1	Context-Based-Query-Tree	166
4.3.2	Single Measure Type: Single Category (SMS-Type)	168
4.3.3	Single Measure Type: Multiple Category (SMM-Type)	171
4.3.4	Multiple Measure Type (MM-Type)	172
4.3.5	Exception Query (E-Type)	173
4.4	The Context-Based-Query-Driven Conceptual Schema Design Framework	175
4.5	Summary	177

<b>5.</b>	<b>IMPLEMENTATION OF CONCEPTUAL SCHEMA DESIGN TO SUPPORT AD-HOC QUERIES</b>	<b>178</b>
5.1	Introduction	178
5.2	The Context-Based-Query-Driven (CBQD) Schema Using Queries	179
5.3	Mapping Query Trees	183
5.3.1	TypeI : Spatial	183
5.3.2	TypeII: Spatial/ Attribute	185
5.3.3	Type III: Contextual attribute	186
5.3.4	Other Query Types	186
5.3.5	Validation and Categorization of CBQG	186
5.3.6	Matching Function	188
5.4	SMS-Type: Single Queries and Mapping	191
5.4.1	Mapping SMS-TypeI Queries	192
5.4.2	Mapping SMS-TypeII Queries	197
5.4.3	Mapping SMS-TypeIII Queries	200
5.5	SMM-Type: Multiple Queries and Mapping	201
5.5.1	Mapping SMM-TypeI Queries	201
5.5.2	Mapping SMM-TypeII Queries	202
5.5.3	Mapping SMM-TypeIII Queries	204
5.6	Mapping Other Query Types	204
5.6.1	Mapping MM-Type Queries	205
5.6.2	Mapping Exception Queries	205
5.7	Reconsidering Rejected Queries	206
5.8	Mapping a CBQG to CBQD	207
5.9	Summary	210
<b>6.</b>	<b>CONTEXT-BASED-QUERY-DRIVEN APPROACH TO CONCEPTUAL SCHEMA DESIGN SYNTHESIS AND EVALUATION</b>	<b>212</b>
6.1	Introduction	212
6.2	Mapping of Schema to Existing SDW Models	213
6.2.1	Spatial Cube using CBQD Schema	213
6.2.2	Mapping to Star-Schema	224
6.3	Reduction Concepts	228
6.3.1	Map and Reduce algorithm	230
6.4	Experimental Approach	234
6.4.1	Evaluation I (Annova test)	240
6.4.2	A Precision-Recall Analysis of the Experimental Results	244
6.5	Summary	248
<b>7.</b>	<b>CONCLUSION</b>	<b>249</b>
7.1	Introduction	249
7.2	Research Questions Revisited	249
7.2.1	Conclusion on the Set of Requirements for the Decision-Making Process in Disaster Management for the Development of a Generic Conceptual Schema Design Framework (Research Objective 1 & 2)	250

7.2.2	Conclusion on the Systematic Design in the Conceptual Schema Generation Method (Research Objective 3)	252
7.2.3	Conclusion on the Systematic Design of the General and Independent Designs (Research Objective 4)	252
7.3	Future Works	253

<b>REFERENCES</b>	<b>255</b>
<b>APPENDICES</b>	

## LIST OF TABLES

<b>TABLE</b>	<b>TITLE</b>	<b>PAGE</b>
1.1	Summary of research problems	19
1.2	Summary of research contributions	25
2.1	Level of Disaster in Malaysia	34
2.2	Differences between an operational database and a data warehouse	45
2.3	Comparative Analysis on Existing Design Approach	109
3.1	Mapping of Research Stages with Research Problem, Research Question and Research Objectives	122
3.2	Summary of subject's studies	123
3.3	Empirical Study Level Research Plan	125
3.4	Development levels for the research plan	129
4.1	Vertex in the Basic Foundation Graph (BFG)	162
4.2	Edge in Basic Foundation Graph (BFG)	162
5.1	CBQD Properties	180
5.2	Possibilities and Corresponding Outcomes	181
6.1	Details of Hardware Used	236
6.2	Details of Software Used	237
6.3	MDX vs Cypher Query	239
6.4	Type of Test Cases	240
6.5	Rubric or Conceptual Schema Properties Evaluation	241

6.6	ANOVA analysis of schema quality	242
6.7	The ANOVA analysis of time execution to retrieve the information	243
6.8	Recall and Precision Evaluation	246

## LIST OF FIGURES

FIGURE	TITLE	PAGE
1.1	Occurrence of reported natural disasters by continent: 2002 – 2011	1
1.2	Emergency Command Centre (ECC)	5
1.3	Spatial data accessibility issues	13
1.4	Research structural processes	26
2.1	The cycle of phases in Disaster Management	31
2.2	General architecture of Data Warehouse	42
2.3	Multidimensional database concepts (non-spatial) data cubes	47
2.4	The multidimensional spatial model	51
2.5	Spatial Dimension	52
2.6	OLAP operators on multidimensional data cube	56
2.7	Illustration of key ideas in the design of SDW	57
2.8	General analysis and decision making process	69
2.9	Generalized Representation of Star Schema	80
2.10	Generalized Representation of Snowflake Schema	81
2.11	Generalized Representation of Fact Constellation Schema	81
2.12	Example of results from aggregation operation of geometric union	82
3.1	Research design	115
3.2	The Research Framework and Development	119



3.3	The Research Operational Implementation and Evaluation Framework	131
3.4	The K-chart of the research implementation and evaluation stage	138
4.1	Requirement and Organizational Formalization Process	141
4.2	Requirements/ Organizational Management Process	143
4.3	An Example of Contextual Schema Graph	150
4.4	A grouping function in CSG	151
4.5	An Example of CS Lattices	153
4.6	Representation of contextual attribute (Da) water level	155
4.7	The Basic Foundation Graph (BFG)	161
4.8	The Basic Foundation Graph (Disaster Management Case Study)	163
4.9	A CBQG from a Case Study	166
4.10	A Context-Based-Query-Tree (CBQT) structure	168
4.11	The Context-Based Query-Driven schema design framework	176
5.1	CBQD Schema Algorithm	182
5.2	Flowchart of Validation and Categorization of CBQG	187
5.3	Flowchart of CBQG Classification	188
5.4	Flowchart of Mapping Process	194
5.5	Pseudocode of CBQG Mapping Process for SMS-TypeI	196
5.6	CBQG of SMS-TypeI	197
5.7	CBQG of SMS-TypeII	200
5.8	CBQG of SMS-TypeIII	200
5.9	CBQG of SMM-TypeI	202
5.10	CBQG of SMM-TypeII	203
5.11	CBQG of SMM-TypeIII	204
5.12	Context-Based-Query-Driven (CBQD) Schema	207

6.1	A CBQD Schema from a Case Study	214
6.2	Dimension Selection Algorithm	220
6.3	An Example of Star Constellation	226
6.4	An Example of Star Galaxy	227
6.5	Pseudocode of Map and Reduce Algorithm	230
6.6	An Example of Minimal Schema	233
6.7	MDX Query Syntax	236
6.8	Neo4j Community Server Start Page	238
6.9	Cypher Query Syntax	239
6.10	Mean quality scores as a function of problem complexity levels and design	242
6.11	Mean time scores as a function of problem complexity levels and design	244

## LIST OF APPENDICES

APPENDIX	TITLE
A	<i>Aide Memoir</i>
B	SMS-TYPEII Queries
C	SMS-TYPEIII Queries
D	SMM-TYPEI Queries
E	SMM-TYPEII Queries
F	SMM-TYPEIII Queries
G	Test Cases for Test Data
H	MDX vs Cypher Language
I	Average Score of Conceptual Schema Quality
J	Average approximate time used to retrieve the information

## LIST OF ABBREVIATIONS

CCC	Contingent Control Center,
CPO	Chief Police Officer
CBQD	Context-Based Query Driven
CBQG	Context-Based Query Graph
CBQT	Context-Based Query Tree
DB	Database
DCC	District Control Centre
DDMRC	District Disaster Management and Relief Committee
DM	Decision Maker
DMRC	Disaster Management and Relief Committee
DOCC	Disaster Operation Controlling Centre
DW	Data Warehouse
ECC	Emergency Command Centre
EMD	Emergency Medical Department
ER	Entity Relationship
ETL	extraction, transformation and load
GIS	Geographical Information System
GKD	geographic knowledge discovery
ICT	Information and Communication Technology

IoT	Internet of Thing
LoD	Level of Detail
MCC	Malaysia Control Centre
NDMRC	National Disaster Management and Relief Committee
NL	Natural Language
NSC	National Security Council
OCPD	Officer in Charge of Police District,
ODS	operational data store
OLAP	On-Line Analytical Processing
OLTP	On-line Transactional Processing
OSC	On-Scene Commander
OSCP	On Scene Control Post
OWB	Oracle Warehouse Builder
R&D	Research and Development
RO	Research Objective
RP	Research Problem
RQ	Research Question
SDB	Spatial Database
SDMRC	State Disaster Management and Relief Committee
SDW	Spatial Data Warehouse
SOLAP	Spatial Online Analytical Processing
SOP	Standard Operating Procedure
SQL	Structured Query Language
UI	User Interface

## LIST OF PUBLICATIONS

Safiza Suhana Kamal Baharin and Nanna Suryana Herman (2007), The Importance of Managing Context and Moving Objects Data for Emergency Management System, In: Marizan Sulaiman, Nanna Suryana Herman, et al. (Eds.); Proceedings of the International Conference on Engineering and ICT (ICEI 2007), Melaka, Malaysia

Edward Verbree and Safiza Suhana Binti Kamal Baharin, (2007), Geo-info to-go, In: Proceedings of the 4th International Symposium on LBS & TeleCartography, Hongkong.

Safiza Suhana Kamal Baharin and Sisi Zlatanova, (2008), Managing Moving Objects and Context Data for Emergency Management System, Journal Advanced Manufacturing Technology, Vol (2). No (1), Malaysia.

Zlatanova, S and Safiza Suhana Kamal Baharin (2009), Optimal Navigation of First Responders using DBMS In: Proceedings of the 4th International Symposium on LBS & TeleCartography, Hongkong.

Safiza Suhana Kamal Baharin, Abdul Samad Shibghatullah and Zahriah Othman (2009), Disaster Management in Malaysia: An Application Framework of Integrated Routing Application for Emergency Response Management System, In International Conference of Soft Computing and Pattern Recognition, Socpar09, Melaka, Malaysia

Khairul Anshar, Nanna Suryana Herman, Zahriah Othman, Safiza Suhana Kamal Baharin, (2015) Different Geospatial Data on Hybrid Map Application. *Journal of Networks, North America*, 10, Aug. 201 5

Suryana, N., Othman, Z., Baharin, S.S.K. The Use of Ontology Merging and Parsing Technique in Handling and Retrieving Large Spatial Data. *International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC)*. (Submitted).

Suryana, N., Othman, Z., Baharin, S.S.K. Large Spatial Data Integration Framework and Semantic Translator Engine in Resolving Semantic and Schematic Heterogeneity. *International Journal of Advanced Engineering Research and Science (IJAERS)*. (Submitted).

# CHAPTER 1

## INTRODUCTION

### 1.1 The importance of Disaster Management: An Overview

Disasters are characterized by the scope of emergency they cause. When managing an emergency exceeds a local entity's capabilities, the situation becomes a disaster. According to statistics reported in the 2012 by International Federation of Red Cross and Red Crescent Societies (IFRC) (IFRC World Disasters Report (2012)), the continent most often hit by natural disasters was Asia, as shown in Figure 1.1. Disasters may occur anytime and anywhere, thus posing great challenges to a nation's government, which acts as a policy maker in order to provide better disaster management.

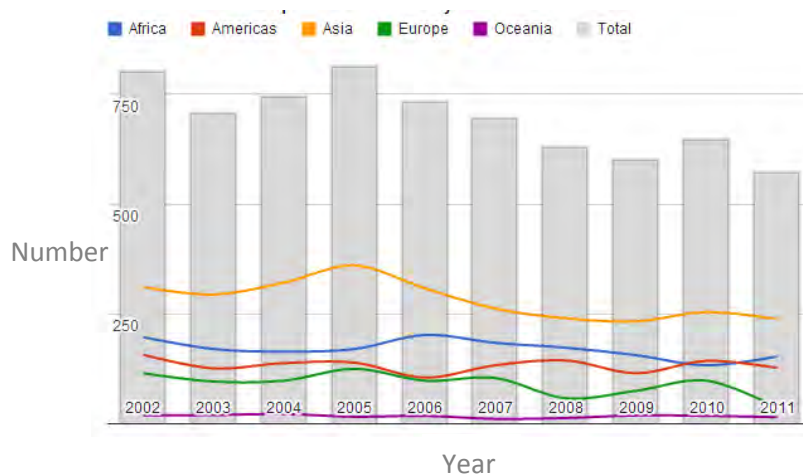


Figure 1.1 Occurrence of reported natural disasters by continent: 2002 – 2011 (*source*: IFRC World Disasters Report, 2012)