An approach for improving decision-making with heterogeneous geospatial big data: an application using spatial decision support systems and volunteered geographic information to disaster management

Flávio Eduardo Aoki Horita

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: ______________________

Flávio Eduardo Aoki Horita

An approach for improving decision-making with heterogeneous geospatial big data: an application using spatial decision support systems and volunteered geographic information to disaster management

Doctoral dissertation submitted to the Instituto de Ciências Matemáticas e de Computação – ICMC-USP, in partial fulfillment of the requirements for the degree of the Doctorate Program in Computer Science and Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Prof. Dr. João Porto de Albuquerque Pereira

USP – São Carlos
May 2017

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados fornecidos pelo(a) autor(a)

Horita, Flávio Eduardo Aoki
H811a An approach for improving decision-making with heterogeneous geospatial big data: an application using spatial decision support systems and volunteered geographic information to disaster management / Flávio Eduardo Aoki Horita; orientador João Porto de Albuquerque Pereira. - São Carlos - SP, 2017.
210 p.

Tese (Doutorado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2017.

1. Decision-making. 2. Heterogeneous Geospatial Data. 3. Spatial Decision Support Systems. 4. Volunteered Geographic Information. 5. Disaster Management. I. Pereira, João Porto de Albuquerque, orient. II. Título.

Flávio Eduardo Aoki Horita

Uma abordagem para melhorar a tomada de decisão com grande volume de dados espaciais heterogêneos: uma aplicação usando sistemas de suporte à decisão espacial e informações geográficas voluntárias na gestão de desastres

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. João Porto de Albuquerque Pereira

USP – São Carlos
Maio de 2017

To my dear love, Marcela S. Mussio

To my parents, Carlos E. H. Horita and Clarice M. A. Horita

To my sister, Camila A. Horita

Para meu grande amor, Marcela S. Mussio

Para meus pais, Carlos E. H. Horita e Clarice M. A. Horita

Para minha irmã, Camila A. Horita

ACKNOWLEDGEMENTS

First of all, I'd like to thank God for all the opportunities that he has offered me, and the strength onto the difficulties that I faced throughout my journey in this project. I truly believe that without my faith, I didn't finish it.

Deep gratitude to my advisor, Prof. Dr. João Porto de Albuquerque that definitely helped me in finishing this project, always pushing me forward, showing the right track, encouraging in the difficult times, and giving important contributions. More than an advisor, he has become a friend with which I could share and improve my knowledge and experience. Even more, he gave me all the conditions for developing this project, and also the opportunity of taking part of the AGORA Research Group.

I'm also thankful to Prof. Dr.-Ing. Bernd Hellingrath from the University of Münster in Germany for giving me the opportunity to learn a lot with his group. All the experiences that I faced in my internship contributed to my project move a step further, as well as the opportunity to learn different perspectives and methodologies of research and to be in touch with other cultures and people.

I'm immensely thankful to my great love, Marcela S. Mussio for her support and patience during all my journey, and also for inspiring me in each conversation that we've. Even when I spent hours explaining my research topic, and I asked her to discuss it with me without she knowing anything about it. She helped me, even more, when I was abroad, talking with me in the hardest periods, and also sharing the great moments.

I wish to express my gratitude to my parents, Carlos E. H. Horita and Clarice M. A. Horita that have taught me all the best things in life, and also for keeping me on the correct track when I was lost or looking for answers. All these steps that I went through were only possible because of their education and efforts. I also would like to thank my sister, Camila A. Horita for the discussions and contributions during the development of this project.

Many thanks are also to my aunt Maria Inês Ohnuma and her family - Daniel Ohnuma, Gustavo Ohnuma, Alfredo Ohnuma, Ana, Erika Alvim, Cristina Guimarães, my uncle Akira Ohnuma (in memoriam), and my little cousins Arthur, Pedro, and Mariana. I'm sincerely grateful for the time that they hosted me in São Carlos, all the conversations, lunches, and dinners that we've have shared. All these moments were wonderful and I owe you all for the rest of my life.

Many thanks to all my family for all the conversations and advice, especially, my grandma, Apparecida Horita (Bá), my uncles Mariangela Aoki, José Roberto Rogero, Cristina

Aoki, Ruy Hizatugu, and Leiko Horita. I'd like to thank my family-in-law, especially, Silvana Mussio and Luiz Mussio who also gave me important and relevant advice on this journey.

Special thanks to my great friend Daniel Link, for the friendship, support, conversations, and encouragement. I can definitely say that his support was essential in many phases of this project.

Thanks to Prof. Dr. Eduardo M. Mendiondo and Prof. Dr. Jó Ueyama for their support and help in the project, contributions onto the papers, and all the brainstorming that we've had. I also learned a lot with both of them.

Special thanks to my group mates (AGORA) Luiz Fernando Assis, Raul Castanhari, Lívia Degrossi, Roberto Rocha, and Sidgley Andrade for all the conversations and contributions to this project. I'm also grateful to the contributions of our research fellows from NIBH, mainly, Jairo Rotava, Danielle Bressiani, Camilo Restrepo, Altair Rosa, Clarissa Camara, Marina Batalini and Maria Clara Fava, and from CEMADEN, in particular, Victor Marchezini.

Thanks to my colleagues and professors from ICMC/USP, in special: Stevão Andrade, Brauner Oliveira, Faimison Porto, Rafael Durelli, Rafael Oliveira, Carlos Damasceno, Kamilla Lyra, Diogenes Dias, Silvana Morita, Wilmax Cruz, Anderson Marcolino, André Endo, Francisco Carlos, Alinne Correa, Valéria Carvalho, Rafael Mantovani, Valdemar Neto, Elias Nogueira, Mauricio Arimoto, Bruno Faiçal, Lina Garcés, Prof. Ellen Francine, Prof. Simone Souza, Prof. Seiji Isotani, Prof. Paulo Masiero, and Prof. Elisa Nakagawa. I've learned a lot with you all.

I also would like to thank my friends in Germany, in special, Diego Siqueira, Matheus Pires, Ajay Kamur, Basheer Sayed, Michael Middelhoff, and Luiz Filipe Araújo, my department mates. Thank you for the coffees, discussions about soccer, and all the funny conversations that we've shared. Finally, I'd like to thank Jim Jones for hosting me as soon as I arrived in Münster and Luciano Sales, my tutor in the city.

Thanks to all members of CEMADEN that have attended to the semi-structured interviews and focus group sessions, which I have conducted at the center as part of this research project.

Thanks to all that have contributed in reviewing any of the articles of this project and of this thesis itself. Your contributions were essential for finishing these texts.

Thanks to the staff and employees of ICMC/SP in Brazil, as well as of the University of Münster in Germany, for their effective and continuous support.

I'm thankful for the financial support provided by the FAPESP-IVA project "Assessment of Impacts and Vulnerability to Climate Change in Brazil and Strategies for Adaptation Options" (Grant no. 2008/58161-1), CAPES (Grant Pró-alertas no. 88887.091744/2014-01 and no. 88887.091743/2014-01), Heidelberg University (Excellence Initiative fund: 2300054, assignment: 7812617), and FAPESP (Grant no. 2008/58161-1, no. 2011/23274-3, and no.

2012/18675-1), and special thanks to CNPq (Grant no. 202453/2014-6) that supported and funded my internship at the University of Münster in Germany.

AGRADECIMENTOS

Primeiramente, gostaria de agradecer a Deus por todas as oportunidades que ele tem me oferecido e toda a força para enfrentar as dificuldades que me foram impostas durante o desenvolvimento deste projeto. Eu realmente acredito que sem fé, eu não teria conseguido!

Muito obrigado ao meu orientador, Prof. Dr. João Porto de Albuquerque que contribuiu para a realização deste projeto, sempre me incentivando, mostrando o caminho certo e oferecendo importantes contribuições. Mais do que um orientador, ele se tornou um amigo com o qual pude compartilhar minhas ideias e conhecimentos. Também, me ofereceu todas as condições para desenvolver esse projeto e a oportunidade de fazer parte do Grupo de Pesquisa AGORA.

Eu gostaria de agradecer ao Prof. Dr.-Ing. Bernd Hellingrath da Universidade de Münster na Alemanha por me oferecer a oportunidade de aprender muito com seu grupo de pesquisa. Todas as experiências que eu enfrentei durante meu estágio contribuíram substancialmente para a melhoria do meu trabalho, além de aperfeiçoar meu conhecimento em outras metodologias de pesquisas e poder conhecer novas culturas e pessoas.

Sou imensamente grato ao meu amor, Marcela S. Mussio, não apenas por todo o suporte e paciência ao longo do desenvolvimento deste projeto, mas também por me inspirar e encorajar em todas as conversas, mesmo quando eu ficava horas explicando minha pesquisa e perguntava incessantemente para ela discuti-lo comigo, sem ter conhecimento no assunto. Me ajudou ainda mais durante meu estágio, sempre me acalmando nos períodos mais difíceis e compartilhando os momentos felizes.

Obrigado aos meus pais, Carlos E. H. Horita e Clarice M. A. Horita que me ensinaram todas as coisas boas da vida e também por me oferecer suporte quando precisava de ajuda. Todos os meus passos somente foram possíveis por causa da sua educação e esforços. Também, agradeço a minha irmã, Camila A. Horita pelas discussões e contribuições ao longo deste projeto.

Muito obrigado a minha tia Maria Inês Aoki Ohnuma e toda a sua familia - Daniel Ohnuma, Gustavo Ohnuma, Alfredo Ohnuma, Ana, Erika Alvim, Cristiana Guimarães, meu tio Akira Ohnuma (in memoriam) e meus priminhos Arthur, Pedro e Mariana. Sou sinceramente grato por todo o tempo que vocês me acolheram em São Carlos, todas as conversas, almoços e jantares que nós compartilhamos. Todos esses momentos foram maravilhosos e eu serei eternamente grato por isso.

Muito obrigado a toda a minha família pelas conversas e conselhos, especialmente, minha avó Apparecida Horita (Bá), meus tios e tias Mariangela Aoki, José Roberto Rogero, Maria

Cristina Aoki, Ruy Hizatugu e Leiko Horita. Também, sou grato a minha segunda família, em especial, Silvana Mussio, Luiz Carlos Mussio, meus sogros, e Dilma Molck, os quais me deram muitos conselhos e ajuda durante o desenvolvimento deste projeto.

Agradeço especialmente ao meu grande amigo Daniel Link, pela sua amizade, suporte, conversas e encorajamento. Eu posso dizer que seu ajuda foi fundamental para que eu finalizasse este projeto.

Obrigado ao Prof. Dr. Eduardo M. Mendiondo e Prof. Dr. Jó Ueyama por todo o suporte a ajuda ao longo do projeto, contribuições em artigos e as discussões enriquecedores que tivemos. Eu também aprendi muito com vocês.

Obrigado aos meus colegas de grupo (AGORA) Luiz Fernando Assis, Raul Castanhari, Lívia Degrossi, Roberto Rocha e Sidgley Andrade por todas as conversas, discussões, conselhos e contribuições neste projeto. As contribuições dos nossos parceiros de pesquisa do NIBH, principalmente, Jairo Rotava, Danielle Bressiani, Camilo Restrepo, Altair Rosa, Clarissa Camara, Marina Batalini, Maria Clara Fava, e do CEMADEN, em particular, ao Victor Marchezini.

Obrigado aos meus colegas e professores do ICMC/USP por todas as conversas, cafés, experiências compartilhadas e festas, em especial: Stevão Andrade, Brauner Oliveira, Faimison Porto, Rafael Durelli, Rafael Oliveira, Carlos Damasceno, Kamilla Lyra, Diogenes Dias, Silvana Morita, Wilmax Cruz, Anderson Marcolino, André Endo, Francisco Carlos, Alinne Correa, Valéria Carvalho, Rafael Mantovani, Valdemar Neto, Elias Nogueira, Mauricio Arimoto, Bruno Faiçal, Lina Garcés, Profa. Ellen Francine, Profa. Simone Souza, Prof. Seiji Isotani, Prof. Paulo Masiero, e Profa. Elisa Nakagawa. Eu aprendi muito com todos vocês.

Também, gostaria de agradecer aos meus amigos da Alemanha, especialmente, Diego Siqueira, Matheus Pires, Ajay Kamur, Basheer Sayed, Michael Middelhoff, e Luiz Filipe Araújo, meus colegas de departamento. Obrigado pelos cafés e todas as conversas que tivemos. Finalmente, gostaria de agradecer ao Jim Jones por ter me hospedado durante meus primeiros dias em Münster e ao Luciano Sales, meu tutor na cidade.

Obrigado a todos os membros do CEMADEN que participaram das entrevistas semi-estruturadas e das sessões de grupos focais, conduzidas no centro como parte deste projeto de pesquisa.

Obrigado a todos aqueles que contribuíram revisando os artigos deste projeto e esta tese. Suas contribuições foram essenciais para a realização destes textos.

Obrigado aos funcionários e pessoal de suporte do ICMC/USP no Brasil e da Universidade de Münster na Alemanha pelo suporte e ajuda.

Sou grato ao financiamento do projeto FAPESP-IVA project “Assessment of Impacts and Vulnerability to Climate Change in Brazil and Strategies for Adaptation Options” (FAPESP no. 2008/58161-1), o projeto Pró-alertas (CAPES no. 88887.091744/2014-01 e no.

88887.091743/2014-01), a FAPESP (Projeto no. 2008/58161-1, no. 2011/23274-3 e no. 2012/18675-1), e agradecimentos especiais ao CNPq (Grant no. 202453/2014-6) por financiar meu estágio na Universidade de Münster na Alemanha.

"On a more human scale, software-intense products have helped cure the sick and have given voice to the speechless, mobility to the impaired, and opportunity to the less able. From all these perspectives, software is an indispensable part of our modern world."

(Grady Booch, 1998)

DECLARATION OF ORIGINAL AUTHORSHIP AND LIST OF PUBLICATIONS

I confirm that this dissertation has not been submitted in support of an application for another degree at this or other teaching or research institution. It is the result of my own work and the use of all material from other sources has been properly and fully acknowledged. Research done in collaboration is also crealy indicated.

Excerpts of this dissertation have been either published or submitted for the appreciation of editorial boards of journals, conferences and workshops, according to the list of publications presented as follows. My contributions to each publication are listed as well.

Publications Resulting from Work on This Thesis

(3) Journal Articles

- **Horita, F. E. A.**; Albuquerque, J. P.; Marchezini, V.; Mendiondo, E. M. "*Bridging the gap between decision-making and emerging big data sources: an application of a model-based framework to disaster management in Brazil*" (HORITA *et al.*, 2017).
  Journal: Decision Support Systems.
  DOI: http://dx.doi.org/10.1016/j.dss.2017.03.001.
  Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with his contributors.

- **Horita, F. E. A.**; Albuquerque, J. P.; Degrossi, L. C.; Ueyama, J.; Mendiondo, E. M. "*Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks*" (HORITA *et al.*, 2015).
  Journal: Computers & Geosciences.
  DOI: http://dx.doi.org/10.1016/j.cageo.2015.04.001
  Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with his contributors.

- **Horita, F. E. A.**; Braga, D. S.; Monteiro, C. D. D. "*The use of multimodal interaction for supporting the production of location-based information*" (HORITA; BRAGA; MONTEIRO, 2016).

Journal: IEEE Latin America Transactions.
DOI: http://dx.doi.org/10.1109/TLA.2016.7587660
Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with his contributors.

(7) Conference Papers

- **Horita, F. E. A.**, Albuquerque, J. P., Marchezini, V. Mendiondo, E. M. "*A qualitative analysis of the early warning process in disaster management*" (HORITA *et al.*, 2016)
  Conference: 13th International Conference on Information Systems for Crisis Response and Management (ISCRAM '16).
  Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with his contributors.

- **Horita, F. E. A.**; Link, D.; Albuquerque, J.P.; Hellingrath, B. "*oDMN: An integrated model to connect decision-Making needs to emerging data sources in disaster management*" (HORITA *et al.*, 2016).
  Conference: 49th Hawaii International Conference on System Sciences (HICSS '16).
  DOI: http://dx.doi.org/10.1109/HICSS.2016.361
  Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with his contributors.

- **Horita, F. E. A.**; Link, D.; Albuquerque, J. P.; Hellingrath, B. "*A Framework for the Integration of Volunteered Geographic Information into Humanitarian Logistics*" (HORITA *et al.*, 2014).
  Conference: 20th Americas Conference on Information Systems (AMCIS '14).
  Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with his contributors.

- **Horita, F. E. A.**; Fava, M. C., Mendiondo, E. M.; Rotava, J.; Souza, V. C.; Ueyama, J., Albuquerque, J. P. "*AGORA-GeoDash: A Geosensor Dashboard for Real-time Flood Risk Monitoring*" (HORITA *et al.*, 2014a).
  Conference: 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM '14).
  Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with his contributors.

- **Horita, F. E. A.**; Assis, L. F. F. G.; Castanhari, R. E. S.; Isotani, S.; Cruz, W. M.; Albuquerque, J. P. "*A gamification-based social collaborative architecture to increase resilience against natural disasters*" (HORITA *et al.*, 2014).
  Conference: X Brazilian Symposium on Information Systems (SBSI '14).

Level of contribution: High - the PhD candidate helped in the definition and evaluation of the conceptual architecture, and in the paper writing.

- **Horita, F. E. A.**; Degrossi, L. C.; Assis, L. F. F. G. ; Zipf, A.; Albuquerque, J. P. "*The use of Volunteered Geographic Information and Crowdsourcing in disaster management: a systematic literature review*" (HORITA *et al.*, 2013).
Conference: 19^{th} Americas Conference on Information Systems (AMCIS '13).
Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with his contributors.

- **Horita, F. E. A.**; Albuquerque, J. P. "*An approach to support decision-Making in disaster management based on Volunteer Geographic Information (VGI) and Spatial Decision Support Systems (SDSS)*" (HORITA; ALBUQUERQUE, 2013a).
Conference: 10^{th} International Conference on Information Systems for Crisis Response and Management (ISCRAM '13).
Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with his contributor.

(2) Workshop Papers

- **Horita, F. E. A.**; Fava, M. C.; Souza, V. C.; Ueyama, J.; Albuquerque, J. P.; Mendiondo, E. M. "*Web-based Tool to deliver useful Real-time Information for Decision-making in the Flood Management*" (HORITA *et al.*, 2014b).
Workshop: Climate Change Adaptation Conference (Adaptation Futures '14).
Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with his contributors.

- **Horita, F. E. A.**; Albuquerque, J. P. "*Providing real-time decision support in flood management using volunteered information*" (HORITA; ALBUQUERQUE, 2013b).
Workshop: 1a Escola Latino Americana de Engenharia de Software (ELA-ES '13).
Level of contribution: High - the PhD candidate is the main investigator and conducted the work together with his contributor.

Other Related Publications

(1) Book Chapter

- Albuquerque, J. P., **Horita, F. E. A.**, Degrossi, L. C., Rocha, R. S., Andrade, S. C., Restrepo-Estrada, C. "*Leveraging Volunteered Geographic Information for improving disaster resilience: Lessons learned with AGORA and future research directions*".

Book: Campelo, C., Bertolotto, M., Corcosan, P. (eds.). Volunteered Geographic Information and the Future of Geospatial Data. Idea Group Inc., (ALBUQUERQUE *et al.*, 2017).
Level of contribution: Medium - the PhD candidate contribute and conducted the work together with his contributors.

(1) Journal Articles

- Assis, L. F. F. G.; Herfort, B.; Steiger, E.; **Horita, F. E. A.**; Albuquerque, J. P. "*Geographical prioritization of social network messages in near real-time using sensor data streams: an application to floods*".
  Journal: Brazilian Journal of Cartography and Brazilian Society of Cartography, Geodesy, Photogrammetry and Remote Sense, (to appear).
  Level of contribution: Low - the PhD candidate helped in the paper writing.

(8) Conference Papers

- Leyh, W.; Fava, M. C.; Abe, N.; Restrepo-Estrada, C.; **Horita, F. E. A.**; Mendiondo, E. M.; Albuquerque, J. P. "*SDI-Node to interlink Information, essential for Disaster Preparedness and Management, with other Linked Open Data*" (LEYH *et al.*, 2016).
  Conference: $13^t h$ International Conference on Information Systems for Crisis Response and Management (ISCRAM '16).
  Level of contribution: Low - the PhD candidate helped in the brainstorming.

- Assis, L. F. F. G.; Behnck, L. P.; Doering, D.; Freitas, E. P.; Pereira, C. P.; **Horita, F. E. A.**; Ueyama, J.; Albuquerque, J. P. "*Dynamic sensor management: extending sensor web for near real-time mobile sensor integration in dynamic scenarios*" (ASSIS *et al.*, 2016).
  Conference: 30^{th} IEEE International Conference on Advanced Information Networking and Applications (AINA '16).
  DOI: http://dx.doi.org/10.1109/AINA.2016.100
  Level of contribution: Medium - the PhD candidate supported the development and evaluation of the work, and helped in the paper writing.

- Assis, L. F. F. G.; Herfort, B.; Steiger, E.; **Horita, F. E. A.**; Albuquerque, J. P. "*Geographical prioritization of social network messages in near real-time using sensor data streams: an application to floods*" (ASSIS *et al.*, 2015).
  Conference: XVI Brazilian Symposium on Geoinformatics (GEOINFO '15).
  Level of contribution: Low - the PhD candidate helped in the paper writing.

- Poiani, T. H.; **Horita, F. E. A.**; Albuquerque, J. P. "*Análise geográfica entre mensagens georreferenciadas de redes sociais e dados oficiais para suporte á tomada de decisões de agências de emergência*" (POIANI; HORITA; ALBUQUERQUE, 2015).

Conference: XVI Brazilian Symposium on Geoinformatics (GEOINFO ‘15).
Level of contribution: Low - the PhD candidate helped in the the paper writing.

- Assis, L. F. F. G.; Herfort, B.. Steiger, E.; **Horita, F. E. A.**; Albuquerque, J. P. "*A geographic approach for on-the-fly prioritization of social-media messages towards improving flood risk management*" (ASSIS *et al.*, 2015).
  Conference: IV Brazilian Workshop on Social Network Analysis and Mining (BraSNAM ‘15).
  Level of contribution: Low - the PhD candidate helped in the the paper writing.

- Link, D.; **Horita, F. E. A.**; Albuquerque, J. P.; Hellingrath, B.; Ghasemivandhonaryar, S. "*A method for extracting task-related Information from social media based on structured domain knowledge*" (LINK *et al.*, 2015).
  Conference: 21^{st} Americas Conference on Information Systems (AMCIS ‘15).
  Level of contribution: Medium - the PhD candidate helped in the analysis of the collected, as well as in the paper writing.

- Rocha, R. S.; Degrossi, L. C.; **Horita, F. E. A.**; Albuquerque, J. P. "*AGORA-PL: Uma Proposta para desenvolvimento de famílias de sistemas colaborativos baseados em VGI para a gestão do risco de inundação*" (ROCHA *et al.*, 2014).
  Conference: Brazilian Symposium on Collaborative Systems (SBSC ‘14).
  Level of contribution: Low - the PhD candidate helped in the development of the conceptual architecture, and in the paper writing.

- Fava, M. C.; Santana, G.; Bressiani, D. A.; Rosa, A.; **Horita, F. E. A.**; Mendiondo, E. M. "*Integration of information technology systems for flood forecasting with hydrid data sources*" (FAVA *et al.*, 2014).
  Conference: 6^{th} International Conference on Flood Management (ICFM ‘14).
  Level of contribution: Low - the PhD candidate supported the development of the conceptual framework and in the paper writing.

RESUMO

HORITA, F. E. A.. **An approach for improving decision-making with heterogeneous geospatial big data: an application using spatial decision support systems and volunteered geographic information to disaster management**. 2017. 210 f. Doctoral dissertation (Doctorate Candidate Program in Computer Science and Computational Mathematics) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

Contexto: Uma tomada de decisão precisa exige informações maisatualizadas para estabelecer a realidade da situação geral. Novas fontes de dados (p. ex., tecnologias vestíveis) tem aumentado a quantidade de dados úteis disponíveis, que agora é chamado de "big data". Isso tem grande potencial para transformar todo o processo de negócio e melhorar a precisão na tomada de decisão. Nesse contexto, a gestão de desastres representa um interessante cenário que depende de "big data" para aprimorar a tomada de decisão. Isso porque, ela tem que lidar com dados fornecidos não apenas por fontes tradicionais (p. ex., sensores estáticos), mas também por fontes emergentes – por exemplo, informações compartilhadas por voluntários locais, ou seja, as informações geográficas de voluntários (VGI). Quando combinadas, essas fontes de dados podem ser consideradas grandes em *volume*, com diferentes *velocidades* e uma *variedade* de formatos. Além disso, uma análise com relação à sua *veracidade* é necessária, uma vez que essas fontes de dados são desconectadas e propensas a erros. Esses são os "4Vs" que caracterizam "big data". **Problema**: Embora todos esses dados abrem novas oportunidades, seu grande volume em conjunto com uma integração inapropriada e uma visualização inadequada podem tornar as informações ignoradas por tomadores de decisão. Isso ocorre, pois, a integração dos dados disponíveis torna-se complicada devido a heterogeneidade intrínseca em suas características (e.g., dados em formatos diferentes). Quando integradas, essas informações frequentemente também não chegam aos tomadores de decisão em uma condição apropriada (p. ex., no formato de visualização adequado). Além disso, não existe uma clara compreensão sobre as necessidades dos tomadores de decisão ou sobre como os dados disponíveis podem ser usados para atender essas necessidades. **Objetivo**: Dessa forma, esta tese de doutorado apresenta uma abordagem para melhorar a tomada de decisões com grande volume de dados espaciais heterogêneos baseada em sistemas de suporte à decisão espacial e informações geográficas de voluntários na gestão de desastres. **Métodos**: Mapeamentos sistemáticos foram conduzidos para identificar lacunas de pesquisa no uso de dados voluntários e sistemas de suporte à decisão na gestão de desastres. Com base nestes estudos, dois projetos de *design science* foram conduzidos. O primeiro deles buscou definir elementos para entender a integração de dados heterogêneos, enquanto o segundo projeto buscou fornecer um melhor entendimento das necessidades dos tomadores de decisão. Também foi conduzido um projeto de pesquisa-ação interinstitucional para definir princípios de projeto que deveriam ser observados para um sistema de suporte à decisão espacial ser efetivo no apoio a tomada de decisão com grande volume de dados espaciais heterogêneos. Uma série de estudos de caso empíricos foram conduzidos para avaliar os resultados desses projetos. **Resultados**:

A abordagem geral então é composta pelos três resultados significantes que foram derivados desses projetos. Em primeiro lugar, uma arquitetura conceitual que especifica a integração de fontes de dados heterogêneas. O segundo elemento é uma estrutura baseada em modelo que descreve a conexão entre a tomada de decisão com as fontes de dados mais adequadas. Com base nessa estrutura, o terceiro elemento consiste em um conjunto de princípios de design que guiam o desenvolvimento de um sistema de suporte à decisão espacial para tomada de decisão com grande volume de dados espaciais heterogêneos. **Conclusão**: Essa tese de doutorado realizou importantes contribuições para a prática e pesquisa. Em resumo, ela define formas para integrar fontes de dados heterogêneos, fornece uma melhor compreensão sobre as necessidades dos tomadores de decisão e ajuda no desenvolvimento de sistemas de suporte à decisão espacial para tomada de decisão com grande volume de dados espaciais heterogêneos.

Palavras-chave: Tomada de Decisão, Dados Espaciais Heterogêneos, Sistemas de Suppote à Decisão Espacial, Informações Geográficas Voluntárias, Gestão de Desastres.

ABSTRACT

HORITA, F. E. A.. **An approach for improving decision-making with heterogeneous geospatial big data: an application using spatial decision support systems and volunteered geographic information to disaster management**. 2017. 210 f. Doctoral dissertation (Doctorate Candidate Program in Computer Science and Computational Mathematics) – Instituto de Ciências Matemáticas e de Computação (ICMC/USP), São Carlos – SP.

Context: Accurate decision-making requires updated and precise information to establish the reality of an overall situation. New data sources (e.g., wearable technologies) have been increasing the amount of available and useful data, which is now called “big data”. This has a great potential for transforming the entire business process and improving the accuracy of decisions. In this context, disaster management represents an interesting scenario that relies on “big data” to enhance decision-making. This is because it must cope with data provided not only by traditional sources (e.g., stationary sensors) but also by emerging sources - for instance, information shared by local volunteers, i.e., volunteered geographic information (VGI). When combined, these data sources can be regarded as large in *volume*, with different *velocities*, and a *variety* of formats. Furthermore, an analysis is required to confirm their *veracity* is required since these data sources are disconnected and prone to various errors. These are the “4Vs” that characterize “big data”. **Gap**: However, although all these data open up further opportunities, their huge volume, together with an inappropriate data integration and unsuitable visualization, can result in information being overlooked by decision-makers. This problem arises because the integration of the available data is hampered by the intrinsic heterogeneity of their features (e.g., their occurrence in different formats). When integrated, this information also often fails to reach the decision-makers in a suitable way (e.g., in appropriate visualization formats). Moreover, there is not a clear understanding of the decision-makers’ needs or how the available data can meet these needs. **Objective**: In light of this, this thesis presents an approach for improving decision-making with heterogeneous geospatial big data based on spatial decision support systems and volunteered geographic information in disaster management. **Methods**: Systematic mapping studies were conducted to identify gaps in research studies with regard to the use of volunteered information and spatial decision support systems in disaster management. On the basis of these studies, two design science projects were carried out. The first of these aimed at defining the elements that are essential for ensuring the integration of heterogeneous data, whereas the second project aimed at obtaining a better understanding of decision-makers’ needs. A cross-organizational action research project was also conducted to define the design principles that should be observed for a spatial decision support system to effectively support decision-making with heterogeneous geospatial big data. A series of empirical case studies was undertaken to evaluate the outcomes of these projects. **Results**: The overall approach thus consists of the three significant outcomes that were derived from these projects. The first outcome was the conceptual architecture that defines the integration of heterogeneous data sources. The second outcome was a model-based framework that describes the connection of decision-making with appropriate data sources. The third outcome is based on the framework and comprises a set of design principles for guiding the development of spatial decision support systems for decision-making with heterogeneous geospatial big data. **Conclusion**: This thesis has made a useful contribution to both practice and research. In short, it defines ways of integrating heterogeneous data sources, provides a better understanding of decision-makers' needs, and supports the development of a spatial decision support system to effectively assist decision-making with heterogeneous geospatial big data.

Keywords: Decision-making, Heterogeneous Geospatial Data, Spatial Decision Support Systems, Volunteered Geographic Information, Disaster Management.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ADR ..... Action Design Research
AGORA .. A Geospatial Open collaboRative Architecture for Building Resilience against Disasters and Extreme Events
AGORA-DS Decision Support Component of AGORA
AGORA-GeoDash GeoDashboard System of AGORA
AGORA-VOS Volunteered Observation Service of AGORA
ANA ..... National Water Agency
AR .... Action Research
BPM ..... Business Process Management
BPMN ... Business Process Management Notation
Cemaden . Center for Monitoring and Early Warning of Natural Disasters
CENAD .. National Center for Disaster Risk Management
CPRM ... Brazilian Geological Survey
CPTEC ... Center for Weather Forecasting and Climate Research
DSR ..... Design Science Research
DSS ...... Decision Support Systems
EM-DAT . The International Database maintained by Center for Research on the Epidemiology of Disasters (CRED)
GIS ...... Geographic Information Systems
IBGE ..... Brazilian Institute of Geography and Statistics
INMET ... National Institute of Meteorology
O&M .... Observation & Measurements
oDMN ... observation-aware Decision Model and Notation
OGC ..... Open Geospatial Consortium
OMG ..... Object Management Group
OSM ..... Open Street Map
SDI ...... Spatial Data Infrastructure
SDSS .... Spatial Decision Support Systems
SLR ...... Systematic Literature Review
SMS ..... Systematic Mapping Study
SOA ..... Service Oriented Architecture
SOS ...... Sensor Observation Service

UAV ..... Unmanned Aerial Vehicle
UNISDR . United Nations International Strategy for Disaster Reduction
URL ..... Uniform Resource Locator
VGI ...... Voluntereed Geographic Information
VTC ..... Volunteer and Technical Communities
WFS ..... Web Feature Service
WMS .... Web Map Service
WSN ..... Wireless Sensor Network

CONTENTS

CHAPTER

1

INTRODUCTION

1.1 Contextualization

Decisions are an intrinsic part of the everyday life of communities, from the time when people wake up at home (e.g., what will my tasks be during the day?) until the time when they go to sleep (e.g., what time do I have to wake up tomorrow?). We are constantly making important decisions which, in some cases, could determine the path we take or the way we conduct our lives. The same applies to the daily running of organizations; their managers and directors are continuously analyzing business variables to choose the best alternative for increasing profits and/or reducing production costs. Decision-making can thus be regarded as one of the oldest research areas in the literature, since it is based on methodologies, tools, and theories that are useful in different scenarios from traffic management to psychology. However, at the same time, all these different applications have turned the area into a complex field, largely because it has to deal with intangible factors (e.g., the psychological perspectives of different individuals) rather than tangible factors (e.g., the particular features of a context and use of technologies) (SIMON, 1977; SAATY, 2008).

Despite this, the available data still play a crucial role in the area, since they influence the outcomes of decisions, for example, if a traffic engineer requires data about the condition of the roads when deciding on what is the appropriate traffic flow. Both tangible and intangible factors affect the success or failure of a decision but a decision-maker still requires suitable data when making an appropriate decision. Otherwise, he/she might simply have to depend on his/er own experience and this might result in a wrong decision and raise questions about his reliability and efficiency (TVERSKY; KAHNEMAN, 1974). In particular, geospatial data have a great potential in this context, because they are based on tangible factors that are always linked to a geographical setting or particular region (BILLA *et al.*, 2006) - for instance, the top-selling products (information) of a region (a particular place) or the volume of customers (information) buying at an individual supermarket (geographical setting).

Since available data is essential for updated and accurate decision-making, there has been a concern among both practitioners and researchers with either developing new technologies and mechanisms for enhancing data collection or improving existing data collection technologies and mechanisms. By using these emerging data sources, "managers can measure, and hence know, radically more about their businesses, and directly translate that knowledge into improved decision-making and performance" (MCAFEE; BRYNJOLFSSON, 2012). For example, supermarket managers are now able to understand their customers' behavioral patterns and thus predict what products are most likely to be bought or booksellers can recommend products based on the customer's clickstream on the sites.

All these data have led to the so-called "big data" that is characterized by four factors (HASHEM *et al.*, 2015): (1) a huge volume of data; (2) traditional and regular databases that are not able to handle the available data; (3) the rapid growth in the volume of data; and (4) the concern in regarding the veracity of the available data. This has been followed by the development of innovative technologies for overcoming the problem of the demand for data storage (e.g., SciDB and NoSQL databases). At the same time, the area of data science has been growing through the development of algorithms and new technologies for data analysis and mining (e.g., Hadoop and distributed analysis), and thus more research has been carried out to cover decision-making with big data (GOPALKRISHNAN *et al.*, 2012; MILLER; MORK, 2013; HASHEM *et al.*, 2015; VIEWEG; CASTILLO; IMRAN, 2014; LINK *et al.*, 2015). For example, Miller & Mork (2013) adopted the concept of a data value chain for a methodology that not only assists data processing but also describes useful technologies that can support all the activities involved in the methodology. Mandviwalla & Watson (2014) described an organization as a mix of "capitals" (human, economic, social, symbolic, and organizational); for example, economic capital includes financial, physical, and manufacturing capital resources. On the basis of this definition, they employed a social media strategy for generating these capitals. Another group of works explores the area of data-driven decision-making (MARSH; PANE; HAMILTON, 2006; POYNTON; CAREY, 2006; NOVAK; PAULOS; CLAIR, 2011; PROVOST; FAWCETT, 2013; TSAI; SUNG; KANG, 2016). Within this group, Duggan (2014) employed natural language processing for querying datasets with the aim of enabling people to make informed decisions. In contrast, Gill, Borden & Hallgren (2014) developed a methodology that defines a set of measures that can be employed to support data-driven decision-making in students' education.

Overall, big data has a great potential for transforming processes within several areas, e.g., healthcare, biology, disaster management, engineering, finance, urban science, business, and eventually, society itself. In this project, disaster management has been taken as a representative scenario of decision-making with big data. This is particularly important and necessary because several natural disasters have affected the communal life of a number of countries (Haiti in 2010, Australia in 2010-11, Japan in 2011, the Philippines in 2013, and Nepal in 2015) and caused serious damage to all of them (Figure 1). These events can be defined as a disruption of the "local capacity" of a community that has several adverse effects (e.g., loss of life, the spread of disease, financial problems, environmental degradation, etc). It also leads to difficult circumstances for their members, and the need for assistance at both a national and international level (MARCELINO, 2007; JHA, 2010). In terms of losses, in 2015, natural disasters incurred financial losses of close to US$100 billion worldwide and caused 23,000 fatalities. 42% of these reported disasters, 24% of the fatalities, and 19% of the financial losses were linked to extreme hydrological events (MUNICHRE, 2015).

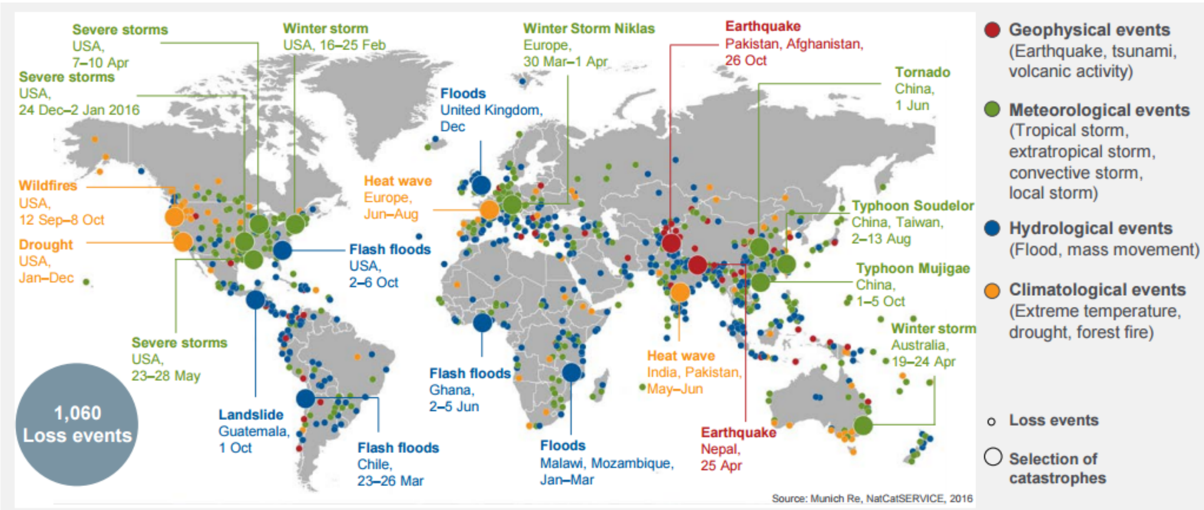


Figure 1 – Map of Disaster 2015.

Munich RE, Geo Risks Research.

In Brazil, natural disasters have also affected different regions of the country and caused severe damage to the infrastructure and the loss of lives. In the period 1991-2012, the number of disasters grew from 773 to almost 4,000 per year (Figure 2) and this number is expected to increase owing to the frequency of climate changes and extreme weather conditions.

In an attempt to reduce the impact of these disasters, governments, official agencies,

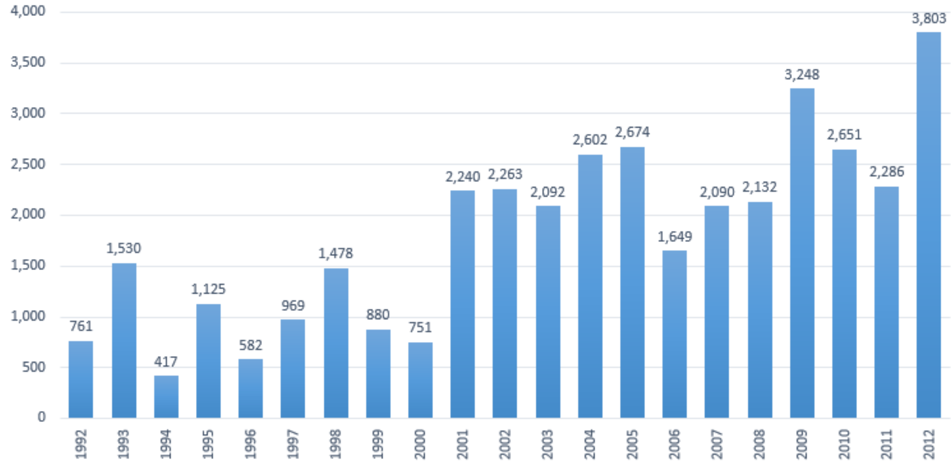


Figure 2 – Number of disaster in Brazil from 1991 to 2012.

Brazilian Atlas of Natural Disasters (1991-2012).

and local communities must carry out relief activities with a view to building some form of resistance against them, i.e., to enable communities to resist, change or adapt to the conditions if there is a disaster (NORRIS *et al.*, 2008). Traditional data sources have been providing useful information about environmental variables; for example, hydrological stations to measure the water level in riverbeds, weather radars to detect the weather conditions, and rainfall gauges to calculate the volume of rainfall (MANSOURIAN *et al.*, 2006; LEE *et al.*, 2008; MARKOVIC; STANIMIROVIC; STOIMENOV, 2009; HUGHES *et al.*, 2011; ASSIS *et al.*, 2016). These data sources have been supplemented by ordinary volunteers who have been providing valuable information about the current situation in affected areas by means of social media platforms (e.g., Twitter), SMS, participatory platforms (e.g., citizen observatories and mobile apps), and collaborative mapping (e.g., OpenStreetMap) (ERSKINE; GREGG, 2012; HORITA *et al.*, 2013; DEGROSSI *et al.*, 2014; ALBUQUERQUE *et al.*, 2015). This kind of information is defined as Volunteered Geographic Information (VGI), which is a collection of digital spatial data produced by individuals and informal institutions, i.e. by ordinary citizens using appropriate tools to gather and disseminate their views and geographical knowledge on the web (GOODCHILD, 2007). For example, Zook *et al.* (2010) showed that the volunteers of the OSM community around the world generated good-quality maps a few hours after the earthquake which struck Haiti in 2010. These maps were essential for supporting the activities of the relief teams. Since the shared volunteered information might be limited, unstructured, and difficult to understand, there is also a group of volunteers who are known as Volunteer and Technical Communities (VTC) or Digital Humanitarians (MEIER, 2014; ALBUQUERQUE; HERFORT; ECKLE, 2016). These are volunteers and professionals spread around the world, whose task is to make sense of the vast amount of social media data, SMS, and imagery captured from satellites and unmanned aerial vehicles to support relief efforts worldwide (MEIER, 2014). For instance, Link, Hellingrath & Ling (2016) used these volunteers to supervise social media messages filtered by a machine learning system. Through this moderator, the system was able to make improvements and provide more valuable and accurate information.

However, although all these data open up new frontiers, its huge volume and the fact that the strategies for data integration and visualization are inappropriate, raise a number of problems that must be overcome to improve decision-making with heterogeneous geospatial big data. The next section provides a problem statement, as well as setting out the motivating factors behind this project.

1.2 Motivation and Problem Statement

As mentioned earlier, decisions are an intrinsic part of in the day-to-day life of different communities and people. These could be related to our personal lives (e.g., going to work by car or bus), which are important to the way our individual lives are conducted and linked to our membership of a community (e.g., walking to work is better for the environment) that is important to its overall well-being. In this manner, the information at hand provides valuable support for decision-making, e.g., one might decide not go to work on foot if it is raining outside. Accurate decision-making thus requires updated and appropriate available data, mainly because they are able to affect the outcomes of decisions. In this context, the motivational factors behind this project are twofold.

First, this project is driven by important challenges that still remain in the literature with regard to the decision-making with heterogeneous geospatial big data, as well as the development of a Spatial Decision Support System (SDSS) that integrates heterogeneous data sources (e.g., VGI and in-situ sensors). The question of integrating heterogeneous data has been widely investigated in the literature. Some studies have employed a spatial data infrastructure as a means of achieving this. Mansourian *et al.* (2006) devised an spatial data infrastructure conceptual model to allow an integrated infrastructure to be established for a different disaster management system. In a similar way, Molina & Bayarri (2011) designed an SDI architecture and developed a web-based application that employs a cognitive approach to allow information sharing. Although the implementation of interoperable standards can be regarded as an important requirement, these works did not use them either for interoperability between the systems or for the integration of data sources. This is addressed by another group of studies, which seek to analyze the use of the open standards of the Open Geospatial Consortium (OGC) to achieve interoperability among different systems. Zhang & Li (2005) stress the importance of OGC open standards - i.e. Web Feature Service (WFS) and Web Map Service (WMS) - for sharing near real-time spatial data over the web. Moreover, Markovic, Stanimirovic & Stoimenov (2009) use OGC open standards, i.e., Sensor Observation Service (SOS), WFS and WMS, for encapsulating sensor and spatial data in the River water Monitoring and Alert System (RWMAS) with a view to detecting and preventing water pollution. As for the question of integrating VGI with other data sources, Wan *et al.* (2014) and Schnebele, Cervone & Waters (2014) describe distinct approaches for integrating authoritative and non-authoritative data with the aim of providing location-based eventful visualization, a statistical analysis and graphic capabilities for the authorities and the public.

Although the integrated information can be of great value in supporting decision-making, it often fails to reach the decision-makers in a suitable way (e.g., in appropriate visualization formats and at the required time) (BARTON; COURT, 2012; VIEWEG; CASTILLO; IMRAN, 2014; DOLIF *et al.*, 2013). As a result, decision-makers are supplied with useless information that still requires extensive knowledge or experience for further data processing. This also makes it difficult to predict the impact that the change of data availability may have on specific tasks, which makes it hard to determine the most appropriate ways of generating new data, since it is virtually impossible to find out if and where there is a lack of information. In view of this, it is essential to fully understand which data source can meet the information needs of the decision-makers so that key items of information can effectively contribute to decision-making and operational excellence.

In the light of this, in recent years, growing attention has been paid to works involved in analyzing the use of data to support decision-making, largely due to the increasing interest in data science, and thus more research has been carried out in this area (MARSH; PANE; HAMILTON, 2006; PROVOST; FAWCETT, 2013; MILLER; MORK, 2013; DUGGAN, 2014). Another group of studies has attempted to analyze the use of different data sources (e.g., the external data server and sensors) to support decision-making in different scenarios (GOPALKRISHNAN *et al.*, 2012; CHEN; CHIANG; STOREY, 2012; HORITA *et al.*, 2014a; WAMBA *et al.*, 2015). Within this group, Taylor *et al.* (2013) devised a software architecture for processing and visualizing data gathered from sensors, while Vera-Baquero, Colomo-Palacios & Molloy (2013) outlined an architecture for integrating data from different organizations, which included a data server for improving the analysis of business performance management.

Social media messages have also been employed to support organizational tasks like marketing trends (MALSBENDER *et al.*, 2013; KURNIAWATI; SHANKS; BEKMAMEDOVA, 2013), disaster management (IMRAN *et al.*, 2014; VIEWEG; CASTILLO; IMRAN, 2014; LINK *et al.*, 2015) or measuring customer satisfaction (ROSEMANN *et al.*, 2012). For instance, Kleindienst, Pfleger & Schoch (2015) integrated social media analytics with the business goals of an organization, by breaking down these goals into critical success factors that make it possible to find out the information requirements, so that they can be combined with the appropriate social media analytics. In summary, the existing literature still lacks research studies that explore the architectural requirements for collecting and integrating heterogeneous data sources and then carrying out the visualization of the integrated information. Furthermore, although these works addressed important issues, they failed to provide a method that is able to provide a better understanding of the relationship between decision-making and its appropriate data, i.e., what is the information needed for decision-making.

Second, since this project has selected disaster management as a representative scenario for decision-making with big data, it is also driven by practical and domain considerations. This is because disasters are expected to increase in frequency in the coming years due to climate change, and thus mitigation and response measures are urgently required to reduce the severity of their impact (NORRIS *et al.*, 2008; MENDIONDO, 2010; KELMAN; GAILLARD; MERCER, 2015). While, the policymakers responsible for disaster risk reduction have been recommending the use of location-based information and the development of social technologies and hazard-monitoring telecommunications systems, these in turn are expected to increase the volume and heterogeneity of the available data. For instance, the Sendai Framework for Disaster Risk Reduction 2015-2030 defines 13 guiding principles and sets out four priorities for disaster risk reduction (UN, 2015; AITSI-SELMI *et al.*, 2015). On the basis of these principles, it is clear that disaster risk reduction requires "a multi-hazard approach and inclusive risk-informed decision-making based on the open exchange and dissemination of disaggregated data, as well as on easily accessible, up-to-date, comprehensible, science-based, non-sensitive risk information, complemented by traditional knowledge". The framework also makes recommendations for addressing these principles - for instance, the provision of location-based disaster risk information to decision makers in an appropriate format by using geospatial information technology. Other important recommendations, which are of importance in this project, include the development of early warning systems, disaster risk and emergency communications mechanisms, social technologies and hazard-monitoring telecommunications systems that should be tailored to the needs of users. In addition, reliable data is required that can make use of space and in situ information, including GIS, as well as information and communications technology to enhance the collection, analysis, and dissemination of data.

In Brazil, the government has released its national framework with the measures that have been adopted for coping with disasters in the country. The National Strategy for Natural Disaster Risk and Response Management that was issued in 2012 aims at reducing the damage caused by disasters, and ensuring the communities safety, and thus saving lives[1]. The strategy was centered on four main areas: preventive measures, mapping and understanding risks, monitoring and early warning, and response, as shown in Figure 3.

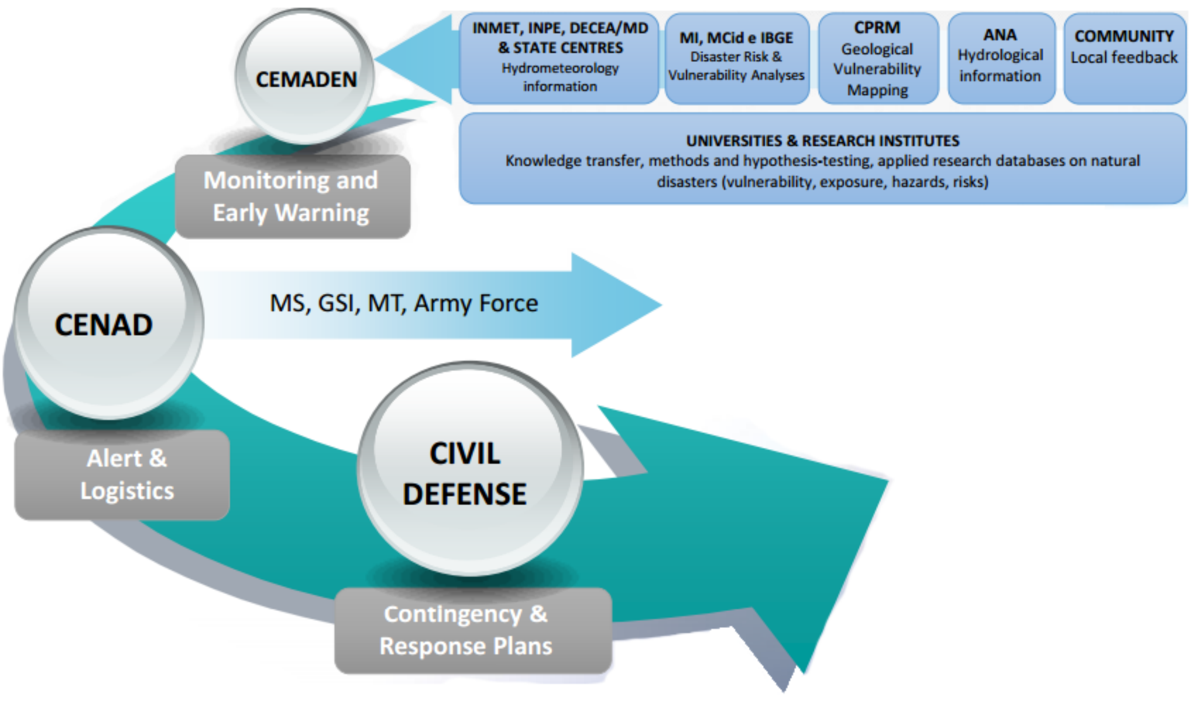


Figure 3 – Brazilian National Strategy for Natural Disaster Risk and Response Management.

Adapted from <https://goo.gl/dTAmbg>.

This project is particularly interested in the "Monitoring and Early Warning" area, which covers the tasks undertaken by the Brazilian National Center for Early Warning and Monitoring of Natural Disasters (Cemaden)[2]. Founded in 2011, the aim of this center is to forecast the occurrence of a natural disaster on the basis of the data provided by the monitoring systems of

[1] <https://goo.gl/Z4tofA>
[2] <http://www.cemaden.gov.br/>

different institutions (e.g., hydrological data from Natioanl Water Agency). It is also responsible for communicating the risk of a disaster to the National Center for Disaster Risk Management (CENAD), which then notifies the response agencies (e.g., Army and Civil Defense). Given high volume of data, the center has been facing structural problems with regard to understanding the decision-making of the monitoring teams, as well as how the available data can meet the target users' requirements. It also needs to adopt an approach that involves integrating all the data provided by the available monitoring systems (e.g., rainfall gauges and hydrological stations).

Finally, all these important motivating factors that have arisen from both the literature and practice have led to the following problem statements, which this project seeks to address.

- **The visualization and integration of heterogeneous geospatial big data sources for supporting decision-making:** most of the essential information to support decision-making is obtained from heterogeneous data sources, and is usually shared without complying with interoperability standards or a relationship being formed with them. As a result, there is a significant amount of useful information that is extremely difficult for responsible decision-makers to integrate and visualize.
- **The understanding of decision-makers' information needs and its connection with data sources:** although existing studies on data-driven decision-making or the use of big data for decision-making have addressed important issues, they have failed to provide a method for establishing a link with the information needed for the decision-making and for providing a representative model that describes this link. Furthermore, they also do not provide a method that can be employed either for improving the decision-making or establishing decision support systems. This is particularly important for providing information that can meet the decision-makers' requirements.

In the light of all these challenges, the next section outlines the objectives of this project, as well as its research questions.

1.3 Objectives

The general goal of this thesis is to develop an approach for improving decision-making with heterogeneous geospatial big data through the use of spatial decision support systems and volunteered geographic information in disaster management. In this context, this thesis seeks to answer the following research question:

RQ) How can the integration of heterogeneous data sources contribute to the improvement of decision-making?

The term “integration” here is understood as the combination of data provided by different sources, as well as the visualization of the integrated data in a simple and unified view. However, the focus of this thesis does not lie either on database theories (e.g., data model, database design, etc) or on data mining/data science techniques but rather on the development (or improvement) of software architectures for dealing with the idiosyncrasies in gathering, combining, sharing and visualizing data. This comprehends the understanding not only of the system requirements (e.g, the essential components and technologies) and the user requirements (e.g., required information, decisions, and activities) but also of how decision support systems should be designed for effectively supporting decision-making (e.g., by presenting the required information at the right time).

In addressing the research question, the main objectives of this project are described as follows:

1. **Review of the existing literature on the use of VGI and SDSS for disaster management.** we first aim at investigating the existing literature with regard to the use of VGI in disaster management; for instance, to understand in which phases of disaster management, volunteered information has been used. Furthermore, we also seek to understand the reasons why volunteered information has been used, as well as the existing types of information that are available. By analyzing the obtained results, it will be possible to find out if there are serious gaps in the literature.

2. **Definition of a conceptual architecture for integrating heterogeneous data sources and supporting decision-making.** we put forward and evaluate a conceptual architecture that integrates data provided by heterogeneous data sources (e.g., in-situ sensors and volunteered information). It is hoped that this might provide more accurate information for supporting decision-making. In evaluating this approach, we intend to employ the architecture in a real-world scenario of flood risk management in São Carlos.

3. **Definition of a model-based framework for describing the connection between decision-making and data sources.** On the basis of the results obtained from the literature review and conceptual architecture, we plan to establish a model-based framework that describes the connection between the decision-makers’ tasks and decisions based on emerging data sources. As a result, this model is expected to define conceptual elements that are essential for establishing this kind of connection, e.g., tasks, decisions, and information requirements. The usefulness and effectiveness of the model is evaluated in a case study for modeling the monitoring environmental variables and early-warning tasks of a Brazilian early warning center.

4. **Development of design principles for SDSS of dynamic environments with heterogeneous geospatial data sources.** With the aid of the model-based framework and conceptual architecture, our aim is to set out and evaluate a set of design principles that must be

observed to enable an SDSS to effectively support decision-making with heterogeneous geospatial big data. Moreover, these principles are essential for meeting the requirements of decision-makers in these kinds of environments. The usefulness and effectiveness of these principles is evaluated in a cross-organizational collaborative research project with a Brazilian early warning center.

The next section describes the overall research methodology, as well as the main contributions that were made throughout this project.

1.4 Research Methodology and Summary of Contributions

In the light of the research question discussed in the previous section, this thesis can be regarded as a combination of exploratory and prescriptive research strategies. The reasons are twofold. First, it aims at understanding both the use of VGI and SDSS for supporting decision-making in disaster management (the representative scenario of decision-making with heterogeneous geospatial big data used for this thesis), which is the objective of exploratory research, i.e., "to gain familiarity with a phenomenon or to achieve new insights into it". Furthermore, it is also concerned with either describing a method for the construction of artifacts or introducing the artifacts themselves. This is the aim of prescriptive research, i.e., research that "provides a description of the method or structure or both for the construction of an artifact" (KOTHARI; GARG, 2004; GREGOR, 2006). It is also worthwhile to mention that the use of the artifacts in practice employs exploratory strategies for supporting its analysis.

This thesis then adopted a mixture of qualitative and quantitative approaches to attain each of the individual objectives. A qualitative approach is concerned with the subjective assessment of data collected by employing empirical methods (e.g., participatory observations and focus group sessions) to understand a particular phenomenon. This assessment is often analyzed through non-statistical methods. In contrast, a quantitative approach involves the collection of data in a quantitative form that can be analyzed by means of rigorous statistical methods (KOTHARI; GARG, 2004; ROBSON, 2011). These approaches also play an essential role in guiding this project by defining the methods and techniques required for the software prototyping, as well as for data collection and analysis.

Figure 4 displays the methods that were employed for achieving each of the objectives and determining the contributions they made. The research methodology was separated into four essential research cycles in which the intermediary tasks involved a rigorous process of design, analysis, and evaluation of outcomes. The purpose of this was not only to consolidate the obtained results but also to ensure they could be generalized.

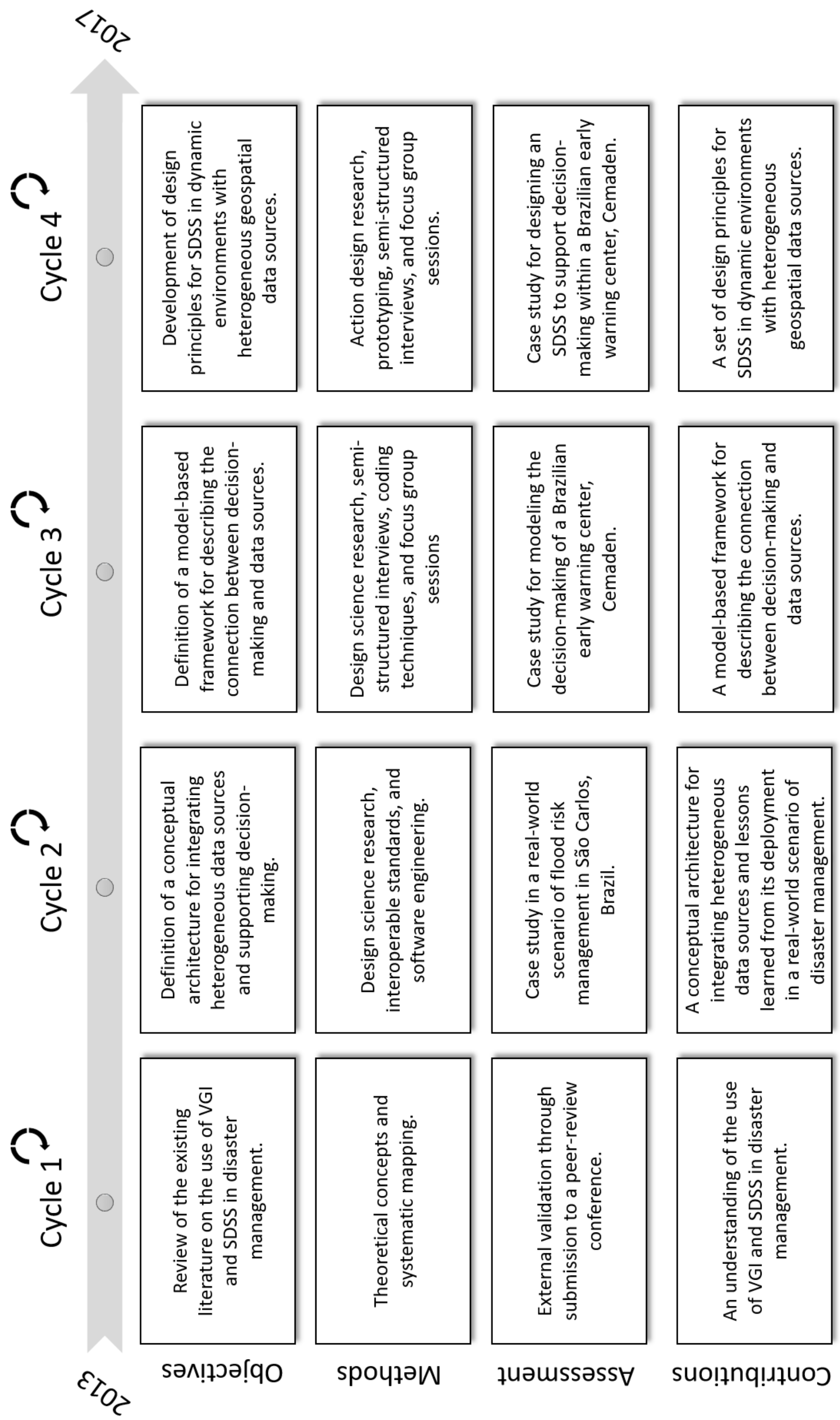


Figure 4 – Research methodology.

The objective of Cycle 1 was to investigate the use of VGI and SDSS for disaster management. This was addressed by carrying out two different Systematic Mapping Studies (SMS). An SMS adopts a systematic and formal approach to identifying, analyzing, and evaluating the existing studies in the literature to give an overview of each topic area or theme (PETERSEN *et al.*, 2008). This raised questions about three essential items of information: a) in which phases of disaster management the volunteered information has been used; b) in what types of disasters it has been used; and c) what types of research methods have been employed in the existing works. The study also helped to understand the potential benefits of adopting VGI in disaster management. In evaluating the obtained results, it followed the recommendations of (BRERETON *et al.*, 2007). Hence, the systematic mapping was externally evaluated through their submission to a peer-review conference, which was later published and presented to the *19th Americas Conference on Information Systems (AMCIS)* (HORITA *et al.*, 2013).

The purpose of the second systematic mapping was to understand how SDSS has been used to support decision-making in disaster management. The study set out by seeking to find out the reasons why volunteered information has been used, and then went on to define the different types of volunteered information. The study provided an overview of how VGI and SDSS can be combined to support decision-making in disaster management. The results obtained in this second systematic mapping were also evaluated externally by being submitted to peer-review conferences (BRERETON *et al.*, 2007). Overall, the main purpose of Cycle 1 was to provide a clear understanding of the opportunities for future research studies with regard to the use of VGI and SDSS in disaster management, i.e., the combined use of VGI and sensor data, as well as the visualization of integrated information (two of the most important gaps in research were found in these SMSs).

In addressing the research opportunities outlined in the previous cycle, Cycle 2 attempted to define a conceptual architecture that can be used for integrating heterogeneous data sources. This architecture also defines a component that assists in the visualization of integrated information. This involved employing a Design Science Research (DSR) method for first designing the artifact (i.e., the conceptual architecture), and then evaluating it. This method is employed to achieve practical solutions through the development of new artifacts, in some cases with a high degree of abstraction (GREGOR; HEVNER, 2013). Furthermore, interoperability standards (e.g., SOS) were adopted to achieve a flexible communication between the layers of the architecture. The evaluation involved designing an SDSS prototype that followed the conceptual architecture and its deployment in a real-world case study of flood risk management in São Carlos, Brazil. This covered the integration of data provided by a WSN and a citizen observatory, as well as the visualization of the integrated information on a digital dashboard. The conceptual architecture developed for this study can be regarded as a valuable contribution to this project, as well as the lessons learned from its deployment in the case study. These results were published in the *Computers & Geosciences* (HORITA *et al.*, 2015).

Further results of the Cycles 1 and 2 also provided evidence that a better understanding of how the available data sources could be used for meeting the information requirements, was important for improving decision-making. Cycle 3 was concerned with defining a model-based framework for describing the connection of decision-making and data sources, which was called observation-aware Decision Model and Notation Framework (oDMN$^+$ Framework). This framework consists of two essential elements. First, a multi-layered model and notation (oDMN$^+$) that describes the conceptual elements for connecting decision-makers' tasks and decisions with data sources. Standard models and notations (e.g., Business Process Model and Notation, Decision Model and Notation, and Observation & Measurements) were used as a basis for the development of oDMN$^+$. Secondly, there was a modeling process, which aims at guiding the way oDMN$^+$ is employed in practice. A DSR was also conducted to support the development of the artifacts (i.e., oDMN$^+$ and the modeling process).

As well as being evaluated, the framework was employed for modeling the tasks of environmental modeling and early-warning at the Brazilian Center for Monitoring and Early Warning of Natural Disasters (Cemaden). In addition, a series of semi-structured interviews and focus group sessions were held with members of the center (SEAMAN, 1999; SOBREPEREZ, 2008). The coding technique was also employed for assisting in the data analysis (SALDAÑA, 2015). The oDMN$^+$ Framework that connects decision-making to data sources is indeed the main contribution to research made this study. The study results were then collated in a paper and submitted to the *Decision Support Systems* (HORITA *et al.*, 2017).

Cycle 4 was based on the outcomes of the previous cycles and aimed at developing and evaluating a set of design principles for guiding the development of an SDSS, and enhancing decision-making with heterogeneous geospatial big data. This involved employing an Action Design Research (ADR) method that "aims to contribute both to practical concerns of people in an immediate problematic situation and to the goals of social science" (SEIN *et al.*, 2011). Cemaden was again adopted as the case study and to provide the organizational settings for the development of the ADR project. Two interactive design and analysis cycles were followed with members of the monitoring control room of the center. First, a set of semi-structured interviews were held for the data collection on the use of an SDSS inside the room. A number of problems emerged from the analysis of collected data and this led to the definition of a research question and a working hypothesis. In the second cycle of the ADR project, these definitions (together with the oDMN$^+$) underpinned the design principles that were used for designing an SDSS prototype that embodies the decision-makers' requirements. The SDSS prototype was then evaluated with members of the monitoring control room through a series of focus group sessions and individual semi-structured interviews. The analysis of the collected data helped to refine and consolidate the design principles and this was a valuable feature of this project since it can indeed support the design of an SDSS that improves decision-making with heterogeneous geospatial data sources in disaster management.

In summary, the overall approach combines all latter three contributions that were made in this project, and has assisted in answering our general research question. First, there is a conceptual architecture that allows the integration of heterogeneous data sources, as well as the lessons learned from its deployment in a real-world scenario of disaster management. Secondly, this is supplemented by a model-based framework that describes the connection of decision-making with appropriate data sources. Finally, there is a set of design principles that must be observed to ensure that an SDSS can effectively support decision-making with heterogeneous geospatial big data. Furthermore, it should be noted that the approach can work in Brazil on the basis of the definition provided by the Sendai Framework for Disaster Risk Reduction 2015–2030, which recommends the development of early warning systems that are tailored to the needs of users and make use of reliable data from local knowledge and monitoring systems.

Since this project is part of a larger research project, the next section outlines the background in which the overall research project is undertaken. It also includes other projects that have been affected by the outcomes of this project.

1.5 Research Context

This project is part of a research project that is named AGORA - **A** **G**eospatial **O**pen collabo**R**ative **A**rchitecture for Building Resilience against Disasters and Extreme Events (ALBUQUERQUE; ZIPF, 2012), which has being developed by an interdisciplinary research group formed by postdocs, research assistants, graduate research assistants, undergraduate assistant and Visiting Scientists and Professors. AGORA is coordinated by Prof. Dr. João Porto de Albuquerque from the Institute of Mathematics and Computer Science (ICMC) of the University of São Paulo (USP) in Brazil and Prof. Dr. Alexander Zipf from the GIScience Research Group of the University of Heidelberg in Germany.

AGORA proposes a collaborative architecture inspired by the Greek *Agorá* (literally, “gathering place”), place in ancient Greek city-states that was the center of artistic, athletic, spiritual, and political life (ALBUQUERQUE; ZIPF, 2012). This place is then considered the birthplace of democracy. In our context, AGORA is used as inspiration to reference a transdisciplinary architecture responsible for gathering organizations and individuals by allowing them to bring their information supply and demands to a common platform using web 2.0 technologies, i.e., AGORA combines a set of heterogeneous data that are provided by sensors, volunteers, and official agencies, and develops mechanisms (e.g., forecast and alert systems, real-time flood monitoring, business processes, and reference models) that could be used to support the decisions of general population and government (including local decision-makers, disaster management, and emergency agencies). Therefore, AGORA will not only provide decision-support interfaces for decision makers but also make the value-added data openly available in a standardized fashion. In this way, AGORA yields a flexible ecosystem that is able to cope with the ever-changing needs of participatory environmental monitoring and resilience building against disasters and extreme events, as displayed in Figure 5.

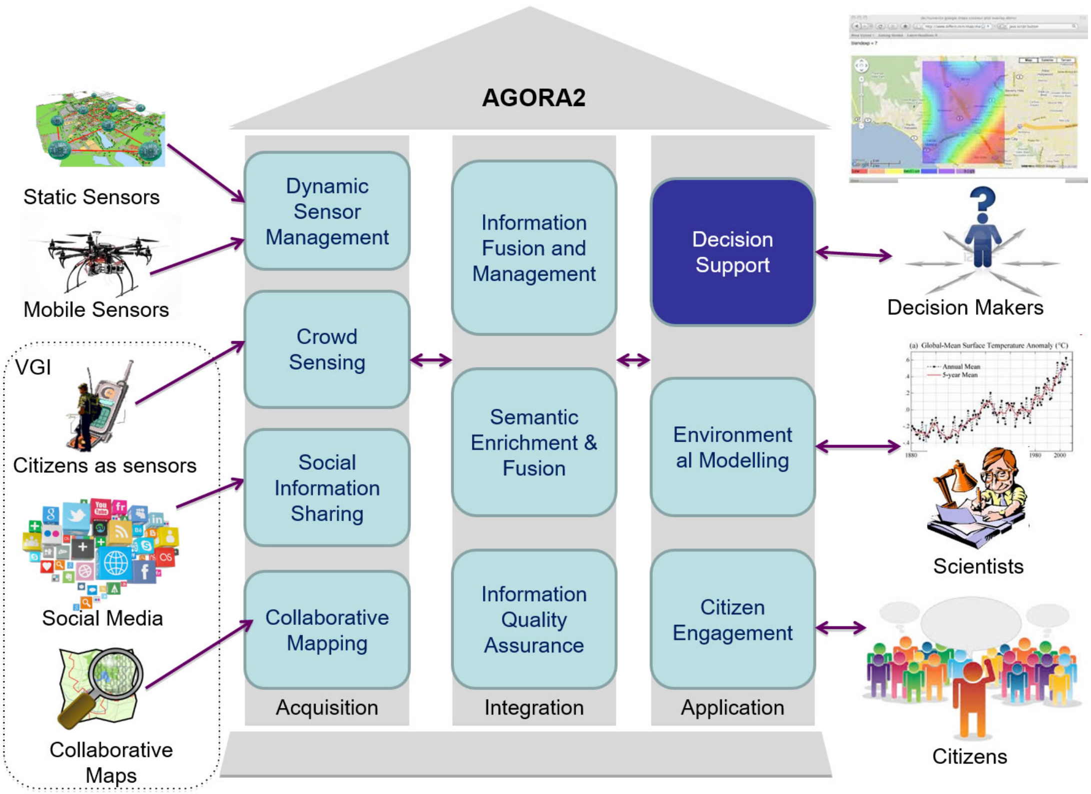


Figure 5 – Global context of this project. Adapted from Albuquerque *et al.* (2017)

For achieving its objectives, AGORA consists of three essential pillars: (1) Acquisition, (2) Integration, and (3) Application. The Acquisition pillar defines a set of components that is responsible for collecting data provided from different data sources (e.g., Unmanned Aerial Vehicle (UAV), collaborative mapping platforms, and social media platforms) and sharing them using standard protocol (e.g., SOS or WFS). All these shared data are integrated into the Integration pillar. This pillar is also responsible for evaluating the quality of data, as well as providing developments for semantic enrichment and information fusion. Furthermore, the Application pillar uses the data provided by Integration pillar with the aim of supporting decision-making of different institutions (e.g., civil defense and emergency agencies), assisting other researchers with more accurate and high-quality data for their works, and finally achieving a better citizen engagement by creating risk maps, and early warning systems, and data collection platforms. Thus, AGORA through this set of pillars and their components is the virtual place which different experts/researchers and other stakeholders could provide relevant issues bringing their multiple perspectives to define problems that must be solved, e.g., to building resilience against flooding, particularly focusing on subsidizing preparation and risk reduction efforts.

The outcomes of this project contribute to the development of the "Decision Support" component from the Application pillar of AGORA. This component aims at designing and developing approaches that can be employed by emergency organizations for supporting (or improving) their decision-making. These approaches might be innovative information visualization techniques, new models for representing real-world elements and their relationships, or modern architectures of information systems. For this, they should take into consideration the integrated and high-quality data provided by heterogeneous data sources (e.g., sensors and volunteered information) and particular requirements of decision-makers. The results obtained here also bring relevant contributions for the following research projects that were associated with this project:

- Prof. Dr. João Porto de Albuquerque – ICMC/USP
  [FAPESP 2012/18675-1] Geospatial Open collaboRative Architecture for Building Resilience against Disasters and Extreme Events (AGORA).
  [MICROSOFT] AGORA-CLOUD: Applying the cloud computing paradigm with the Azure platform to support A Geospatial, Open collaboRative Architecture for building resilience against disasters and extreme events.
  [CAPES 12065-13-7] A Geospatial Open collaboRative Architecture for Increasing Resilience against Floods.
  [CAPES 88887.091744/2014-01] CAPES Pró-Alertas CEPED-USP.

- Prof. Dr. Eduardo Mario Mendiondo – EESC/USP
  [FAPESP 2008/58161-1] Assessment of Impacts and Vulnerability to Climate Change in Brazil and Strategies for Adaptation Options (FAPESP-IVA).

- Prof. Dr. Jó Ueyama – ICMC/USP
  [CIA^2] Building Smart Cities: From Instrumentation of Environments to Development of Application, RIO (Monitoring Urban Rivers).

- Prof. Dr. Carlos Eduardo Pereira – PPGEE/UFRGS
  [Universal CNPq 477499/2012-0] Sensor Network Using Wireless and Unmanned Aerial Vehicles in Support Systems Search and Rescue Operations in Disaster Situations.

The next section presents the organization of this thesis and the structure of its chapters.

1.6 Thesis Outline

The remainder of this thesis is structured as follows. Chapter 2 provides an overview of the essential underlying concepts of this thesis. Disaster management is first introduced through a description of its main phases and related tasks. Following this, the foundations of SDSS are set out, i.e., the concept of decision support systems, the architecture and characteristics of an

SDSS, and the digital dashboards. The chapter also characterizes big data and VGI, as well as introducing the business models that are employed in this project and describing their core notations.

Chapter 3 provides a review of the existing literature on the use of VGI and SDSS for disaster management. This consists of two distinct systematic mappings. First, an investigation was conducted with the aim of understanding the use of VGI and crowdsourcing in disaster management. After this, another investigation was carried out to collect evidence about the use of SDSS for disaster management. These systematic mappings were essential for detecting any gaps in the existing literature.

Chapter 4 introduces a conceptual architecture for integrating heterogeneous data sources and visualizing them in an SDSS. It employs interoperable standards to ensure the integration of information, and thus makes the architecture flexible enough to support the inclusion of different features. Furthermore, the integrated information is displayed in a digital dashboard that can support decision-making. The lessons learned from the employment of the architecture in a real-world scenario of flood risk management in São Carlos are also described.

Chapter 5 introduces the observation-aware Decision Model and Notation Framework (oDMN^{+} Framework), which comprises two essential elements: (1) a multi-layered model and notation for describing the connection between decision-makers' tasks and decisions with data sources; and (2) a modeling process that defines a set of activities for obtaining conceptual elements from decision-makers and employing the oDMN^{+} in practice. There is also a detailed examination of the results obtained from the employment of the framework for modeling the tasks of environmental modeling and early-warning of a Brazilian early warning center.

Chapter 6 describes a set of design principles for guiding the development of SDSS for decision-making with heterogeneous geospatial big data. Two interactive design and analytical cycles were carried out in the context of monitoring and early-warning within a Brazilian emergency agency. The results obtained from the evaluation in practice are also described.

Chapter 7 summarizes the final considerations of this project, as well as outlining the main contributions that were achieved throughout the project, and making some recommendations for future work.

CHAPTER

2

BACKGROUND

2.1 Overview

This chapter presents the main background concepts that are essential for the development of this thesis. Section 2.2 presents the definition, phases, and activities of the disaster management. Section 2.3 introduces the main concepts around the spatial decision support systems. This section also presents the digital dashboards. Section 2.4 describes the definitions of big data, as well as its characterization and association with heterogeneous geospatial data. Section 2.6 introduces the business models that are employed in this project and also describes their core notations.

2.2 Disaster Management

During the last years, it has been seeing a growing in the occurrence of natural disasters around the world, those that affected Haiti, Chile, and Pakistan in 2010, Queensland in 2010-11, Japan and Christchurch in 2011, the USA in 2012, and the Philippines in 2013 are some of them. The damages caused to the Philippines are estimated in close to 7.000 deaths and $USS2.86 billion, it was the worst and deadliest storm that affected the country (data from EM-DAT[1]).

A natural disaster occurs when a community is struck by a hazard and financial, social, and structure losses and impacts are so large that exceeds the ability of the affected community or society to cope using its own resources (LONGUEVILLE *et al.*, 2010b). Besides, it can be described as a combination of three elements: exposure to a hazard, the current vulnerability, and capability to deal with the impacts (UNISDR, 2009), i.e., the disaster is determined by a natural hazard (e.g., a floods) that affected and brought several damages for an exposure community (e.g., buildings located close to flooding critical sides); vulnerability can be defined by the social,

[1] The International Disaster Database, Centre for Research on the Epidemiology of Disasters - CRED. http://www.emdat.be/.

structural, and economic characteristic of population susceptible to damaging effects (e.g., poor drainage system); and the prevention infrastructures (e.g., evacuation routes) (IBRD/IDA, 2013).

These growing occurrence has highlighted the needed to adopt measures to increase the power of resilience, change, and adaptation of communities affected (MENDIONDO, 2010; BAHARIN; SHIBGHATULLAH; OTHMAN, 2009; POSER; DRANSCH, 2010; NORRIS *et al.*, 2008). In this context, the disaster management presents as an important alternative to achieve this resilience and, as a consequence, avoid or, at least, reduce the impacts caused by natural disasters (BAHARIN; SHIBGHATULLAH; OTHMAN, 2009). In general, it is a continuous process composed of a number of different activities executed before, during and after an event and separated into four main phases: mitigation, preparation, response, and recovery, as shown in Figure 6.

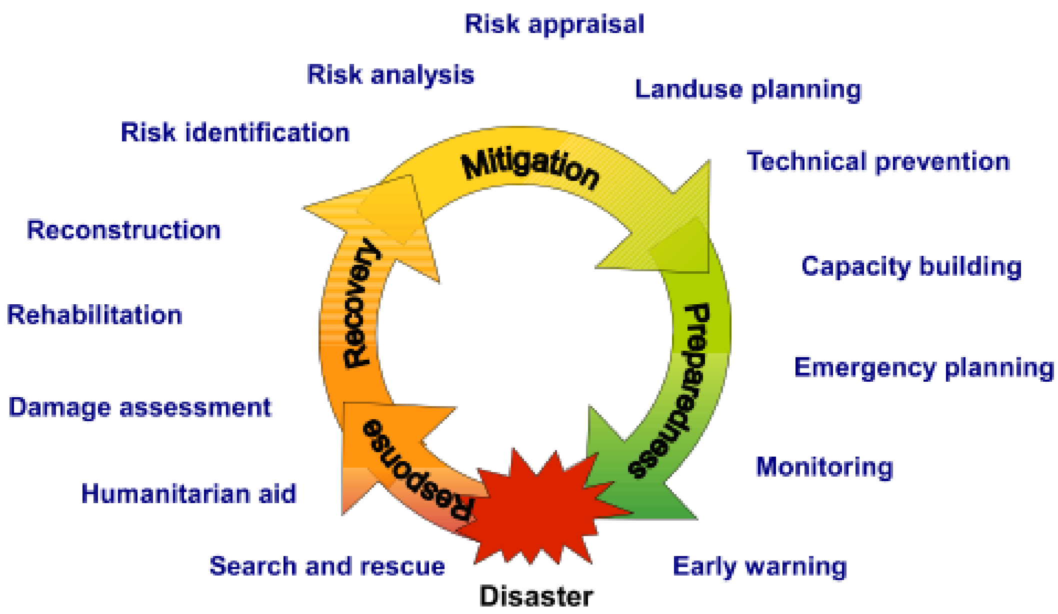


Figure 6 – Disaster management.

Adapted from Poser & Dransch (2010).

Since each phase of the disaster management consist of different activities, these are described following:

1. **Mitigation:** activities of this phase are focused on three objectives (POSER; DRANSCH, 2010; VIVACQUA; BORGES, 2012): (a) reducing the likelihood of future natural disasters, (b) minimizing the vulnerability of communities, and (c) reducing the impacts of future events. The effectiveness and success of these activities depend on of several available information like risk maps, thresholds, the condition of buildings, and forecasting. Some examples of activities in this phase are technical prevention and land use planning.

2. **Preparedness:** in the phase, several activities that aim to reduce damages caused by a disaster are performed. It also aims at planning and defining the appropriate response

actions (POSER; DRANSCH, 2010; VIVACQUA; BORGES, 2012). Some activities of this phase are monitoring and early-warning.

3. **Response:** this phase deploys activities in the aftermath of a disaster (POSER; DRANSCH, 2010). These aim to manage the impact of an event, protect communities, capital works (AHMAD; SIMONOVIC, 2006). Furthermore, reactions activities are executed, such as evacuation and humanitarian actions (e.g., procurement and resource allocation). Other activities are humanitarian aid, search and rescue, and evacuation measures.

4. **Recovery:** in this phase, the activities aim to support the affected communities in returning to its normal life (AHMAD; SIMONOVIC, 2006). These are then focused on repairing, reconstructing, and recovering from the impacts caused by a disaster (POSER; DRANSCH, 2010). Long-term policies and mitigation strategies can also be defined with the aim of making the community better prepared in case of new events (SIMONOVIC, 1999). Therefore, this phase could last for days, weeks, months or even many years, depending to the impacts caused by the event.

The monitoring of different variables (e.g., structural, environmental, and social) in disaster management is particularly crucial for reducing the impacts of a disaster. This is because emergency institutions can identify a potential event through the available systems and then issue an early warning to the disaster relief institutions, such as civil protection, humanitarian teams or ground teams. In Brazil, the National Plan of Disaster Management and Response to Natural Disasters defines that these two tasks (i.e., monitoring and early warning) are in the scope of the activities performed by Cemaden. The center uses systems from different organizations for supporting in these tasks (e.g., INPE or ANA) apart of having its own systems. Other used systems are the collaborative platforms (e.g., citizen observatories (DEGROSSI *et al.*, 2014) and mobile apps (LINK *et al.*, 2015)) that could supplement the traditional systems with updated and accurate data (GOODCHILD, 2007).

2.3 Spatial Decision Support Systems

2.3.1 Decision Support Systems

The Decision Support System (DSS) is a computer system that aims to support the decision-making with semi-structured or unstructured problem solving, i.e., this system provides technological resources to aid in the structured portion of a decision but it still needs the evaluation of a decision-maker to deal with the semi-structured or unstructured part (SHIM; FONTANE; LABADIE, 2002). Because of this, it was born using theoretical concepts of business decision-making (e.g., information management) and technological works (e.g., visualization tools and optimizing databases) but along with the year's others, disciplines have also contributed to their improvement, e.g., knowledge management and artificial intelligence (POWER, 2004; HOSACK *et al.*, 2012).

In terms of supporting decision-making, during the intelligence phase in which the decision-maker collects information needed and observes existing problems or new opportunities, the DSS provides technological resources that make easy the analysis of this information (e.g., visualization tools) (ER, 1988). Through the expert models or analytic techniques defined at DSS, it can support the decision-maker in examining the discovered problem or opportunity in the design phase (MORA *et al.*, 2003; SPRAGUE; CARLSON, 1982). This model normally is composed by decision alternatives, criteria, and the numerical relationship between the variables (MORA *et al.*, 2003). Finally, along of the next two phases (Implementation and, more recently, Learning), DSS makes possible the analysis of consequences, monitoring, and recording of results, and the understanding of the potential next problem/opportunities (MORA *et al.*, 2003). Summary, the following list of the characteristics of DSS are described by Sprague & Carlson (1982):

- This system tends to be aimed at the less well structured, underspecified problem that upper-level managers typically face;
- This system attempts to combine the use of models or analytic techniques with traditional data access and retrieval functions;
- This system specifically focuses on features that make them easy to use by non-computer people in an interactive mode;
- This system emphasizes flexibility and adaptability to accommodate changes in the environment and the decision-making approach of the user.

To achieve these characteristic, the DSS is composed of at least three important components: (i) a database management capabilities with access to internal and external data (ii) powerful modelling functions accessed by a model management system, and (iii) an user-friendly and understandable user interface designs that enable interactive queries, reporting, and graphical charts (SHIM; FONTANE; LABADIE, 2002). In this context, the interactivity and integration at the technological resources provided are extremely important because the decision-makers must be able to use models, data analysis, query and reporting tools, and online analytical processing (OLAP) in real time (HOSACK *et al.*, 2012). On another hand, this fact also contributes to the use of this information systems by non-specialists users. Together, these components are able to support the decision-making through a set of technological resources, such as real-time analytics, performance management, providing a usable view of what the data contains, and ensuring that decision-makers can gain insights from their ad hoc questions (HOSACK *et al.*, 2012).

Because of this variety of resources and potential application, these systems could be categorized in Personal Decision Support Systems (PDSS), Group Support Systems (GSS), Negotiation Support Systems (NSS), Intelligent Decision Support Systems (IDSS), Knowledge Management-Based DSS (KMDSS), Data Warehousing (DW), and Enterprise Reporting and Analysis Systems (see (ARNOTT; PERVAN, 2008)). Thus, they have been used for several decision-making purposes: water resources management (GONGQUAN; KEYAN, 2011; GU; TANG, 2000), logistics services (WANG; LI, 2013), emergency management (NEVILLE *et al.*, 2013), and organization management (SAMBAMURTHY; DESANCTIS, 1990).

Additionally, in flood management, DSS can support the decision-making tasks performed in their phases, e.g., when you need to perform simulations and create scenarios to analyze the use of land around the river channel, to create and execute evacuation routes, shelters and distribute strategically reservoirs, build dikes and barriers, to map disease occurrence to assist in its control, and optimize the targeting of available resources during the emergency management phase. However, in all these cases, the geographic setting is important as it helps in better visualization and analysis of related variables (BILLA *et al.*, 2006). Because of this, SDSS are important to incorporate the spatial context in the DSS (RUSHTON, 2001).

2.3.2 Characterizing Spatial Decision Support Systems

The geographic position or context are frequently used by people when they are making decisions (RUSHTON, 2001), for example, we choose the faster alternative during the selection of route or, in case of a disaster the identification in a geographic context, of resources closer to the affected area are important in the definition of its allocation because lives could depend on of this material. When this fact is extended to organization management, several others geographic components can be observed and analyze until decision-making, the customer's distribution, marketing criteria, and logistics allocation (CROSSLAND; WYNNE; PERKINS, 1995). Additionally, in all of these cases, the information systems play an important role not only to storage the spatial information with their relationship but also to process and display them in a visualization tool, these are some of the features presented in the Geographic Information Systems (GIS).

GIS are systems that provide features to store, manipulate, analyze, and display information geographically distributed (DENSHAM, 1991; MITTEN; PARSONS, 2011). For this, they are normally structured using a database, able to storage spatial data, and a visualization tool in which it is possible to display the information stored, explore among the datasets, present graphically and numerically the selected areas, and identify locations within a targeted environment which meet specific criteria (CROSSLAND; WYNNE; PERKINS, 1995). However, when these systems are used to support decision-making associated to geographical variables, Densham (1991) highlighted some important challenges. First, the GIS databases were developed to support the only cartographic display, this fact makes difficult the use of analytical modeling. In second, these systems did not present mechanisms which make easy and flexible the data analysis by the decision-makers. Finally, the decision-making strategies take in account the particularities in the interpretation of the decision-maker, i.e., they normally define different values for variables and relationships, however, the GIS were not designed to support this frequent variation.

In this context, the Spatial Decision Support Systems (SDSS) are designed to help one or more users in supporting decision-making in an unstructured or semi-structured geographical context (BILLA *et al.*, 2006; KEENAN, 2003). These systems emerged from the combination of the functionality of Geographic Information Systems (GIS) with the processes and methods of Decision Support Systems (DSS) (RUSHTON, 2001). This integration happened on the one hand, from the limitation of researchers using spatial data on DSS and, on the other hand, the efficiency of GIS to store and manage geographic data, however, without supporting mechanisms to help in decision-making (HOSACK *et al.*, 2012). Figure 7 presents the common architecture of SDSS.

The SDSSs are typically composed of (1) an enabled database system for managing spatial data, (2) a mathematical or expert model to assist in predicting the outcomes of decisions (e.g., location-allocation algorithms and multi-criteria decision), and (3) a graphical user interface to display tabular reports to assist in decision-making (DENSHAM, 1991; CHANG; WAN; LEI, 2010) as it was appointed, these systems represent the integration of GIS and DSS (through the models management system). Among of its application, urban analysis (STOLLBERG; ZIPF, 2008; RUIZ *et al.*, 2012; SOLTANI; MONAVARI; MAHINY, 2011), urban traffic planning (SANTOS; COUTINHO-RODRIGUES; ANTUNES, 2011; ATKINSON; CANTER, 2011), disease control (KELLY *et al.*, 2012), reaction and prevention of accidents (GUOXIANG; MAOFENG, 2010; SCHOOLEY *et al.*, 2010), disaster management (YONGSONG; SIUMING; KWOKKIT, 2010; CHEN; MORA; OUYANG, 2011) and flood management (SULAIMAN *et al.*, 2012; LEVY *et al.*, 2007; BILLA *et al.*, 2006) are highlighted.

As presented in the next section, it has grown the use of digital dashboards as information systems, which support the decision-making of top executives presenting the most important information needed to their context. Besides, the use of geographic information is also presented as a GeoDashboard (HORITA *et al.*, 2014a).

2.3.3 Digital Dashboards

A Digital Dashboard is a type of information system that is available for use on a single screen. It provides the most important management information needed to support strategic activities, such as goal setting, planning and forecasting, and tracking performance in strict accordance with organizational objectives (FEW, 2006; LIANG; MIRANDA, 2001). In addition, it can be understood as a type of information system that is usually used to support decisions at the strategic level of organizations (HOSACK *et al.*, 2012). In these kinds of decisions, the systems still handle a large number of parameters and relationships but also attempt to alleviate the effects of any unknown or shifting parameters and relationships (HOSACK *et al.*, 2012).

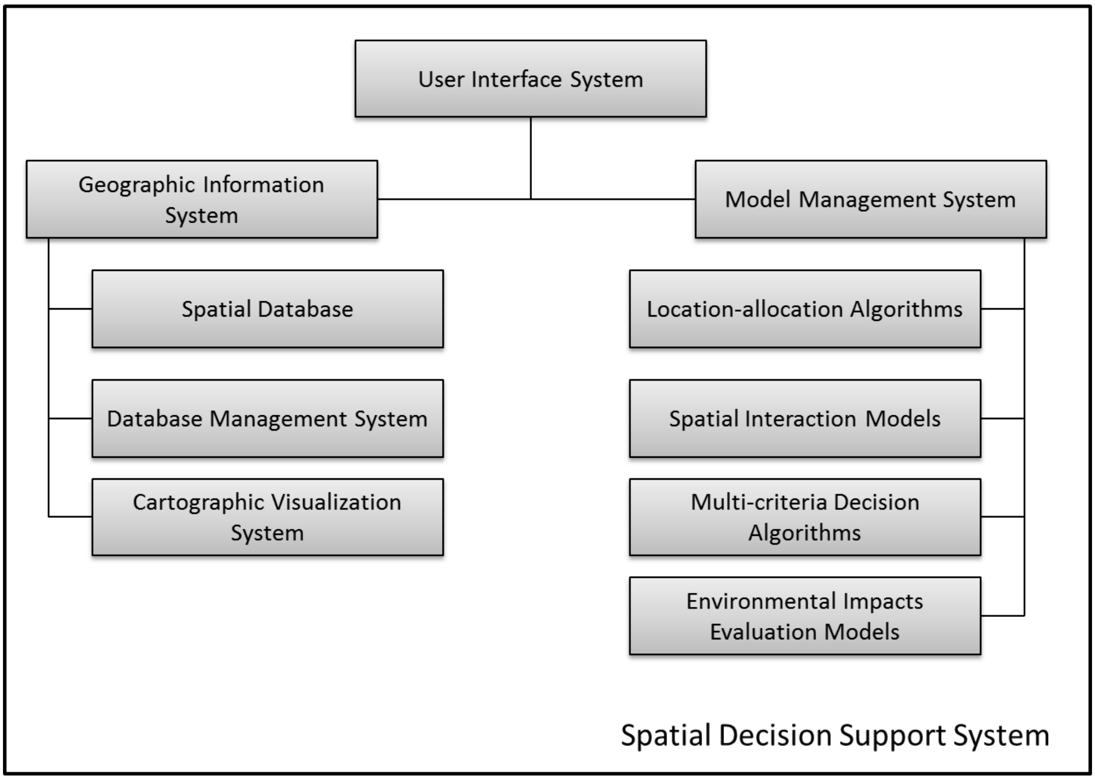


Figure 7 – Architecture of SDSS.

Adapted from Rushton (2001).

There are several essential features that these digital dashboards must include to achieve their objectives. First of all, they must be customizable and adaptive so that they can carry out the specific objectives of their users e.g., the managers. Furthermore, they are characterized by the use of drill-down capabilities that provide quick access and allow them to identify the root cause of a problem (VASILIU, 2006). Finally, they must provide an overview of everything that is currently going on in the business by means of simple charts, indicators, and forecasting capabilities (VASILIU, 2006; LIANG; MIRANDA, 2001).

Examples of digital dashboards can be found in several application fields (MAHENDRAWATHI; PRANANTHA; UTOMO, 2010; MITCHEL; RYDER, 2013). With regard to disasters, Zheng *et al.* (2010) produced a set of dynamic dashboards to provide a quick view of the user's interests, i.e., the aim is to predict the information that is most needed by different users and to display that information directly. These dashboards are a part of a project which seeks to provide disaster recovery information for a community network and improve collaboration and information exchange. Bharosa *et al.* (2010) developed a dashboard that is used for data visualization of key performance indicators for multi-agent emergency preparedness. With this end in mind, they undertook an action research project with the aim of evaluating the indicators of simulation game performance with 36 relief agency managers. The results of the evaluation showed that most of the participants were satisfied with the dashboards and intended to make practical use of them. Horita *et al.* (2014a) detail a GeoDashboard (Figure 8), which processes data streams from WSN and makes them available in the form of a set of indicators.

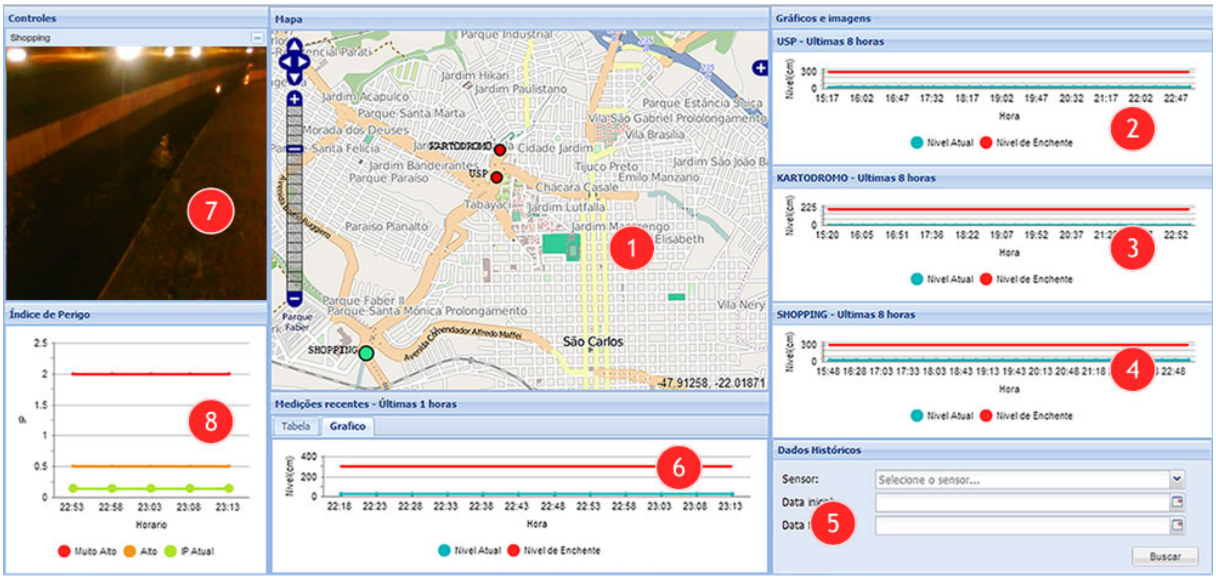


Figure 8 – Example of GeoDashboard for flood risk monitoring.

Adapted from Horita et al. (2014a).

In order to achieve the decision-making in flood risk monitoring, the GeoDashboard displays the information needed using indicators (Elements 2, 3, 4, and 8) as well as a simple photo (Element 7) taken in the monitored critical area and a geospatial map. Through the manipulation of the georeferenced points in the map (red circles), the decision-maker can analyze the current state of the environmental variables at that monitored location (HORITA *et al.*, 2014a). Thus, this dashboard aims to make faster the decision-making in flood risk monitoring.

2.4 Big Data

The interest in the generation of data and their analysis is very old in the area of computer science and software technology. Practitioners and researchers always have looked to the generated data with the aim of gaining insights and supporting their decision-making, e.g., for identifying a new product or understanding the shopping behavior of customers. In short, all this interest has passed by three major phases (HURWITZ; NUGENT; HALPER, 2013). Firstly, the generation of structured data about daily activities of companies (i.e., transactions operations) through the simple enterprise resource planning systems in the 70s; for example, all the customers' shop of a product at a supermarket. With the emerging of the Internet, the second phase comprehended mainly the generation of unstructured data, such as audio, videos, and images, and not only simple documents or texts in the 90s. Personal blogs and corporative websites have been used in this context; however, due to the complexity of the data integration, its management and analysis have become more complicated. This has become even more complex when further sources of data have emerged. This led to the last phase, which now refers to a huge volume of generated data for different objectives through distinct data sources; for example, data collected via social media platforms, wearable technologies, and in-situ sensors for traffic monitoring. In parallel, this progress in data generation and analysis was followed by the evolution of existing technologies. From the local relational database management systems (RDBMS) that use SQL for querying few megabytes of data, through the analysis of unstructured data (e.g., images and audio), until the recent use of high sophisticated technologies (e.g., parallel processing, distributed file systems, and NoSQL) for analyzing an increasing volume of data.

The term "big data" raised in this context and has multiple definitions. For instance, Hashem *et al.* (2015) defined big data as "a set of techniques and technologies that require new forms of integration to uncover largely hidden values from large datasets that are diverse, complex, and of a massive scale". Other works defined big data as a group of a large volume of data that requires capabilities that go beyond the technology's capability to store, manage, process, and analyze data effectively (MANYIKA *et al.*, 2011; GROUP, 2011). With this background, there is a common understanding that the available data has a great potential for improving decision-making and changing the business process in different business (GOPALKRISHNAN *et al.*, 2012; WAMBA *et al.*, 2015). Organizations, however, would require a solid and appropriate capacity for collecting, storing, processing, and analyze huge volume of data (DUMBILL, 2013). In this context, several processes for processing this massive volume of data have been proposed in the literature. Hurwitz, Nugent & Halper (2013) defined a process consisted by the following five activities: capture, organize, integrate, analyze, and act. Furthermore, a more detailed process is provided by Misra *et al.* (2014) that is composed of seven tasks: collect, manage, measure, consume, store, and govern. Although these processes might appear appropriate and in a certain way simply, their activities have several intrinsic issues that could turn to a negative outcome. Data integration is one of these crucial issues. This happens because as mentioned before, big data also refers to unstructured data that might demand an intensive, overwrought, and complex preprocessing activity. If it is not done properly, a wrong result could be achieved.

Because of the potential for gaining valuable insights from data and supporting decision-making, these big data have attracted further attention in the past few years due to the growing interest in data science, and thus more research has been carried out in this area (PROVOST; FAWCETT, 2013; DUGGAN, 2014). Another group of studies seeks to analyze the use of different data sources (e.g., the external data server and sensors) to support decision-making in different scenarios (GOPALKRISHNAN *et al.*, 2012; HORITA *et al.*, 2014a; WAMBA *et al.*, 2015). Within this group, Taylor *et al.* (2013) devised a software architecture for processing and visualizing data gathered from sensors while Vera-Baquero, Colomo-Palacios & Molloy (2013) outlined an architecture for integrating data from different organizations which included a data server for improving the analysis of business performance management. Social media messages have also been employed to support organizational tasks, such as marketing trends (MALSBENDER *et al.*, 2013; KURNIAWATI; SHANKS; BEKMAMEDOVA, 2013), disaster management (IMRAN *et al.*, 2014; DOLIF *et al.*, 2013; VIEWEG; CASTILLO; IMRAN, 2014) or measuring customer satisfaction (ROSEMANN *et al.*, 2012). For instance, Mandviwalla & Watson (2014) described an organization as a mix of capitals (human, economic, social, symbolic, and organizational); for example, economic capital includes financial, physical, and manufactured capital resources. Based on this, they took a social media strategy for generating these capitals.

Several definitions have been used for describing to what Big Data refer to. Likewise, some existing works also characterized these data by the intersection of the following four Vs:

- **Volume:** This characteristic indicates that big data refers to a large volume of data that can be provided by different data sources and are continuously growing in scale. This also makes difficult the management and storage of such data since organizations might be dealing with terabytes, petabytes, or even exabytes of generated data (HASHEM *et al.*, 2015). For example, there is an estimation that the digital universe will increase from 130 exabytes to around 40,000 exabytes up to 2020 (GANTZ; REINSEL, 2012).

- **Variety:** All big data are very often generated in different formats, which can be either structured or unstructured (O'LEARY, 2013). Structured data are more easily handled because they do not require further processing for deriving their meaning, e.g., the water level of a river or the product bought by a customer at a store. In contrast, structured data are more heterogeneous and require further processing, i.e., they need to be characterized or tagged somehow, otherwise, their meaning is unknown (GANTZ; REINSEL, 2012); for instance, audio and photos shared in social media platforms. For gaining insights from big data, an organization must efficiently deal with this variety of formats, as well as support an appropriate data integration.

- **Velocity:** Since data can be provided by different sources, they tend to be shared in distinct time frames. For example, social media messages can easily reach all the world in a manner of seconds while weather satellite images are shared every 12 hours. Moreover, the velocity of big data is an important factor for business because it may affect data analysis and data processing, as well as could make decision-maker faster (MCAFEE; BRYNJOLFSSON, 2012). For example, faster insights may provide an obvious competitive advantage to Wall Street analysts.

- **Veracity:** Multiple data sources can be used for generating data from official monitoring sensors to personal blogs. Therefore, there is an intrinsic necessity in assessing continuously the veracity of the data. This basically aims at ensuring that those data reflects the truth, and thus can be effectively used for supporting decision-making (HURWITZ; NUGENT; HALPER, 2013).

This definition of “4Vs” is widespread in the literature and practice although other “Vs” are constantly used for characterizing big data; for instance, Chen, Mao & Liu (2014) indicate that big data should be also characterized in regarding their value, i.e., the insights generated from big data are only valuable if they provide an information that was not known by the organization.

Through the use of GPS, GIS, and remote sensing, the geospatial data is a valuable and nowadays a very important type of big data. It is because such data can be potentially employed for better understanding the overall environmental, urban systems, and communities behavior.

2.5 Volunteered Geographic Information

The emergence of Web 2.0 and evolution of mobile devices have become the basis for the emergence of a new paradigm, where users in general (i.e., citizens) become established as producers of data and information (NIKO *et al.*, 2011). Interestingly, all these generated data and information in many cases, are more detailed and of a higher quality than those provided by official organizations (GOODCHILD, 2007; ELWOOD, 2008; LONGUEVILLE *et al.*, 2010b; OSTERMANN; SPINSANTI, 2011). In this context, Heipke (2010) proposed the term “crowdsourcing” for this phenomenon, which involves content production being carried out by a third party that is assigned to intelligence and knowledge. It is based on the experience of volunteers, who are independent in the way they use their free time and are located in remote and diverse areas.

In a similar context, but more closely linked to geographical issues, Goodchild (2007) coined the term Volunteered Geographic Information (VGI) to name this phenomenon which was defined as a collection of digital spatial data produced by individuals and non-formal institutions, i.e., by ordinary citizens using appropriate tools to gather and disseminate their views and geographical knowledge on the web. As a result, these volunteered data have a high potential to expand and qualify the amount of information available about the events and experiences of the community members to performance their activities (COLEMAN *et al.*, 2009).

VGI have been employed by several technological platforms (e.g., Wikimapia, OpenStreetMap, Mechanical Turk, CloudCrowd, and Wikipedia), as well as to assist decision-makers from a different context in the performance their activities, e.g., marketing trends (KLEINDIENST; PFLEGER; SCHOCH, 2015) or urban and communities analysis (BUGS *et al.*, 2010). A study in the USA shows that 35% of these users create content and put it online, and 26-34% of them share this content (FLANAGIN; METZGER, 2008). Furthermore, all these data have been also of a great value for supporting decision-making in disaster management (HORITA *et al.*, 2013; HAWORTH; BRUCE, 2015; GRANELL; OSTERMANN, 2016). A group of existing studies is focused on analyzing the use of social media platforms (e.g., Twitter and Flicker) in disaster management (AHMED, 2011; KAEWKITIPONG; CHEN; RACTHAM, 2012; ERSKINE; GREGG, 2012; FUCHS *et al.*, 2013; GUAN; CHEN, 2014; ALBUQUERQUE *et al.*, 2015).

Collaborative mapping platforms such as OpenStreetMap and Wikimapia also emerged as an important type of volunteered information that could support in disaster management (ZOOK *et al.*, 2010; DORN; VETTER; HÖFLE, 2014). Furthermore, the development of frameworks, as well as information systems for processing and sharing the volunteered data is the objective of another group of works (GOODCHILD; GLENNON, 2010; NIKO *et al.*, 2011; HORITA *et al.*, 2014a; DEGROSSI *et al.*, 2014; HORITA *et al.*, 2015; HAWORTH; BRUCE, 2015). The potential of VGI was showed in the aftermath of the 2010 Haitian Earthquake when volunteers have mapped the affected areas on OSM using satellite imagery as a basis for analyzing. Figure 9 displays a comparison of the mapped area before and after the earthquake in Haiti.

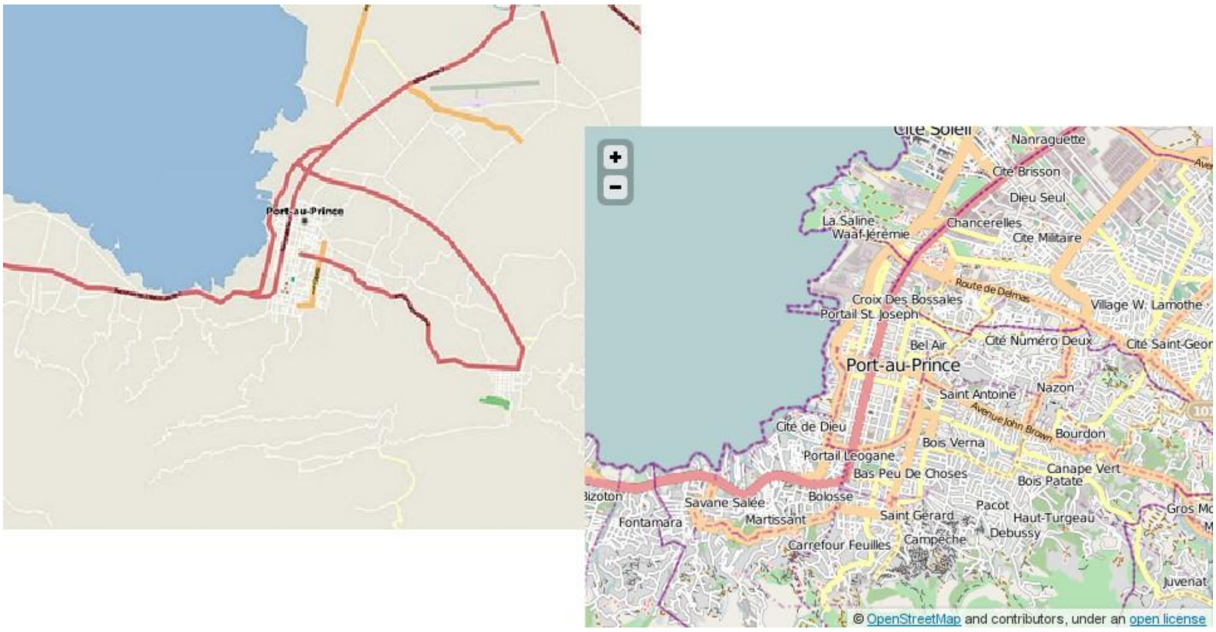


Figure 9 – A comparison of the mapped area before and after the earthquake in Haiti.

However, this huge data volume brings several problems related to its quality (ELWOOD, 2008). The quality of volunteered information can be assessed by its accuracy and credibility (POSER; DRANSCH, 2010). Flanagin & Metzger (2008) relate quality to accuracy, i.e., the degree of similarity between the information produced by volunteers in the technological platform to data presented in the real world. On the other hand, Bishr & Janowicz (2010) associate quality to credibility, i.e., an information confidence (social and reliable bond) created between the information found by the consumer and those defined by its creator. These authors extend this notion to a geographical dimension, in which the volunteers' location influences in confidence of his observations as well as temporal dimension, in which the time between different observations also affects your confidence in Bishr & Janowicz (2010).

Besides this, the information abundance also highlighted another challenge with their structuring, storage, and dissemination, i.e., how to define a Spatial Data Infrastructure (SDI) to address these requirements (BISHR; JANOWICZ, 2010; GENOVESE; ROCHE, 2010). Finally, Doan, Ramakrishnan & Halevy (2011) raised others important challenges. Firstly, the understanding of how these systems recruit volunteers and maintain their assistance. Secondly, what different kinds of support can users provide and how they can be combined. Finally, although this huge amount of data is useful to obtain knowledge, it causes several problems with regard to its quality (ELWOOD, 2008; GOODCHILD; GLENNON, 2010).

2.6 Business Models and Notations

Business processes can be defined as "a chain of functionally connected activities using information and communication technologies, which lead to a closed outcome providing a measurable benefit for a customer" (DUMAS *et al.*, 2013). Its elements are depicted in Figure 10.

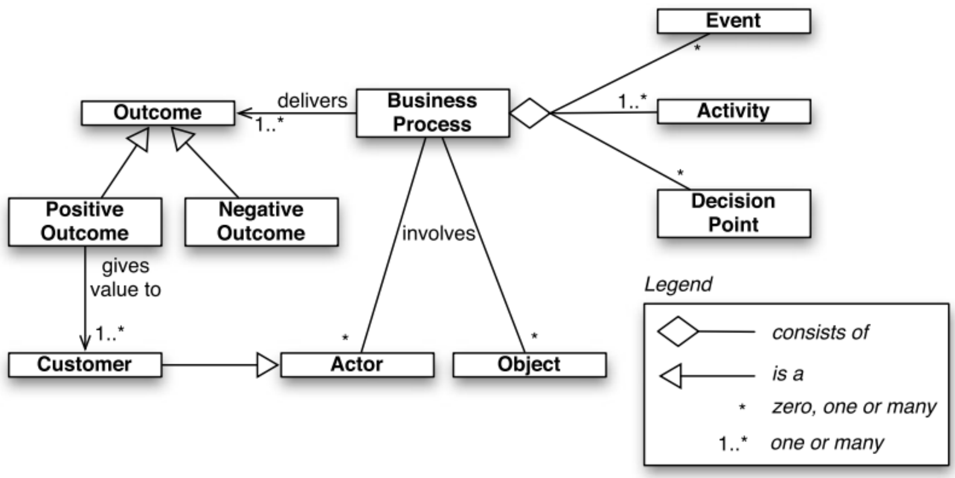


Figure 10 – Elements of a business process. Based on (DUMAS *et al.*, 2013).

The Business Process Model and Notation (BPMN) is a standard model and notation that was designed by the Object Management Group (OMG) for modeling these processes. It defines a set of conceptual elements that are an abstraction of real-world elements and their relationships (OMG, 2013), e.g., the "conceptual element activity" represents a single unit of work (e.g., inventory checklist). A business model consists of a set of events, activities, or decision points (e.g., ANDs and ORs) that can be interconnected with the aim of following a logical sequence. It also involves actors or objects (e.g., data objects). For providing a common understanding, BPMN defines graphical notations for each of these elements that also aim at providing reusability and easy information sharing among the different stakeholders (see Figure 11). Figure 12 displays an example of a business model for security check at the airport.

The starting point of Figure 12 is the collection of the boarding pass. This triggers the execution of the first activity "Proceed to security check", i.e., the passenger is allowed to proceed for security check only if he/she has the boarding pass. After this first activity, two activities are executed in parallel "Pass security screening" and "Pass luggage screening", i.e., the passenger has to pass through the personal security screening, as well as luggage screening.

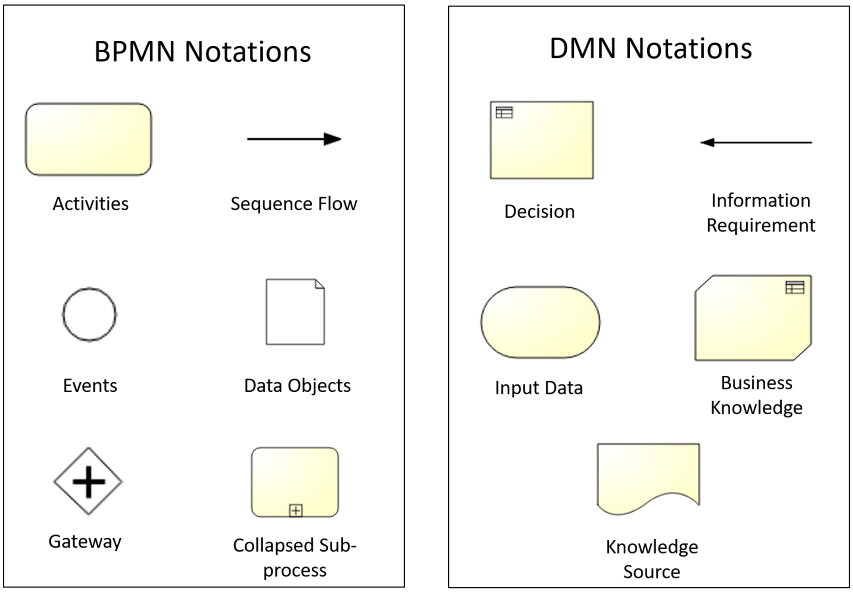


Figure 11 – BPMN and DMN Notations.

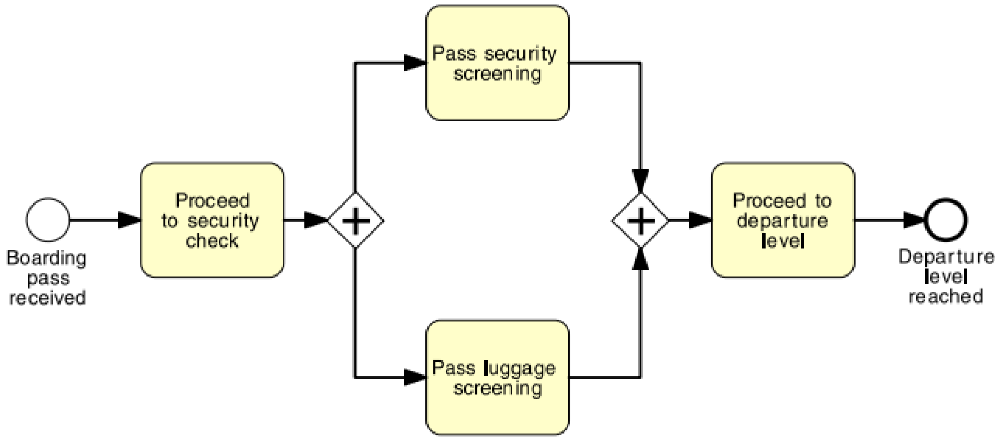


Figure 12 – Example of a business process modeled using BPMN. Based on Dumas *et al.* (2013).

Once both activities are concluded, the last activity is triggered "Proceed to departure level". This means that the passenger is allowed to go to the departure level only if he/she concludes the personal and luggage screening. In short, the business model is a means for better understanding the existing activities, the logical sequence flow, and related objects. BPMN has been applied in different areas, such as customer services (SAEEDI; ZHAO; SAMPAIO, 2010), disaster management (JANSEN; LIJNSE; PLASMEIJER, 2010; SACKMANN; HOFMANN; BETKE, 2013), and business management (ELVESÆTER *et al.*, 2010). Sackmann, Hofmann & Betke (2013) proposed an extension to BPMN with the aim of including elements to represent place-related information like water hydrants or ambulances. Although process modeling notations are valuable to represent organizational activities that involve decision-making, they do not include an explicit consideration of the decisions involved.

To fill this gap, OMG has recently released the Decision Model and Notation (DMN), which is designed to establish a relationship between business processes and decision-making. This provides conceptual elements (e.g., business rules, input data, and required information) for modeling decisions (OMG, 2014). Figure 13 depicts an example of a business process in which an activity is connected to a decision (i.e., "Decide order acceptance" activity).

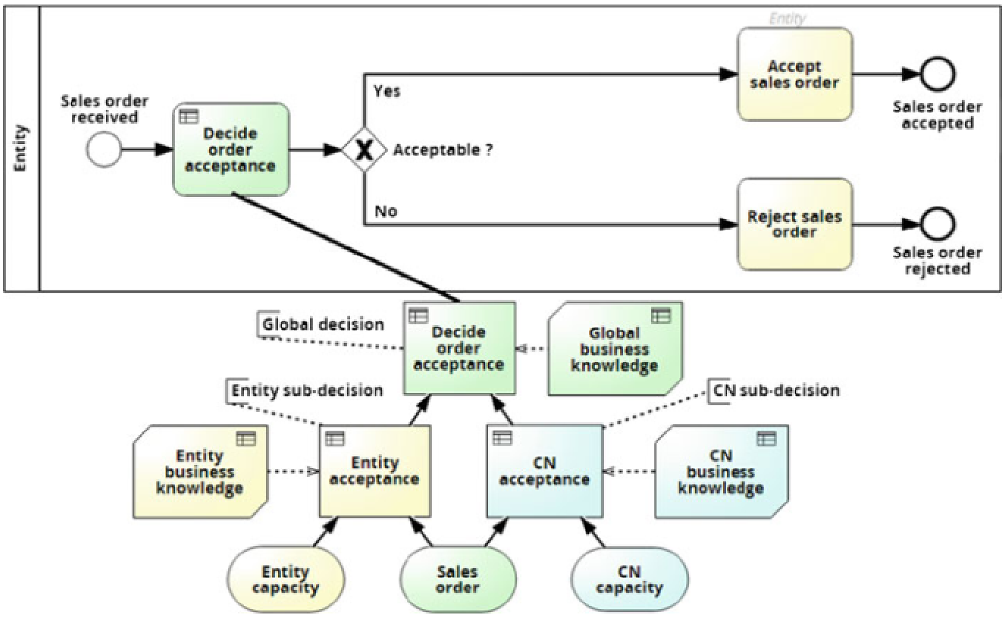


Figure 13 – Example of a business process and a decision modeled using BPMN and DMN. Based on Biard *et al.* (2015).

The business process is displayed in the top of the figure, while the decision in the bottom part of the figure. In the business process, a specific activity, named "Decide order acceptance", is associated with the "Decide order acceptance" decision. This in turn requires the outcome of two other decisions ("Entity acceptance" and "CN acceptance") that are analyzed through a specific business knowledge ("Global business knowledge"). The result of this business knowledge defines the outcome of the decision. Some first examples of studies that have employed DMN can be found in the literature (AA *et al.*, 2015; BAZHENOVA; WESKE, 2015; JANSSENS; SMEDT; VANTHIENEN, 2016). Biard *et al.* (2015) highlights the fact that DMN has a great potential for modeling decisions in collaborative networks. Furthermore, Batoulis *et al.* (2015) described a methodology, which aims to derive a DMN model from business processes, and adapts the original process model by replacing the decision logic accordingly. DMN and related studies are certainly an important step for providing a further understanding of decision-making in organizations. However, this model notation does not take into account the kind of data sources that could provide the modeled information requirements to support decision-making. It only models the values of an information requirement in terms of the necessary input data, e.g., "the price of a product is U$ 28,00".

2.7 Final Remarks

In this chapter, the disaster management was presented. This management defines a set of activities to be performed before, during, and after an event, in other words, it is composed of four phases: mitigation, preparedness, response, and recovery. In all of them, the geographic information is important to help with better analysis of environmental variables. Because of this, they must to integrated, updated, and faster for supporting the decision-making of emergency agency. The SDSS is also presented as the information systems, which aim to support unstructured and semi-structured decisions related to a geographical context. These systems use mathematics models and decision visualization resources to interpret the geographical information and support the decision-making. Furthermore, the big data was characterized by the intersection of “4Vs”: volume, veracity, velocity, and variety. When associated with a geographic location, these data have a potential for improving decision-making of different organizations; however, at the same time, they must follow a strict and solid process for supporting the collection, storage, processing, and analysis of this amount of data. Finally, the business process and notation were introduced as a valuable alternative for describing the business process, as well as enabling a common understanding of the process among all stakeholders.

With this background, the next chapter presents two literature reviews that were executed for analyzing the use of volunteered information and SDSS in disaster management.

CHAPTER 3

VGI AND SDSS IN DISASTER MANAGEMENT

3.1 Overview

During the last year, several researchers have been adopting a systematic and systematic approach for identifying, analyzing, and evaluating the existing studies in the literature in regarding of specific research questions. This is called "Systematic Literature Review (SLR)" (BUDGEN *et al.*, 2008). The SLR is a means of evaluating and interpreting all the studies available in the literature about research questions, area, or phenomenon of interest (KITCHENHAM; CHARTERS, 2007; DYBA; DINGSOYR; HANSSEN, 2007). Conducting systematic research follows a sequence of well-defined and rigorous methodological steps, in accordance with a previously predetermined protocol (BIOLCHINI *et al.*, 2005).

According to Kitchenham & Charters (2007), an SLR comprises three steps: planning, conduction and results. In the planning phase, the research questions are defined together with the SLR protocol. This protocol is the review guide in which is not only established all its details but also the conduction and analysis process. It also aims at eliminating possible execution bias and researchers' influences, as well as to make viable the reproduction of review. In the conducting phase, the planning is executed in the electronic databases and a selection made of the returned studies. Last but not least, the returned studies are evaluated and the data extracted.

In the same manner, the Systematic Mapping Study is a review approach that has been using by software engineering researchers to obtain relevant studies in the literature. However, different from SLR, the systematic mapping study aims to analyze relevant studies in order to give an overview of topic area or theme (PETERSEN *et al.*, 2008). Then, the state of evidence in specific topics can be investigated using an SLR (PETERSEN *et al.*, 2008; BUDGEN *et al.*, 2008).

This chapter is organized as follows: in the Section 3.1.1 is presented the protocol, the conduction, and result in the analysis of a systematic mapping study, which aims to analyze the use of VGI in disaster management. Further, Section 3.1.2 details the planning, conduction, and analysis of a systematic mapping study executed to analyze the use of SDSSs for disaster management.

3.1.1 Systematic Mapping Study A - Volunteered Geographic Information and Disaster Management

Planning and Protocol

This systematic mapping study aims to identify relevant studies available in the literature which use information from volunteers as a source of information to achieve resilience in disaster management. In this context, the following research goal was defined:

GO1: How can the use of information from volunteers in order to achieve the resilience in disaster management?

As appointed by Kitchenham & Charters (2007), research questions are defined through analyzing four items identified as PICO (Population, Intervention, Comparison, Outcome). Here, it is defined only the Population, Intervention, and Outcome items, mainly because there is not a comparison between some approaches. Below, there is the definition of the three remain items:

1. **Population:** the population of this review is natural disaster management;
2. **Intervention:** we are analyzing the use of information provided by volunteers;
3. **Outcome:** the general outcome is the identification of the use of VGI in disaster management.

Taking these definitions into account, the secondary questions presented below, aim to help in the answering of the main research question:

- **RQ1.1:** In which phases of disaster management has information from volunteers been used?
- **RQ1.2:** In what types of disasters is information from volunteers used?
- **RQ1.3:** What types of methodologies are employed in research that use information from volunteers for disaster management?

The four phases of disaster management defined by Poser & Dransch (2010) were used as a benchmark to determine at what stage RQ1.1 is included. In the case of RQ1.2, we used the types outlined by Jha (2010) to identify the category, type, and subtype of disaster employed in this chaper. To answer RQ1.3, we used the information provided by the reviewed studies themselves about the research methodology employed, e.g., the literature review, case studies, surveys, experiments, and action research. For this review, the search process was performed using electronic databases presented in Table 1, these are the most relevant databases in the context of Software Engineering and Information Systems.

Table 1 – Electronic databases of Systematic Mapping Study A.

Name	**URL**
IEEExplore	http://ieeexplore.ieee.org/
ACM Digital Library	http://dl.acm.org
Scopus	http://www.scopus.com
ISI Web of Science	http://www.isiknowledge.com
AISeL	http://aisel.aisnet.org/
ScienceDirect	http://www.sciencedirect.com
SpringerLink	http://www.springerlink.com

In this review, we used a subjective search string definition, the terms were defined using the domain knowledge and past experiences of the reviewers (ZHANG; BABAR; TELL, 2011). Besides, its structure was composed by relevant keywords and their synonyms, these terms were separated respectively by the boolean operators “AND” and “OR”, as shown in Figure 14.

In order to support the selection of relevant studies that are really associated with the research questions, this review defined a set of include (IC) and exclude criteria (EC) as presented below:

- **IC1:** The study uses information from volunteers in disaster management;

- **EC1:** The study is not write in English and Portuguese;
- **EC2:** The study is not available for downloading;
- **EC3:** The study is duplicate;
- **EC4:** The study is poster or tutorial;
- **EC5:** The study uses information from volunteers but is not related to disaster management;
- **EC6:** The study is related to disaster management but did not use information from volunteers;

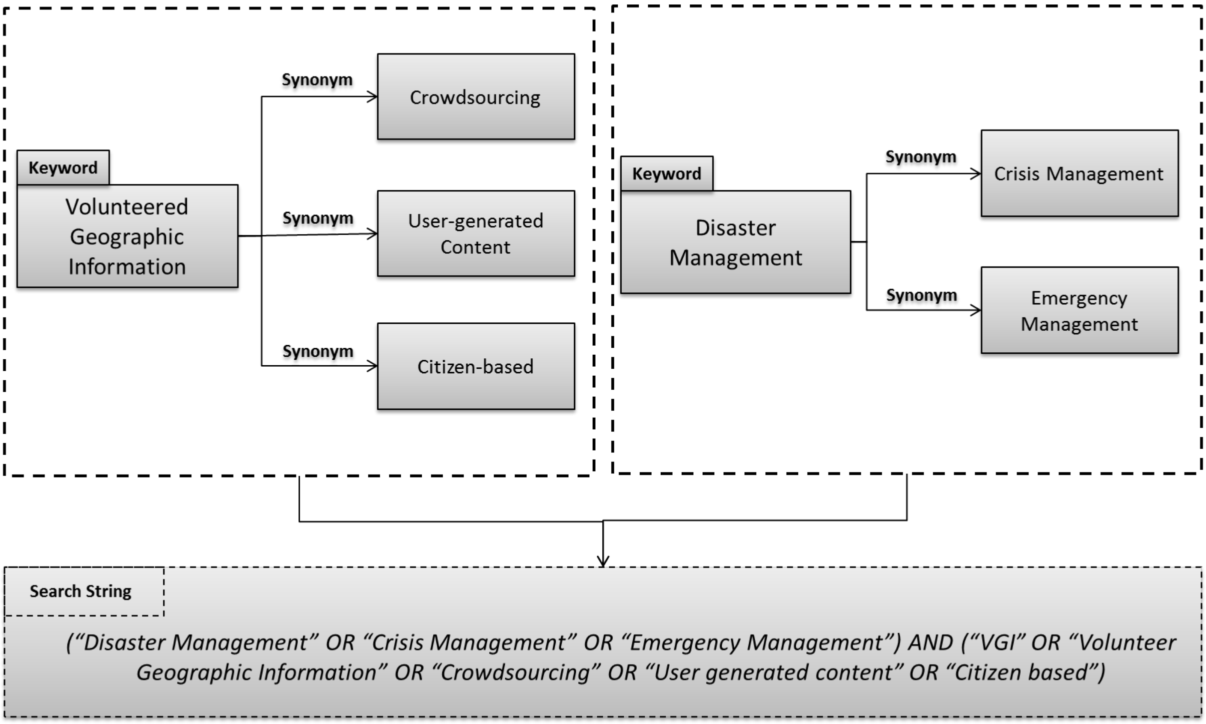


Figure 14 – Search string for Systematic Mapping Study A.

In this review, the following conduction process was defined. Initially, the search string was performed in electronic databases with the aim of identifying papers that are important for this study (Step 1). Next, it was assessed the title and abstract and applied pre-defined inclusion and exclusion criteria to each paper, and thus obtained a set of relevant papers (Step 2). Following this, the full text of each of these papers was obtained and its contents were critically evaluated. In doing so, the studies were read in full and the inclusion and exclusion criteria applied (Step 3). For this step, a form was used to extract the data and integrate the researcher's answers in the best way possible. Lastly, the main data were summarized and the relevant studies categorized. All this process aims to identify the most relevant studies in order to aid answering the research question proposed for this review.

Finally, the quality of each selected studies is performed. For this, it was analyzed five criteria that seek to measure the existence of fundamental concepts for the context area. For each question was applied the following scale of values: Yes (Y) = 1 point, No (N) = 0 point, Partly (P) = 0.5 point. Next, these values were summed, resulting in the quantification of the quality of each study. The list of questions is shown in Table 2.

Table 2 – List of quality criteria of Systematic Mapping Study A.

ID	Description
QC1	Is there a description of the context in which the research was performed?
QC2	Is there a clear statement of the research objectives?
QC3	Is the background section detailed in accordance with the research context?
QC4	Are the results reported clearly?

All the elements previous described were used to compose the protocol of review. Next, it is presented in detail the results obtained during the conduction phase.

Conduction

During the conduction phase, the search string defined in the protocol was performed in the pre-determined electronic databases in order to obtain the most important studies. In Step 1, a total of 320 studies were returned for a period of time from January/2007 to January/2013, SpringerLink had the highest return, 183 studies, (57% in total) while IEEExplore returned 8 studies, 3 studies at ACM, 27 at Scopus, 8 at ISI Web of Science, 36 at AISeL, and 55 at ScienceDirect. Figure 15 displays these data in detail.

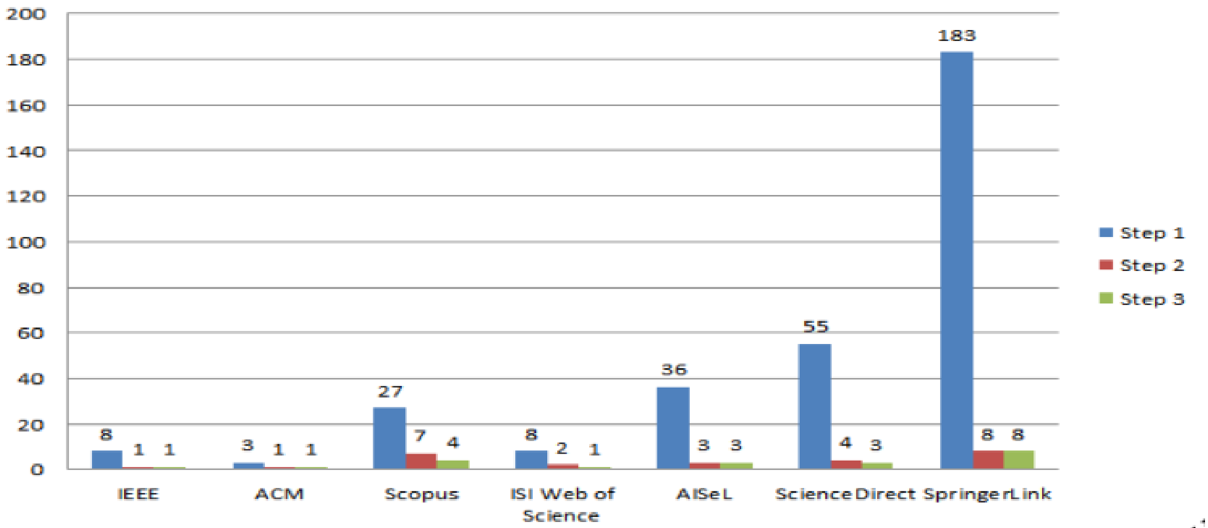


Figure 15 – Studies returned separated by phases and electronic database.

All these 320 studies returned had their title and abstract analyzed as well as the application of inclusion and exclusion criteria, a total of 26 studies remained (Step 2). Among them, ISI Web of Science and ACM Digital Library showed an index of 25% (2 out of 8) and 33% (1 out of 3) in the selected papers. During Step 3, the paper developed by Ahmed (2011), although only a work-in-progress, provided support for the application of social media for disaster management, and was thus maintained among the relevant papers. Moreover, at this stage, some papers were excluded, since they do not use or apply VGI for disaster management, are related to the crowdsourcing carried out by know-how specialists, and are related to models and strategies for collecting information available on the Internet. Finally, the 21 studies remained were included and used to summarize the data in order to answer the research questions.

This review was conducted by two reviewers, one coordinator, and two others specialists supervisors with a well-known background in disaster management and VGI. In the entire search process, the reviewers performed their tasks in parallel (e.g., application of include and exclude criteria, and the extraction of data from studies selected) coordinated by the coordinator which aims to help in case of divergences and problems. Besides, in some cases, the supervisors were requested to aid in the problem of which the coordinator was not able to solve. In the next section will be presented the analysis and discussion around the studies selected.

Description of Studies Selected

The analysis and discussion process were performed taking account the 21 studies selected in the previous phases through a simple extract form developed at the GoogleDocs[1]. However, before the presentation of the results obtained, it will be present a brief description of the general characteristic of the studies selected. Figure 16 presents the number of studies selected along the years per electronic database.

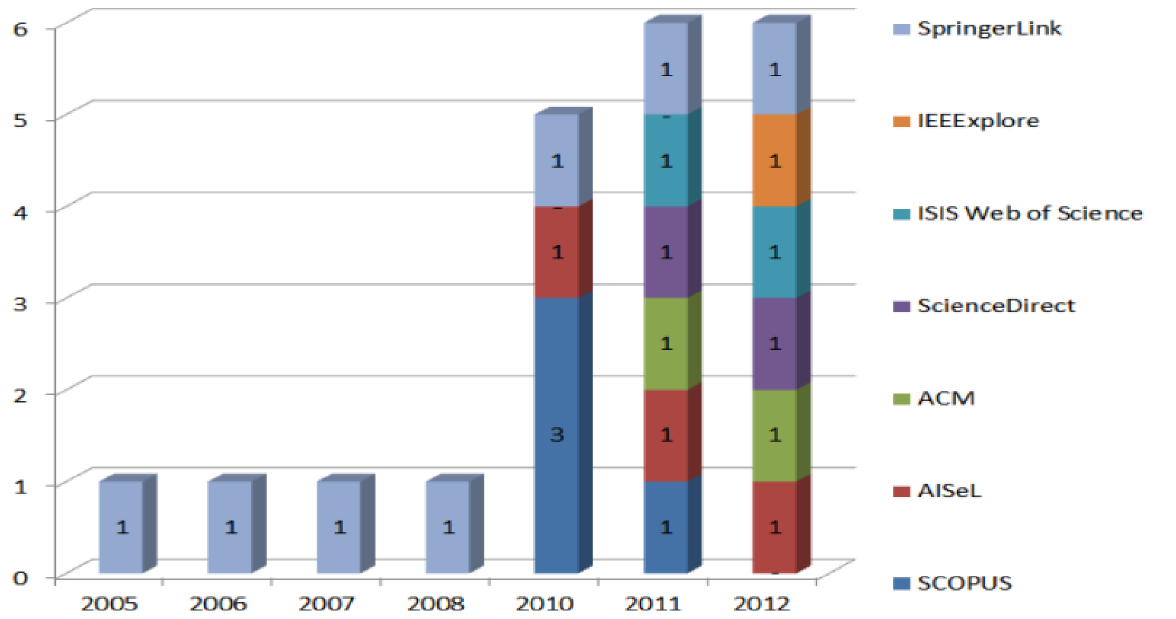


Figure 16 – Studies selected along the years per electronic database.

As we can see, the studies developed in the first three years were published in SpringerLink. There is not study in 2009 but there is a growing in the number of studies developed (from 1 in 2008 to 5 in 20010, and 6 in 2011-12). Besides, the number of studies presented in the last two years is equally distributed between the electronic databases. The SpringerLink is the database with more studies selected (7 in 21) following by Scopus (4 in 21). Table 3 displays the information of studies selected, in column 1 is shown the identification of each study while column 2 displays its reference. Column 3 presents the quality score of the respective study and column 4 its electronic database. Finally, column 5 presents the type of publication: (C) conference or (J) journal.

Based on the information presented on Table 3, 28,57% of the studies selected were published in a conference while others 71,43% in a journal. According to quality assessment, 71.42% of the studies selected were classified as Excellent (from 4 to 3), 19.05% as Medium

[1] https://drive.google.com/

Table 3 – Studies selected in Systematic Mapping Study A.

ID	Reference	Score	Database	Type
P1	(PEARCE, 2005)	3	SpringerLink	J
P2	(CHEN; LIU; CHAN, 2006)	3.5	SpringerLink	J
P3	(SCHAFER; GANOE; CARROLL, 2007)	3	SpringerLink	J
P4	(IKEDA; SATO; FUKUZONO, 2008)	2	SpringerLink	J
P5	(HUANG; CHAN; HYDER, 2010)	2.5	SpringerLink	J
P6	(SINNAPPAN; FARRELL; STEWART, 2010)	4	AISeL	C
P7	(POSER; DRANSCH, 2010)	4	Scopus	J
P8	(GOODCHILD; GLENNON, 2010)	4	Scopus	J
P9	(LONGUEVILLE *et al.*, 2010a)	4	Scopus	J
P10	(SCHADE *et al.*, 2011)	3	SpringerLink	J
P11	(SAVELYEV *et al.*, 2011)	3	ACM	C
P12	(ROCHE; PROPECK-ZIMMERMANN; MERICSKAY, 2011)	3	Scopus	J
P13	(YATES; PAQUETTE, 2011)	4	ScienceDirect	J
P14	(AHMED, 2011)	2	AISeL	C
P15	(NIKO *et al.*, 2011)	3.5	ISI Web of Science	J
P16	(POHL; BOUCHACHIA; HELLWAGNER, 2012)	3.5	ISI Web of Science	C
P17	(VIVACQUA; BORGES, 2012)	3	ScienceDirect	J
P18	(WEAVER; BOYLE; BESALEVA, 2012)	1	Scopus	C
p19	(ADAM; SHAFIQ; STAFFIN, 2012)	1	IEEExplore	J
P20	(KAEWKITIPONG; CHEN; RACTHAM, 2012)	4	AISeL	C
P21	(MUNRO, 2012)	3	SpringerLink	J

(from 3 to 2), and only 9.52% as Bad (from 2 to 0). This fact evidences the relevance and credibility of the studies selected.

Analysis and Discussion

In this section, we will summarize the data outlined in the previous section and show some indicators that were identified and defined to address the research questions.

RQ1: In which of the disaster management phases has information from volunteers been used?

With regard to the use of VGI in the different phases of disaster management, we identified a strong predominance of works covering the response phase (11 out of the 21 selected papers). Three other papers discussed the mitigation phase and three were also found in the preparation phase. Figure 17 shows these papers distributed per year.

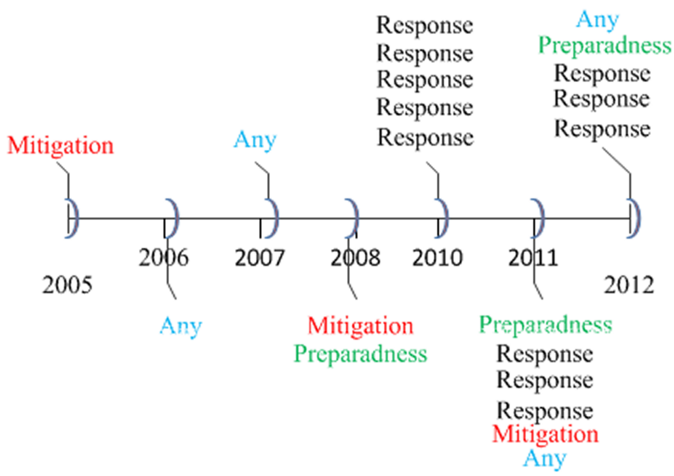


Figure 17 – Studies per disaster type and years.

Moreover, in general, the studies of Ahmed (2011), Schafer, Ganoe & Carroll (2007), Chen, Liu & Chan (2006), and Kaewkitipong, Chen & Ractham (2012) were carried out in more than one stage of the disaster management. Savelyev *et al.* (2011) and Adam, Shafiq & Staffin (2012) do not identify which phase of disaster management their research was concerned with.

RQ2: In what types of disasters was the information from volunteers used?

Figure 18 displays the disasters types where VGI is being used with more frequency. As can be seen, the occurrence of floods, earthquakes, fires and storms are very common.

As can be seen in Figure 18, almost 50% of the research in the field entails carrying out activities for managing disasters related to fire and floods, and another 30% were related to storms and earthquakes. These facts can be explained by the growing number of these types of disasters around the world. Some research has been conducted in countries like Taiwan, Japan, Thailand, and South Korea that are seriously affected by tectonic instability which causes these types of disasters. Moreover, on the basis of data collected during the SLR, it was revealed that there is still a gap in the research on data involving the use of volunteers for disasters in Latin America, which have recently experienced severe floods, fires, and storms.

RQ3: What are the types of methodologies used for research that aims to use information from volunteers for disaster management?

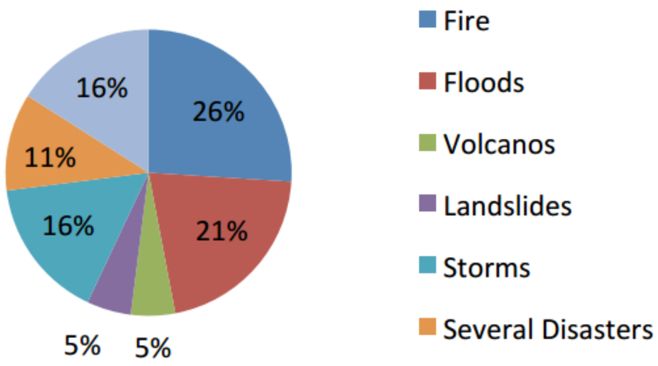


Figure 18 – Types of disaster addressed in the studies selected.

Table 4 shows the quantification of each of the key research methodologies employed by the different papers. As disaster management is an area with a strong field of applications, the majority of the studies analyzed used methods for conducting case studies (12 papers). The design of new processes, techniques, and methods following the design science approach was also used by a significant number of works (7 papers). Only a few papers have conducted a literature review to explore papers in the area so far (2 papers).

Table 4 – Methodology of studies selected in Systematic Mapping Study A.

Methodology	**# of Studies**	%
Case Study	12	57,14%
Literature Review	2	9,52%
Design Science	7	33,34%
Total	21	100,00%

The conclusions of this review are that the use of VGI in disaster management is growing, with a significant growth in the number of publications over the last three years (2010-2012). Moreover, it was found that the predominant research area was disaster response, whilst fewer studies were devoted to mitigation and preparedness, and no study was found dealing with recovery. This can be explained by the fact that response is the most visible part of disaster management and is also more likely to attract the attention of volunteers. However, there is a challenge for researchers which is how to advance knowledge about methods that include VGI in mitigation and preparedness activities such as risk analysis and early warning systems, as well as in the recovery phase of communities, by helping them reorganize their routine, and create mechanisms to prevent disasters from happening again in the future.

This research also showed that VGI is commonly used to manage floods and fires. The prevailing media for sharing VGI was found to be social media (i.e., Twitter, Facebook, Youtube) and mobile devices. Interestingly, very few of the reviewed papers address VGI platforms like Ushahidi, Elva, OpenStreetMap, and Wikimapia; this offers an interesting avenue to explore in future research. Furthermore, apart from the fact that most current studies are based on case studies, there is also a broad field for conducting research that employs other methodologies. Action research seems particularly suited to this area, since it seeks to solve current practical problems (e.g., establishing resilience against disasters with the aid of information from volunteers) while expanding scientific knowledge (e.g., learning how can we effectively use information from volunteers to tackle disasters).

3.1.2 Systematic Mapping Study B - VGI, Decision-Making and Disaster Management

Planning and Protocol

The main objective of this systematic mapping study is to identify relevant studies that address the use of VGI as a source of information to increase the quality of decision-making processes at the disaster management. In order to achieve this objective, the following research goal were established:

GO1: How can information from volunteers help on the decision-making process used in the disaster management?

In general, research questions are formulated using the analysis of four items identified as PICO (Population, Intervention, Comparison, Outcome) (KITCHENHAM; CHARTERS, 2007). In this systematic mapping study, there is not a comparison of approaches because of this, the item comparison is not presented. Following, the aspects used to elaborate the research questions are detailed:

1. **Population:** Once this review focuses on identifying ways to use information from volunteers to decision-making in disaster management, the population is composed of citizens located in a risk zone and official agencies;

2. **Intervention:** In this review, we are analyzing models, methods, techniques, tools, and process used to integrate, qualify, and share information from volunteers;

3. **Outcome:** the main objective of this review is to identify resources used to make information from volunteers available for help on the decision-making at the disaster management. We also expect to identify the information needed to assist in each phase of disaster management and how to retain, integrate and share them.

Besides this, we also defined some secondary questions that are presented below:

- **RQ1.1:** In which purpose was the information from volunteers used?

- **RQ1.2:** What is the type of volunteers' information?

In order to establish the search strategy, it was defined the source selection criteria; sources list; studies language; and keywords and their related terms. Table 5 presents the electronic databases defined for this systematic mapping, these databases are efficient to conduct systematic reviews in the context of Software Engineering (DYBA; DINGSOYR; HANSSEN, 2007). Scopus has been added since it is considered the largest database of abstracts and citations (KITCHENHAM; CHARTERS, 2007). Besides this, AISeL also has been added because it has important papers related to information systems.

Table 5 – Electronic databases of Systematic Mapping Study B.

Name	**URL**
IEEExplore	http://ieeexplore.ieee.org/
ScienceDirect	http://www.sciencedirect.com
Scopus	http://www.scopus.com
ISI Web of Science	http://www.isiknowledge.com
AISeL	http://aisel.aisnet.org/

In relation to studies languages, only primary studies written in English were considered, mainly because most of the research in Computer Science has been reported in this language.

The process used to define the search string used in this systematic mapping starts with the identification of the main terms associated with the main research question. Next, the synonyms of each main terms were identified aiming to obtain all relevant studies. Finally, the synonyms were joined using the boolean operator "OR" while the main terms using "AND". The search string and its terms are shown in Figure 19.

After this, the systematic mapping also defined some important include or exclude criteria aiming to identify primary studies that are relevant and help to answer the research questions. Thus, the inclusion criteria of this systematic mapping are:

- **IC1:** The study uses information from volunteers to help with the decision-making process at the disaster management;
- **IC2:** The study presents some resource to make a decision using information from volunteers for the disaster management.

Additionally, the exclusion criteria established are:

- **EC1:** The study is not write in English;
- **EC2:** The study is not available for downloading;
- **EC3:** The study is duplicate;

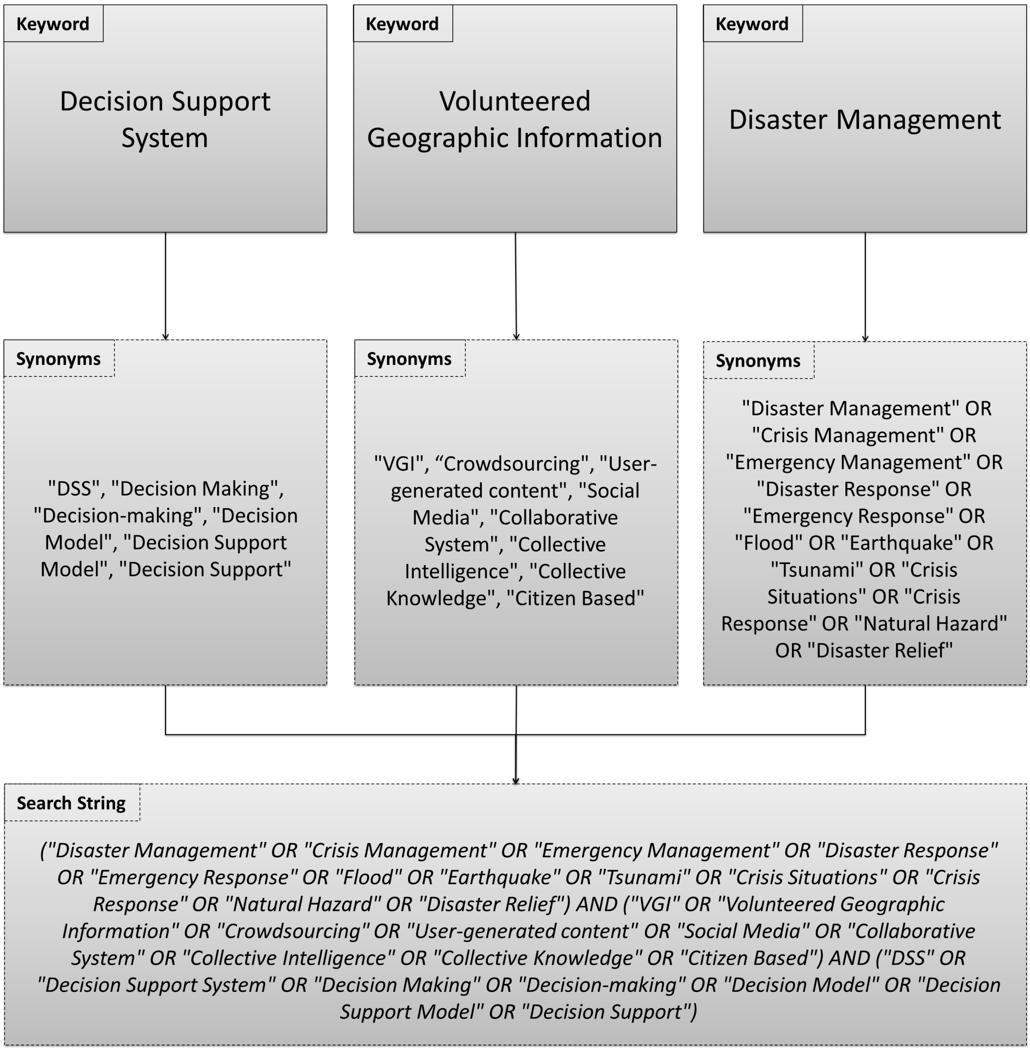


Figure 19 – Search string for Systematic Mapping Study B.

- **EC4:** The study did not archived more than 2 on the quality index;
- **EC5:** The study is not related to information from volunteers;
- **EC6:** The study is not related to disaster management;
- **EC7:** The study is not related to support decision;
- **EC8:** The study is a previous version of a more complete study about the same research;

In this systematic mapping, the selection and evaluation of primary studies were being performed in three essential steps:

1. **Initial Selection:** First, it is selected only studies obtained by applying the search string in the electronic databases defined for this review.

2. **Second Selection:** The studies returned in the previous step will be evaluated according to its title and abstract, and a list containing only those considered potentially relevant should be generated. In this step, the include and exclude criteria are also applied.

3. **Third Selection:** Finally, the studies included are full read and evaluated. The inclusion and exclusion criteria will be applied, refining a number of resulting papers. An automated tool is used aiming to facilitate the information extraction.

In addition, the quality of each selected studies is performed. For this, it was analyzed five criteria that seek to measure the existence of fundamental concepts for the context area. For each question was applied the following scale of values: Yes (Y) = 1 point, No (N) = 0 point, Partly (P) = 0.5 point. Next, these values were summed, resulting in the quantification of the quality of each study. Thus, studies that do not reach a minimum score defined by the exclusion criterion EC4 was dropped from the set of studies analyzed. The list of questions is shown in Table 6.

Table 6 – List of quality criteria of Systematic Mapping Study B.

ID	Description
QC1	Is there a clear statement of the research objectives?
QC2	Were the results evaluated in an unbiased manner?
QC3	Are the results reported clearly?
QC4	Is there a clear background which supports the research?
QC5	Is the method used for validation of the study clear?

Conduction

This systematic mapping was conducted by two researchers under the supervision of a specialist supervisor from May/2013 to June/2013. For this, it was used the definitions presented in the protocol, detailed in the previous section. In this conduction phase, we adapted the search string defined in the protocol according to the peculiarities of each electronic database, in exception the IEEExplore. The search string applied at IEEEXplore had to be separated in two others sub-strings, since it does not allow searching in the abstract and title in the same string. Figure 20 presents in detail the steps of the selected process as well as their results.

As we can see in Figure 20, the initial selection was performed in the electronic databases defined in the protocol applying the search string the IEEExplore returned 12 studies, 31 at AISeL, 7 at ISI Web of Science, 47 at Scopus, and 8 at ScienceDirect (Step 1). All these 105 studies had their respective title and abstract analyzed and were applied to include and exclude criteria, in the second selection (Step 2). In the third selection, 26 studies were selected for full-text review but only 13 of them were used for information extraction and analysis (Step 3). Figure 21 presents the number of studies selected along of the phase.

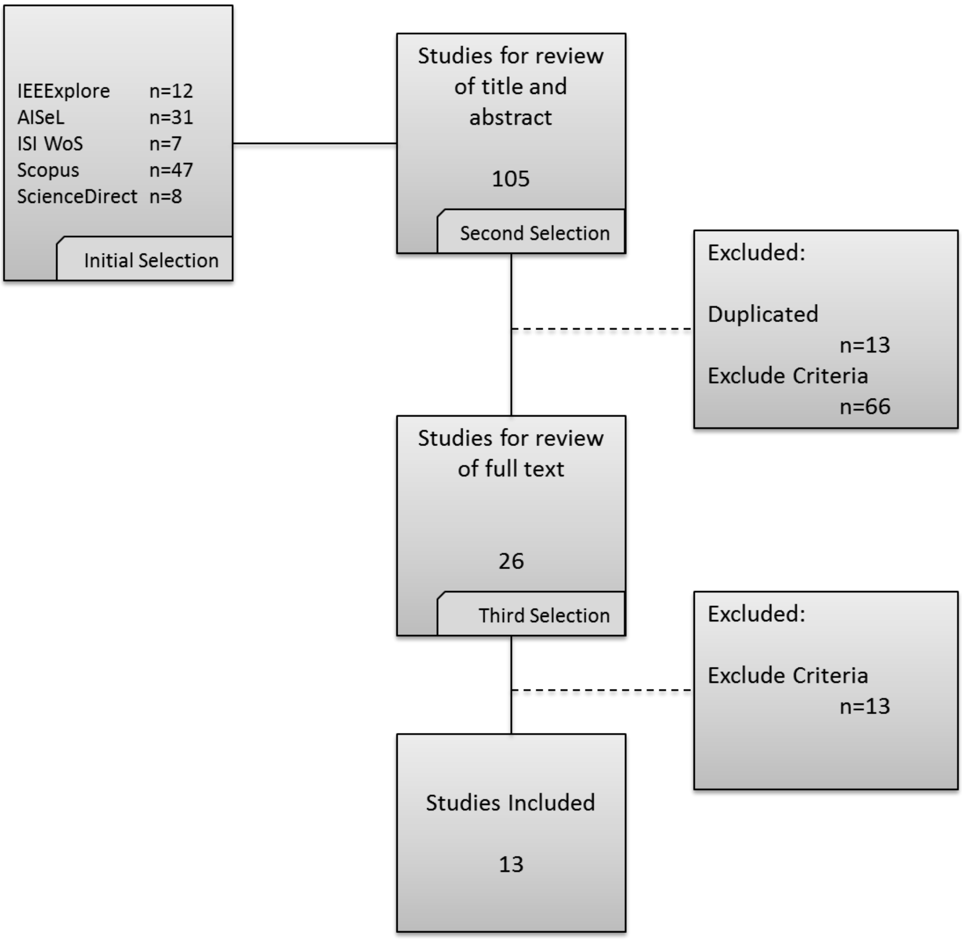


Figure 20 – Selection process.

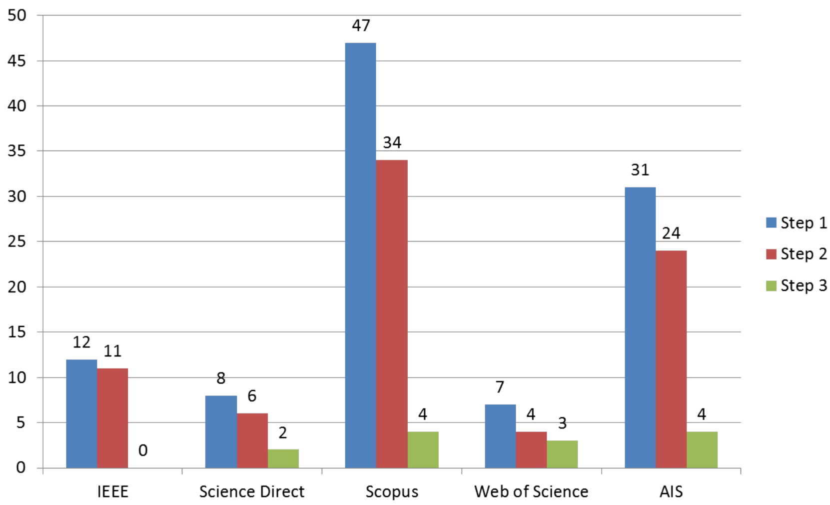


Figure 21 – Studies returned separated by phases and electronic databases.

With the exception of IEEExplore, the others databases returned at least one relevant paper. Among of them, close to 42% were from ISI Web of Science (3 out of 7), 25% were from Science Direct (2 out of 8), 12,90% were from AISeL (4 out of 31), and 8,51% were from SCOPUS (4 out of 31). In the first database, this high degree of percentage could be explained because of the low number of studies returned in the initial selection. This fact could highlights that both synonyms used and the refinement process of the string have to be evaluated aiming to achieve the most relevant papers for this systematic mapping.

Description of Studies Selected

The 13 studies selected in the final selection were equally distributed along of the database (1 per year) with the exception of 2010 and 2012. The first study was published in 2010, after that, there is an interest growing process in the use of VGI for decision-making in disaster management from 1 in 2010 to 6 in 2012. In 2013, the low number could explain because this systematic mapping was conducted until the end of the year. Figure 22 presents the number of studies selected along the years per electronic database.

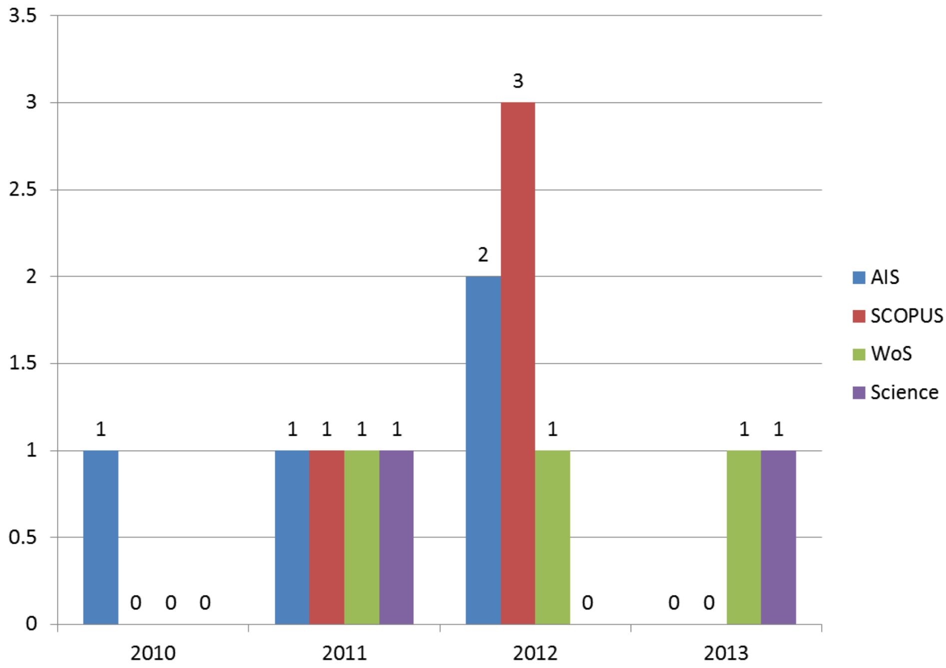


Figure 22 – Studies per years.

SCOPUS and AISeL have the same number of studies selected (4 studies), they are followed by ISI Web of Science with 3 studies and, at last, Science Direct with 2 studies. The number of studies at SCOPUS and AISeL could be explained due to the fact that SCOPUS is a well-known integrated repository while AISeL is an important repository from Information Systems (IS) research community having a lot of studies related to DSS and IS for decision-making. Table 7 displays the main information of the studies selected.

The quality assessment of this systematic mapping achieve some relevant index, all of the studies selected scores index over the 3 points in the evaluation having 46,15% classified as Excellent (5 points), close to 69% as Good (4.5 points), and 30,76% were classified as Medium (4). This fact evidences the relevance and credibility of the studies selected. Besides, 69,23% of the studies selected were published in a conference while others 30,77% in a journal. This fact also evidences that this research field still immature and further researches are needed.

Analysis and Discussion

In this section, the data presented in the previously are analyze aiming to answer the research questions proposed for this systematic mapping.

RQ1.1: In which purpose was the information from volunteers used?

The studies selected in this systematic mapping highlight the usage of information provided by volunteers for five purposes in disaster management:

- **Event Forecasting:** Due to the huge number of data shared by volunteers in social media, they have been used to aid to forecast events before or during it strikes an area (HASHIMOTO *et al.*, 2012). This detection could help in efficiently managing a disaster situation, e.g., identifying the real-time situation in the area (POHL; BOUCHACHIA; HELLWAGNER, 2012).
- **Situation Awareness:** To identify the situation awareness presented in the affected areas are essential to support the activities of decision-makers. Social media are becoming an important information input to support situational assessment (to produce awareness) in all domains (MACEACHREN *et al.*, 2011).

Table 7 – Studies selected in Systematic Mapping Study B.

ID	Reference	Score	Database	Type
P1	(OH; KWON; RAO, 2010)	5	AISeL	Conference
P2	(AHMED, 2011)	3	AISeL	Conference
P3	(MACEACHREN *et al.*, 2011)	5	SCOPUS	Conference
P4	(HASHIMOTO; KUBOYAMA; SHIROTA, 2011)	3	ISI Web of Science	Conference
P5	(YATES; PAQUETTE, 2011)	5	Science Direct	Journal
P6	(ERSKINE; GREGG, 2012)	4	AISeL	Conference
P7	(KAEWKITIPONG; CHEN; RACTHAM, 2012)	4	AISel	Conference
P8	(POHL; BOUCHACHIA; HELLWAGNER, 2012)	5	SCOPUS	Conference
P9	(HASHIMOTO *et al.*, 2012)	4.5	SCOPUS	Conference
P10	(GOTTUMUKKALA *et al.*, 2012)	4	SCOPUS	Conference
P11	(RIJCKEN *et al.*, 2012)	4	ISI Web of Science	Journal
P12	(WIDENER; HORNER; METCALF, 2013)	5	ISI Web of Science	Journal
P13	(KUMAR; HAVEY, 2013)	5	Science Direct	Journal

- **Prevention:** The prevention activities are important to reduce the communities' vulnerability to a disaster (RIJCKEN *et al.*, 2012). In this context, the information provided by volunteers using collaborative maps (e.g., OpenStreetMap) or social media has been used to identify the current state of environmental variables and, as a consequence, aid to decision-making in preparedness phase (e.g., planning the land use at risk areas).

- **Evacuation:** During an evacuation, the current state of the environmental variables are essential to guarantee its efficiency and accuracy (GOTTUMUKKALA *et al.*, 2012). For example, if there is a tree blocking the road as soon as decision-makers (or even the community) known this fact, they will have more time to analysis an alternative route. In this context, the information shared by volunteers located in the affected areas have been helping to update these environmental variables (e.g., traffic flow).

- **Supply Chain:** The processes of supply chain during disaster relief efforts have several unstable variables as well as a set of environmental factors associated. Because of this, the proliferation of social media creates research opportunities to find the best ways to integrate both the efforts of relief agencies and communities' actions in the event of large-scale disasters (KUMAR; HAVEY, 2013).

Figure 23 presents the distribution of these purpose of usage in the phases of disaster management. The mitigation and recovery were removed because there is not any study which uses information from volunteers to support the activities of these phases. Besides, Hashimoto *et al.* (2012) did not specify in which phase of disaster management they are developing their analysis while Gottumukkala *et al.* (2012) conducted their study focusing on activities of preparedness and response phases.

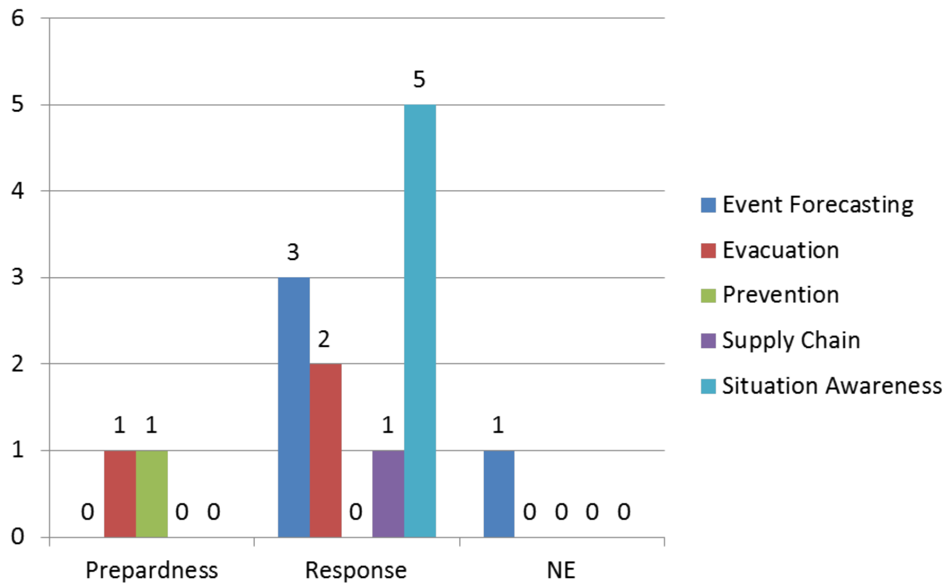


Figure 23 – The use of VGI per disaster management phase.

Figure 23 also evidenced that the number of studies conducted for the response phase is higher in comparison to those performed for preparedness. This fact could be explained because this phase is a phase that has received a lot of attention of volunteers (HORITA *et al.*, 2013). Besides, most of the interesting analysis of communities reaction in case of disaster are performed in this phase. For example, we could understand how the community deals with the routes alternatives in case of evacuation or how the members close to an affected area share the information with their peers.

RQ1.2: What is the type of volunteers' information?

By following the classification proposed by Albuquerque *et al.* (2016), Table 8 presents the separation of the number of information from volunteers used to support decision-making in disaster management. Because of the volume of data produce, social media is the type most founded of volunteers' information (11 studies) followed by one study on collaborative mapping and crowd sensing.

Table 8 – Number of types of volunteers' information used in Systematic Mapping Study B.

Type	**Number of studies**	%
Social Media	11	84,62%
Collaborative Mapping	1	7,69%
Crowd Sensing	1	7,69%
Total	13	100,00%

The main conclusions of this systematic mapping are that the use of VGI for support decision-making in disaster management, even needing further researches is a research field that has been growing over the last four years. In this manner, the usage of collaborative mapping and collaborative analysis are both promising research areas, mainly because of the intensive use of OpenStreetMap and Ushahidi. Moreover, the development of approaches focused on activities conducted during the preparedness (e.g., early warning and planning of resource allocation) still remain in the literature.

Due to the time to conduct this systematic review, some essential threats could be highlighted. The final number of studies selected is really low and could not be used to generalize the results. The inclusion of more electronic databases and synonyms could solve this fact and increase the quality of the results. Furthermore, during the extraction process, some information described in the study were difficult to analyze and understand. To achieve this problem, the opinion of one more reviewer could be essential to reduce the biases and understanding problems.

3.2 Final Remarks

This chapter described the conduction of two systematic mappings that aimed to analyze the use of VGI in disaster management and its usage to support decision-making along of disaster management phases. For this, it was used both the SLR process defined at Kitchenham & Charters (2007) and Systematic Mapping Study process presented by Petersen *et al.* (2008). In conclusion, there is a growing usage of VGI as a source of information to disaster management as well as its usage to support some decision-making activities. However, further research still needs, mainly those which aim to analyze the use of VGI to support resource allocation and recommendation of alerts.

Based on the findings of these two systematic mappings, the next chapter presents a conceptual architecture which aims at integrating heterogeneous data sources.

CHAPTER

4

A CONCEPTUAL ARCHITECTURE THAT INTEGRATES HETEROGENEOUS DATA SOURCES

4.1 Overview

Recent floods have shown the damage that such disasters can cause to the economy and citizens of a country (KAEWKITIPONG; CHEN; RACTHAM, 2012; MERZ *et al.*, 2012). This is particularly the case in Brazil, where frequent serious flooding accounts for 54% of the disaster events of the past few years (IBGE, 2014). In view of this, flood risk management has become a critical issue. Timely and accurate information can greatly assist the emergency agencies involved in flood risk management. However, the continuous monitoring of the potential risks of flood hazards require precise estimates of the risks incurred that are based on the observation of rainfall and water levels in local regions (JHA; BLOCH; LAMOND, 2012).

WSN have emerged as an alternative approach which can provide updated information for water resource management at a relatively low deployment cost (ALBUQUERQUE *et al.*, 2013). Although this approach has been successfully employed for different situations (LEE *et al.*, 2008; HUGHES *et al.*, 2011; SEAL *et al.*, 2012; SHUKLA; PANDEY, 2014), it requires a considerable effort to ensure it works effectively (PATEL; KAUSHIK, 2009). In addition, WSNs often fail to provide data from several parts of the riverbed since there is a lack of an appropriate station in what are called “ungauged areas”. Running parallel with this, another valuable source of information is VGI. This enables ordinary citizens who reside in high-risk areas, to provide information through various devices (e.g., smartphones) (GOODCHILD, 2007; LONGUEVILLE *et al.*, 2010a; ROCHE; PROPECK-ZIMMERMANN; MERICSKAY, 2011).

The combination of WSN and VGI can act as a mutual support system for achieving effective flood risk management. However, it raises challenges which are threefold: (1) dealing with distinct formats (e.g., photos vs numeric values) at different levels of measurements (e.g., water level gauges vs citizen's perceptions), (2) ensuring interoperability among data providers, and (3) conveying the integrated information in a single and understandable way. This chapter therefore aims to tackle these challenges by presenting a conceptual architecture that integrates information provided by heterogeneous data sources (i.e., WSN and VGI). This architecture is employed for developing the AGORA-DS, a Spatial Decision Support System (SDSS) for supporting decision-making in flood risk management. The architecture comprises the following: (a) the acquisition layer responsible for defining available data sources, (b) the integration layer designed to integrate the data and make them available in compliance with interoperable standards, and (c) the decision-support layer, the purpose of which is to provide a web-based decision support tool for visualizing the integrated information to support decision-making.

This chapter is structured as follows. Section 4.2 describes AGORA-DS and sets out its conceptual architecture. On this basis, Section 4.3 describes the deployment used for analyzing AGORA-DS. Finally, Section 4.4 discusses the lessons learned from this deployment and the results obtained. The limitations of this project are also described in this section.

This chapter summarizes the main results of the paper "*Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks*", published in the journal Computers & Geosciences (HORITA *et al.*, 2015).

4.2 AGORA-DS: Description and Architecture

The AGORA-DS was built on the conceptual architecture shown in Figure 24. This architecture is based on the architecture proposed by Horita *et al.* (2013) and consists of the following layers: the acquisition layer, the integration layer, and the decision-support layer. The next sections will outline these layers in detail.

4.2.1 Acquisition layer

This layer seeks to encapsulate the idiosyncrasies of data sources, and provide the appropriate technological resources (e.g., interfaces or web services) to disclose their data to the integration layer. This encapsulation is beneficial in several ways such as improving scalability and maintainability, reducing the impact of a change in data sources, and allowing new data sources to be included in a flexible way.

In this manner, two adapters were employed owing to the difference in the data structure and external sources of communication. First, there is a sensor adapter which is responsible for defining a standardized and easy way to receive data from the WSN. Since it is an input interface, the sensor adapter also converts the received data to an Observation and Measurements (O&M) specification (the XML format used to describe an observation of interoperable standards) (OGC,

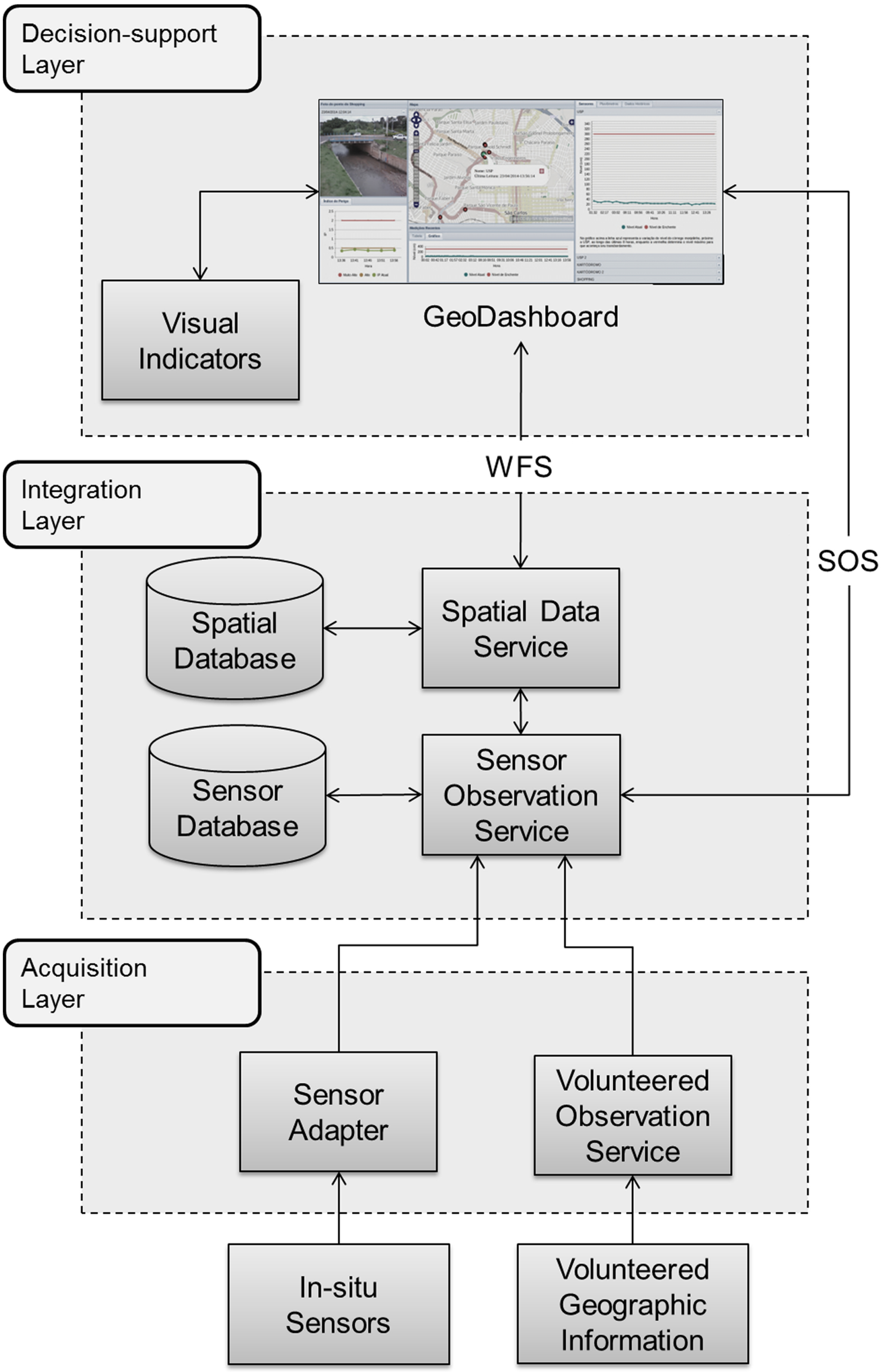


Figure 24 – Conceptual architecture of AGORA-DS

2013). This adapter conveys the data to the integration layer through the *InsertObservation* operation of the Sensor Observation Service (SOS) (OGC, 2012).

The second adapter called Volunteered Observation Service is based on our previous work (DEGROSSI *et al.*, 2014) and handles the information provided by volunteers. This service interacts with the citizen observatory with the goal of collecting relevant information, e.g., the reports provided by volunteers. After this, it also executes the *InsertObservation* operation for conveying the data to the integration layer.

4.2.2 Integration Layer

The purpose of this layer is to define mechanisms to receive, store, and share data provided by the acquisition layer. This is achieved by adopting the SOS, which defines interfaces both for receiving the data (via *InsertObservation* operation) and sharing them (via *GetObservation* operation). It also stores the received data in a sensor database (see data model in OGC (2012)). The SOS then supports the integration of distinct formats of data as well as the interoperability of data sources.

Furthermore, this layer sets up the Spatial Data Service which aims to convert data collected in SOS to the geospatial service standards, Web Feature Service (WFS). WFS provides resources which allow the creation, modification, and querying of geographical features[1] on the Internet rather than working with it on the file level (e.g., raster or shapefile) using File Transfer Protocol (FTP) (OGC, 2014). We have considered WFS instead of Web Map Service (WMS) because it supports complex query operations regarding geographic features that result in low-time for data processing on the client-side (ZHANG; LI, 2005). In addition, the Spatial Data Service also defines a simple database for storing all geographic information.

Therefore, the information can be shared with the decision-support layer by means of two specific operations, as shown in Figure 25. First, the *GetFeature* operation is processed at WFS and aims to return a document to the decision-support layer containing a set of available geographic features (OGC, 2014). Second, the *GetObservation* operation provides access to integrated information by means of the spatial, temporal, and thematic filtering which will be used by the visual indicators (JIRKA *et al.*, 2012). This operation needs a set of parameters (see details in OGC (2012)) and its response is based on O&M specifications (OGC, 2013). These operations allow then to the decision-support layer receives independently geographic information and integrated information.

4.2.3 Decision-support Layer

The decision-support layer defines elements for enabling decision-makers to interpret information in a more efficient and effective way. In the scope of this project, they are a web-based decision support tool and visual indicators.

The web-based decision support tool displays the integrated information provided by the integration layer for supporting flood risk management. This “web-based decision support tool” is hereby called Geodashboard. The term “dashboard”, well known in the field of business

[1] A geographical feature can be defined as an object that is an abstraction of a real phenomenon (OGC, 2014), e.g., in our case sensors data and volunteer information.

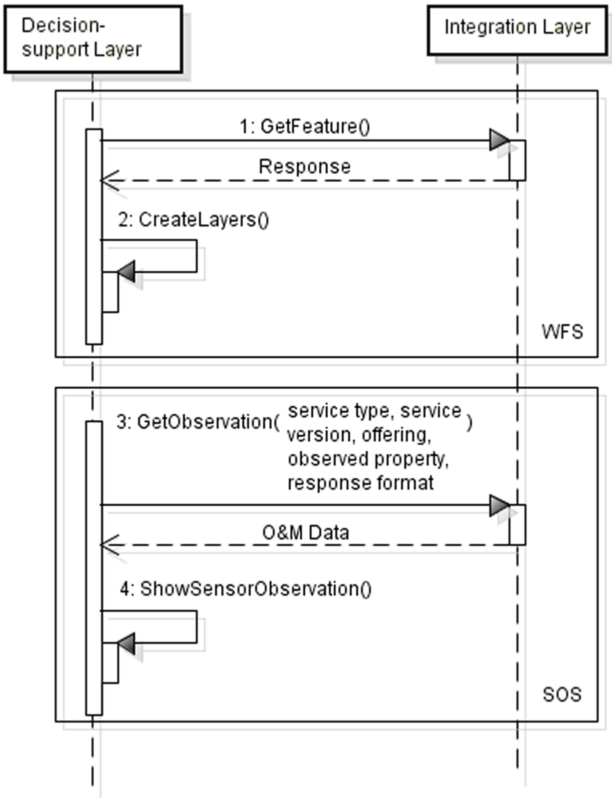


Figure 25 – The interaction with the decision-support layer

analytics, refers to an information system which aims at providing the most important information needed for supporting decision-making at different organizational levels (FEW, 2006; LIANG; MIRANDA, 2001). When geospatial factors are also important for supporting decision-making (e.g., location of customers and transport routes), it can be called “Geodashboard” (HORITA *et al.*, 2014a).

Moreover, the visual indicators were designed to show the information collected in the acquisition layer which has been processed by expert models or not, e.g., the water level at a particular period of time or the vulnerability of communities who live nearby the river. Additionally, the Geodashboard displays a simple chart for each in-situ sensor with the aim of supporting the analysis of its water flow and the limit for flooding. Finally, in the context of this project, a photo taken by in-situ sensor at the critical area of the flood site, also composes the current version of the GeoDashboard.

4.3 AGORA-DS: Deployment and Analysis

4.3.1 Study Area

The study area is located in São Carlos/SP, a city that is 230 km from São Paulo in Brazil; it has a high population density and is in a region that is frequently affected by floods. Because of this, a flood risk management of the river catchments of the city is carried out in this study area to analyze the deployment of AGORA-DS. Figure 26 shows the creek of Monjolinho, Santa Maria Madalena, and Tijuco Preto, all located in the town center of São Carlos.

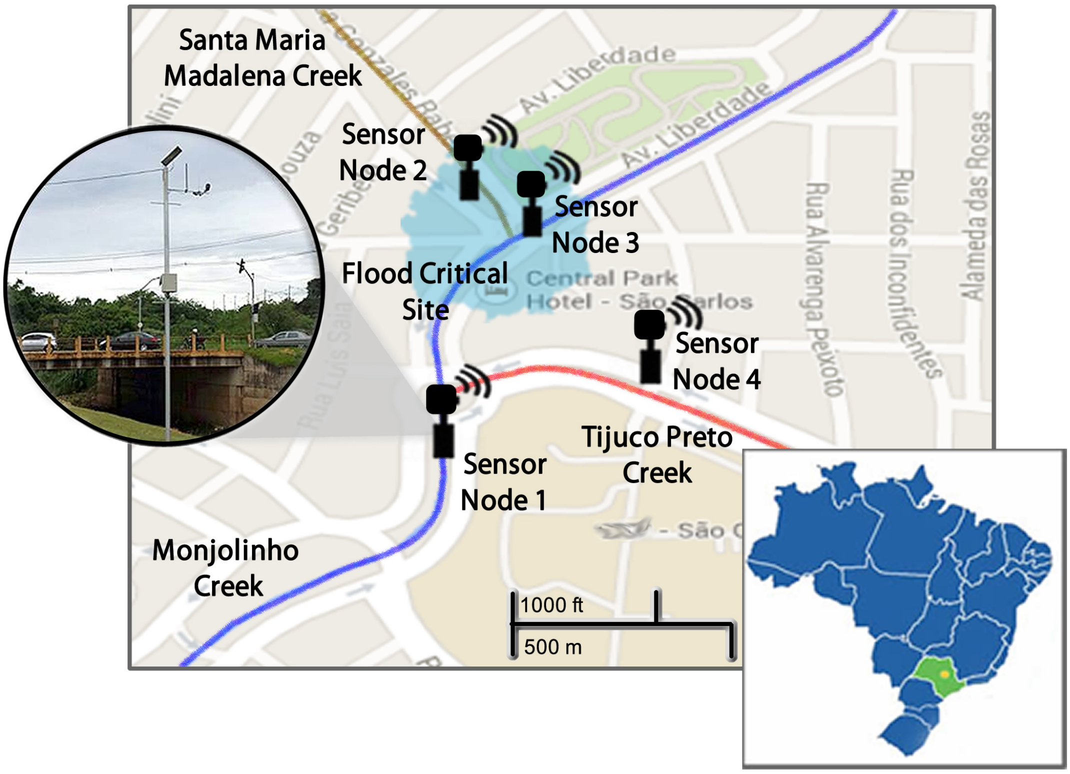


Figure 26 – The study area in São Carlos/SP, Brazil

Adapted from Horita et al. (2014a)

4.3.2 Deployment

A critical flood zone (the blue shaded area in Figure 26) is also highlighted at the intersection of the Santa Maria Madalena and Monjolinho creeks, a region with a large number of inhabitants living in residential buildings and houses and often affected by floods (MENDES; MENDIONDO, 2007; BARROS; MENDIONDO; WENDLAND, 2007; PERHOVAZ, 2010; HORITA *et al.*, 2014a). Owing to these problems, the area is often monitored by a WSN installed along the riverbed (HUGHES *et al.*, 2011; DEGROSSI *et al.*, 2013; HORITA *et al.*, 2014a).

Furthermore, citizens had provided reports related to the water level in different parts of the study area, as a result of the crowdsourcing-based approach developed by (DEGROSSI *et al.*, 2014). This approach defines four interpretation mechanisms for helping volunteers to have a better understanding of environmental variables, i.e., water level ruler, multi-color band, puppet in the shape of a human figure, and general tags rather than the other strict mechanisms (e.g., the water level is low or overflowing). They were also used to categorize the citizens' reports (e.g., "flooded area"). Following this, these mechanisms are employed in a web-based citizen observatory - also known as "Flood Citizen Observatory", which is built as an instance of the

Ushahidi crowdsourcing platform[2]. Figure 27 displays the main interface of the Flood Citizen Observatory.

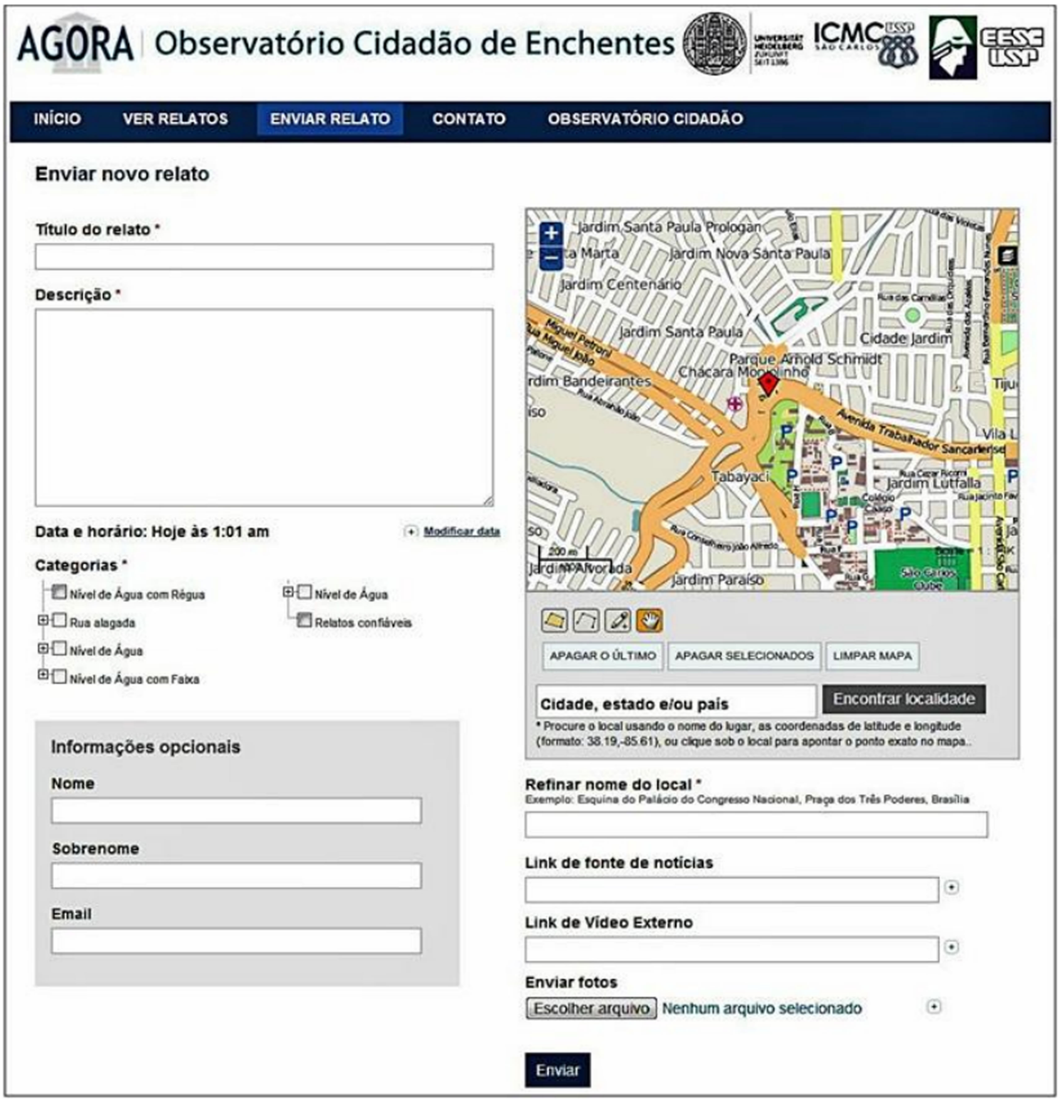


Figure 27 – Flood Citizen Observatory (DEGROSSI *et al.*, 2014)

All reports provided by citizens are evaluated out manually by the Flood Citizen Observatory's administrator before making them visible on the Web or accessible through the Ushahidi REST API. This evaluation is important for improving credibility and quality of the shared volunteered information, although a more complex assessment is beyond the scope of this project. Finally, Figure 28 shows that the Volunteered Observation Service obtains the categories and requests citizens' reports after the Flood Citizen Observation collected them from the volunteers. This communication is addressed by using the Ushahidi REST API. The *InsertObservation* operation is used afterwards for sharing the citizens' information with the integration layer.

The data provided by the data sources mentioned above are later shared with the integration layer. The sensor adapter was developed on this layer using Java Server Pages (JSP)

[2] http://www.ushahidi.com/

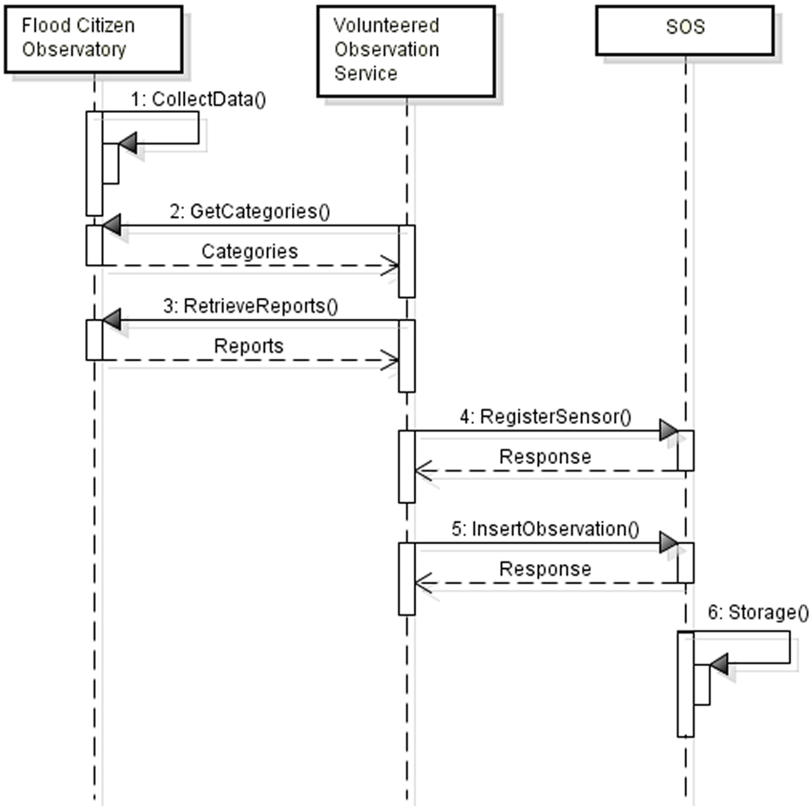


Figure 28 – Interaction with the Flood Citizen Observatory

while the volunteered observation service used Java language. The 52north framework[3] was used as an SOS implementation (JIRKA *et al.*, 2012). Both databases, sensor and spatial, used PostgreSQL as its database management system and had an extension to the geographic database (PostGIS). The Spatial Data Service was developed by means of Java language, together with the GeoServer[4].

Finally, the GeoDashboard was implemented on the decision-support layer with the aid of ExtJS[5] with OpenLayers[6] which are used for structuring the geospatial data and displaying the visual indicators. This dashboard displays both the source of the information on the map, the volunteered information collected by the Flood Citizen Observatory and the in-situ sensors. Furthermore, the expert model adopted in this project is called the "Hazard Index" (ROTAVA; MENDIONDO; SOUZA, 2013). This index represents vulnerability loss related to human stability in flood flows. It was defined based on the works proposed by Jonkman & Penning-Rowsell (2008) and HR-Wallingford (2006) that show how flood depth and flood velocity waters are combined for dynamical body equilibrium. This index is then related to human instability due to two physical mechanisms: moment instability (toppling) and friction instability (sliding).

3 <http://52north.org/>
4 http://geoserver.org/
5 <http://www.sencha.com/products/extjs/>
6 <http://openlayers.org/>

4.3.3 Deployment Analysis

The deployment analysis of AGORA-DS aims at gathering and classifying evidence of its efficiency in the real scenario of supporting flood risk management in São Carlos. This involves adopting two approaches: (1) drawing on evidence which shows the success of data integration and interoperability between WSN and VGI, and (2) using evidence to show the usefulness of this kind of integrated information for supporting flood risk management.

4.3.3.1 Interoperability and Data Integration

As mentioned previously, the interoperability between the data sources was achieved through SOS. This service is based on commonly agreed standards and has been employed in different contexts to facilitate the integration of data. The provided data — hereby named "Observation" — is stored following an appropriate data model (see in OGC (2012)). Within this data model, the table observation centres the all relevant information associated with a specific observation, i.e., its date, observed value, the procedure used, and *observedProperty*. The procedure refers to the source which provided the observation, and *observedProperty* could be water level or temperature. Therefore, an evidence of the interoperability between data sources and data integration might be given when their respective data are stored into this table.

Figure 29 evidenced the storage of some observations provided by WSN and VGI from the application of AGORA-DS in the table observation of SOS database. Due to the possibility of having distinct types of data, i.e., text or numeric, two different fields are used for observed value. This was especially useful for storing the complementary information provided by volunteers (e.g., "0.2 m").

	date text	observedValue text	observedValue numeric	procedure character varying(100)	observedProperty character varying(100)
1	05-16-2014 10:06:07		32.97	KARTODROMO 2	GAUGE HEIGHT
2	05-16-2014 10:07:00		0.00	USP 2	GAUGE HEIGHT
3	05-16-2014 10:07:00	A altura da água é de 0,30m.		FLOOD CITIZEN OBSERVATORY	níveldeáguacomrégua
4	05-16-2014 10:07:07		31.40	KARTODROMO 2	GAUGE HEIGHT
5	05-16-2014 10:08:00		0.00	USP 2	GAUGE HEIGHT
6	05-16-2014 10:08:07		32.97	KARTODROMO 2	GAUGE HEIGHT
7	05-16-2014 10:09:00		0.00	USP 2	GAUGE HEIGHT
8	05-16-2014 10:09:07		32.97	KARTODROMO 2	GAUGE HEIGHT
9	05-16-2014 10:09:41		21.04	USP	GAUGE HEIGHT
10	05-16-2014 10:10:00		0.00	USP 2	GAUGE HEIGHT
11	05-16-2014 10:10:00	KART 2 - 0.3m		FLOOD CITIZEN OBSERVATORY	níveldeáguacomrégua
12	05-16-2014 10:10:07		32.97	KARTODROMO 2	GAUGE HEIGHT
13	05-16-2014 10:10:44		0.00	KARTODROMO	GAUGE HEIGHT
14	05-16-2014 10:11:00		0.00	USP 2	GAUGE HEIGHT
15	05-16-2014 10:11:00	Kartódromo 1 : 25cm		FLOOD CITIZEN OBSERVATORY	níveldeáguabaixo
16	05-16-2014 10:11:00	0.2 m		FLOOD CITIZEN OBSERVATORY	níveldeáguacomrégua

Figure 29 – Observations stored in the observation table of the SOS database of AGORA-DS

Moreover, the *observedProperty* considered for each observation provided by the Flood Citizen Observatory corresponds with the categories defined in the platform (see Figure 28). While for the WSN, it depends on the measurement of the water level reported by the sensor, i.e., "GAUGE_HEIGHT".

Additionally, other evidence which illustrates that there is interoperability between WSN and VGI is the result of the *GetObservation* operation of the SOS. This is because this operation retrieves observations stored in the SOS in accordance with the O&M specifications (OGC, 2012). One observation from the application of AGORA-DS can be seen in Listing 1. The structured XML provides the same information about the observation as what is stored in the observation table for each observation. The observation displayed in Listing 1 is thus linked to those on line 12 of Figure 29. The information about the reports was conveyed by the comments on the XML, i.e., the date was May 16, the observed property was water level ruler (which in Portuguese is “níveldeáguacomrégua”), the procedure was “Flood Citizen Observation”, and the observed value reported was “the water level was 0.3m”.

Source code 1: Return of *GetObservation* operation

```

<om:ObservationCollection ... >
 <gml:boundedBy>...</gml:boundedBy>
   <om:member>
     <om:CategoryObservation ... >
       <om:samplingTime>
         <gml:TimeInstant>
           <!-- date -->
           <gml:timePosition>2014-05-16T10:10:00.000Z</gml:timePosition>
         </gml:TimeInstant>
       </om:samplingTime>
     <om:procedure xlink:href="urn:ogc:object:feature:
           Sensor:Ushahidi:observatóriocidadão"/>

       <!-- observedProperty -->
       <om:observedProperty xlink:href="urn:ogc:
           def:phenomenon:OGC:1.0.30:níveldeáguacomrégua"/>
         <om:featureOfInterest>
           <sa:SamplingPoint ... >
             <gml:description>NOT_SET</gml:description>

             <!-- procedure -->
             <gml:name>FLOOD CITIZEN OBSERVATORY</gml:name>
             <sa:sampledFeature ... />
             <sa:position ../>
           </sa:SamplingPoint>
         </om:featureOfInterest>
         <om:domainFeature ... />
           <!-- observedValue -->
           <om:result codeSpace="null">KART 2 - 0.3m</om:result>
     </om:CategoryObservation>
 </om:member>
</om:ObservationCollection>
```

Thus, both sources of evidence outlined above - the observations stored in the SOS database and the response of the *GetObservation* operation - highlight the interoperability between WSN and VGI and the integration of data provided by these data sources. It should be noted that AGORA-DS is fully compliant with the interoperable standards. These ensure the reuse of available data by other information systems, allow the inclusion of new data sources and provide easier access to data for the next layer (or potential data consumers).

Finally, Figure 30 also presents an evidence of the integration of heterogeneous data given by the Geodashboard. In the same view, the Geodashboard displays the numeric values about water flow (esp. the raw data and calculated hazard index) and a photo provided by the in-situ sensor, called "SHOPPING". Thus, this integrated visualization may give more information about the overall situation with the aim of supporting the analysis of decision-makers.

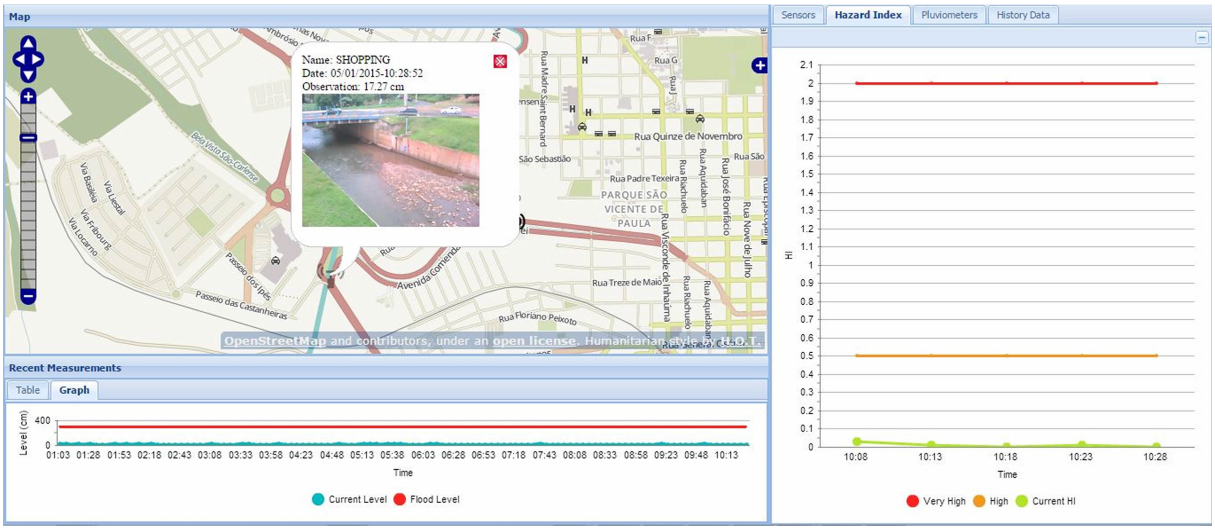


Figure 30 – Integrated visualization of heterogeneous data

4.3.3.2 Supporting Flood Risk Management: Lessons Learned

The use of integrated information gathered via WSN and VGI makes it possible to assess the data collected by the in-situ sensors and to provide data about poorly gauged or ungauged areas. In the case of data assessment, VGI supports the decision-makers' analysis to provide data from the same location as the in-situ sensors, e.g., the citizens reports that the water level at a specific location of riverbed is 50cm on the Water Level Ruler while the sensors state that it is 10cm. This might be because the sensor is broken, needs recalibration or requires a change of battery.

On May 16, 2014, this fact was witnessed in the study area. At 09:28 by using the Flood Citizen Observatory, a citizen reported that the water level at the specific location monitored by sensor node 4 – called USP 2 – measured 30cm on the water level ruler mechanism, although the visual indicator of the respective sensor was measuring 0 cm, as shown in Figure 31. After this was checked manually, we found that the sensor had become clogged with mud from the river.

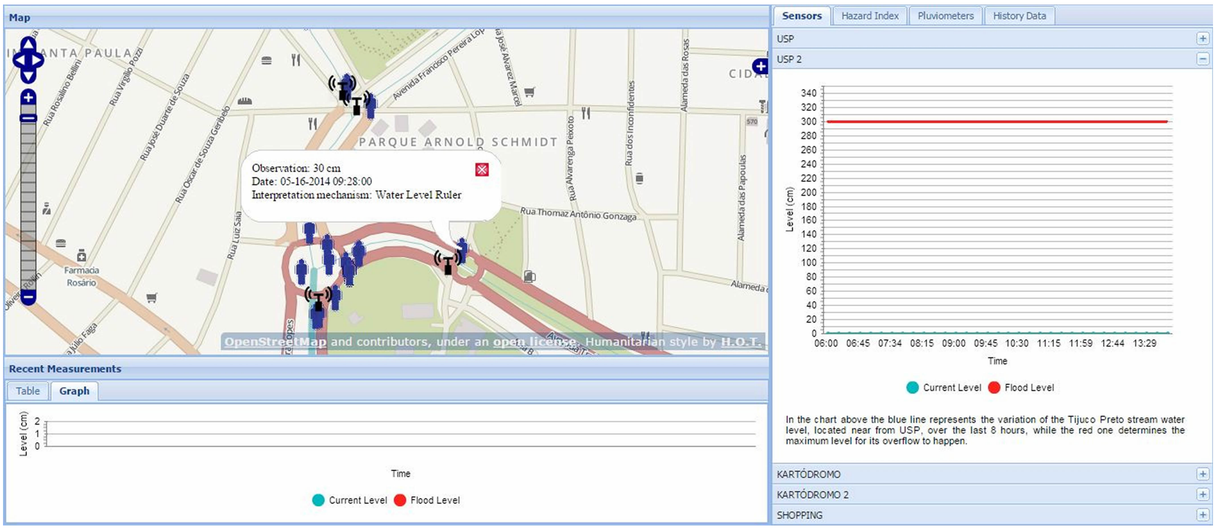


Figure 31 – Citizen's report used for assessing the sensor data

On the same day, citizens had reported the current situation of the riverbed by resorting to the Flood Citizen Observatory. Figure 32 illustrates an example of these reports on the Geodashboard that was provided by a volunteer. The general tag mechanism of the Flood Citizen Observatory was used by the volunteer to estimate the water level as "Normal". As can be seen from the chart on the right-hand side of Figure 32, this volunteered information is coherent with the data collected by the in-situ sensor called "Kartódromo 2", which lies closer to the area observed by the volunteer. This situation therefore shows that VGI might expand effectively the coverage of monitored areas.

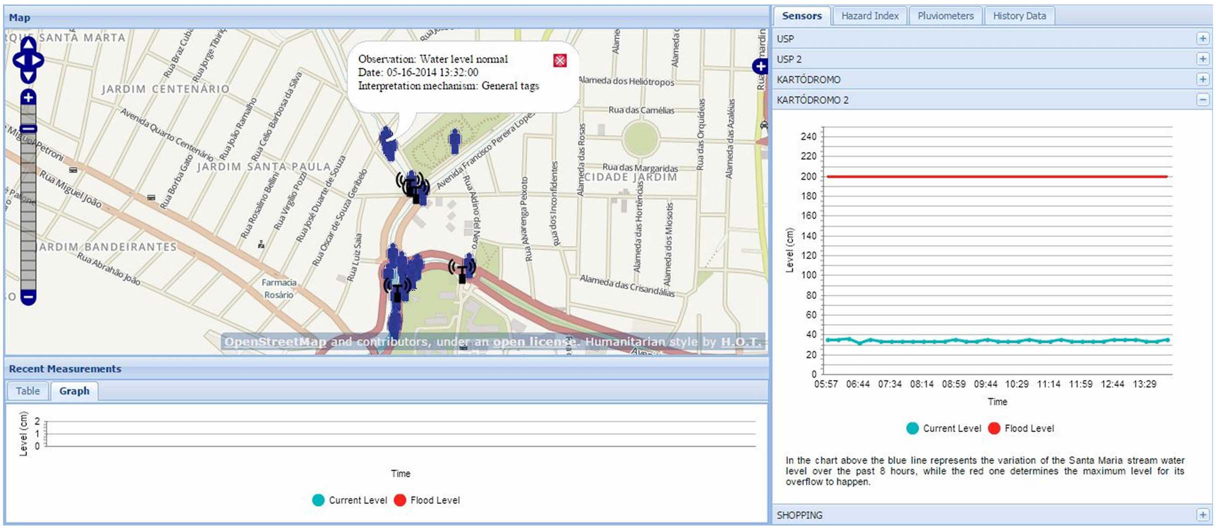


Figure 32 – Citizen's report from the ungauged area

4.4 Discussion

The deployment described in the previous section provides evidence of the effective use of AGORA-DS for combining heterogeneous data sources, and for supporting decision-making in flood risk management.

It can be seen that the integration of heterogeneous data sources provides more complete, accurate, and updated information about the situation in the affected areas. Although official data has provided useful information, they still need to be supplemented by other information to estimate the overall situation, e.g., it is difficult to determine the situation outside the area covered by in-situ sensors. These findings about the advantages of using of distinct data sources are in line with those of previous studies (SCHNEBELE; CERVONE; WATERS, 2014; WAN *et al.*, 2014). However, the approaches adopted by these studies only address the question of integration without complying with any interoperable standards. As a result, this might restrict the possibility of including additional data sources. In contrast, in our study we employed these standards so that the data sources could be included in a more flexible way.

In our approach, the integration of distinct data sources is facilitated by adapters responsible for gathering heterogeneous data and converting them to uniform data formats and interoperable standards, as defined by the integration layer. This is because most of the data sources have their own data structures and employ distinct formats (e.g., photo and numeric values) for the same observed property (e.g., the water level), and a different periodicity of data provision. Regarding the data structure, they can be well structured (e.g., water level provided by in-situ sensors), semi-structured when some complementary information is needed (e.g., categories of the citizen observatory), or unstructured when inference techniques are necessary (e.g., social media). Thereby, this chapter builds upon and extends previous works that use adapters for integrating geosensor data (BROERING *et al.*, 2010; BROERING; JIRKA; FOERSTER, 2010; KOGA; MEDEIROS, 2012). This is performed by using adapters for encapsulating the different data sources, converting the available data to standardized specifications (i.e., the O&M scheme), and then sending them to the appropriate web service (e.g., SOS). In this way, we were able to handle the discrepancies and idiossyncratic features in the data.

Additionally, the advantages of employing a dashboard to convey the integrated information to the decision-makers should also be considered within this project. Unlike information systems with various functions that may produce information overload, dashboards concentrate all the information needed on a single screen by means of understandable visual indicators. This can thus be regarded as an extension of previous studies (ZHANG; LI, 2005; MARKOVIC; STANIMIROVIC; STOIMENOV, 2009; MOLINA; BAYARRI, 2011) in so far as it enables the information to be analyzed and makes clear the amount of integrated information that is available.

Finally, the lessons learned from the deployment in a real scenario have confirmed that

AGORA-DS brings relevant contributions to flood risk management. This was made evident by the analysis of the integrated data provided by in-situ sensors and citizens' reports. This adds to existing experimental studies, which indicate that citizen observations of river levels can be considered statistically equivalent to data acquired from in-situ sensors (DEGROSSI *et al.*, 2014; MOREIRA; DEGROSSI; ALBUQUERQUE, 2015). However, to the best of our knowledge, this is the first study to evaluate the integration, visualization, and usefulness of heterogeneous data sources (i.e.,WSN and VGI) using a real-world deployment rather than simulated scenarios. Thus, the findings indicate that the approach described herein can effectively leverage volunteered information to complement data from sensors.

4.5 Final Remarks

This chapter has outlined AGORA-DS, an SDSS designed to support the decision-making of emergency agencies by combining WSN and VGI. It was implemented based on a conceptual architecture, which employs interoperability standards. Results of our deployment in a real scenario showed that AGORA-DS is able to combine information provided by WSN and VGI and then display it in a way that can be useful for flood risk management. This integration can support decision-making of emergency agencies by improving maintainability and assessment of WSN, as well as providing data from poorly gauged or ungauged area through VGI. They also show that well-known interoperable standards are appropriate for handling both this integration and interoperability of heterogeneous data sources. These standards ensure that available data can be used again; they also make it easier to include new data sources and obtain access to integrated data for potential data consumers.

Furthermore, it has become clear with results of the study that decision-making can be improved through a better understanding of which data sources are more appropriate for providing the information required by a decision. Therefore, the next chapter presents a model-based framework that describes the connection of decision-makers' tasks and decisions with data sources.

CHAPTER 5

A MODEL-BASED FRAMEWORK THAT CONNECTS DECISION-MAKING TO DATA SOURCES

5.1 Overview

With the widespread adoption of smartphones, social media platforms, and wearable technologies, there has been not only an increase in the amount of available data but also a proliferation of new data sources. All of these “big data” have the potential to transform the entire business process (GOPALKRISHNAN *et al.*, 2012; WAMBA *et al.*, 2015), as well as to provide greater support for decision-making in different contexts, such as business management and marketing (KOŚCIELNIAK; PUTO, 2015; YOU *et al.*, 2015). However, a remaining challenge lies on how to align decision-making within the organization with the data sources, e.g., how to determine, which data sources could provide useful information for assessing market trends? The reason for this challenge is that despite the fact that the available data could be of great value in supporting decision-making, they often fail to reach the decision-makers in a suitable way (VIEWEG; CASTILLO; IMRAN, 2014). As a result, decision-makers are supplied with useless information that still requires extended knowledge or experience for further data processing. This also makes it difficult to predict the impact that a change of data availability may have on specific tasks, since it is virtually impossible to find out if and where there is a lack of information.

In light of this challenge, the existing standards and notations from business modeling (i.e., BPMN and DMN) could potentially be used in this context. However, these fail to clearly represent the connection between the information required to make a decision and the data sources. To address this need, we have supplemented the existing standards and notations (i.e., BPMN and DMN) by conducting DSR to develop and evaluate the oDMN$^+$ Framework, a novel approach to connect decision-making to useful data sources. This framework consists of two essential features: (1) oDMN$^+$, a general, multi-layered model and notation for systematically mapping decision-makers' tasks and decisions to data sources, and (2) a modeling process for systematically employing oDMN$^+$ in practice, i.e., for modeling the specific features of particular decision-making scenarios.

This chapter is structured as follows. Section 5.2 describes the methodology employed in this research that forms the basis of the oDMN$^+$ Framework set out in Section 5.3. Section 5.4 describes the case study used for evaluating the framework and finally Section 5.5 discusses the lessons learned from this case study and contributions achieved in this research.

This chapter summarizes the main results of the paper "*Bridging the gap between decision-making and emerging big data sources: an application of a model-based framework to disaster management in Brazil*", published in the journal Decision Support Systems (HORITA *et al.*, 2017).

5.2 Research Design and Methods

This research project rests on the hypothesis that a model artifact can be used as an effective means of describing the connection of the decision-makers' tasks and decisions with the data sources. For this reason, we carried out a DSR method for designing and evaluating an artifact, as shown in Table 9. We understand an artifact following Hevner *et al.* (2004) as "a thing that can be transformed into a material (e.g., model) or process (e.g., method)". This method is a particularly effective means of developing and evaluating new artifacts that can solve identified organizational problems (HEVNER *et al.*, 2004). In this chapter, the artifact designed is the oDMN$^+$ Framework, which is composed of two features: (1) the oDMN$^+$ multi-layered model and notation, and (2) a modeling process. The Hevner's guidelines were adopted for ensuring the quality and rigor of our DSR method.

The left-hand side of Table 9 shows activities involved in the DSR method while the right-hand side displays their descriptions. It is worth mentioning that this chapter itself can be regarded as the "communication of research", since it fulfills Hevner's Guideline 7 - Communication of Research. The next sections provide more information about each activity in turn.

5.2.1 Activity 1: Research Problem

As explained earlier, the existing literature still lacks research studies concerned with creating a conceptual representation that interconnects all the elements necessary for establishing a link between the data sources and the decision-makers' tasks and decisions, i.e., deriving the information requirements from the decision-makers' tasks and decisions, and then linking these requirements to the emerging big data sources. The review of the state-of-the-art also provides evidence of a lack of research studies where the modeling of decisions and data sources are embedded in their modeling process; the existing works are restricted to the elicitation and

Table 9 – DSR method.

Activities	Description
Act 1. Research problem	The problem of this research lies in: (a) providing a modeling notation that is able to fill the gap between the decision-making and the data sources, and (b) creating a modeling process that supports analysts in this task. Thus, these problems lead to the following research question: How can the decision-makers' tasks be connected to emerging big data sources?
Act 2. Design as an artifact	The development of the oDMN$^+$ Framework, which is composed of two features: (1) oDMN$^+$, a multi-layered model and notation, and (2) a modeling process.
Act 3. Design evaluation	The connection of the decision-makers' tasks with data sources through a case study at Cemaden, as well as empirical evaluation of the generated model diagrams with members of Cemaden in focus group sessions.
Act 4. Research contributions	The contributions here are threefold: (1) oDMN$^+$, a multi-layered model and notation, (2) a modeling process, and (3) lessons learned from the employment of the framework in a case study.
Act 5. Research rigor	This work has theoretical foundations in well-known standard models from the area of Business Process Modeling (BPM), and techniques of qualitative analysis, tested analytical features, and empirical evaluation methods.
Act 6. Design as a search process	This research extends our previous work and thus represents the second cycle of improvements on the framework.

modeling of the business processes. The problem of this research thus lies in: (a) providing a modeling notation that is able to fill the gap between the decision-making and the data sources, and (b) creating a modeling process that supports analysts in this task. Thus, this problem leads to the following research question: How can the decision-makers' tasks and decisions be connected to emerging big data sources?

5.2.2 Activity 2: Design as an Artifact

The artifact designed in this research is a framework called oDMN$^+$ Framework which comprises two essential features: (1) the **oDMN$^+$**, a multi-layered model and notation for describing the connection between decision-makers' tasks and decisions with data sources; and (2) a **modeling process** that defines a set of activities for obtaining conceptual elements from decision-makers and employing the oDMN$^+$ in practice.

5.2.3 Activity 3: Design Evaluation

For evaluating the oDMN$^+$ Framework, we employed the designed framework for modeling a case study at Cemaden in Brazil. This scenario was chosen because, as mentioned before, it is a notable scenario of decision-making within a "big data" context, which is often characterized by dimensions such as volume, velocity, variety and veracity. Therefore, the linking between decision-making and data sources makes clear the relationship between activities and decisions and is able to point out the information and decision rules required by each decision.

Two activities were adopted for this evaluation. Firstly, the employment of oDMN$^+$ in the case study at Cemaden has produced a set of model diagrams including the business process diagrams, and a model with the decisions, information requirements and input data, together with the data sources. Hence, the model diagrams generated by our framework provide a first source of evidence about the effectiveness of the oDMN$^+$ Framework for connecting the decision-makers' tasks and decisions with data sources. Additionally, the generated model diagrams were also evaluated together with members of Cemaden. For doing so, we conducted focus group sessions with the aim of gathering feedback from practice on the effectiveness of the framework, as well as at identifying recommendations for improvements and/or gaining new insights (see Appendix B). The focus groups also provided evidence on whether the modeled business process, decision, and data sources were implemented into the context of a Brazilian early-warning center.

5.2.4 Activity 4: Research Contribution

The contribution of this research is threefold: (1) the oDMN$^+$, a multi-layered model and notation that describes the linking of the decision-makers' tasks and decisions with data sources, (2) a modeling process used for employing the oDMN$^+$ and putting it in practice, and (3) the lessons learned from employing the framework in a case study at Cemaden. The results obtained in this research also contribute to practice by providing a better understanding of the linking between decision-making with data sources in a Brazilian early-warning center.

5.2.5 Activity 5: Research Rigor

Scientific rigor in design research entails the effective use of theoretical foundations and appropriate research methodologies (HEVNER *et al.*, 2004). The use of standard models and notation is widespread in the area of business process modeling, e.g., BPMN has been supporting the modeling of different business processes from procurement companies to transport and management in humanitarian logistics. Additionally, DMN bridged the gap between the modeling of decisions and their application to tasks involving business processes. In this research, both BPMN and DMN were the main theoretical foundations in the development of the multi-layered model and notation of the oDMN$^+$ Framework. Afterwards, these were supplemented by an extension of the O&M, which provides an interface between the observations provided by data sources and the input data of DMN. Altogether these standards comprise the background of the multi-layered model and notation of the oDMN$^+$ Framework.

Regarding the methods used for developing this research, the UML was used for designing the oDMN$^+$ metamodel because it is a well-known and widespread standard language in both practice and research. Furthermore, the case study was performed in a real-case scenario following a protocol that was elaborated together with practitioners and researchers from Cemaden with solid experience in empirical research. The analysis of the case study employed the coding technique devised by Saldaña (2015) and was based on the analytical features proposed by Horita *et al.* (2016). The guidelines proposed by Peterson (2000) were employed in the construction of the instrument for the questionnaire survey, which was the basis for the evaluation of the generated model diagrams in the case study. Finally, the focus group sessions were conducted, guided, and facilitated by a moderator that followed a conduction protocol that was elaborated in collaboration with practitioners. These sessions were particularly important because they allowed in-depth discussions among all the participants, increasing the richness of the information gained (RUNESON; HÖST, 2008).

5.2.6 Activity 6: Design as a Search Process

In our previous work, we designed an integrated model and notation for connecting decision-making with emerging data sources (HORITA *et al.*, 2016). An application in the 2015 Nepal Earthquake not only provided evidence of the effectiveness of this integrated model and notation but also showed some points that required improvements. Taking these points into account, the current research is an extension of the previous model that also includes an evaluation carried out in a different setting. As a result, the previous model was improved and a new artifact was designed.

5.3 oDMN$^+$ Framework: a Framework for connecting Decision-making with Data Sources

This section introduces the oDMN$^+$ Framework by outlining its essential features: the oDMN$^+$ and the modeling process.

5.3.1 oDMN$^+$: a Multi-layered Model and Notation

Since oDMN$^+$ aims at describing the connection between the decision-makers' tasks and decisions with data sources, its main features have been separated in accordance with their functional objectives. This resulted in a layer-based structure consisting of three layers (Figure 33): (1) the business process layer, (2) the decision layer, and (3) the data layer.

To start with, the business process is defined in the business process layer. The modeling of these elements can be carried out using BPMN. After the business process modeling is complete, all the activities linked to a decision are modeled in the decision layer using DMN. In the data layer, the data sources used for providing the required information for a decision are modeled using an extended version of the O&M (OGC, 2013). Figure 33 displays the relationship among these layers and the employed modeling languages.

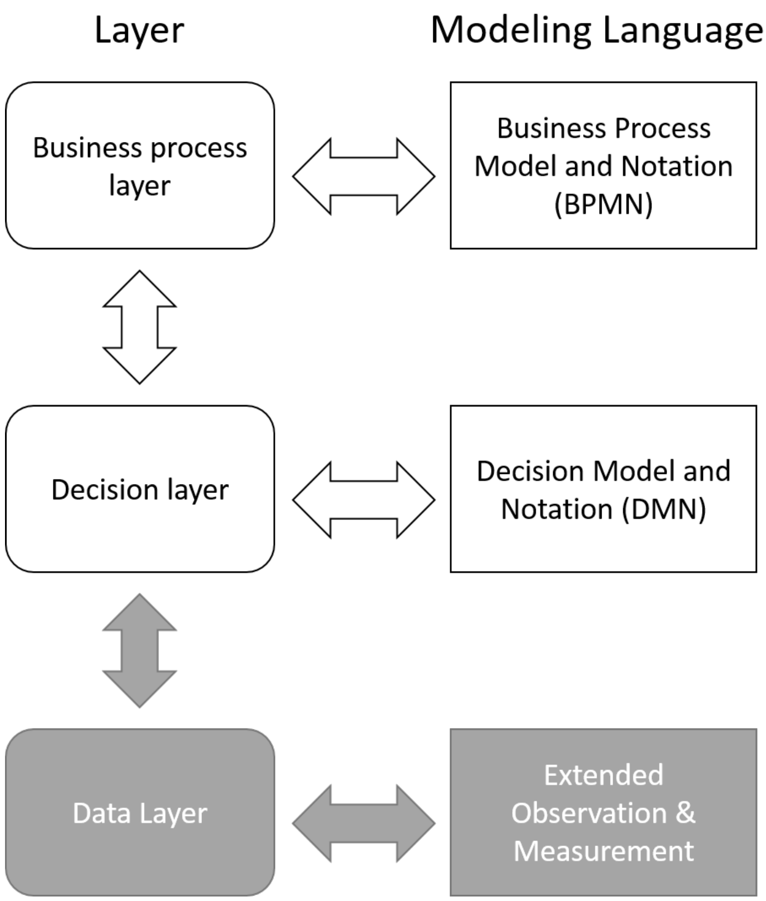
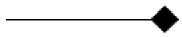


Figure 33 – Conceptual architecture of oDMN$^+$. *In the figure, elements evidenced in gray are those proposed in this research, while the others already exist in the literature.*

Just as the model notations BPMN and DMN aim at providing reusability, common understanding, and easy information sharing among the different stakeholders, oDMN$^+$ supplements these notations by including a graphical representation of the connection of decision layer with a new data layer, as shown in Table 10.

In order to give a first idea of the practical use of oDMN$^+$, Figure 34 displays an example that makes use of the model to represent the decision-making involved in the request of a new product. This decision consists of the selection of a supplier for buying a specific product and takes into account the reputation of the supplier and its price for the product. On the top of Figure 34, there is a business process and its tasks that are modeled using the BPMN notation. Within this, a specific task ("Define the appropriate supplier") involves a decision ("Appropriate supplier"), which is modeled on the bottom using the DMN notation and Extended O&M

Table 10 – New graphical representation included by oDMN$^+$.

Elements	Description
 Data Source	Represents the data source that may provide an input data (or: observation)
 Associated Data Source	Indicates the associated data source to an input data (or: observation)

notation. The decision requires an item of information (“Product data”), as well as the outcome of another decision (“reputation of supplier”). It also employs a decision table (“suitable supplier rules”) for processing the required items of information. The value of these items is thus provided by an external data source (“Supplier report”).

With regard to the modeling of decisions, DMN separates the components into two categories: the decision requirement (the shapes in white) and decision logic (the shapes in light gray), Extended O&M is part of the first category (decision requirement). By means of the decision logic level, the required components of a decision are defined at the topmost table on the right-hand side (“Appropriate supplier”). The items of information provided by the input data and their values are defined in the middle table (“Product data”) while the decision rules of the business knowledge are determined at the table on the left-hand side (“Appropriate supplier rules”). Finally, the observations provided by the data source are shown in the table at the bottom. The data source combines observations with properties, which in turn are related to the properties defined for the decision.

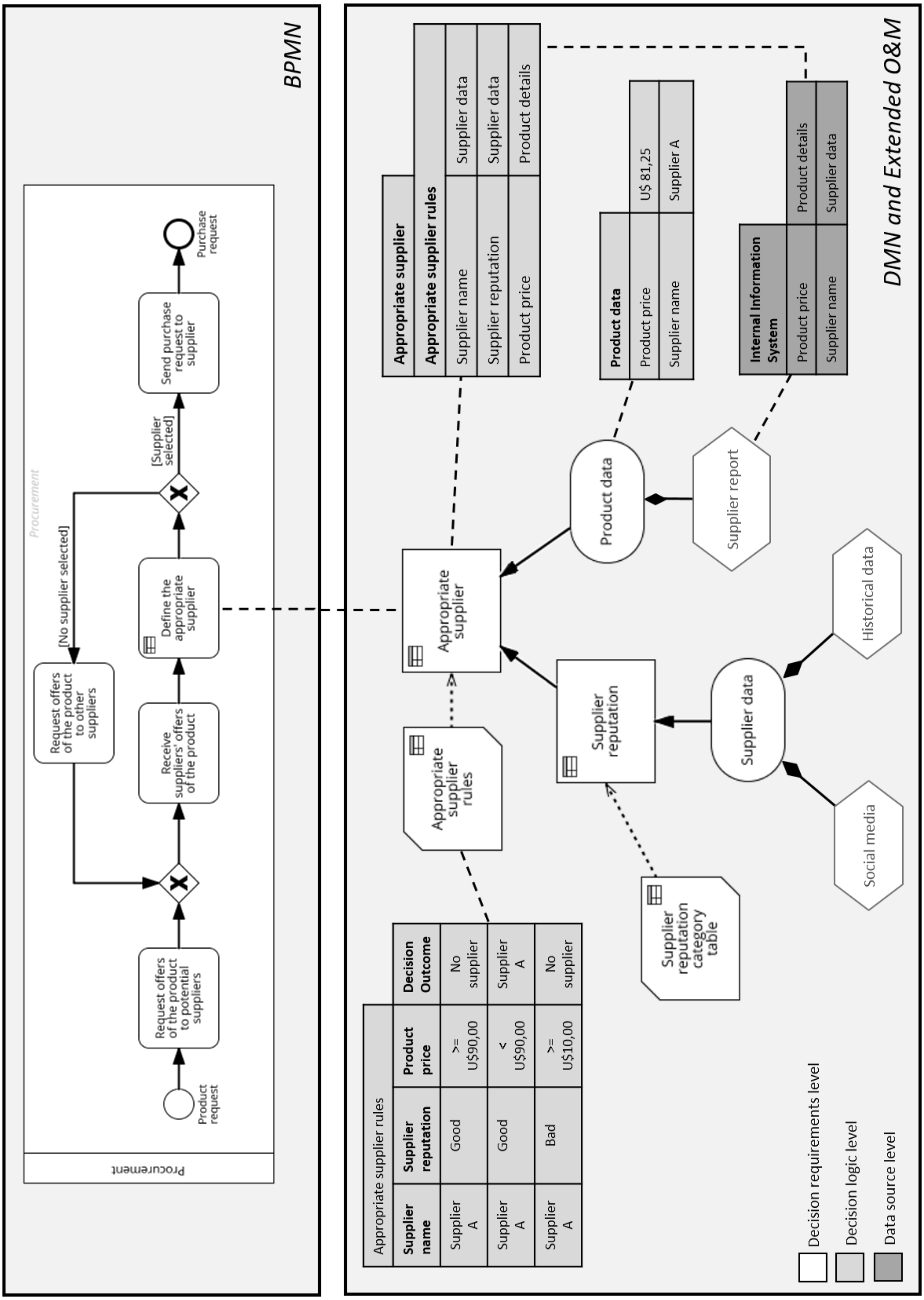


Figure 34 – Example of usage.

The graphical representation is followed by the development of a metamodel that describes its elements and their relationships. Therefore, the oDMN$^+$ also includes a metamodel in UML for describing the connection between the decision-makers' tasks and decisions and the data sources, as shown in Figure 35. In conform to the conceptual architecture (Figure 33), the elements of the oDMN$^+$ metamodel were separated in the three layers: business process, decision, and data. All new elements proposed in this research are highlighted in gray, i.e., the *AssociatedDataSource*, *DataSource*, *TypeOfDataSource*, and the arrows connecting the decision and data layers. The other elements, depicted in white, were proposed before as part of BPMN, DMN, and O&M. For simplification, only the elements of BPMN and DMN that are related to the oDMN$^+$ were presented in Figure 35.

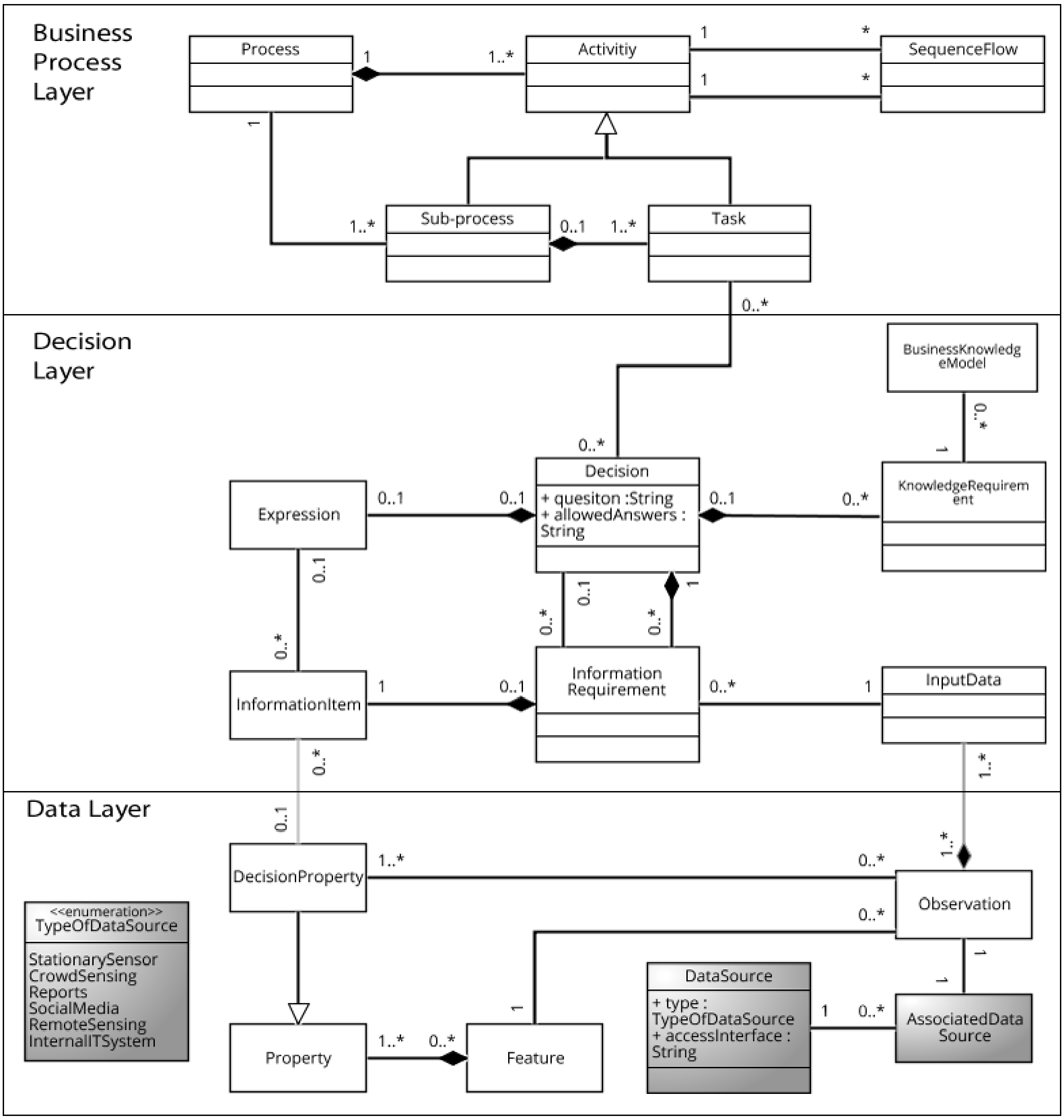


Figure 35 – oDMN$^+$ metamodel.

5.3.2 The Modeling Process

The practical use of oDMN^{+} consists of identifying conceptual elements of a particular application context and translating them into a model diagram. To support this, we propose a modeling process (Figure 36), which defines a systematic sequence of five activities for employing oDMN^{+}.

The aim of Activity 1, **plan data collection**, is to produce a protocol that defines the method for data collection and the analytical features for guiding the modeling process in practice. The method for data collection provides the basis for gathering the data in this context together with its decision-makers. The analytical features are key factor for determining the conceptual elements of the application context (e.g., the activities, the decisions, or sequence flows) (HORITA *et al.*, 2016). Because of this, they are later used as a basis for the definition of the coding scheme. Activity 2, **execute fieldwork**, implements the previously defined instruments of data collection, e.g., conducting interviews with decision-makers within an application context.

After the fieldwork, the collected data are used in Activity 3, **data analysis and classification**, which consists of employing a coding technique for supporting the analysis and classification of the data. Coding is an analytical technique employed for extracting quantitative data from qualitative datasets, such as the transcriptions of interviews (SALDAÑA, 2015). In the first place, the analytical features defined in the data collection protocol are used for deriving the coding categories, which will comprise the overall coding scheme. This scheme is employed in the analysis of each dataset of the collected data. In other words, when the modeling analyst finds a reference to a particular category during the analysis of a dataset, he assigns it to that particular category. If a reference does not fit any available category, the modeling analyst can either create a new category or refine one that already exists. In this way, the coding scheme is refined and improved so that it only represents the relevant categories. As a result, all the coded data will be used as input for Activity 4, **model elements**.

Based on the coded data, the modeling of elements is achieved by four tasks, each of which is responsible for producing a particular model diagram. The business process modeling should be first defined. This comprehend the analysis of all the references assigned to the "Activity" and "Sequence Flow" categories of the coding scheme. After this, the data sources that supply useful data for decision-making should be identified from the analysis of the "Data sources" category. Having obtained the data sources, a catalogue of the information requirements is compiled after reviewing all the references to the "Information Requirement" category. In the last task, the decisions are first listed by conducting an analysis of the activities. After this, both the business knowledge and its input data are planned in a way that provides a solid background for the decision-making. Altogether, these will be used in the final activity for interconnecting the modeled elements.

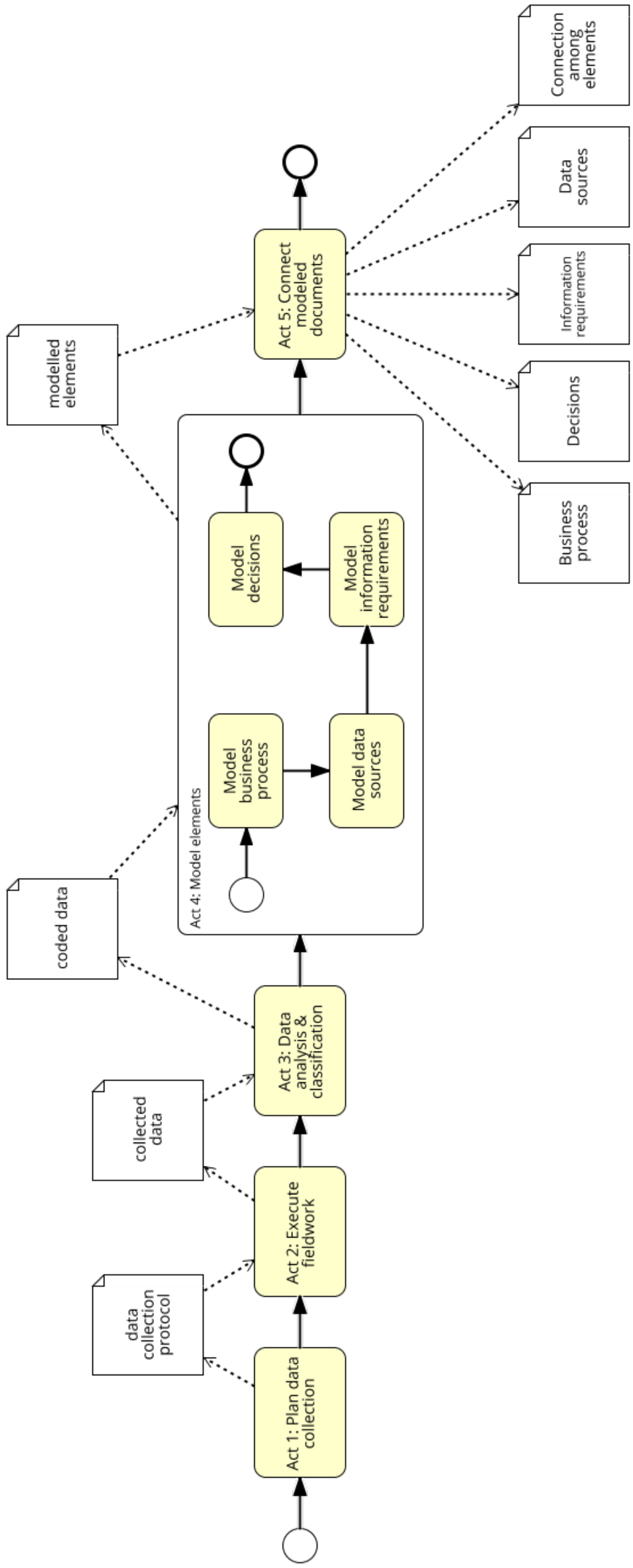


Figure 36 – Modeling process.

The final aim of Activity 5, **connect modeled elements**, is to interconnect all the modeled elements that were found in the previous activity. This means attributing the activities to the appropriate decisions, and then linking these decisions to their required data, by defining the required information, which this required data belongs to, and ultimately determining the data sources that might provide the required information. This activity then summarizes all the modeled elements in specific model diagrams. The next section presents the evaluation of the oDMN$^+$ Framework at Cemaden.

5.4 Case Study

This section gives more information about how the oDMN$^+$ Framework is employed in the case study at Cemaden.

5.4.1 The Cemaden

In Brazil, preventive countermeasures have been taken to mitigate loss and damage, as well as to improve the coping strategies of communities against floods, droughts, and landslides. One of these countermeasures was to set up Cemaden (in 2011), which is a branch of the Brazilian Ministry of Science, Technology, Innovation, and Communications (MCTIC, in Portuguese). Among Cemaden's activities, the development, implementation and operation of monitoring systems for natural disasters play a crucial role in the issuing of warnings of imminent natural disasters to the National Center for Disaster Risk Management (CENAD, in Portuguese), and thus support disaster management in Brazil.

The current monitoring system of Cemaden consists of a large-scale sensor network, which provides data about precipitation, and describes the type of weather (e.g., rain) for different locations. This sensor network has been supplemented with different crowdsourcing projects with a view to increasing the volume, type, and quality of the available data. Among these projects, the "Pluviômetros nas Comunidades" is a project [1] that seeks to involve vulnerable communities in the monitoring of environmental variables. This means that people collect data from semi-automated rainfall gauges installed in the region of these communities and share them with the center via a web-based system. Around 1,000 of these gauges have been installed in different communities in Brazil. Furthermore, Cemaden works in collaboration with several institutions such as the National Water Agency (ANA, in Portuguese), the Brazilian Geological Survey (CPRM, in Portuguese), and the National Institute of Meteorology (INMET, in Portuguese). These provide further data about weather conditions, risk mapping, and environmental variables, which thus supplement the existing data of the center. All these data are displayed on a video wall located in a monitoring control room that is used by members of Cemaden to assist them in decision-making (Figure 37), e.g., identifying areas requiring attention, deciding on whether to

[1] http://www.cemaden.gov.br/pluviometros/

issue a warning, and defining the level of a warning. These members are divided into different monitoring teams that work around the clock in six-hour shifts. These teams comprise five to seven members, and include at least one specialist in each of the following areas of expertise: hydrology, meteorology, geology, and disaster management. In addition to the video wall, each member has an individual working station where they can analyze particular information on their own, e.g., a geologist may want to analyze data provided by geological agencies (e.g., CPRM), while a hydrologist is more interested in data from water resource agencies (e.g., ANA).

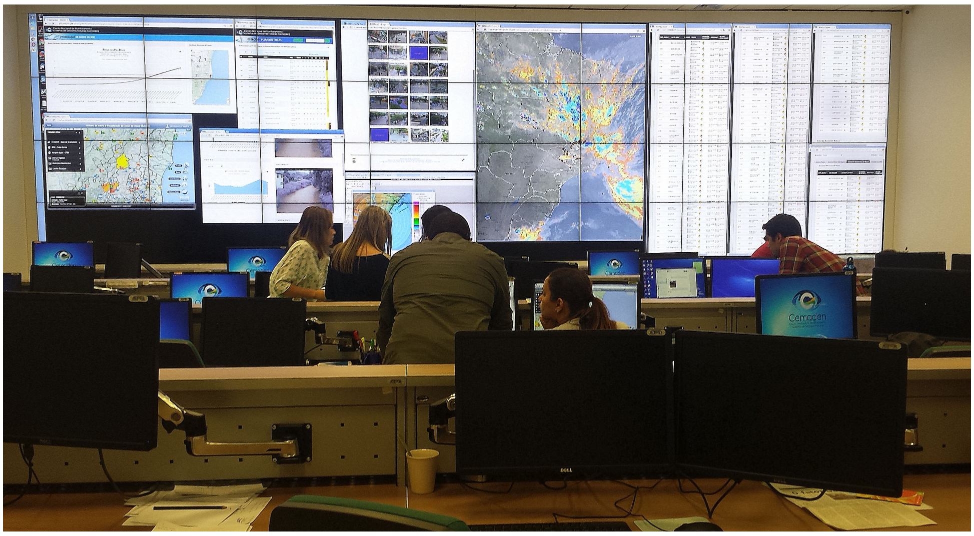


Figure 37 – The monitoring control room of Cemaden.

However, all these data raise further challenges for the work in the monitoring room of Cemaden. To start with, there are structural issues regarding the data that is supplied; for example, the interoperability of data sources is hampered by the fact most of them do not comply with any common standard. Furthermore, the real-time data processing and integration are complex because the data comes in different formats (e.g., photos and numeric values) and at different velocity/update rates (e.g., once an hour or every 5 minutes). The decision-making of the monitoring teams does not follow a strictly-defined business pattern. This makes decision-making more empirical and based on tacit knowledge, as well as making it difficult to understand a) the information which is required, b) the appropriate data sources, c) the tasks that must be performed, and d) the corresponding decision-makers. These challenges thus make the decision-making carried out in the monitoring room of Cemaden a largely ad hoc process.

5.4.2 The Modeling Process at Cemaden

In Activity 1, as a part of the data collection protocol (see Appendix A), we defined how participatory observation and interviews with decision-makers could be employed in this case. These were chosen because, participatory observation allows data to be obtained during the decision-makers' daily activities while, interviews make it possible to find out the personal perspectives of the decision-makers.

In Activity 2, the data was collected by one of the authors on January 19th-22nd, 2016 and February 1st, 2016 at the Cemaden's headquarters in São José dos Campos, Brazil. In these periods, 88 warnings were issued from the monitoring room to vulnerable communities.Participatory observation, with a limited degree of interaction by the researcher, was first carried out as an instrument for gathering data about the day-to-day activities and interactions of the subjects. This meant that the observer was only regarded as a researcher, and did not interact with the subject or interfere with the subjects' activities (RUNESON; HÖST, 2008).

Additionally, individual and face-to-face interviews were also conducted to obtain data about the individual business activities of the subjects (e.g., activities they carried out or data sources that were used). This included 10 semi-structured interviews with members of the monitoring room all of whom completed a questionnaire that was prepared in the data collection protocol. Since the interviewees were working within strict time constraints, the interviews took no more than 30 minutes, while the participatory observations lasted for two hours, as summarized in Table 11.

Table 11 – Summary of the methods of data collection.

<table>
<tr><th>#</th><th>Method</th><th>Subjects</th><th>Period (min)</th><th>Objective</th></tr>
<tr><td>1</td><td>Direct observation</td><td>-</td><td>120</td><td>Analyze activities in the monitoring room</td></tr>
<tr><td>2</td><td>Direct observation</td><td>-</td><td>60</td><td>Analyze the shift of a monitoring team and the use of available systems</td></tr>
<tr><td>3</td><td>Direct observation</td><td>-</td><td>150</td><td>Analyze the communication in the monitoring room</td></tr>
<tr><td>4</td><td>Interview</td><td>2 Meteorologist</td><td>35/45</td><td rowspan="4">Collect data of the subject's daily activities</td></tr>
<tr><td>5</td><td>Interview</td><td>2 Hydrologist</td><td>21/46</td></tr>
<tr><td>6</td><td>Interview</td><td>2 Geologist</td><td>30/33</td></tr>
<tr><td>7</td><td>Interview</td><td>4 Disaster Mgmt Spec</td><td>48/22/33/30</td></tr>
</table>

After the data had been collected, we started on Activity 3. The interviews were then transcribed verbatim and a coding scheme was adopted for the analysis and classification that was based on the analytical features proposed by Horita *et al.* (2016). This analysis was supported by the NVivo Data Analysis Software[2]. An in-depth analysis of the content defined by each coding category was then undertaken. For example, the coding category "Open Warning" and "Execute Group Meeting" - sub-elements of the coding category "Activity" – became activities of the business process while the coding category "Satellite" - sub-element of the coding category "Stationary Sensors" – was turned into a data source. In the end, the coded data were generated.

[2] http://www.qsrinternational.com/what-is-nvivo

The coded data were used in Activity 4 for modeling the conceptual elements of the case study, i.e., we modeled the business process of the monitoring room, its derived decisions, information requirements, and data sources. This activity was supported by the Signavio Modeling Platform[3]. Finally, we concluded with Activity 5 in which we decided to model the connection between the elements by means of a general diagram editor, since the Signavio Platform does not provide a means of including the customized features.

5.4.3 The Modeling Results of Cemaden

This section conducts an analysis of the model diagrams generated by this application of the oDMN$^+$ Framework.

5.4.3.1 Modeled Business Process

With the aid of the coded data, we began with the modeling of the business process using BPMN. Figure 38 displays a compressed version of the business process diagram generated, which represents the process of issuing a warning that is executed inside the monitoring room of Cemaden - an expanded version is displayed in Figure 40.

The business process starts when there is a new shift of a monitoring team, which takes place every 6 hours starting at midnight. In the first activity, the meteorologists of the team are responsible for analyzing the weather conditions in the country for the next 12 hours and communicating all the areas under surveillance, i.e., those areas that are susceptible to an adverse situation such as a forecast of heavy rainfall. After these areas of surveillance have been determined, the conditions of the rivers and mountains are continuously analyzed so as to be able to predict events involving landslides or floods. When a potential disaster has been detected, the whole team will gather together for further discussions about the potential event (e.g., if the hydrologist is not absolutely certain about the occurrence of the event). In case the available data is not enough for determining the event and the warning cannot be opened, further data have to be gathered. Otherwise, the warning can be opened with all the data so that evidence can be provided to forecast the potential event. Finally, the disaster management specialist revises the warning and communicates it to CENAD.

[3] http://www.signavio.com/

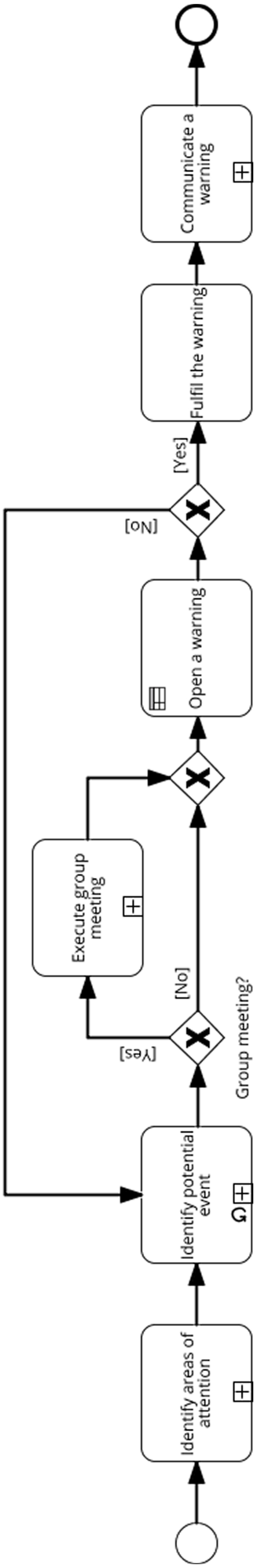


Figure 38 – Compressed version of the business process of issuing a warning.

5.4.3.2 Modeled Decisions

It should be mentioned that all the tasks based on a decision, are highlighted in the business process, e.g., the “Open a warning” task marked on the top-left of the rectangles (see Figure 38). Following this, this decision is expressed in greater detail by showing the connection between business knowledge, required information, and data sources, as shown in Figure 39.

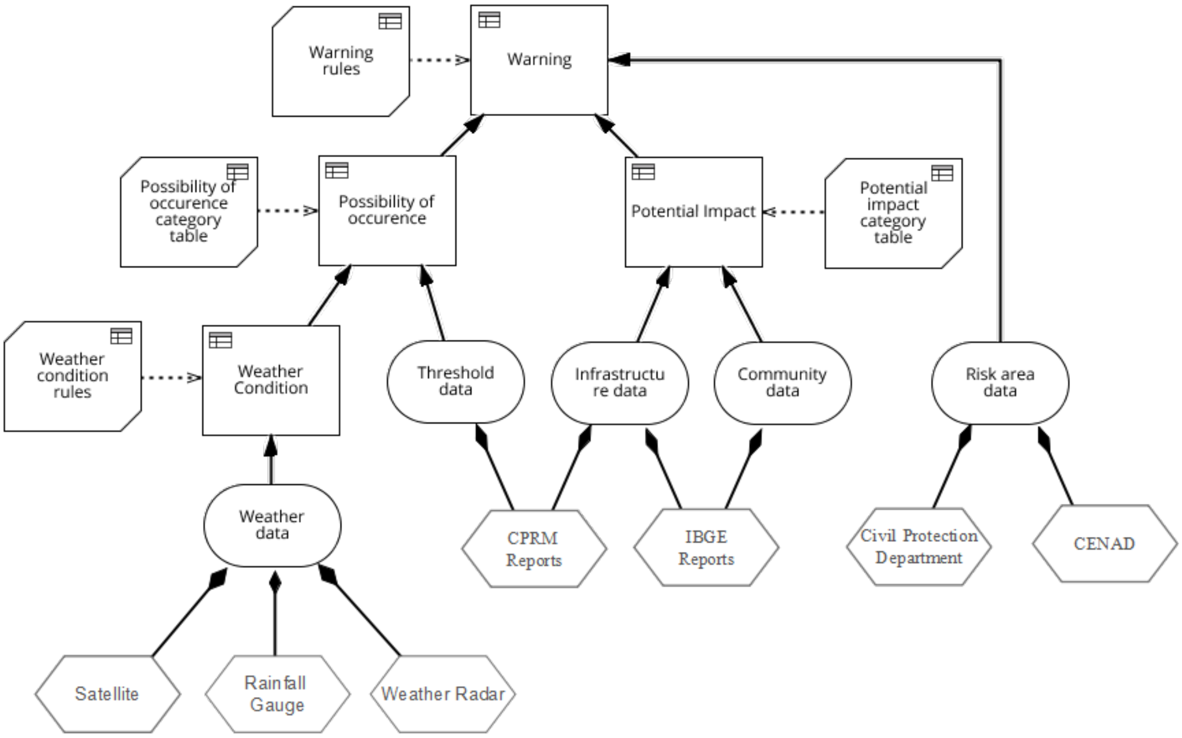


Figure 39 – Decision of opening a warning.

The decision requirements level of the DMN and Extended O&M were used for modeling the way these elements were interconnected. The “Warning” decision employs a “business knowledge” (Warning rules and regulations) and requires information from two other two decisions (“Possibility of Occurrence” and “Potential Impact”). These decisions require a particular business knowledge (“Possibility of occurrence” category table and “Potential impact” category table) for processing the required information, whereas the “Possibility of Occurrence” also requires information from another decision (“Weather Condition”), which employs further business knowledge (“Weather condition” rules). The Weather data, Threshold data, Infrastructure data, Community data and Risk data have been defined as the input data for these decisions. These input data, in turn, can be provided by different data sources. For example, satellites, rainfall gauges, and weather radar are suitable data sources for supplying the required information about weather conditions. Similarly, data about thresholds and the infrastructural facilities of an area can be gathered from the CPRM reports. By means of the DMN’s decision logic level, it is also possible to model more detailed information about the required business knowledge, as well as the information requirements, as shown in Figure 41.

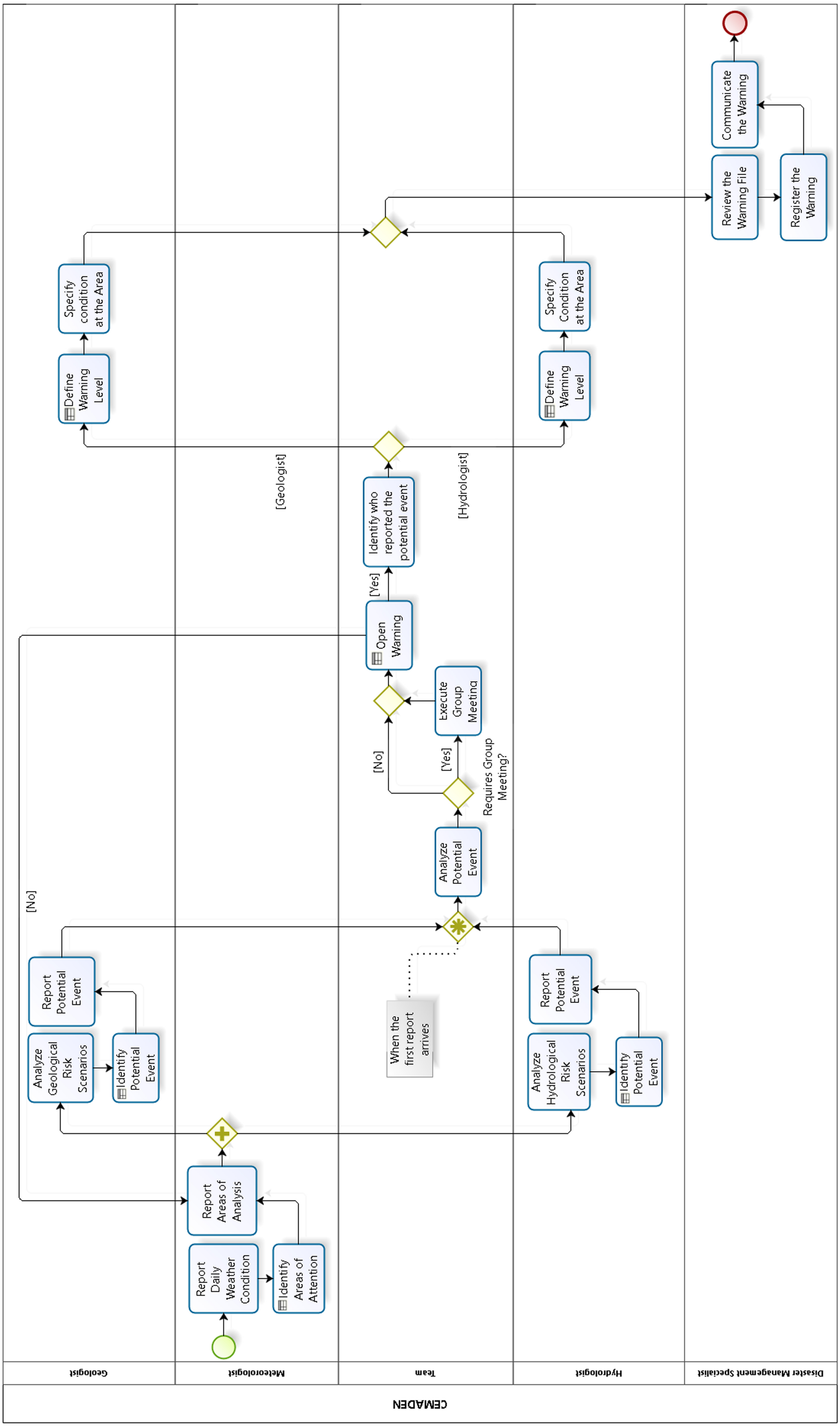


Figure 40 – Expanded version of the business process of issuing a warning.

Potential Impact	
Potential impact category table	
Infrastructure of buildings	Condition of buildings
Estimated number of vulnerable inhabitants	Population density

Figure 41 – Data of the decision “Potential Impact”.

Figure 41 displays the data for the “Potential Impact” decision that employs a business knowledge (“Potential impact” category table) and requires two items of information (the infrastructure, i.e., buildings and houses, and the estimated number of vulnerable inhabitants). The definition of the potential impact is achieved by a cross-comparison of the items of information, as displayed in Figure 42. For example, when the number of buildings is unsuitable for a particular area and its estimated number of vulnerable inhabitants is very high, the potential impact on the area is very high.

	Potential Impact		
	Infrastructure of buildings	Estimated number of vulnerable inhabitants	Potential Impact
	Inadequate, Adequate	*Low, Moderate, High*	
1	Inadequate	Low	Moderate
2	Inadequate	Moderate	High
3	Inadequate	High	Very High
4	Adequate	Low	Regular
5	Adequate	Moderate	Moderate
6	Adequate	High	High

Figure 42 – Business knowledge of the decision “Potential Impact”.

The cross-comparison of items of information takes into account their specific values in a particular area, e.g., the infrastructure with regard to buildings is unsuitable at the city. These items of information are some of the items that comprise the information requirements catalogue. This link between decisions about items of information and their values are defined by the input data. Figure 43a shows the value of the item of information “Situation of buildings” as unsuitable while Figure 43b shows that the estimated number of vulnerable inhabitants is “moderate”.

Since the items of information are provided by specific data sources, such as official reports, Figure 44 outlines data sources that include their items of information. For example, reports from CPRM may provide information about the location of the houses and estimated number of vulnerable inhabitants in the area. Furthermore, these items of information are also linked to decision properties like population density, the condition of the buildings, and the rainfall threshold. This association thus highlights the connection between the information requirements of decisions (“Infrastructure with regard to buildings”) and items of information (“Location of houses”) provided by the data sources (CPRM Reports).

Infrastructure of buildings	
Situation of buildings	*Inadequate*

(a) Infrastructure of buildings.

Population density	
Estimated number of vulnerable inhabitants	*500*

(b) Population density.

Figure 43 – Input data of the decision "Potential Impact".

CPRM Reports	
Situation of houses at the area	Condition of the buildings and houses
Threshold	Rainfall Threshold

(a) CPRM Reports.

IBGE Reports	
Situation of houses at the area	Condition of the buildings and houses
Estimated number of vulnerable inhabitants	Population density

(b) IBGE Reports.

Figure 44 – Data sources.

The interconnection of all the modeled elements for the "Open a warning" decision is displayed in Figure 45. After the employment of the oDMN$^+$ Framework, the generated model diagrams were evaluated at Cemaden and their results are presented in the next section.

5.4.4 Evaluation of the Generated Model Diagrams at Cemaden

Two focus group sessions were held on August 23rd, 2016 at Cemaden (see Appendix B). The analysis of the collected data revealed that all the participants had concerns about their business processes and decision-making, mainly because they lacked an official "action protocol". In view of this, it was stated that the business process diagrams we achieved should be turned into the official "action protocol" for all the monitoring teams, since they defined the essential tasks carried out in the monitoring room. This is indeed a crucial issue as was constantly being mentioned; for example, one of the participants pointed out that *"it is a huge problem here that we do not have a business process because you do not have a clear description of the boundaries or of your activities." (Disaster management specialist B)*. It is even more important since decision-makers often remarked that they feel more confident when they are following a standard response protocol for action. This provides evidence that our approach is useful both for standardizing the decision-making of all the monitoring teams and for making decision-makers more confident about their decisions.

This statement is in line with another area of discussion. The participants supported the idea that the decision-making could be speeded up by the generated model diagrams, since they would serve as guidelines for taking actions. The reasons are that *"decision-makers will be trained to know how the existing processes and decisions should be carried out, and thus*

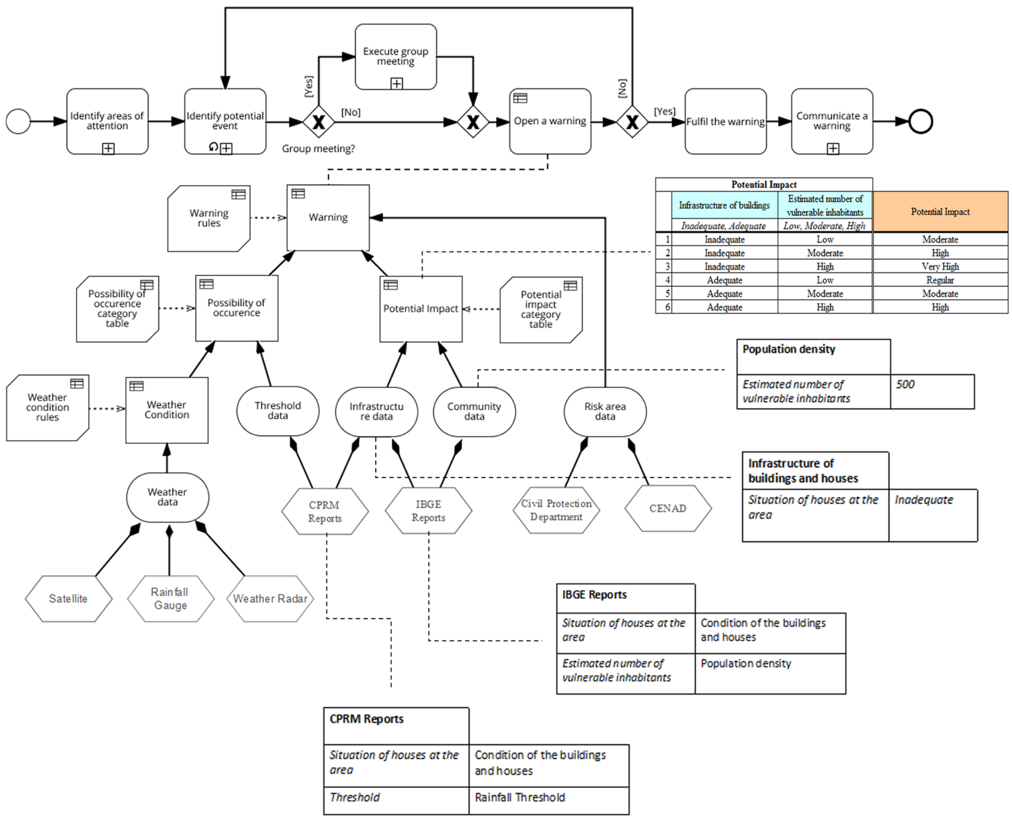


Figure 45 – The interconnection of all the elements for the “Open a warning” decision.

be prepared for making better decisions or even improving them” (Geologist B). In this way, they will know what their activities are and what information they have to look for. However, it was noted that these model diagrams should provide guidelines for action but not constrain the activities of the monitoring team owing to the intrinsic dynamics of disaster management.

This is in line with another argument that was raised during the sessions. According to the participants, the status of a warning could fall into one of the following phases: (a) Analysis; (b) Opened; (c) Kept; (d) Ceased; and, (e) Under Review. When a warning is “kept”, it can either remain at the same warning level (e.g., moderate), increase (e.g., go from moderate to high), or decrease (e.g., go from very high to moderate). The participants then stated that a definition of business processes for each of these phases would be of great value to the monitoring team. It would mean that further activities could be detected; for example, in the “In Review” phase, the warning could be analyzed in terms of its quality, while in the “Kept” phase, a disaster management specialist could investigate the occurrence of events reported by the media in the affected area. In this case, new required information and data sources for decision-making would be made available.

The identification of emerging data sources that would be useful for decision-making was another important result of the generated model diagrams. Existing sensors still require supplementary data to enable them to estimate the overall situation in a vulnerable area, mainly because of the limited type of data that these sensors share and the restricted coverage of the monitored area. As a result, when asked about the modeled data sources, the participants stated that the information provided by ordinary citizens is of particular value for understanding the situation in their area. All this information can support the analysis of a warning that is examined when it is opened in the monitoring room, as one participant pointed out:

> "*This information [from social media] is more useful for the monitoring of a warning in a place where it was opened." (Disaster management specialist A).*"

However, social media messages "*should not be determining factors that directly influence decision-making*" (Geologist A), since there are serious concerns about their quality and reliability. In contrast, messages sent by official organizations such as civil protection services or the police department, are useful because they can report on current incidents in the area, e.g., which areas are more damaged or what are the evacuation areas. Following this discussion, one participant mentioned the possibility of using the National Information System of Disasters (S2ID) as a useful data source for providing historical data about disasters. This would support the activities of the disaster management specialists, especially those that are concerned with assessing the degree of vulnerability of an area. Hence, the analysis of the generated model diagrams has enabled practitioners to find alternative data sources that can be adopted to meet the requirements of the decision-makers. The decision-makers were also able to establish which tasks require further information to ensure better decision-making (e.g., the use of the historical data of S2ID). Furthermore, owing to the intrinsic dynamics of this study, finding the most useful data sources for particular decisions plays an important role in providing more accurate information as well as for speeding up the decision-making.

Another statement often made by the participants was the need to continuously update the reference values of the information parameters that comprise a business rule per location. This can be attributed to the intrinsic changes of the reference variables, which affect decision-making over a period of time, e.g., the number of inhabitants or buildings in a particular region. The environmental conditions in different locations are also worth noting, e.g., the flood threshold changes from the Central region of São Paulo to the Serrana region of Rio de Janeiro. This led to a valuable discussion about how to address this issue.There was a common agreement that the reference values of the information parameters should not be defined when modeling the decisions, largely because they will often change depending on the nature of the location. Instead, a reference table could be created to link the information parameters of a business rule with their respective reference values, e.g., a rainfall decision table (business rule) could establish that the moderate level of rainfall (information parameter) for Rio de Janeiro is 50mm per hour (reference value), while in Recife it is 30mm per hour. Likewise, a hydrologist raised an interesting comment about information regarding changing parameters when making decisions:

> "*It is not only a matter of quantitative factors but also of parameters; other parameters should be taken into account and not only the water level" (Hydrologist A).*"

Hence, another important outcome of the generated model diagrams that was raised during the discussions held in the focus group sessions was determining the information parameters for decision-making. These parameters and reference values should be used to define the decision rules, which in turn "*are of significant value for supporting the decision-making*" (Meteorologist A). This is even more important when involved in critical tasks which need to be discussed in depth by all the members of the monitoring team. In effect, the analysis of the generated model diagrams during the focus group sessions showed that these diagrams can allow decision-makers to establish the information parameters for their decision-making, as well as how the business rules should be formulated.

Finally, the model diagrams generated by our framework enabled the participants to determine which tasks were most urgent; for example, the "Open a warning" and "Identify potential events" tasks. These tasks require further discussion by the monitoring team both about the decision-making tasks and the conditions for using data sources, the reliability and type of used data, and the decision rules that are employed. Evidence of this is provided by the following statements:

> "*When the team gathers there is a decision-making meeting [to open a warning], which is a critical task. And, identifying the nature of the potential event is a task that requires more discussion. I have often asked the other members: 'Is it going to rain?' because I was not confident enough about the data that I received." (Geologist A).*"

> "*These critical decisions ["Open a warning" and "Identify potential events"] are affected by different variables, such as the existence of data in the area, the temporal scale of existing data, and the condition of some of the equipment." (Meteorologist A).*"

These statements showed that the model diagrams generated by our framework have allowed decision-makers to identify both non-critical and urgent tasks that could affect their decisions. This also helps the monitoring teams to understand their business process.

In summary, the model diagrams generated by our approach proved to be a valuable means of identifying alternative data sources that could more effectively meet the requirements of the decision-makers (e.g., reports from official agencies). The fixing of the information parameters, as well as the definition of a reference table that would underpin the analysis of these parameters, were also made possible through the generated model diagrams. Additionally, by conducting an analysis of these generated model diagrams, the decision-makers were able to determine which data sources and information had the greatest influence on their decisions. These results are particularly important on account of the criticality and dynamicity of the decision-making that takes place within disaster management.

5.5 Discussion

This section discusses the theoretical and practical implications in regarding the three contributions of this research.

5.5.1 A Multi-layered Model and Notation for the Connection of Decision-making with Data Sources

Based on well-known standard and notations, this research developed the oDMN$^+$ that links the decision-makers' tasks and the emerging big data sources. Results obtained from the case study at Cemaden have provided evidence that oDMN$^+$ successfully achieved its objective, which revealed that the oDMN$^+$ Framework is successful in supplementing the existing standard notations (e.g., BPMN (OMG, 2013), DMN (OMG, 2014)) for making a connection between decision-making and data sources. Moreover, this chapter adds to existing work on the question of making a connection between decision-making and data sources (MALSBENDER *et al.*, 2013; KURNIAWATI; SHANKS; BEKMAMEDOVA, 2013; MANDVIWALLA; WATSON, 2014; KLEINDIENST; PFLEGER; SCHOCH, 2015) because it provides the Extended O&M, which could represents the abstraction of the data sources and connects it to the information required to make a decision. It also supports the modeling of this required information of decisions, which can be understood as a relevant contribution to previous research works (DOLIF *et al.*, 2013; VIEWEG; CASTILLO; IMRAN, 2014; KLEINDIENST; PFLEGER; SCHOCH, 2015). This is because these works are only focused on analyzing the necessity of considering required information in decision-making while, at the same time, they do not provide a systematic way of describing the connection of decision-making and its required information. In contrast, the generated model diagrams act as guidelines both to enable decision-makers to recognize all the useful information for specific tasks and to determine the impact of information for the existing tasks.

These findings can also assist in supporting and improving the use of the analytical models proposed in previous studies (SAADATSERESHT; MANSOURIAN; TALEAI, 2009; SIMONOFF *et al.*, 2011; CHEN *et al.*, 2012). This can be done by pointing out the data sources that can provide the information required by decision-makers. For example, social media platforms like Twitter might provide information about the consumers' attitudes and intention in a specific area. Hence, this data can be used in supporting the planning of marketing actions in the area. It is thus expected that decision-making can be improved because decision-makers are provided with more valuable information. Finally, oDMN$^+$ adds to the previous research on metamodeling for disaster management (OTHMAN; BEYDOUN, 2013; SACKMANN; HOFMANN; BETKE, 2013) and decision-making (BAZHENOVA; WESKE, 2015; JANSSENS; SMEDT; VANTHIENEN, 2016) by describing all variables associated with decision-making.

5.5.2 A Modeling Process for supporting the Connection of Decision-makers' Tasks with the Emerging Data Sources

The development of a model and notation that links decision-makers' tasks with the emerging data sources is necessary for describing and understanding the relationship among the existing concepts and variables. This however should be followed by a modeling process, which guarantees the effectiveness of the model and notation. Therefore, another important contribution achieved in this chapter is the presentation of a modeling process that aims at employing the oDMN$^+$ in practice.

Previous studies on business process modeling (BECKER; ROSEMANN; UTHMANN, 2000; SANTORO; BORGES; PINO, 2010; ANTUNES *et al.*, 2013) are limited to provide guidelines for eliciting conceptual elements and modeling the business process of an organization. This chapter thus complements these works by proposing and evaluating a modeling process that is not only focused on modeling business processes but also adds the modeling of variables related to decision-making (e.g., information requirements and business knowledge) and also establishes their links with data sources. In summary, all the layers (i.e., business process, decision-making, and data) of an organization can be modeled with oDMN$^+$, instead of only modeling the business process layer.

5.5.3 Lessons Learned and Implications of the Case Study

The lessons learned from the case study can be understood as the third significant factor of this research, which has important implications for practice and research on decision-making with big data. In the first place, results of the study suggest that practitioners and researchers should not only focus on the information requirements when making a decision, but also go beyond this and understand how current data sources can be used to meet those requirements. This is particularly important because decision-making often involves a large volume of data that are shared in different formats and supplied by different sources. In view of this, a clear understanding of the data sources that affect decisions could improve decision-making either by speeding it up or by allowing people to acquire information that is more accurate. In this way, this research supplements existing studies that are confined to studying the use of big data in decision-making without recognizing the importance of data sources (MILLER; MORK, 2013; WAMBA *et al.*, 2015; ZHOU; FU; YANG, 2016). Moreover, the study can support the task of "specifying the authorized source" as described in the decision-making process proposed by KoŚcielniak & Puto (2015) when the decisions of an organization are based on big data. In this task, practitioners should identify which authorized data sources may affect their decision-making. For example, the analysis of social media messages (a data source) could form the basis of an assessment of customer intentions or market trends (required information) with regard to a given product. On the basis of this information, the organization could decide which alternatives are more appropriate for a new marketing campaign (decision).

The findings of this research also suggest that practitioners and researchers need to be aware that decisions and their requirements (e.g., business rules, information requirements, and data sources) have an impact on the flexibility of organizations, i.e., on their capacity to adapt to new conditions and situations. In the case of decision-making that relies on big data sources, this is even more challenging because new sources can emerge in real-time, while existing sources may be rendered ineffective. As a result, the business rules and information requirements should be able to adapt to this new set of data sources. When addressing this issue, the findings of our case study imply that it is very important to design mechanisms to support a flexible system for matching information requirements with the corresponding reference values and thresholds. This finding makes it easier to understand the multiple dimensions of flexibility in the modeling of a business process that was provided by Albuquerque *et al.* (2015), insofar as it adds new dimensions of flexibility that correspond to different data sources, decisions, information requirements, and business rules.

Another factor is that the results obtained in this research suggest that decision support systems in the context of big data, provide features that allow end-users to control the visualization of data sources of interest. This is particularly important since the required information and data sources often change; for example, decisions about resource allocation require data about available modes of transport, which are not required when defining a marketing strategy. In addition to controlling the visualization of data sources, the findings suggest that big data can be integrated in terms of their geographic location. This research thus supplements existing studies by adopting an approach that involves multiple levels of visualization for displaying big data (VIEWEG; CASTILLO; IMRAN, 2014; IMRAN *et al.*, 2014). Furthermore, the required information for the decisions modeled with our approach can be used to improve the quality and size of the existing ontologies within disaster management, such as Humanitarian eXchange Language (HXL)[4]. This can be regarded as a significant extension of previous works (TAYLOR *et al.*, 2013; KLEINDIENST; PFLEGER; SCHOCH, 2015) that offers a basis for the creation of innovative analytical algorithms or better ontologies for big data. Moreover, our approach can support the design of "information services" and "analytical services" as in the conceptual

[4] http://hxlstandard.org/

architecture of Demirkan & Delen (2013) for service-oriented DSS. In these services, our results suggest that a supplementary service (or component) should be provided for managing the connection between a decision and its business rules, information requirements, down to its related data sources. As a result, it is expected that service-oriented decision support systems can provide information that is better suited to the needs of the decision-makers.

5.6 Final Remarks

This chapter has set out the oDMN$^+$ Framework with the aim of connecting the decision-makers' tasks with data sources by means of two essential features: (a) the oDMN$^+$, a multi-layered integrated model and notation, and (b) a modeling process to employ oDMN$^+$ in practice. oDMN$^+$ describes the connection of decsion-makers' tasks and decisions, determines the required information of these decisions, and identifies sources that might provide this required information. Since oDMN$^+$ relies on several conceptual elements in the context of the application, the modeling process defines a set of activities for employing oDMN$^+$. Thus, this process will ensure that all these conceptual elements are obtained and useful model diagrams are produced that can support decision-makers.

The employment of oDMN$^+$ in a case study involving early warning decision-making at a Brazilian early warning center (Cemaden) provides evidence that the oDMN$^+$ Framework is able to connect decision-making to data sources, and thus align business processes to data sources. This is because decision-makers can understand which information is required for each task, and then find out what information is lacking or predict the impact of changes in information. Results of this application also showed that oDMN$^+$ effectively describes the connection between decision-makers' tasks with data sources by modeling conceptual elements from decision-makers. However, this was only possible after the Extended O&M had been defined, since it seeks to model the relationship between the input data and data sources (i.e., from the DMN to the data sources), at the bottom of Figure 35. The evaluation results indeed provided evidence of the effectiveness of the modeling process in obtaining conceptual elements from the decisions-makers in the context of an application.

Since oDMN$^+$ provides a suitable description of how decision-making can be connected with data sources, it has a potential for developing better DSS. Therefore, the next chapter describes a set of design principles that supports the design of an SDSS for dynamic environments with heterogeneous geospatial big data; for instance, the decision-making in disaster management.

CHAPTER

6

DESIGN PRINCIPLES FOR GUIDING THE DEVELOPMENT OF AN SDSS FOR DECISION-MAKING WITH HETEROGENEOUS GEOSPATIAL DATA

6.1 Overview

Emerging sources of spatial big data such as social media and crowdsourcing platforms have been changing the way of organizations make their business spatial analysis and decision-making. These data sources could expand the understanding and efficiency of business decisions in different scenarios, e.g. marketing analysis, disease epidemiology (CARROLL *et al.*, 2014), and disaster monitoring (GOODCHILD, 2007; BISHR; JANOWICZ, 2010; SCHNEBELE; CERVONE; WATERS, 2014; KOŚCIELNIAK; PUTO, 2015; YOU *et al.*, 2015). Existing works have focused on the development of location-based analytic features that are based on the integration of data provided by these new data sources. For example, Albuquerque *et al.* (2015) present an approach that analyzes the geographic correlation of data provided by official in-situ sensors with those provided by crowdsourcing platforms for retrieving a meaningful information. In contrast, Dominkovics *et al.* (2011) present the development of a web-based application that aims at supporting decision-making by presenting information on spatial density maps. Although these works have overcome important issues in the literature and practice, they often provide a spatial algorithm or generated data that do not match decision-makers' needs and become useless for supporting decision-making (DOLIF *et al.*, 2013; VIEWEG; CASTILLO; IMRAN, 2014).

In this sense, there have been a growing interest in the area of data-driven decision support, which follows a tendency to provide more accurate and appropriate data to decision-making (GOPALKRISHNAN *et al.*, 2012; PROVOST; FAWCETT, 2013; MALSBENDER *et al.*, 2013; KURNIAWATI; SHANKS; BEKMAMEDOVA, 2013; DUGGAN, 2014; HORITA *et al.*, 2014a; WAMBA *et al.*, 2015). Within the existing works, Kleindienst, Pfleger & Schoch (2015) integrated social media analytics and the business goals of an organization by breaking down these goals into critical success factors that make it possible to find out the information requirements, so that these requirements can be combined with the appropriate social media analytics. Furthermore, the oDMN$^+$ metamodel that was presented in Chapter 5 can be even regarded as a work that could support in this context. Nevertheless, these works do not describe a way of incorporating requirements of decision-makers from dynamic environments with heterogeneous geospatial big data in the development of a decision support system.

This chapter, therefore, presents a set of design principles that would guide the development of SDSS for decision-making with heterogeneous geospatial big data. These were derived from a two-cycles action design research project, which was performed in the context of monitoring and early warning within a Brazilian emergency agency.

The remaining of this chapter is structured as follows. Section 6.2 describes the research method that was employed for the development of this design principles. Section 6.3 details the action design research, e.g., the cycles, stages, and main outcomes of each stage. Finally, Section 6.4 presents the design principles and discussions their implications for research and practice.

6.2 Research Method

The action research method "aims to contribute both to practical concerns of people in an immediate problematic situation and to the goals of social science" (SEIN *et al.*, 2011). It then makes possible developing solutions to a practical problem that are at same time a theoretical knowledge of value to a research community. Action researchers are then supplied with different cycles and interactive processes for achieving the goal of linking practice and theory (BASKERVILLE; WOOD-HARPER, 1996; LAU, 1997; BASKERVILLE, 1999; DAVISON; MARTINSONS; OU, 2012). Susman & Evered (1978) proposed the most known action research process, which is composed of five activities: (1) diagnosing (identifying the problems and working hypotheses), (2) action planning (planning actions), (3) action taking (implementing the planned actions and collecting data), (4) evaluating (analyzing the collected data against the working hypotheses), and (5) specifying learning (identifying the lessons learned).

Since information systems research has an aim of making theoretical contributions that can also assist in solving organizations' problem (SEIN *et al.*, 2011), the use of action research has widespread among the IS research projects over the last two decades (BASKERVILLE; WOOD-HARPER, 1996; LAU, 1997; BASKERVILLE, 1999; DAVISON; MARTINSONS; OU, 2012; BROOKS; ALAM, 2015; SPAGNOLETTI; RESCA; SÆBØ, 2015). Within these works, Sein *et al.* (2011) proposed the action design research method for supporting the building IT artifacts in an organizational context and learning from the intervention while addressing a problematic situation. It also aims to guarantee the research rigor and controlled evaluation without losing the essence of the action research and design science research (HEVNER *et al.*, 2004). For doing this, the action design research method consists of four stages: (1) problem formulation; (2) building, intervention, and evaluation; (3) reflection and learning; and, (4) formalization of learning. This research method was adopted here in a cross-organizational collaborative research project, which aims at developing and evaluating design principles for an SDSS that takes into account the decision-makers' requirements, and supports decision-making in dynamic environments with heterogeneous spatial big data (e.g., disaster management). Here, design principles are understood as a body of knowledge that guides the development of an artifact (WALLS; WIDMEYER; SAWY, 1992; MARKUS; MAJCHRZAK; GASSER, 2002), e.g., a conceptual architecture for SDSS. The next section presents the context of monitoring and early-warning of Cemaden (see Section 5.4.1) that was adopted as the organizational settings for the action design research project.

6.2.1 The Action Design Research Project at Cemaden

The action design research project at Cemaden followed the sequential stages that were defined by Sein *et al.* (2011), as shown in 46. While the actions were adapted for the particular context of the center. In Stage 1, the research question and working hypothesis are defined from the research gaps evidenced through analysis of the existing literature, as well as the collected data at the center. The iterative cycles of design and implementations and evaluation of the interventions are performed in Stage 2. While the outcomes of these two stages are analyzed in Stage 3. For example, lessons learned from the analysis of data collected with decision-makers. Finally, Stage 4 summarizes the result obtained in Stage 3, which is the design principles of SDSS for dynamic environments with heterogeneous geospatial big data.

The next section introduces the action design research project and gives more detail about the stages, actions, and results obtained.

6.3 Design of an SDSS for Dynamic Environments with Heterogeneous Big Geospatial Data

The action design research project in December 2015 with discussions about the necessity of improving the comprehension of the daily activities into the monitoring control room at the Cemaden, as well as how the available systems have been used for such activities. Three stakeholders were identified: the members of the monitoring control room (hereby called operators), and practitioners of the center (e.g., its researchers) and the action design research researchers (e.g., researchers from the USP) that together composed the action design research team. After that, two interactive design and analysis cycles were conducted by the action design research team in collaboration with operators, as shown in Figure 47. The first cycle began at the end of

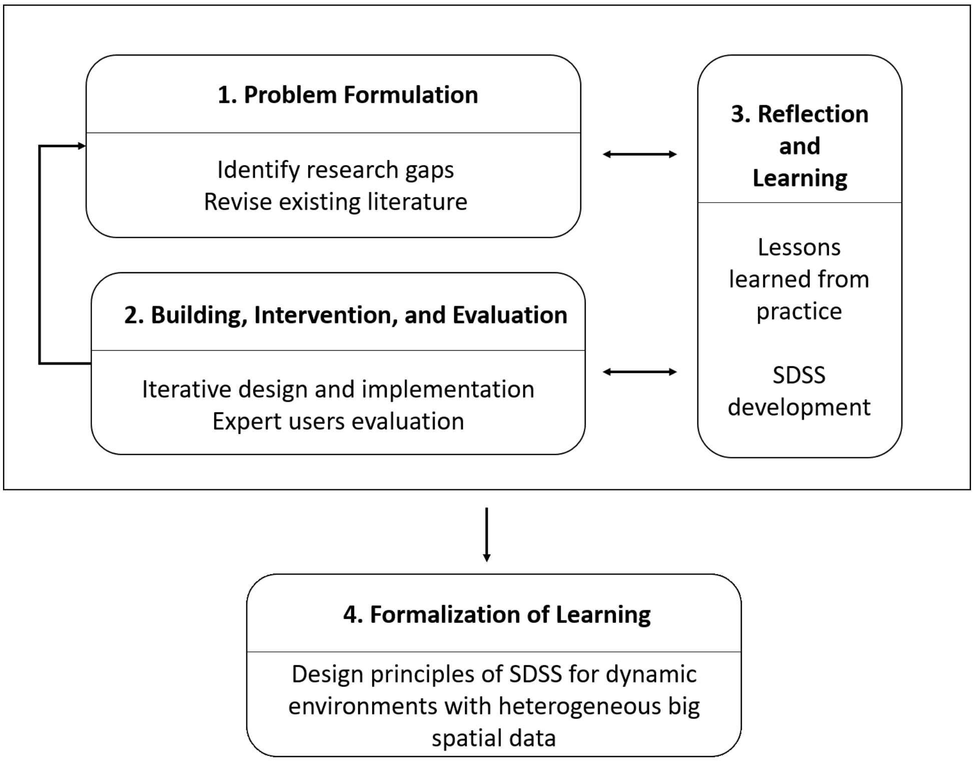


Figure 46 – Stages and actions of action design research. Based on Sein *et al.* (2011).

January 2016 and lasted for six months. In the middle of August, the action design research team was ready to start the second cycle, which continued until the middle of December 2016.

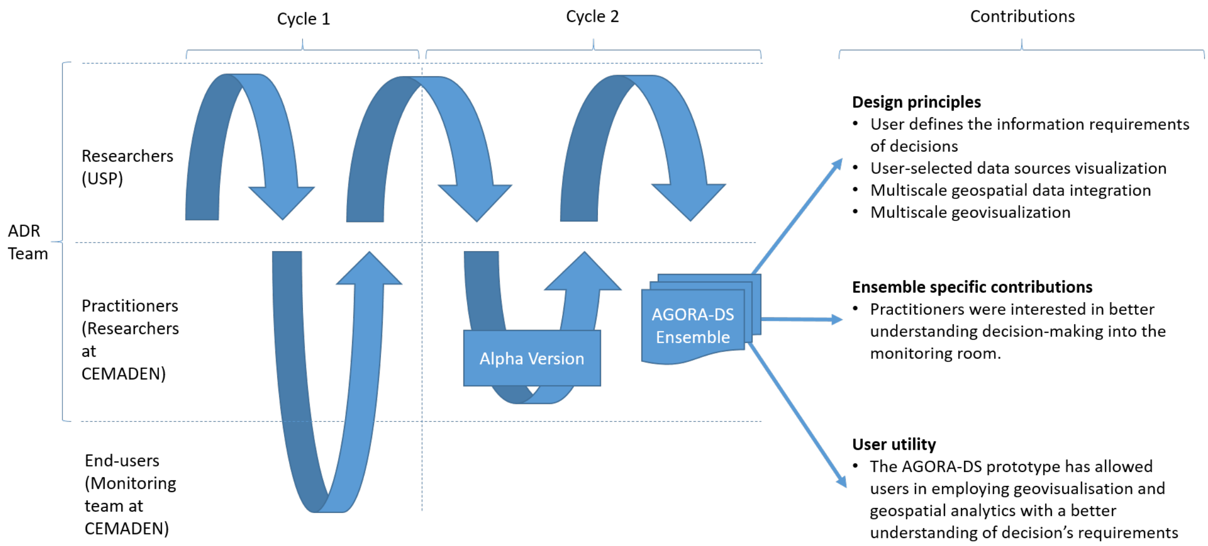


Figure 47 – The interactive cycles of the action design research project. Based on Sein *et al.* (2011).

The right-hand side of the Figure 47 shows the ensemble artifacts generated after the employment of the action design research. These artifacts bring specific contributions to the three stakeholders groups that are shown on the left-hand side of Figure 47 - the researchers from the University of São Paulo (USP), the practitioners, and the end-users. The next sections presents each cycle in more detail.

6.3.1 First Cycle

In this first cycle, the action design research team was focused on identifying potential problems in the daily activities of the monitoring control room of Cemaden. Therefore, the data used for supporting the formulation of a problem was collected by one of the authors on January 19th-22nd, 2016 and February 1st, 2016 at the Cemaden's headquarters in São José dos Campos, Brazil. In these periods, 88 warnings were issued from the monitoring control room to vulnerable communities: 9 on Jan 19th, 63 on Jan 20th, 15 on Jan 15th, and 1 on Feb 1st. Participatory observation, with a limited degree of interaction by the researcher, was first carried out as an instrument for gathering data about the day-to-day activities and interactions of the subjects. This meant that the observer was only regarded as a researcher, and did not interact with the subject or interfere with the subjects' activities (RUNESON; HÖST, 2008). Additionally, individual and face-to-face interviews were also conducted to obtain data about the individual business activities of the subjects (e.g., the activities they carried out or the data sources that were used). This included 10 semi-structured interviews with members of the monitoring control room – comprising 2 geologists, 2 hydrologists, 2 meteorologists, and 4 disaster management specialists – all of whom completed a questionnaire that was prepared in the data collection protocol. Since the interviewees were working within strict time constraints, the interviews took no more than 30 minutes, while the participatory observations lasted for 2 hours. In summary, the methods of data collection in this cycle were twofold: (a) direct observation of the daily activities of the members in the monitoring room, and (b) 10 semi-structured interviews with these members.

6.3.1.1 Stage 1: Problem Formulation

Based on the collected data, we identified three main problems that were negatively affecting the activities in the use of SDSS into the monitoring room. First, members of the monitoring control room do not have a clear understanding about the information required for decision-making (Problem 1). This is mostly due to the dynamic variables intrinsic to the different situations, e.g., a member could make a better decision if he knew the threshold of a monitored region; however, in other regions, this information is useless because he needs to know the condition of buildings instead. A typical statement of participants during the interviews sessions was:

> *"There are several occurrences of floods in areas with a lack of appropriate data, e.g., thresholds". (Employee, during interview sessions).*

As a result, many data that are provided for decision-making become useless, since they do not fulfill the decision-makers' requirements. While, on the other hand, more data are required for overcoming the inexistent of useful data, e.g., from unmonitored areas.

> "*We confirm that the water level decrease in the riverbed but did the flood cause any problem? We still lack of more information for making decisions". (Employee, during interview sessions).*

> "*There is a municipality that has several rainfall data but there is not information about the threshold of the area. In other cases, the threshold is known but the rainfall gauge is not properly working". (Employee, during interview sessions).*

Second, there is a notable delay in the decision-making of the monitoring control room caused by the necessity of analyzing data from different systems (Problem 2). This happens because the available data is stored in disconnected datasets due to their different sharing formats. As a result, it hampers not only the data integration but also makes that decision-makers analyze individually each data set in their available system, which takes more time that analyzing all these data into a unified system.

> "*When we confirm [in the system] that there is a situation of risk, you can access the database of Center for Weather Forecasting and Climate Research (CPTEC) and analyze the satellite images". (Employee, during interview sessions).*

> "*We do not have this information in the system; therefore, we have to use data from other sources, e.g., National Agency of Water (ANA)." (Employee, during interview sessions).*

Finally, there is a remarkable necessity in continuously changing the decisions' variables and measurement values per location (Problem 3). It occurs due to the intrinsic modifications along on the time on the variables that affect a decision like number of inhabitants, geology into a riverbed, or degradation of the environment. Furthermore, the environmental conditions in different locations also changes, e.g., a flood threshold changes from the Center region of São Paulo to the Serrana Region of Rio de Janeiro.

> “*The number of people and buildings in the municipality is very dynamic. After a data collection, if there is a rainfall event, the risk situation would change and the people that were not before in a risk situation, can be now exposed”. (Employee, during the focus group sessions).*”

> “*These thresholds should be continuously revised every two years, or after an important event.” (Employee, during the focus group sessions).*

> “*A warning of land slide to the Serrana Region of Rio Grande do Sul is completely different from the Serrana region of Rio de Janeiro because the buildings are stronger than those in the favelas.”(Employee, during the focus group sessions).*

6.3.1.2 Stage 3: Reflection and Learning

Once the problems were identified from practice, they were adopted by action design research team for driving the definition of both a research question and a working hypothesis, at the “building, intervention, and evaluation” stage. These were defined after ongoing reflection on the identified problems and analysis of all transcriptions of interviews and focus group sessions. The deep interventions or evaluations actions were not included at this stage because the action design research team first envisioned an initial discussion, as well as a research design before starting such actions. Nevertheless, these sessions of reflections and analysis of collected data have worked as a preliminary evaluation with the specialists that composed the action design research team in order to save time with useless interventions or interventions that are not viable to the Cemaden. Hence, the action design research team came up with the following research question:

RQ) How can an SDSS be designed for supporting decision-makers in dynamic environments to make sense of heterogeneous geospatial big data?

To tackle this research question, the working hypothesis of action design research team lies on the development and evaluation of appropriate design principles for guiding the design of SDSSs that take into account the decision-makers’ requirements and support decision-making in dynamic environments with heterogeneous geospatial big data. Such principles thus were expected to be a relevant theoretical contribution, which could both provide guidance to developers in the development of new SDSS for the application domain.

After defining the research question and working hypothesis, the action design research team was ready for moving forward with the deep interventions and evaluations. This was performed in a new action design research cycle that is described in the next section.

6.3.2 Second Cycle

The second cycle of the action design research project aimed to develop artifacts that could be used to test the proposed hypothesis, and thus to answer the proposed research question. These artifacts may also solve the problems identified from practice.

6.3.2.1 Stage 2: Building, Interventions, and Evaluation

Starting by the proposed hypothesis, we first revised the literature with a view on existing works that addressed the gap between decision-makers' requirements and data sources at the "building, intervention, and evaluation" stage of this second action design research cycle. The aim was to understand how data provided by different data sources could fulfill the information requirements of decision-makers in disaster management. Results however showed that most of the works in the topic still fail to establish a link with the information needed for the decision-making (GOPALKRISHNAN *et al.*, 2012; CHEN; CHIANG; STOREY, 2012; HORITA *et al.*, 2014a; IMRAN *et al.*, 2014; WAMBA *et al.*, 2015). This mismatch not only made the existing systems counterproductive to fulfill decision-makers' requirements but also resulted in useless features or produced data by existing systems (DOLIF *et al.*, 2013; VIEWEG; CASTILLO; IMRAN, 2014). Based on these findings, the action design research team expanded the research for covering not just the concerns specific to Cemaden, but rather, the class of field problems: SDDSs that takes into account the decision-makers' requirements in dynamic environments to make sense of heterogeneous geospatial big data.

Theoretical framework. The action design research team then envisioned that the decision-makers' requirements modelled by means of standard models could support in the definition of system requirements. Hence, the DMN that is released by the OMG, provides conceptual features (e.g., business rules, input data, and required information) for modeling decisions (OMG, 2014). Furthermore, the OGC and the ISO defined the O&M standard, which provides an abstract view of observations that originate from various data sources, as well as being able to integrate the data sources to the requirements of the information (OGC, 2013). The observation-aware Decision Model and Notation (oDMN) standard defines additional conceptual features (e.g., observation and property) that can be used for the linking of DMN with O&M (HORITA *et al.*, 2016). Although these standard model and notations address important issues, they do not provide features for modelling spatial data. This is particular important since decisions' variables (e.g., required information or the definition of business rules are not similar at different locations. In this context, Dominkovics *et al.* (2011) proposed a common data model that can be used to describe the elements and association of spatial analytics features. This data model followed the naming conventions and types for the set of features objects and properties that were defined by the ESRI Company, the world leader supplier of GIS applications. Together, the DMN, O&M,

oDMN, and common data models provided the basis for the definition of a theoretical framework that describes decision-makers' requirements for dynamic environments with heterogeneous spatial big data. This framework is displayed using UML class diagram in Figure 48.

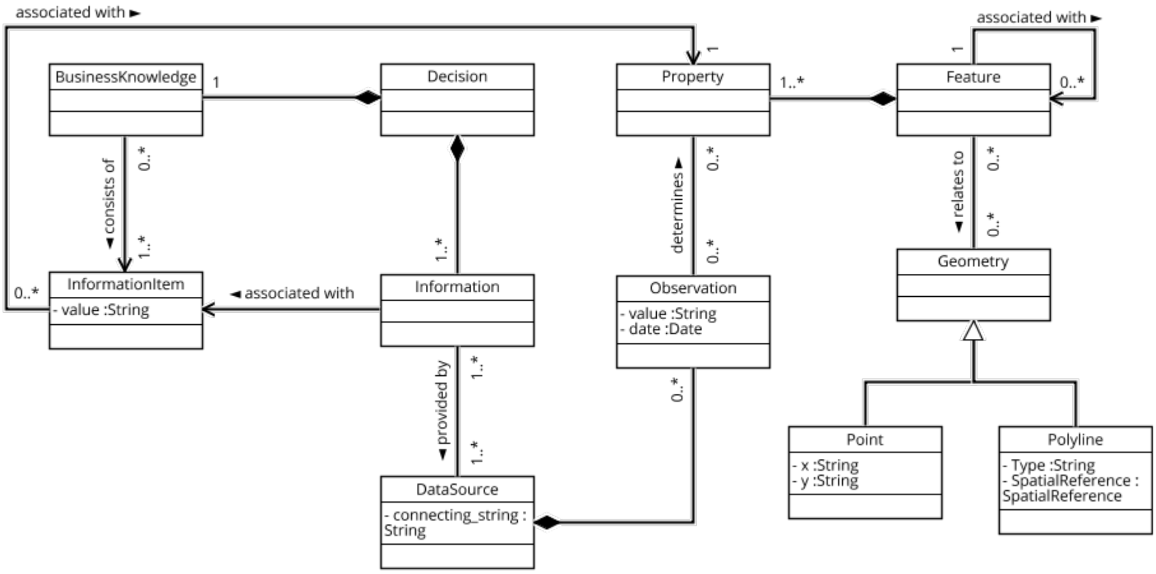


Figure 48 – Theoretical framework.

In the description of decision-makers' requirements, the theoretical framework starts by defining that a **decision** requires a set of **information**, which forms the basis of input information items for a **business knowledge** (e.g., decision rule). For example, an operator responsible for issuing warnings to vulnerable communities (decision) may compare the level of water into the riverbed (information item) against the flood threshold (information item) for checking if it is crossing the defined threshold for the area (business knowledge). Then, each required information may be provided by specific **data sources** (e.g., rainfall gauges), which store several **observations** (e.g., several measurements of water level). Since **features** are understood as an abstraction of a real-world phenomenon (e.g., a river) (OGC, 2013), these observations can determine the value of a **property** of a feature (e.g., water level). Simultaneously, this property associates with the information item of the decision logic, e.g., the flood threshold (information item) is associated with the water level (property) of a river (feature). Furthermore, a feature can be used for representing a real-world object on a map. Its geographic position is then determined by a **geometry object** that can be a simple **point** with the traditional coordinates (x and y) or a **polyline** that could represent one single line or a polygon (several connected lines). Features can be also associated with other features of real-world objects; for example, a feature "road" is part of another feature "city", as well as a feature "state" consists of a group of several features "municipality").

The theoretical framework indeed is important because it gives a focus on describing the variables that are associated with decision-making. This also makes clear not only the information required but also how information can be combined in order to provide a more accurate input information for a decision that has to be taken. Consequently, it provides a theoretical basis for describing a solution to the raised problems from practice in regarding the understanding about information required (Problem 1), and provide a solid basis for supporting the integration of different systems (Problem 2). Given the problem of continuously changing variables and measurement values of decisions per location (Problem 3), the framework draws on the linking of information items to a property of a feature, i.e., the value of an information item is associated with a property of a feature. This creates the possibility of changing the value among different locations, e.g., a decision rule may define that the rainfall threshold (information item) of the volume of rainfall (property) of the Metropolitan Region of São Paulo (feature) is 60 mm, while, the threshold for the Serrana Region of Rio de Janeiro could change to 40 mm.

System requirements. With a basis of the theoretical framework, the action design research team understood that a better input of information to support decision-making in dynamic environments depends on the understanding of all its information requirements and how they could be combined according to the location of interest. This is particularly important since decision rules and their required information change on different locations, and indeed such understanding may play an important role in supporting decision-makers. For example, it could provide an overview of the lack of data to fulfill the required information or which data sources can be used for gathering information. The data integration might be also impacted, once the decision rules detail how the required information can be combined in order to provide the information that better matches decision-makers' requirements. Furthermore, the linking of the observations with a feature, which is associated with a geographic location, may allow decision-makers to adapt their decision rules and required information according to a location. In turn, this supports location analytics, as well as multi-scale location analytics. For example, an operator can set up that a flood threshold in Rio de Janeiro is 40 cm, while in São Paulo is 60 cm. Finally, the action design research team summarized these identified associations into the following systems requirements:

1. Decisions should consist of information requirements and a decision rule.

2. A decision rule should consist of a set of information items that must be associated with an information requirement required by the decision.

3. Information items should require an analytical value, which supports data analytics.

4. Features of interest should be associated with a set of specific properties, which aims at characterizing and describing them.

5. Information items of decision rules should be associated with properties of a feature of interest, which in turn supports the understanding of the lack data.

6. Since a property can characterize features of interest, its descriptive value should be determined by the observations.
7. Information items should be thus associated with appropriate data sources.
8. The association of information with data sources should be flexible for allowing the inclusion (or exclusion) of data sources, as well as the changing of existing data sources.
9. A connection string should be required for allowing the authentication with the data source.
10. A data source should consist of a group of individual observations (also known as data).
11. Observations should require a date of creating and a descriptive value.
12. Once information items are associated with properties of a feature of interest, this should allow the data integration based on their associated location.
13. Decision rules provide an understanding of how information items should be combined.
14. A feature of interest is associated with a geographic location, which in turn could be represented either by an individual point or by a polyline (i.e., square or triangle).
15. A point should require two essential information for determining its geographic location: “x” and “y”, while a polyline requires a type and the spatial reference.
16. Feature of interest may be associated with other features of interest, which in turn should allow a multiscale geovisalization.

Design principles. All these system requirements then drove for the definition of four design principles: (1) user defines the information requirements of decisions; (2) user-controlled configuration of data sources with a mapping to required information; (3) location-based data integration; and, (4) multiscale geovisualization. The principle of *user defines the information requirements of decisions* defines that only that information required by a decision is displayed in the system. The principle of *user-controlled configuration of data sources with a mapping to required information* determines that appropriate data sources should be assigned to the information required by a decision. The principle of *location-based data integration* defines that a geographic location should be used as a basis for supporting the integration of data. Finally, the principle of *multiscale geovisualization* determines that decisions’ information requirements as a basis for displaying the data. Table 12 details the relationship between the design principles and requirements from theory.

Table 12 – Relationship the design principles and system requirements.

Design principles	Description	Requirements from theory
DP1: User defines the information requirements of decisions	A decision should be defined into the system according to its required information. *However, the user should be able to define explicitly and flexibly what are the information required to make specific decisions.*	• Decisions should consist of information requirements and a decision rule. • A decision rule should consist of a set of information items that must be associated with an information requirement required by the decision. • Information items should require an analytical value, which supports data analytics. • Features of interest should be associated with a set of specific properties, which aims at characterizing and describing them. • Information items of decision rules should be associated with properties of a feature of interest, which in turn supports the understanding of the lack data. • Since a property can characterize features of interest, its descriptive value should be determined by the observations.

Continue on next page

Table 12 – Summary of the methods of data collection (continued).

Design principles	**Description**	**Requirements from theory**
DP2: User-controlled configuration of data sources with a mapping to required information	Data sources should be explicitly assigned to the information requirements of a decision. *However, the user should be able to flexibly define parameters for authenticating with data sources (e.g., connection string).*	• Information items should be thus associated with appropriate data sources. • The association of information with data sources should be flexible for allowing the inclusion (or exclusion) of data sources, as well as the changing of existing data sources. • A connection string should be required for allowing the authentication with the data source. • A data source should consist of a group of individual observations (also known as data). • Observations should require a date of creating and a descriptive value.

Continue on next page

Table 12 – Summary of the methods of data collection (continued).

Design principles	Description	Requirements from theory
DP3: Location-based data integration	The system should provide features for integrating heterogeneous data sources. *However, the integration should be done based on a geographic location.*	• Once information items are associated with properties of a feature of interest, this should allow the data integration based on their associated location. • Decision rules provide an understanding of how information items should be combined.
DP4: Multiscale geovisualization	Geovisualization should use the decisions' information requirements as a basis for displaying the data. *However, users should be able to perform a multiscale geovisualizaiton of data since it might provide a "big picture" of the current situation before a deep analysis in most critical locations.*	• A feature of interest is associated with a geographic location, which in turn could be represented either by an individual point or by a polyline (i.e., square or triangle). • A point should require two essential information for determining its geographic location: "x" and "y", while a polyline requires a type and the spatial reference. • Feature of interest may be associated with other features of interest, which in turn should allow a multiscale geovisalization.

IT-Dominant Intervention. Having these design principles defined, we began the intervention by adopting an IT-Dominant approach for the next steps of the "building, intervention, and evaluation" stage. This approach defines that an IT artifact is "continuously instantiated and repeatedly tested through organizational intervention and subjected to participating members' assumptions, expectations, and knowledge" (SEIN *et al.*, 2011). We first intended to implement a simple add-on that could be incorporated into the Cemaden's existing information system, SALVAR, but we were unable to do so because of organizational issues and unforeseen changes. Therefore, we designed a new version of SDSS prototype, named AGORA-DS, which is based on and similar to the existing IS. This prototype can be regarded as a new version of our previous works (HORITA *et al.*, 2014a; HORITA *et al.*, 2015). It now has incorporated a set of features that were defined as a way of embodying the design principles. Figure 49 shows AGORA-DS architecture using the notation from UML's component diagram.

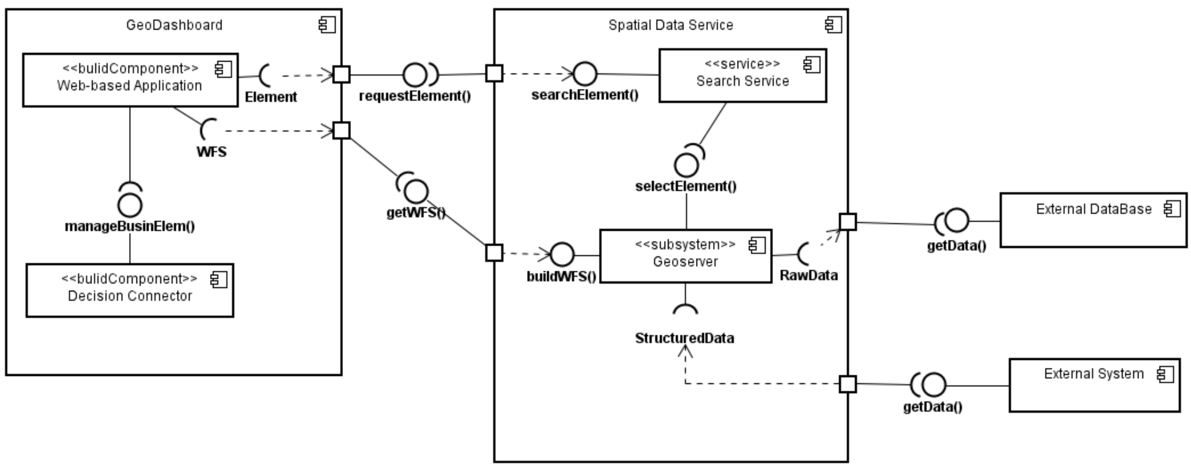


Figure 49 – AGORA-DS architecture that embodies the design principles.

The architecture consists of two main components with complementary responsibilities. First, the **Spatial Data Service**, in the middle, aims at handling the data provided by different sources. This means providing interfaces to collect the data, storing them into appropriate databases, and developing interfaces for external systems to access the data. For achieving these tasks, this main component consists of two sub-components: (a) the *Geoserver* is responsible for storing and sharing the spatial data; and, (b) the *Search service* that performs queries over the geoserver database. In the Geoserver, the data can be shared through the geospatial service standards, WFS. We have used WFS instead of WMS because it supports complex query operations regarding geographic features that result in low-time for data processing on the client-side (ZHANG; LI, 2005). The *Search service* can perform queries based on temporal and thematic criteria specified by users. Temporal search parameter determines a period in terms of starting and end dates. Thematic search parameter uses values provided by the users for comparing attributes of the database; for example, municipality or ids. Searches that combine both parameters are also allowed, e.g., all data of Rio de Janeiro (thematic) in the period from

Jan 2015 to Jan 2016 (temporal). In brief, the processing of data in order to obtain their spatial location allowed us to operationalize the design principle of "Location-based data integration" (DP3).

The second main component of the architecture is the **GeoDashboard**, on the left-hand side, that is responsible for providing an interaction interface with the users. For doing this, it defines two sub-components: the *Web-based application* and the *Decision connector*. The web-based application displays the data in a simple visualization dashboard. The term "dashboard", well known in the field of business analytics, refers to an information system which aims at providing the most important information needed for supporting decision-making (FEW, 2006; LIANG; MIRANDA, 2001). When geospatial factors are also important for supporting decision-making (e.g. location of customers and transport routes), it can be called "GeoDashboard" (HORITA *et al.*, 2014a). Furthermore, the *Decision connector* aims to allow the configuring the system for managing the linking of a decision with its information requirements, and then to associated data sources.

Based on these components, AGORA-DS was developed using a micro-services approach. In this approach, each component has become an individual service that accomplishes a specific task (e.g., collect data from external sources) and might provide an interface for communication. These services were then placed online in individual Docker containers [1], while the communication was done through RESTful calls and geospatial service standards (e.g., WFS). The web-based application was developed using AngularJS together with OpenLayers for handling geospatial data. All other services were developed using Java as the programming language. The configuration data of the business connector component were stored in a NoSQL databased called MongoDB. This was chosen because it makes the development easier than relational databases which require defining complex structures of data, even for simple features.

The AGORA-DS was developed primarily for operationalizing the design principles of "user defines the information requirements of decisions" (DP1), "user-controlled configuration of data sources with a mapping to required information" (DP2) and "multiscale geovisualization" (DP4) by means of features embedded in its web-based application. First, apart from displaying important data, this web-based application provided a particular feature that allowed users to set up a decision. This configuration is associated with the business connector component, which manages the linking of a decision with its information requirements, and then to associated data sources. Figure 50 displays the screen related to this feature in the web-based application.

The web-based application of the AGORA-DS also provides a feature, which allows users to configure all the information associated with a specific location (Figure 51), e.g., the analytical value with its reference value, indicative color, and observable property. Once the user configures this information, the system uses the analytical value as an information item of a business rule of the location and employs it for analyzing all the available data related to

[1] <http://www.docker.com>

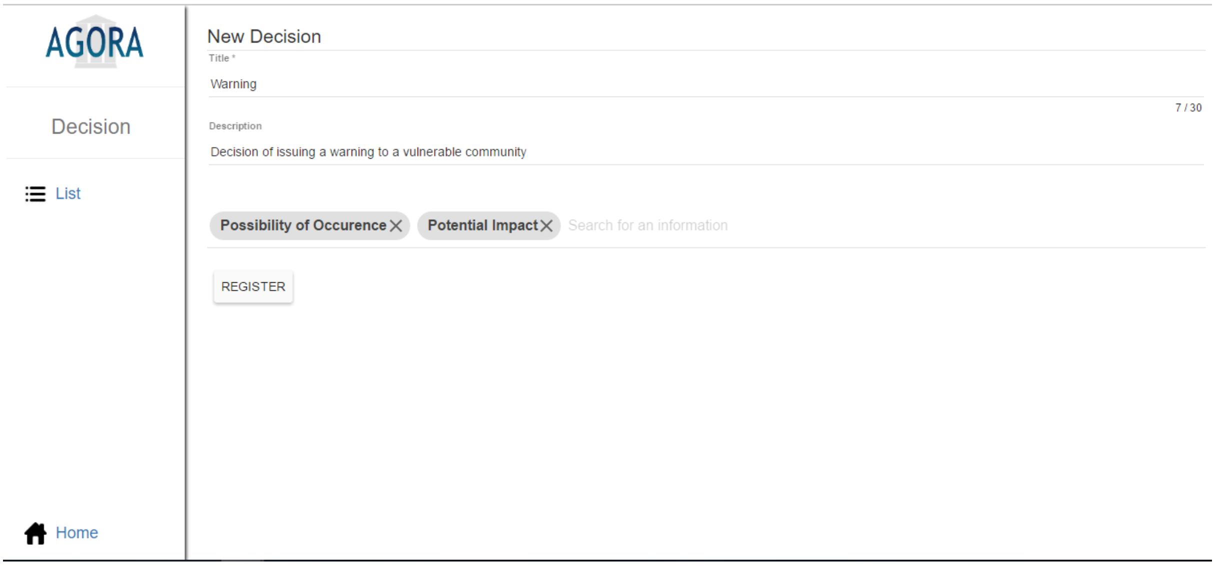


Figure 50 – “Decision setting” feature of the web-based application of the AGORA-DS.

the observable property. For example, if data referring to the accumulated value of a rainfall (observable property) over the last 24 hours (reference value) reaches 40mm (analytical value) the maker associated with the location on the map will be highlighted in red (indicative color).

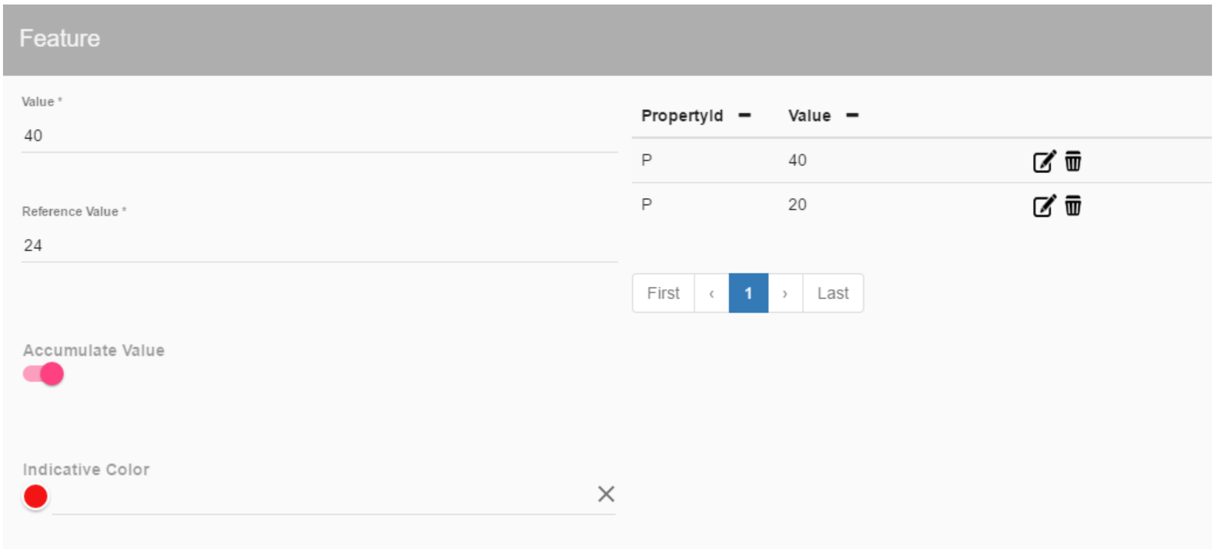


Figure 51 – “Location setting” feature of the web-based application of the AGORA-DS.

In case of configuring more than one analytical value for the same (or different) observable property (e.g., an analytical value for rainfall of 40mm and another for water level of 4cm), the marker on the map will be colored by the first observable value that reaches the analytical value.

Finally, the web-based application of the AGORA-DS supports a geographic drill-down, which uses the information configured for a location to color a feature on the map. In doing so, users can move from more consolidated layers to detailed layers (e.g., from regions to states) by focusing on an area of interest (e.g., a specific city or area). Figure 52 displays an example of this feature in Brazil.

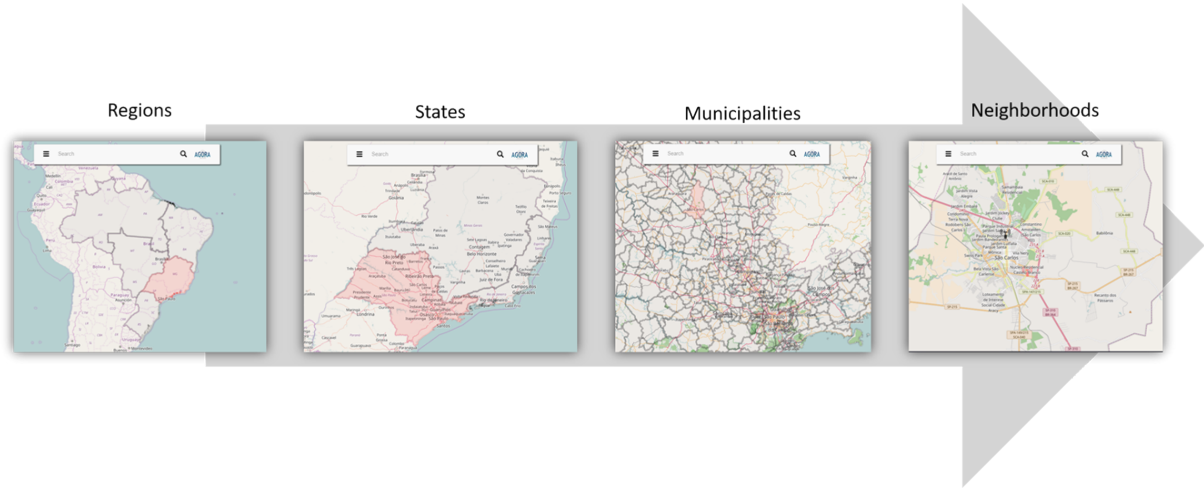


Figure 52 – “Drill-down geospatial analysis” feature of the web-based application of the AGORA-DS.

After the principles were implemented by the features of the web-based application of AGORA-DS, we were able to evaluate in practice the usefulness of the design principles for supporting decision-making in dynamic environments with heterogeneous big spatial data sources.

Evaluation. The evaluation of an IT-Dominant intervention might happen in two phases: (a) early phase, by means of lightweight interventions in a limited organizational context; and (b) mature phase when the artifact is included in a wider organizational setting. Due to time constraints and organizational issues, the scope of this work included an evaluation in an early phase.

In order to understand both the evaluation process and the evaluation object, we adopted the Conceptual Practice Model of Evaluation proposed by (GOLDKUHL; LAGSTEN, 2012). This model defines six important concepts to support the evaluator in structuring and better understanding the variables of his or her evaluation process. A set of questions was proposed for supporting in the definition of each of the evaluation concepts; for example, the questions “What should be evaluated?” and “What should we say something about?” might be useful to determine the evaluation object. Table 13 summarizes the evaluation concepts related to early evaluation of the proposed IT-Dominant intervention.

Table 13 – Evaluation concepts related to early evaluation of the proposed IT-Dominant intervention.

Evaluation Concept	**Specification**
Object	The overall evaluation object is the spatial decision support system that embodies the design principles, mainly the developed features, i.e., the location settings, decision settings, and drill-down geospatial analysis.
Purpose	This evaluation aims at analyzing and evaluating the usefulness of the proposed principles for supporting decision-making in dynamic environments with heterogeneous spatial data sources.
Conceptual Base	Main conceptual basis: • Data-driven decision-making (NOVAK; PAULOS; CLAIR, 2011; GOPALKRISHNAN *et al.*, 2012; MALSBENDER *et al.*, 2013; KURNIAWATI; SHANKS; BEKMAMEDOVA, 2013; WAMBA *et al.*, 2015; KLEINDIENST; PFLEGER; SCHOCH, 2015). • Spatial decision support systems (DENSHAM, 1991; RUSHTON, 2001; SUGUMARAN; SUGUMARAN, 2007). • Early-warning systems (Chatfield & Brajawidagda, 2013; De León et al., 2006; Picozzi et al., 2015). • Spatial features (Dominkovics et al., 2011).

Continue on next page

Table 13 – Evaluation concepts related to early evaluation of the proposed IT-Dominant intervention (continued).

Evaluation Concept	**Specification**
Criteria	This evaluation takes criteria-based strategy of IT-System in use for exploring the usefulness of the design principles in the design of an SDSS to support the activities of a decision-maker in dynamic environments with heterogeneous geospatial big data sources. This strategy means that the evaluation is performed according to some predefined general criteria and that the object of evaluation is the IT-system in use (CRONHOLM; GOLDKUHL, 2003). Thus, the criteria adopted for the evaluation are the usefulness and effectiveness of the design principles in handling heterogeneous geospatial big data.
Data	Measurement from the following tasks: • Exploring the feature “location settings”, • Exploring the feature “decision settings”, • Exploring the feature “drill-down spatial analytics”. Data collected from decision-makers (i.e., participants) and evaluation designer regarding the following criteria (proposed design principles): 1. User defines the information requirements of decisions, 2. User-controlled configuration of data sources with a mapping to required information, 3. Location-based data integration, 4. Multiscale geovisualization.

Continue on next page

Table 13 – Evaluation concepts related to early evaluation of the proposed IT-Dominant intervention (continued).

Evaluation Concept	Specification
Procedure	Initially, the AGORA-DS is introduced to the participant of the evaluation through a demonstration session. All details about each developed features are described and presented. This session is then followed by a semi-structured interview with the participant in order to gather their impressions/opinions in regarding of the features. Employed methods: • Demonstration session, • Semi-structure interviews.
Evaluate [Evaluator]	Evaluators: • Evaluator designer, • Decision-makers of a monitoring control room of Cemaden (participants).
Evaluation Result	Details of the results obtained with this evaluation will be mainly summarized through articles, as well as case study reports, which are expected to be published in well-known journals and conferences.

Continue on next page

Table 13 – Evaluation concepts related to early evaluation of the proposed IT-Dominant intervention (continued).

Evaluation Concept	Specification
Use [Recipient]	This evaluation has three main recipients. First, decision-makers from the monitoring control room of Cemaden (end-users) because their activities would be improved with a more valuable and useful data analysis system. Second, the design principles are of particular interest to the researchers since they provide important contributions to the research field. Finally, these principles compose an ensemble artifact that would support developers interested in the development of SDSS for dynamic environments with heterogeneous geospatial big data. This ensemble artifact also comprises the SDSS architecture that embodies the design principles.

Based on the evaluation concepts defined in the Conceptual Practice Model of Evaluation method, we were ready for conducting the evaluation of the AGORA-DS (see Appendix C). In this manner, a total of six individual evaluation sessions were carried out by one of the authors from 28th-30th November 2016 with members of the Cemaden in São José dos Campos, Brazil. Only those members that work in the monitoring control room of the center and had to deal with a disaster before, were able to participate in the evaluation. During these sessions, the evaluator detailed the research context and the previous works, introduced and demonstrated all the features existing in the AGORA-DS, and then interviewed the participant.

The results of the evaluation were then used as a basis for the reflecting and learning stage that is described in the following.

6.3.2.2 Stage 3: Reflection and Learning

Through the analysis of the results obtained in the "building, intervention, and evaluation" stage, the objective here was to understand how the AGORA-DS and its embedded design principles might affect the practice into the monitoring control room of Cemaden. The evaluation results provided positive implications of our intervention.

Firstly, the analysis of spatial data using the drill-down feature showed as a relevant alternative that could make easier the decision-making into the monitoring room. The possibility of establishing first the overall situation of an area, and then focusing on those locations that require more attention was considered a valuable asset with regard to support decision-making with heterogeneous spatial data sources. Together with the indicative colors, it would make faster the decision-making because the user does not have to go through all the data sources, he would analyze those that were marked with the most representative color. For example, as it was remarked in the following comments:

> "*Since we are monitoring the whole country, when a problem is evidenced in the southeast region, you just click on it and check the problem... You can make a faster decision because you are not looking all the data, you already have the indicative color showing the locations that need attention.*"

> "*For me, it is interesting [the drill-down analysis]; For example, when you move from the global to the regional by selecting the southeast region, and then when you click on São Paulo [a state in the southeast region], it is shown those municipalities in an alert condition. Overall, I will not loose time going through all the available sensors, it will be indicated on the map those that reached the threshold.*"

Likewise, the practitioners seemed to value the fact of making decisions based on geospatial data. It is because the visualization in a geographic format can be provided a better contextual analysis of those displayed in a table form. As a result, decision-makers might have better information at hand, and thus a faster decision-making can be made. For explaining this fact, a geologist provided the following example:

> "*When I see a rainy event at a mountainous area, I know the hydrology issues are faster of those in a plane region. So, if the system highlights the area close to the Serrana Region of São Paulo in red, for me, I already know that a warning should be issued. The same should be done when it is highlighted a location at the Coastal Region of São Paulo.*"

The management of data sources was another advantage that was mentioned by the practitioners; for example, the inclusion of new sources of data for supporting the analysis of vulnerable areas. This is particularly important because data sources can arise from previously unmonitored areas or when sensors that are more accurate are installed in monitored areas. Furthermore, several practitioners appreciated the possibility of selecting the information that they want to visualize in the system, mainly because it would not overload the system with useless (or not interesting) information. A geologist expressed these facts in the following way:

> "*If the system allows me removing the information that I do not want. For me, it is interesting because the system is not overloaded; for example, there are a lot of layers here [i.e., in the current information system in use at Cemaden] that I do not need or I have never used it before.*"

Additionally, a recurrent comment by the practitioners was that the decision-making in the monitoring control room is always associated with a location of interest. This happens because the required information for making a decision (e.g., flood threshold) changes from place-to-place. Hence, the integration of data based on the location is particularly valuable, as a meteorologist remarked when using the the AGORA-DS:

> "*To define the decision-making per regions is interesting because critical thresholds are very different among the regions. Therefore, a decision-making in the Southeast and South regions of Brazil is different from the Northeast. For example, in the Coast Region of São Paulo I would issue a warning when the volume of rainfall reaches 100mm but, in the Northeast, this value would be 40mm.*"

Similarly, several practitioners mentioned that a location-based analysis using the web-based application of the AGORA-DS was easier because they could understand the overall situation in an area. Furthermore, the web-based application might allow practitioners to identify potential problems in locations with conflicting data sources; for example, when sensors are indicating the possibility of an event and other sensors, which are located in the same area, are indicating a contrary situation. A geologist used the following example to explain this fact:

> "*The location is essential... Today, we issued a warning to Nova Friburgo, and among 5 to 6 stations, only one station was indicating a very high possibility of an event at the north region of the city... Because, sometimes, the members of civil protection are located in the city center that is sunny; however, in the north region, where there is a risk area, it is raining and they are not aware of it. Then, it would make easier the analysis of areas because you can see geographically the elements.*"

Together, the location-based data integration, the definition of decisions' requirements, and the management of data sources are particularly important for expanding the coverage of monitored areas in the monitoring room, and making faster and more informed decisions.

In addition to the mentioned positive implications, the analysis of the results obtained in the "building, intervention, and evaluation" stage also highlighted points in which our intervention can be improved. For instance, the principle of user-controlled data sources configuration with a mapping to required information was expected to allow users in managing easily their data sources. Thus, they would able to visualize only the most interesting data sources for supporting their decision-making. However, the feedback from one geologist highlights that most of the decision-makers in the monitoring control room might not have a knowledge or experience that could allow them to establish a communication with a data source, e.g., to gather the required information for setting up a RESTful call.

> "*When I have to configure a data source is difficult. I have to put a path here and the server could be offline...In my opinion, it is a tricky task, which requires a knowledge that I do not have and other users do not have...Nobody here is from information technology...We should not do this configuration but instead the software developers.*"

Furthermore, all the required information might not be provided by the institution due to other reasons (e.g., typos), as well as the existing information was not updated when a new version of the data sources has been set up. These issues pose challenges and could take a time that members of the monitoring control room do not have. Likewise, the use of a new data sources might require a prior bilateral agreement between the supplier (e.g., ANA) and consumer (e.g., Cemaden) due to legal issues about the use of provided data. This is beyond the scope of the monitoring control room and would be done by other departments of the center.

Another point of discussion was the time spending for configuring all the points of interest (features) one-by-one. The decision-makers in the monitoring control room of Cemaden have available in their own sensor network almost 6,000 sensors between hydrologic stations, rainfall gauges, and weather radar, to configure all these sensors would take a while to be done. This could become even more complex and arduous when we consider the crowdsourcing platforms, as well as data sources from other agencies (e.g., ANA). During the interviews, a meteorologist suggested that the "location setting" feature of the AGORA-DS should be more flexible and allow users in setting up the information not only in the sensor level but also in the region, municipality, and state levels:

> "*The configuration of sensors is interesting. However, in my opinion, the interface can be changed by selecting more than one sensor, or all the sensors of a municipality or a region of a state.*"

The quote above suggests that once a decision-maker configures the analytical value for a specific region of a municipality. This value would be assumed as an information item of a business rule by all the sensors located at this location. In turn, it can be employed for analyzing the data related to the indicated observable property. For example, a decision-maker determines that the accumulated volume of a rainfall (observable property) over the last 24 hours (reference value) at the North region of Rio de Janeiro might be dangerous when reaches 40mm (analytical value). Then, all the rainfall gauges located at this region assume this value (40mm) as an information item of their analysis. Indeed, the principle of multiscale location-based analytics still has a potential for further improvements.

6.3.2.3 Stage 4: Formalization of Learning

In the final stage, results obtained during the action design research project should be formalized into relevant lessons learned. This means to understand and describe the outcomes of the proposed interventions (the ensemble artifacts) into a more general class of solutions, i.e. to achieve the generalization of the outcomes. It can be achieved through a three levels approach (SEIN *et al.*, 2011): (1) the generalization of the problem, (2) the generalization of the solution, and (3) the definition of the intervention outcomes as a set of design principles.

Here, as mentioned before, we are addressing the particular class of field problems of SDDSs that takes into account the decision-makers' requirements in dynamic environments to make sense of heterogeneous geospatial big data. Based on the analysis of the reflection and learning results, the AGORA-DS can be definitely regarded as a representative of the class of SDSS. Likewise, this resulted in the definition of a set of design principles that can be employed for creating other instances of this class of solutions. These principles are described in the following section.

6.4 Discussions

This research aims at developing and evaluating design principles for SDSSs that embodies decision-makers' requirements of dynamic environments with heterogeneous geospatial big data. For doing so, a two-cycle action design research project was conducted within Cemaden.

In the first cycle, the action design research team was focused on identifying potential problems in the daily activities of the monitoring control room of Cemaden. Semi-structured interviews were conducted with members of that room, as well as participatory observation with a limited degree of interaction by the researcher. The evaluation of collected data raised important challenges of the application context; for example, there is a notable delay in the decision-making that is caused by the necessity of analyzing data from several different monitoring systems. These challenges provided a basis for defining both the main question and the working hypothesis of this research, which in turn were the motivations for the second cycle of action design research project.

In the second cycle, a set of initial design principles were developed: the principles of (a) user defines the information requirements of decisions, (b) user-controlled configuration of data sources with a mapping to required information, (c) location-based data integration, and (d) multiscale geovisualization. These principles were based on a theoretical framework that combines theories from supplementary areas, such as business process, decision-making, and location-based analytics. An SDSS prototype was implemented for embodying these four principles, named AGORA-DS. This allowed us to evaluate in practice their effectiveness in addressing the identified problems. The analysis of its outcomes showed that our intervention with the proposed design principles raised both positive implications for practice and research, as well as some points that still require further improvements. Therefore, the design principles were revised and refined with the aim of better reflecting these new requirements, and thus providing relevant contributions for design theory, as well as important implications for practice and research.

6.4.1 Design Theory of an SDSS for Dynamic Environments with Heterogeneous Geospatial Big Data

This section details both the design principles and the revised principles that were developed in the action design research project.

6.4.1.1 Principle of User defines the Information Requirements of Decisions

The principle of *user defines the information requirements of decisions* determines that a decision should be defined into the system according to its required information as it provides a better understanding of its variables. The reason of this principle was to define which information is required and essential for making a specific decision. By allowing a user to defining this relationship, a system becomes flexible enough for addressing the intrinsic dynamicity of an application context. On the other hand, decisions-makers have a clear overview about the lack of data to fulfill the required information.

The evaluation in practice by means of the AGORA-DS showed that this principle is particularly relevant because it could reduce the information overload that can be caused by the volume of big data, i.e., only the required information is displayed.

6.4.1.2 Revised Principle of User-controlled Configuration of Data Sources with a Mapping to Required Information

The principle of *user-controlled configuration of data sources with a mapping to required information* specifies that data sources should be configured for fulfilling decisions' requirements.

This principle was motivated by the desire to provide a clear understanding about which data sources are more appropriate for supporting a decision-making. Furthermore, the mapping with required information would provide a basis for decision-makers develop new data collection methods or identify lack of appropriate sources of information.

The user evaluation, as implemented in the AGORA-DS, indicated that the users have found this principle useful for their decision-making. The possibility of managing data sources for the analysis in particularly valuable in the application context because new sources can emerge in real-time, while existing sources may be rendered ineffective. Similarly, the inclusion of data sources acts as a supplementary source of information, which thus expands the coverage of the monitored areas or provides more accurate information in locations of existing data sources. On the other hand, the technological configuration of the communication with a data source should be understood as beyond the scope of the end-users. It is because most of them might not have a knowledge for achieving such configuration or a prior bilateral agreement should be established.

Given these points, it has become clear that the decision-makers are not interested in configuring the data sources but rather in choosing from a pre-determined list, which data sources should be visualized on the system, as a geologist remarked *"I would have the possibility of changing the data sources. This is interesting."*. In light of this, we suggest that the principle of user-controlled data sources configuration with a mapping to required information be renamed to "user-selected data sources visualization". This means that SDSS should allow a user in selecting the data sources of which he wants to visualize data from a pre-determined list of available data sources. Therefore, users still have a clear understanding about the most useful data sources but are not responsible for configuring the data source.

6.4.1.3 Revised Principle of Location-based Data Integration

Since available data might be often associated with a particular point of interest, the principle of *location-based data integration* specifies that this location should be used as a common characteristic for achieving the integration. This design principle was motivated by the fact that there is a remarkable necessity in continuously changing the decisions' variables and measurement values per location. Through the AGORA-DS, this principle was evaluated in practice. The results obtained showed that users found this principle interesting and it could enhance their decision-making because they still have different information requirements among the locations.

However, the evaluation results also highlighted an important point for improving this principle. The system should allow users to configure the data integration for individual locations, as well as for a group of locations (e.g., a neighborhood). This is because a decision-maker might be interested in a more general level of analysis (e.g., a municipality) rather that a more specific level (e.g., a road). In this manner, we suggest that the principle of location-based data integration be refined to read “multiscale geospatial data integration”. By supporting a configuration in different spatial scales, the systems addresses the problem raised in practice and should provide a flexible configuration of data integration for users. For instance, a decision-maker could determine that the analytical value of rainfall for the whole State of Pernambuco is 60mm; however, due to particular geological issues in some neighborhoods of Recife, the analytical value is 30mm of rainfall. In other words, a user is able to configure the system from a general location (e.g., a state) to a more specific location (e.g., a region of a municipality).

6.4.1.4 Principle of Multiscale Geovisualization

The principle of *multiscale geovisualization* specifies that a system should display the data by means of a multiscale visualization approach. On the one hand, this aims at making easier the analysis of geospatial big data by filtering them into more consolidated levels of visualization. While, at the same time, it might make faster the decision-making in dynamic environments because modifications in the information of more specific levels are reflected on more general levels of visualization. For instance, a traffic manager can easily identify those neighborhoods that require more attention when he is analyzing the “big picture” of the current situation of his municipality.

The user evaluation, as it was implemented in the AGORA-DS, revealed that this principle is valuable for supporting the analysis of big data. It is because it provides a consolidated overview of the available data instead of a huge volume of individual data. This in turn makes faster and more practical the decision-making since users do not waste time by analyzing thousands of data.

Furthermore, the principles of multiscale geovisualization and revised version of the location-based data integration seemed to embody valuable features that could make easier and faster the analysis of big data. In the same manner, it has a great potential for addressing the necessity of analyzing data from different systems. Figure 53 summarizes the design theory for SDSS that embodies decision-makers’ requirements for supporting decision-making in dynamic environments with heterogeneous geospatial big data.

6.4.2 Implications for Practice and Research

Three important implications for practice and research were also identified from the development of design principles and their implementation in practice. To start with, the theoretical framework that was developed based on standard models from business process and geospatial analytics (Figure 48) can be regarded as another valuable contribution of this research. Although more research is still needed, we believe that our findings with this framework contribute to a better understanding of the relationship among the stakeholders, business tasks, information, and data. This in turn is important because it allows practitioners and researchers in describing the associations among the data, which might affect the analysis of data provided by different sources.

Problems from practice

- There is not a clear understanding about the information required for decision-making.
- Delayed decision-making caused by the necessity of analyzing different data sources.
- Necessity of continuously changing the decisions' variables and values per location

Requirements for SDSS of dynamic environments

- Decisions should consist of information requirements and a decision rule.
- A decision rule should consist of a set of information items that must be associated with an information requirement required by the decision.
- Information items should require an analytical value, which supports data analytics.
- Features of interest should be associated with a set of specific properties, which aims at characterizing and describing them.
- Information items of decision rules should be associated with properties of a feature of interest, which in turn supports the understanding of the lack data
- Since a property can characterize features of interest, its descriptive value should be determined by the observations.
- Information items should be thus associated with appropriate data sources.
- The association of information with data sources should be flexible for allowing the inclusion (or exclusion) of data sources, as well as the changing of existing data sources.
- A connection string should be required for allowing the authentication with the data source.
- A data source should consist of a group of individual observations (also known as data).
- Observations should require a date of creating and a descriptive value.
- Once information items are associated with properties of a feature of interest, this should allow the data integration based on their associated location.
- Decision rules provide an understanding of how information items should be combined.
- A feature of interest is associated with a geographic location, which in turn could be represented either by an individual point or by a polyline (i.e., square or triangle).
- A point should require two essential information for determining its geographic location: "x" and "y", while a polyline requires a type and the spatial reference.
- Feature of interest may be associated with other features of interest, which in turn should allow a multiscale geovisalization.

Design principles for SDSS of dynamic environments with heterogeneous geospatial data sources

- User defines the information requirements of decisions
- User-selected data sources visualization
- Multiscale geospatial data integration
- Multiscale geovisualization

Figure 53 – A design theory of an SDSS for supporting decision-making in dynamic environments with heterogeneous geospatial big data.

For instance, a supplier reputation can be determined through the analysis of social media and historical database. The results obtained with the framework also determine that available data should be always related to one or more observable property, which aims at characterizing the data, e.g., the value (observable property) of a product is U$ 28,00 (data). Therefore, a specific point of interest (e.g., a road) consists of a set of observable properties (e.g., speed limit and traffic condition) that in turn could be measured by determined data (e.g., 40km/h and heavy traffic). It is worthwhile to mention that the data that measure a property can change along the time, e.g., when a major of the municipality decides to change the speed limit from 40km/h to 50km/h.

The findings of this research also suggest that practitioners and researchers should understand how these data would be useful for fulfilling their information requirements before starting the use of geospatial big data. On the one hand, this aims at avoiding the visualization of information that is not required for decision-making and thus might become useless. On the other hand, it could save time that would be spent on the processing a huge volume of data that might not answer target users' questions. These findings are in line with previous works that are focused on analyzing the use of big data in decision-making (DOLIF *et al.*, 2013; VIEWEG; CASTILLO; IMRAN, 2014; GOPALKRISHNAN *et al.*, 2012; TAYLOR *et al.*, 2013; WAMBA *et al.*, 2015). More importantly, they lack in describing a design process for decision support systems that address these issues.

Furthermore, when designing an SDSS for dynamic environments with heterogeneous big data sources, it might require a view from the socio-technical perspective. This means that software designers should employ a three-step cycle for building an IT-artifact, i.e., a) they first have to understand the decision-makers' requirements (in regarding the data visualization and information requirements); b) interventions are proposed and experienced by end-users in their working environments; c) finally further improvements can be proposed from the lessons learned. In this manner, the design principles could act as guidelines for supporting in the design process for the particular context of dynamic environments with heterogeneous big data source. Likewise, these principles could also set up a new research agenda for design research in the area of SDSS applied to dynamic environments.

6.5 Final Remarks

This chapter has presented a set of design principles that are focused on SDSS for disaster management. However, they are able to support the design and development of SDSS for other dynamic environments with heterogeneous geospatial data sources (e.g., traffic management). These were derived from a two-cycle action research project with joint meetings, semi-structured interviews, and focus group sessions within Cemaden.

Results of our empirical evaluation indicated that the design principles are able to support the design of SDSS that uses a high volume of heterogeneous geospatial data for dynamic environments (e.g., in scenarios with ad hoc decisions). Since an action research aims at addressing a practical problem as a goal of a research project, the contributions provided to both practice and research are manifold. Firstly, the consolidated view of the overall situation provided by the SDSS prototype was mentioned by the decision-makers of the evaluation context as an essential feature that would save time in the data analysis, and thus make faster the decision-making. Furthermore, the overall situation established through the integration of data provided by distinct and heterogeneous data sources, such as hydrological stations, weather radars, and rainfall gauges, is relevant for establishing a more accurate "big picture". The user evaluation also highlighted that the configuration of required information impacts positively decision-making because only the required information is displayed to the decision-makers. The SDSS prototype itself provides a useful contribution to practice because it provides IT-based artifacts that allow the development of better SDSS while, on the other hand, it improves the data analysis of end-users. Finally, the design principles could also set up a new research agenda for design research in the area of SDSS applied to dynamic environments.

CHAPTER

7

CONCLUSIONS

Advances in information technology, such as more sophisticated smartphones, improvement of GPS, emergence of cloud computing, and Internet of Things (IoT) have opened up further opportunities for organizations and communities to broaden their understanding of these areas. For example, wearable devices can be used for remote health monitoring and emergency notification systems (CHEN; MAO; LIU, 2014; ZHOU; FU; YANG, 2016). However, while all these data have a great potential for making improvements, they also produce an overwhelming flow of heterogeneous “big data” in either structured or unstructured formats, which can hamper the decision-making (HASHEM *et al.*, 2015). This was particularly highlighted in the representative scenario of disaster management by the literature reviews and empirical studies that were conducted throughout this project. For instance, the available big data might not be shared in the right format or be easily accessible. They are also very dispersed, disconnected, and prone to errors, and as a result, be of no value in supporting tasks of the decision-makers (VIEWEG; CASTILLO; IMRAN, 2014; DOLIF *et al.*, 2013).

In this context, this project can be characterized as being at a low level of solution maturity (the integration of heterogeneous geospatial big data and connection between decision-making with data sources) and a high level of application domain maturity (the use of big data for supporting decision-making). The combination of these two features is defined by Gregor & Hevner (2013) as “Improvement” in their Knowledge Contribution Framework, i.e., “the development of new solutions for knowledge problems”. The knowledge contributions thus consist of the development (or improvement) of new (or existing) solutions to a well-known problem. For a better means of recognizing and describing these “contributions”, we relied on another distinction drawn by Gregor & Hevner (2013) between descriptive (denoted by Ω), which represents the knowledge of natural phenomena and the laws and regular patterns of phenomena (“what”), and prescriptive (denoted by Λ), which consists of the knowledge of human-built artifacts (“how”). A Λ knowledge can be represented by the following types: constructs, models, methods or instantiations. Overall, this project provides key contributions to both types of knowledge (as summarized in Figure 54) as they advance the current state of the art with regard to the use of heterogeneous geospatial big data to ensure more effective decision-making.

Since disaster management was selected as a representative scenario for decision-making based on heterogeneous geospatial big data, previous studies were important because they characterize the use of VGI and SDSS in disaster management and open up further lines of research in the field. In this manner, a conceptual architecture that integrates heterogeneous data sources was developed and evaluated in the context of flood management. This architecture and lessons learned from the evaluation served as a basis for a better understanding of the integration of heterogeneous data sources (in the Ω knowledge). Moreover, the existing literature on the standard models and notations was drawn on to establish a model-based framework that connects decision-making with data sources.

This consists of two elements: an integrated model and notation (oDMN$^+$) and a modeling process. The employment of the framework for modeling the early-warning decision-making of a Brazilian early-warning center (Cemaden) has also led to Ω knowledge with regard to understanding decision-making with heterogeneous big data and the connection of decision-making to data sources. This knowledge was then employed for defining the design principles that must be adhered to when designing an SDSS to effectively support decision-making with heterogeneous geospatial big data. An evaluation of end-users within Cemaden also provided a better understanding of decision-making with heterogeneous geospatial big data, as well as the design of an SDSS as a socio-technical construction, i.e., ensuring equal consideration is given to technical and human factors in the design process. In addition, it involved the users in determining the quality required for improvements in working-life (MUMFORD, 2006). These knowledge sources were supplemented with information about the application environment, which was essential to characterize the research domain, identify any existing problems, and raise the research question. It was also important for setting out guidelines for the assessment of the acquired knowledge. Furthermore, key elements dependent on human capabilities were also defined for the project. These capabilities were essential to determine the research skills necessary for carrying out the activities, as well as for ensuring the success of the project.

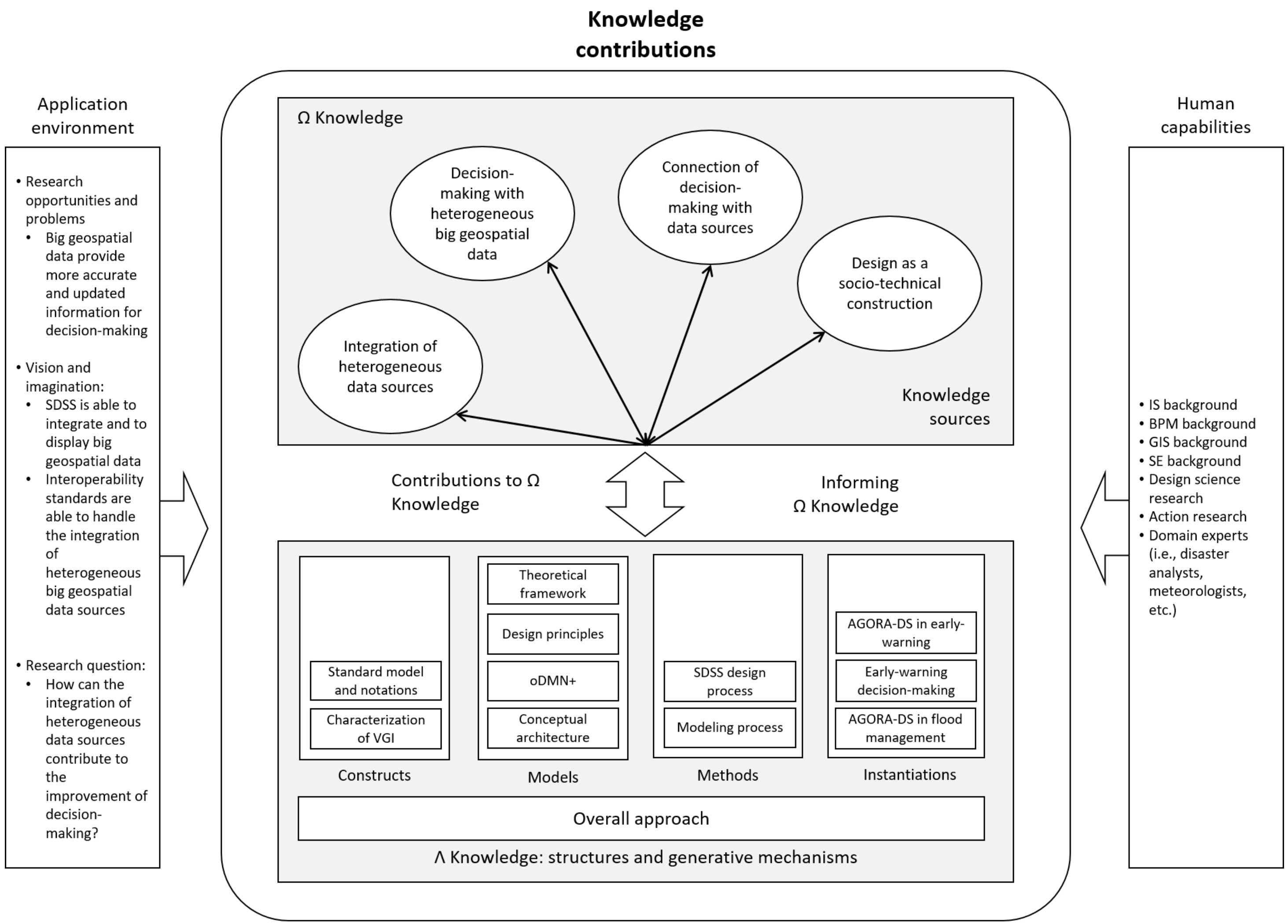


Figure 54 – Summary of Knowledge Contributions. Adapted from Gregor & Hevner (2013).

Overall, all this acquired knowledge has proved to be an important addition to the existing theoretical literature with regard to the integration of heterogeneous data sources, the connection of decision-making with data sources, and the design of an SDSS to effectively assist decision-making with heterogeneous geospatial big data (MANSOURIAN *et al.*, 2006; MARKOVIC; STANIMIROVIC; STOIMENOV, 2009; WAN *et al.*, 2014; BARTON; COURT, 2012; HORITA *et al.*, 2013; VIEWEG; CASTILLO; IMRAN, 2014; KLEINDIENST; PFLEGER; SCHOCH, 2015; WAMBA *et al.*, 2015); for example, by introducing oDMN$^+$. This also has serious implications for practice; for example, the conceptual architecture that supports the integration of heterogeneous data sources, as well as the flexible and simple inclusion of new data sources. Furthermore, this acquired knowledge provides a solid basis for conforming to the guiding principles that were defined in the Sendai Framework (UN, 2015). For example, oDMN$^+$ introduced in Chapter 5 is suitable for providing a better understanding of the target users' needs, while the design principles introduced in Chapter 6 are required for the design of early warning systems and disaster risk and emergency communications mechanisms. Likewise, these principles are of great assistance to the Brazilian National Strategy for Natural Disaster Risk and Response Management and Cemaden, for example, by outlining some important lessons that should be learnt when integrating the heterogeneous data provided by available monitoring systems.

This chapter is structured as follows: the key objectives of this thesis outlined above are summarized and revised in Section 7.1. Section 7.2 describes some of the limitations of this thesis and discusses how they can be overcome. Moreover, it also introduces some future lines of research that can be explored on the basis of the obtained results.

7.1 Revisiting the Thesis Contributions

This section summarizes and revises the contributions of this thesis.

- **Characterization of the use of VGI in disaster management.** we investigated the existing literature with the aim of filling in some of the remaining gaps, as well as anticipating future lines of research (see Section 3.1.1). Results obtained showed that the collaborative platforms (e.g., citizen observations) for supporting disaster management still require more research. Furthermore, they also provide evidence that more empirical research is needed for better understanding of how this volunteered information can be used by end-users; for example, by conducting an action research project within an emergency service.

- **Characterization of the use of SDSS for supporting decision-making in disaster management.** we conducted a systematic mapping study that sought to determine how the VGI has been used to assist decision-making in disaster management (see Section 3.1.2). The results highlighted the gaps in research and practice, e.g., it was showed that the volunteered information has been providing valuable assistance in several tasks, such as forecasting events and understanding the need for situational awareness; however, further research is still required in the area of early warning systems, as well as the monitoring of environmental variables.

- **Definition of a conceptual architecture that integrates heterogeneous data sources.** A conceptual architecture has been set out here for integrating heterogeneous data sources (see Section 4.2) that was based on the layer-based architecture designed by Horita *et al.* (2013). The novelty of the architecture was that it combined interoperability standards and data source adapters for handling the intrinsic heterogeneity of the data sources. Our idea with this was to provide a solid basis for understanding how heterogeneous data sources can be integrated.

- **Lessons learned about the integration of heterogeneous data sources in disaster management.** The new conceptual architecture was implemented and deployed to support decision-making in a real-world scenario for flood risk management (see Section 4.4). The generated system aimed at integrating data provided by a WSN (see Hughes *et al.* (2011)) and a volunteered data collected via a citizen observation platform (see Degrossi *et al.* (2014)). The results obtained in the case study provided evidence that the conceptual architecture is useful both for integrating heterogeneous data sources and also for sharing the integrated data, as well as visualizing them within a simple web-based decision support system.

- **Establishment of an integrated model and notation that describes the connection of decision-making and data sources.** We designed an integrated model and notation, named oDMN^{+}, which describes a) how business tasks are connected to decisions, then b) which information is required by them, and finally c) what data source could be used for providing this required information (see Section 5.3.1). It was based on a standard model and notations (e.g., BPMN, DMN, and O&M), as well as on the model established by Horita *et al.* (2016). This particularly helped in understanding how to make better use of the available data sources that provide useful information for decision-making. It can also assist in locating those activities where there is a lack of information, as well as finding the available data sources that can help.

- **Development of a modeling process for modeling the connection of decision-making and data sources.** A modeling process was devised to support the employment of oDMN^{+} in practice (see Section 5.3.2). This defines five activities for systematically deriving conceptual elements from an application, e.g., business processes together with their activities, related decisions, their information requirements, and the used data sources. This modeling process was expected to assist in establishing a connection between decision-making and data sources.

- **Lessons learned from planned decision-making within a Brazilian early warning center.** We carried out a case study in a Brazilian early warning center for modeling the decision-making that is required for issuing warnings to vulnerable communities about impending disasters. The modeling process for employing oDMN$^+$ (see Section 5.4) was used for this. The results of this case study showed that oDMN$^+$ is able to connect the decision-makers' tasks and decisions with the data sources. Furthermore, they showed that the modeling process successfully obtains conceptual elements from the decisions-makers in the context of an application, as well as enabling the oDMN$^+$ to be employed in practice.

- **Development of design principles for SDSS to support decision-making with heterogeneous geospatial big data.** A set of design principles was formulated to assist the design of an SDSS which could be applied to decision-making with heterogeneous geospatial big data (as discussed in Chapter 6). These were derived from a cross-collaborative action research project that was conducted in collaboration with a Brazilian early warning center. The purpose of these principles was threefold: (a) to guide the development of SDSS, (b) to embody the decision-makers' requirements within the system, (c) and to improve decision-making with heterogeneous geospatial big data.

- **Empirical evaluation of the design principles in disaster management.** Within the scope of the ADR project, we evaluated the design principles with members of the monitoring control room of a Brazilian early warning center (see Chapter 6). The results provide solid evidence that the principles effectively achieve their objective, i.e., to improve decision-making with heterogeneous geospatial big data.

7.2 Limitations and Future Work Lines

The main limitations of this thesis can be summarized as follows:

- **Limitations in the systematic mapping studies.** Both SMSs were only able to identify quite a small number of selected papers (i.e., 21 and 13 primary studies). This can be attributed to the fact that i) we did not include more synonyms in the search strings used, ii) or include manual searches in particular conference proceedings, or iii) supplement the search with any other sources apart from the results achieved from the search engines. Furthermore, the SMSs should be updated in order to identify further studies that might be published from the time of reviews (i.e., 2013 and 2014) until 2017. Identified results could highlight new trends of research, as well as new limitations and challenges.

- **Case studies.** oDMN$^+$ can be regarded as an extension of previous work, which also aimed at increasing the generalization of the model and notation. The scope of the evaluation was limited and there is still room for more case studies, especially in other research areas, e.g., business management and marketing analysis.

- **Evaluation of the design principles in a wider organizational setting.** the design principles generated after Cycle 2 of the action research project were already refined since they had been based on an evaluation that was carried out in a limited organizational setting. The next stage in the project would be to implement the revised and refined design principles in the SDSS prototype and carry out a further evaluation in a wider organizational setting, e.g., an evaluation of users in everyday work for a 10-week period. This might provide further insights into the usefulness of the principles, as well as the way they can be further refined.

- **Pre-processing and processing of data** while the development of conceptual architecture and design principles, the pre-processing and processing of data should be understood as being beyond the scope of this research project. Therefore, it was assumed that all data shared by the heterogeneous big data sources have been already filtered and pre-processed. This research work also did not focus on the missing of data or unavailability of data.

Furthermore, several issues that were raised while carrying out this research project were not fully explored because of time constraints, organizational matters, or because they were beyond its scope. These are particularly related to the question of making improvements (e.g., generalizing the results more widely) or acting as a means of addressing the limitations described above. In this manner, these questions can guide the course of future work in the field.

- The employment of the conceptual architecture to assist decision-making in other domains; for instance, business, marketing, and transport. Furthermore, additional “adapters” (see Chapter 4) can be included to allow the integration of other sources of information, e.g., collaborative mapping platforms.

- Since the use of a model and notation requires an understanding of the metamodel, the use of a supporting tool would make the tasks of modeling easier in practice, such as either a standalone tool or a plug-in for existing tools (e.g., Signavio or Bonita). This could also be of value in future studies.

- oDMN$^+$ should be employed in further case studies. In the same manner, the modeling process should be employed for modeling the decision-making in other application contexts. These further case studies could not only seek to improve the generalization of the achieved results but also determine whether there are any missing conceptual elements.

- In the area of disaster management, oDMN$^+$ should be also employed for modeling decision-making of different levels of the command chain; for example, activities of relief teams on the ground. Having modeled activities of existing chains, it would be possible to analyze the impact of a wrong decision on the related activities on the command chains.

- Another future line of inquiry could focus on the use of oDMN^{+} for explaining the relationship between people, their tasks, and available information. This is a potential means of enhancing the social web or semantic web by improving social media analytics or creating task-related ontologies.

- Future studies should also concentrate on employing the design principles in designing SDSS for other dynamic environments with heterogeneous data sources apart from disaster management; for example, traffic flow or the transport system in a smart city. This also opens up interesting opportunities for helping to design "systems-of-systems", which could also be improved by using the conceptual architecture.

REFERENCES

AA, H. van der *et al.* Integrated process and decision modeling for data-driven processes. In: *Proceedings of the 3rd International Workshop on Decision Mining & Modeling for Business Processes (DeMiMoP)*. Innsbruck, Austria: [s.n.], 2015. p. 1–12. Cited on page 69.

ADAM, N.; SHAFIQ, B.; STAFFIN, R. Spatial Computing and Social Media in the Context of Disaster Management. **Intelligent Systems, IEEE**, v. 27, n. 6, p. 90–96, 2012. Citations on pages 77 and 78.

AHMAD, S.; SIMONOVIC, S. P. An Intelligent Decision Support System for Management of Floods. **Water Resources Management**, v. 20, n. 3, p. 391–410, 2006. Cited on page 57.

AHMED, A. Use of Social Media in Disaster Management. In: **Proceedings of the 33th International Conference on Information Systems (ICIS)**. Shanghai, China: [s.n.], 2011. p. 1–11. Citations on pages 65, 75, 77, 78, and 86.

AITSI-SELMI, A.; EGAWA, S.; SASAKI, H.; WANNOUS, C.; MURRAY, V. The sendai framework for disaster risk reduction: Renewing the global commitment to people's resilience, health, and well-being. **International Journal of Disaster Risk Science**, v. 6, n. 2, p. 164–176, 2015. Cited on page 42.

ALBUQUERQUE, J.; HERFORT, B.; ECKLE, M. The tasks of the crowd: A typology of tasks in geographic information crowdsourcing and a case study in humanitarian mapping. **Remote Sensing**, v. 8, n. 10, p. 1–22, 2016. Cited on page 40.

ALBUQUERQUE, J. P.; ECKLE, M.; HERFORT, B.; ZIPF, A. European handbook of crowdsourced geographic information. In: ____. [S.l.]: Ubiquity Press, 2016. cap. Crowdsourcing geographic information for disaster management and improving urban resilience: an overview of recent developments and lessons learned, p. 309–321. Cited on page 88.

ALBUQUERQUE, J. P.; HERFORT, B.; BRENNING, A.; ZIPF, A. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. **International Journal of Geographical Information Science**, v. 29, n. 4, p. 1–23, 2015. Citations on pages 40, 65, 130, and 133.

ALBUQUERQUE, J. P.; HORITA, F. E. A.; DEGROSSI, L. C.; ROCHA, R. S.; ANDRADE, S. C.; RESTREPO-ESTRADA, C. Volunteered geographic information and the future of geospatial data. In: ____. [S.l.]: Idea Group Inc, 2017. cap. Leveraging Volunteered Geographic Information for improving disaster resilience: Lessons learned with AGORA and future research directions. Citations on pages 20, 27, and 51.

ALBUQUERQUE, J. P.; UEYAMA, J.; MENDIONDO, E.; CARAMORI, V. Using Wireless Sensor Networks for Urban Flood Risk Adaptation in Brazil. In: **Proceedings of the Geoinformatik**. Heidelberg, Germany: [s.n.], 2013. p. 1–3. Cited on page 91.

ALBUQUERQUE, J. P.; ZIPF, A. Collaborative information systems for disaster management: Building resilience against disasters by combining participatory environmental monitoring and vulnerability communication. In: **Alumi seminar natural hazards - research on natural disasters, civil defense, disaster prevention, and aid**. Teresópolis, Rio de Janeiro, Brazil: [s.n.], 2012. p. 71–74. Cited on page 50.

ANTUNES, P.; SIMOES, D.; CARRIĆO, L.; PINO, J. A. An end-user approach to business process modeling. **Journal of Network and Computer Applications**, v. 36, n. 6, p. 1466–1479, 2013. Cited on page 129.

ARNOTT, D.; PERVAN, G. Eight key issues for the decision support systems discipline. **Decision Support Systems**, v. 44, p. 657–672, 2008. Cited on page 59.

ASSIS, L. F. F.; HERFORT, B.; STEIGER, E.; HORITA, F. E.; ALBUQUERQUE, J. P. de. A geographic approach for on-the-fly prioritization of social-media messages towards improving flood risk management. In: **Proceedings of the 4th Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)**. Recife, Brazil: [s.n.], 2015. p. 1–12. Cited on page 21.

ASSIS, L. F. F. G.; BEHNCK, L. P.; DOERING, D.; FREITAS, E. P.; PEREIRA, C. E.; HORITA, F. E. A.; UEYAMA, J.; ALBUQUERQUE, J. P. Dynamic sensor management: Extending sensor web for near real-time mobile sensor integration in dynamic scenarios. In: **Proceedings of the 30th International Conference on Advanced Information Networking and Applications (AINA)**. [S.l.: s.n.], 2016. p. 303–310. Citations on pages 20 and 40.

ASSIS, L. F. G.; HERFORT, B.; STEIGER, E.; HORITA, F. E. A.; ALBUQUERQUE, J. P. de. Geographical prioritization of social network messages in near real-time using sensor data streams: an application to floods. In: **Proceedings of the XVI Brazilian Symposium on GeoInformatics (Geoinfo)**. Campos do Jordão, Brazil: [s.n.], 2015. p. 26–37. Cited on page 20.

ATKINSON, S. F.; CANTER, L. W. Assessing the cumulative effects of projects using geographic information systems. **Environmental Impact Assessment Review**, v. 31, n. 5, p. 457–464, 2011. Cited on page 60.

BAHARIN, S.; SHIBGHATULLAH, A.; OTHMAN, Z. Disaster management in Malaysia: An application framework of integrated routing application for emergency response management system. In: **Proceedings of the 2009 International Conference of Soft Computing and Pattern Recognition (SOCPAR)**. Washington, USA: [s.n.], 2009. p. 716–719. Cited on page 56.

BARROS, R. M.; MENDIONDO, E. M.; WENDLAND, E. Cálculo de áreas inundáveis devido a enchentes para o plano diretor de drenagem urbana de São Carlos (PDDUSC) na bacia escola do córrego do Gregório. **Revista Brasileira de Recursos Hídricos**, v. 12, n. 2, p. 5–17, 2007. Cited on page 96.

BARTON, D.; COURT, D. Making advanced analytics work for you. **Harvard Business Review**, v. 90, n. 10, p. 1–7, 2012. Citations on pages 41 and 168.

BASKERVILLE, R. L. Investigating information systems with action research. **Communications of the Association for Information Systems (AIS)**, v. 2, n. 3, p. 1–33, 1999. Cited on page 134.

BASKERVILLE, R. L.; WOOD-HARPER, A. T. A critical perspective on action research as a method for information system research. **Journal of Information Technology**, v. 11, p. 235–246, 1996. Cited on page 134.

BATOULIS, K.; MEYER, A.; BAZHENOVA, E.; DECKER, G.; WESKE, M. Extracting decision logic from process models. In: **Proceedings of the 27th International Conference on Advanced Information Systems Engineering (CAiSE)**. Stockholm, Sweden: [s.n.], 2015. p. 349–366. Cited on page 69.

BAZHENOVA, E.; WESKE, M. Deriving decision models from process models by enhanced decision mining. In: **Proceedings of the 3rd International Workshop on Decision Mining & Modeling for Business Processes (DeMiMoP)**. Innsbruck, Austria: [s.n.], 2015. p. 1–12. Citations on pages 69 and 129.

BECKER, J.; ROSEMANN, M.; UTHMANN, C. Guidelines of business process modeling. In: AALST, W. V.; DESEL, J.; OBERWEIS, A. (Ed.). **Business Process Management: Models, Techniques, and Empirical Studies**. Germany: Springer Berlin Heidelberg, 2000. p. 30–49. Cited on page 129.

BHAROSA, N.; MEIJER, S.; JANSSEN, M.; BRAVE, F. Are we prepared? experiences from developing dashboards for disaster preparation. In: **Proceedings of the 7th International Conference on Information Systems for Crisis Response and Management (ISCRAM)**. Seattle, USA: [s.n.], 2010. p. 1–6. Cited on page 61.

BIARD, T.; MAUFF, A. L.; BIGAND, M.; BOUREY, J.-P. Separation of decision modeling from business process modeling using new “decision model and notation” (dmn) for automating operational decision-making. In: **Proceedings of the 16th IFIP WG 5.5 Working Conference on Virtual Enterprises Risks and Resilience of Collaborative Networks (PRO-VE)**. Albi, France: [s.n.], 2015. p. 489–496. Citations on pages 27 and 69.

BILLA, L.; SHATTRI, M.; MAHMUD, A. R.; GHAZALI, A. H. Comprehensive planning and the role of SDSS in flood disaster management in Malaysia. **Disaster Prevention and Management: An International Journal**, v. 15, n. 2, p. 233–240, 2006. Citations on pages 37, 59, and 60.

BIOLCHINI, J.; MIAN, P. G.; NATALI, A. C. C.; TRAVASSOS, G. H. **Systematic Review in Software Engineering.** Rio de Janeiro, Brazil, 2005. RT-ES 679/05, Systemas Engineering and Computer Science Department/UFRJ. Cited on page 71.

BISHR, M.; JANOWICZ, K. Can we Trust Information ? – The Case of Volunteered Geographic Information. In: **Proceedings of the Workshop Towards Digital Earth: Search, Discover and Share Geospatial Data**. Berlin, Germany: [s.n.], 2010. p. 1–6. Citations on pages 66 and 133.

BRERETON, P.; KITCHENHAM, B. A.; BUDGEN, D.; TURNER, M.; KHALIL, M. Lessons from applying the systematic literature review process within the software engineering domain. **Journal of Systems and Software**, v. 80, n. 4, p. 571–583, 2007. Cited on page 48.

BROERING, A.; FOERSTER, T.; JIRKA, S.; PRIESS, C. Sensor bus: An intermediary layer for linking geosensors and the sensor web. In: **Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research Application**. Washington, USA: [s.n.], 2010. p. 1–8. Cited on page 103.

BROERING, A.; JIRKA, S.; FOERSTER, T. The sensor bus - integrating geosensors and the sensor web. In: **Proceedings of the 2nd Open Source GIS UK Conference (OSGIS)**. Nottingham, UK: [s.n.], 2010. p. 1–4. Cited on page 103.

BROOKS, L.; ALAM, M. S. Designing an information system for updating land records in bangladesh: Action design ethnographic research (ADER). **Information Systems Frontiers**, v. 17, n. 1, p. 79–93, 2015. Cited on page 134.

BUDGEN, D.; TURNER, M.; BRERETON, P.; KITCHENHAM, B. Using mapping studies in software engineering. In: **Proceedings of the 20th Annual Meeting of the Psychology of Programming Interest Group (PPIG)**. Lancaster, UK: [s.n.], 2008. p. 195–204. Cited on page 71.

BUGS, G.; GRANELL, C.; FONTS, O.; HUERTA, J.; PAINHO, M. An assessment of public participation gis and web 2.0 technologies in urban planning practice in canela, brazil. **Cities**, v. 27, n. 3, p. 172–181, 2010. Cited on page 65.

CARROLL, L. N.; AU, A. P.; DETWILER, L. T.; FU, T. chieh; PAINTER, I. S.; ABERNETHY, N. F. Visualization and analytics tools for infectious disease epidemiology: A systematic review. **Journal of Biomedical Informatics**, v. 51, p. 287–298, 2014. Cited on page 133.

CHANG, K.; WAN, S.; LEI, T. Development of a spatial decision support system for monitoring earthquake-induced landslides based on aerial photographs and the finite element method. **International Journal of Applied Earth Observation and Geoinformation**, v. 12, n. 6, p. 448–456, 2010. Cited on page 60.

CHEN, A. Y.; MORA, F. Peña; OUYANG, Y. A collaborative GIS framework to support equipment distribution for civil engineering disaster response operations. **Automation in Construction**, v. 20, n. 5, p. 637–648, 2011. Cited on page 60.

CHEN, H.; CHIANG, R. H. L.; STOREY, V. C. Business intelligence and analytics: From big data to big impact. **MIS Quarterly**, v. 36, n. 4, p. 1165–1188, 2012. Citations on pages 42 and 140.

CHEN, L.; LIU, Y.; CHAN, K. Integrated community-based disaster management program in taiwan: A case study of shang-an village. **Natural Hazards**, v. 37, n. 1-2, p. 209–223, 2006. Citations on pages 77 and 78.

CHEN, M.; MAO, S.; LIU, Y. Big data: A survey. **Mobile Networks and Applications**, Secaucus, NJ, USA, v. 19, n. 2, p. 171–209, 2014. Citations on pages 65 and 165.

CHEN, X.; KWAN, M.-P.; LI, Q.; CHEN, J. A model for evacuation risk assessment with consideration of pre- and post-disaster factors. **Computers, Environment and Urban Systems**, v. 36, n. 3, p. 207–217, 2012. Cited on page 128.

COLEMAN, D. J.; GEORGIADOU, Y.; LABONTE, J.; OBSERVATION, E.; CANADA, N. R. Volunteered Geographic Information : the nature and motivation of producers. **International Journal of Spatial Data Infrastructures**, v. 4, p. 332–358, 2009. Cited on page 65.

CRONHOLM, S.; GOLDKUHL, G. Strategies for information systems evaluation – six generic types. **Electronic Journal of Information Systems Evaluation**, v. 6, n. 2, p. 65–74, 2003. Cited on page 152.

CROSSLAND, M.; WYNNE, B.; PERKINS, W. Spatial decision support systems: An overview of technology and a test of efficacy. **Decision Support Systems**, v. 14, n. 3, p. 219–235, 1995. Cited on page 59.

DAVISON, R. M.; MARTINSONS, M. G.; OU, C. X. J. The roles of theory in canonical action research. **MIS Quarterly**, v. 36, n. 3, p. 763–786, 2012. Cited on page 134.

DEGROSSI, L. C.; ALBUQUERQUE, J. P.; FAVA, M. C.; MENDIONDO, E. M. Flood Citizen Observatory: a crowdsourcing-based approach for flood risk management in Brazil. In: **Proceedings of the 26th International Conference on Software Engineering and Knowledge Engineering (SEKE)**. Vancouver, Canada: [s.n.], 2014. p. 1–6. Citations on pages 27, 40, 57, 66, 93, 96, 97, 104, and 169.

DEGROSSI, L. C.; AMARAL, G. G.; VASCONCELOS, E. S. M.; ALBUQUERQUE, J. P.; UEYAMA, J. Using wireless sensor networks in the sensor web for flood monitoring in Brazil. In: **Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management (ISCRAM)**. Baden Baden, Germany: [s.n.], 2013. p. 458–462. Cited on page 96.

DEMIRKAN, H.; DELEN, D. Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. **Decision Support Systems**, v. 55, n. 1, p. 412–421, 2013. Cited on page 131.

DENSHAM, P. J. Geographical information systems: principles and applications. In: ____. **Geographical information systems: principles and applications**. [S.l.]: Longman Scientific & Technical, 1991. cap. Spatial Decision Support System, p. 403–412. Citations on pages 59, 60, and 151.

DOAN, A.; RAMAKRISHNAN, R.; HALEVY, A. Y. Crowdsourcing systems on the World-Wide Web. **Communications of the Association for Computing Machinery (ACM)**, New York, NY, USA, v. 54, n. 4, p. 86–96, 2011. Cited on page 66.

DOLIF, G.; ENGELBRECHT, A.; JATOBÁ, A.; SILVA, A. J. D. da; GOMES, J. O.; BORGES, M. R.; NOBRE, C. A.; CARVALHO, P. V. R. de. Resilience and brittleness in the ALERTA RIO system: a field study about the decision-making of forecasters. **Natural Hazards**, v. 65, n. 3, p. 1831–1847, 2013. Citations on pages 41, 64, 128, 133, 140, 163, and 165.

DOMINKOVICS, P.; GRANELL, C.; PÉREZ-NAVARRO, A.; CASALS, M.; ORCAU, À.; CAYLÀ, J. A. Development of spatial density maps based on geoprocessing web services: application to tuberculosis incidence in barcelona, spain. **International Journal of Health Geographics**, v. 10, n. 1, 2011. Citations on pages 133 and 140.

DORN, H.; VETTER, M.; HÖFLE, B. Gis–based roughness derivation for flood simulations: A comparison of orthophotos, lidar and crowdsourced geodata. **Remote Sensing**, v. 6, n. 2, p. 1739–1759, 2014. Cited on page 66.

DUGGAN, J. The case for personal data-driven decision making. **Proceedings of the VLDB Endowment**, v. 7, n. 11, p. 943–946, 2014. Citations on pages 38, 42, 63, 133, and 134.

DUMAS, M.; ROSA, M. L.; MENDLING, J.; REIJERS, H. A. Introduction to business process management. In: ____. **Fundamentals of Business Process Management**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. p. 1–31. Citations on pages 27, 67, and 68.

DUMBILL, E. Making sense of big data. **Big Data**, 2013. Cited on page 63.

DYBA, T.; DINGSOYR, T.; HANSSEN, G. K. Applying systematic reviews to diverse study types: An experience report. In: **Proceedings of the 1st International Symposium on Empirical Software Engineering and Measurement (ESEM)**. Madrid, Spain: [s.n.], 2007. p. 225–234. Citations on pages 71 and 81.

ELVESÆTER, B.; PANFILENKO, D.; JACOBI, S.; HAHN, C. Aligning business and it models in service-oriented architectures using bpmn and soaml. In: **Proceedings of the 1st International Workshop on Model-Driven Interoperability (MDI)**. Oslo, Norway: [s.n.], 2010. p. 61–68. Cited on page 68.

ELWOOD, S. Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS. **GeoJournal**, v. 72, n. 3-4, p. 173–183, 2008. Citations on pages 65, 66, and 67.

ER, M. C. Decision support systems: a summary, problems, and future trends. **Decision Support Systems**, v. 4, n. 3, p. 355–363, 1988. Cited on page 58.

ERSKINE, M.; GREGG, D. Utilizing Volunteered Geographic Information to Develop a Real-Time Disaster Mapping Tool: A Prototype and Research Framework. In: **Proceedings of the 2012 International Conference on Information Resources Management (Conf-IRM)**. Vienna, Austria: [s.n.], 2012. p. 1–14. Citations on pages 40, 65, and 86.

FAVA, M. C.; SANTANA, G. M. D.; BRESSIANI, D. A.; ROSA, A.; HORITA, F. E. A.; SOUZA, V. C. B.; MENDIONDO, E. M. Integration of information technology systems for flood forecasting with hybrid data sources. In: **Proceedings of the 6th International Conference on Flood Management**. São Paulo, Brazil: [s.n.], 2014. p. 1–11. Cited on page 21.

FEW, S. **Information Dashboard Design: the effective visual communication of data**. [S.l.]: O'Reilly Media, 2006. Citations on pages 60, 95, and 148.

FLANAGIN, A. J.; METZGER, M. J. The credibility of volunteered geographic information. **GeoJournal**, v. 72, n. 3-4, p. 137–148, 2008. Citations on pages 65 and 66.

FUCHS, G.; ANDRIENKO, N.; ANDRIENKO, G.; BOTHE, S.; STANGE, H. Tracing the german centennial flood in the stream of tweets: First lessons learned. In: **Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information**. Orlando, USA: [s.n.], 2013. p. 31–38. Cited on page 65.

GANTZ, J.; REINSEL, D. **The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East - United States**. [S.l.], 2012. Cited on page 64.

GENOVESE, E.; ROCHE, S. Potential of VGI as a resource for SDIS in the north/south context. **Geomatica**, v. 64, n. 4, p. 2010, 2010. Cited on page 66.

GILL, B.; BORDEN, B. C.; HALLGREN, K. **A Conceptual Framework for Data–Driven Decision–Making**. [S.l.], 2014. 4993/20634 (15). Cited on page 38.

GOLDKUHL, G.; LAGSTEN, J. Different roles of evaluation in information systems research. In: **Workshop on IT Artefact Design and Workpractice Intervention**. Barcelona, Spain: [s.n.], 2012. p. 1–13. Cited on page 150.

GONGQUAN, L.; KEYAN, X. Research on water resources decision support system integrated with facilities. In: **Proceedings of the 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)**. Jilin, China: [s.n.], 2011. p. 2507–2510. Cited on page 59.

GOODCHILD, M. F. Citizens as sensors: the world of volunteered geography. **GeoJournal**, v. 69, n. 4, p. 211–221, 2007. Citations on pages 40, 57, 65, 91, and 133.

GOODCHILD, M. F.; GLENNON, J. A. Crowdsourcing geographic information for disaster response: a research frontier. **International Journal of Digital Earth**, v. 3, n. 3, p. 231–241, 2010. Citations on pages 66, 67, and 77.

GOPALKRISHNAN, V.; STEIER, D.; LEWIS, H.; GUSZCZA, J. Big data, big business: Bridging the gap. In: **Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications (BigMine)**. Beijing, China: [s.n.], 2012. p. 7–11. Citations on pages 38, 42, 63, 105, 133, 134, 140, 151, and 163.

GOTTUMUKKALA, R.; ZACHARY, J.; KEARFOTT, B.; KOLLURU, R. Real-time information driven decision support system for evacuation planning. In: **Proceedings of the 2012 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support**. [S.l.: s.n.], 2012. p. 206–209. Citations on pages 86 and 87.

GRANELL, C.; OSTERMANN, F. O. Beyond data collection: Objectives and methods of research using {VGI} and geo-social media for disaster management. **Computers, Environment and Urban Systems**, v. 59, p. 231–243, 2016. Cited on page 65.

GREGOR, S. The nature of theory in information systems. **MIS Quarterly**, v. 30, n. 3, p. 611–642, 2006. Cited on page 46.

GREGOR, S.; HEVNER, A. R. Positioning and presenting design science research for maximum impact. **MIS Quarterly**, Minneapolis, MN, USA, v. 37, n. 2, p. 337–356, 2013. Citations on pages 48, 165, and 167.

GROUP, G. **Big Data**. 2011. <http://www.gartner.com/it-glossary/big-data/>. Cited on page 63.

GU, J.; TANG, X. Designing a Water Resources Management Decision Support System: An Application of the WSR Approach. **Systemic Practice and Action Research**, v. 13, n. 1, p. 59–70, 2000. Cited on page 59.

GUAN, X.; CHEN, C. Using social media data to understand and assess disasters. **Natural Hazards**, v. 74, n. 2, p. 837–850, 2014. Cited on page 65.

GUOXIANG, L.; MAOFENG, L. SARGIS: A GIS-Based Decision-making Support System for Maritime Search and Rescue. In: **Proceedings of the 2010 International Conference on E-Business and E-Government (ICEE)**. Guangzhou, China: [s.n.], 2010. p. 1571–1574. Cited on page 60.

HASHEM, I. A. T.; YAQOOB, I.; ANUAR, N. B.; MOKHTAR, S.; GANI, A.; KHAN, S. U. The rise of "big data" on cloud computing: Review and open research issues. **Information Systems**, v. 47, p. 98–115, 2015. Citations on pages 38, 63, 64, and 165.

HASHIMOTO, T.; KUBOYAMA, T.; CHAKRABORTY, B.; SHIROTA, Y. Discovering Topic Transition about the East Japan Great Earthquake in Dynamic Social Media. In: **Proceedings of the 2012 IEEE Global Humanitarian Technology Conference (GHTC)**. Seattle, Washington, USA: [s.n.], 2012. p. 259–264. Citations on pages 86 and 87.

HASHIMOTO, T.; KUBOYAMA, T.; SHIROTA, Y. Rumor analysis framework in social media. In: **Proceedings of the TENCON**. Bali, Indonesia: [s.n.], 2011. p. 133–137. Cited on page 86.

HAWORTH, B.; BRUCE, E. A review of volunteered geographic information for disaster management. **Geography Compass**, v. 9, n. 5, p. 237–250, 2015. Citations on pages 65 and 66.

HEIPKE, C. Crowdsourcing geospatial data. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 65, n. 6, p. 550–557, 2010. Cited on page 65.

HEVNER, A. R.; MARCH, S. T.; PARK, J.; RAM, S. Design science in information systems research. **MIS Quarterly**, v. 28, n. 1, p. 75–105, 2004. Citations on pages 106, 108, and 135.

HORITA, F. E.; ALBUQUERQUE, J. P.; MARCHEZINI, V.; MENDIONDO, E. M. Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in brazil. **Decision Support Systems**, v. 97, p. 12–22, 2017. Citations on pages 17, 49, and 106.

HORITA, F. E.; ASSIS, L.; CASTANHARI, R. E.; ISOTANI, S.; CRUZ, W. M.; ALBUQUERQUE, J. P. de. A gamification-based social collaborative architecture to increase resilience against natural disasters. In: **Proceedings of the X Brazilian Symposium on Information Systems (SBSI)**. Londrina, Brazil: [s.n.], 2014. p. 399–410. Cited on page 18.

HORITA, F. E. A.; ALBUQUERQUE, J. P. An Approach to Support Decision-Making in Disaster Management based on Volunteer Geographic Information (VGI) and Spatial Decision Support Systems (SDSS). In: **Proceedings of the 10^{th} International Conference on Information Systems for Crisis Response and Management (ISCRAM)**. Baden-Baden, Germany: [s.n.], 2013. p. 301–306. Cited on page 19.

____. Providing real-time decision support in flood management using volunteered information. In: **Proceedings of the 1^{st} Escola Latino Americana de Engenharia de Software (ELA-ES)**. Rio de Janeiro, Brazil: [s.n.], 2013. p. 1–1. Cited on page 19.

HORITA, F. E. A.; ALBUQUERQUE, J. P.; DEGROSSI, L. C.; MENDIONDO, E. M.; UEYAMA, J. Development of a spatial decision support system for flood risk management in brazil that combines volunteered geographic information with wireless sensor networks. **Computers & Geosciences**, v. 80, p. 84–94, 2015. Citations on pages 17, 48, 66, 92, and 147.

HORITA, F. E. A.; ALBUQUERQUE, J. P.; MARCHEZINI, V.; MENDIONDO, E. M. A qualitative analysis of the early warning process in disaster management. In: **Proceedings of the 13^{th} International Conference on Information Systems for Crisis Response and Management (ISCRAM)**. Rio de Janeiro, Brazil: [s.n.], 2016. p. 1–10. Citations on pages 18, 109, 114, and 118.

HORITA, F. E. A.; BRAGA, D. S.; MONTEIRO, C. D. D. The use of multimodal interaction for supporting the production of location-based information. **IEEE Latin America Transactions**, v. 14, n. 7, p. 3496–3502, 2016. Cited on page 17.

HORITA, F. E. A.; DEGROSSI, L. C.; ASSIS, L. F. G.; ZIPF, A.; ALBUQUERQUE, J. P. The use of volunteered geographic information (VGI) and crowdsourcing in disaster management: a systematic literature review. In: **Proceedings of the 19th Americas Conference on Information Systems (AMCIS)**. [S.l.: s.n.], 2013. p. 1–10. Citations on pages 19, 40, 48, 65, 88, 92, 168, and 169.

HORITA, F. E. A.; FAVA, M. C.; MENDIONDO, E. M.; ROTAVA, J.; SOUZA, V. C.; UEYAMA, J.; ALBUQUERQUE, J. P. AGORA-GeoDash: A geosensor dashboard for real-time flood risk monitoring. In: **Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM)**. University Park, USA: [s.n.], 2014. p. 304–313. Citations on pages 18, 42, 60, 61, 62, 63, 66, 95, 96, 133, 134, 140, 147, and 148.

HORITA, F. E. A.; FAVA, M. C.; SOUZA, V. C.; UEYAMA, J.; ALBUQUERQUE, J. P.; MENDIONDO, E. M. Web-based tool to deliver useful real-time information for decision-making in the flood management. In: **Proceedings of the Climate Change Adaptation Conference (Adaptation Futures)**. Fortaleza, Brazil: [s.n.], 2014. p. 1–1. Cited on page 19.

HORITA, F. E. A.; LINK, D.; ALBUQUERQUE, J. P.; HELLINGRATH, B. odmn: An integrated model to connect decision-making needs to emerging data sources in disaster management. In: **Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS)**. Kauai, Hawaii, USA: [s.n.], 2016. p. 2882–2891. Citations on pages 18, 109, 140, and 169.

HOSACK, B.; HALL, D.; PARADICE, D.; COURTNEY, J. A Look Toward the Future: Decision Support Systems Research is Alive and Well. **Journal of the Association for Information Systems**, v. 13, n. 5, p. 316–340, 2012. Citations on pages 58 and 60.

HR-WALLINGFORD. **R&D Outputs: Flood Risks to People: Phase 2 Project Record**. [S.l.], 2006. Technical Report, FD2321/PR. Cited on page 98.

HUANG, C.; CHAN, E.; HYDER, A. Web 2.0 and internet social networking: A new tool for disaster management? - lessons from taiwan. **BMC Medical Informatics and Decision Making**, v. 10, n. 1, p. 1–5, 2010. Cited on page 77.

HUGHES, D.; UEYAMA, J.; MENDIONDO, E.; MATTHYS, N.; HORRA, W.; MICHIELS, S.; HUYGENS, C.; JOOSEN, W.; MAN, K.; GUAN, S. A middleware platform to support river monitoring using wireless sensor networks. **Journal of the Brazilian Computer Society**, v. 17, n. 2, p. 85–102, 2011. Citations on pages 40, 91, 96, and 169.

HURWITZ, J.; NUGENT, A.; HALPER, F. **Big Data for Dummies**. Ed. 1. [S.l.]: For Dummies, 2013. ISBN 1118504224. Citations on pages 62, 63, and 64.

IBGE. **Perfil dos Municípios Brasileiros 2013**. [S.l.], 2014. Technical Report. Cited on page 91.

IBRD/IDA. **Building Resilience: Integrating Climate and Disaster Risk into Development**. [S.l.], 2013. Technical Report. Cited on page 56.

IKEDA, S.; SATO, T.; FUKUZONO, T. Towards an integrated management framework for emerging disaster risks in japan. **Natural Hazards**, v. 44, n. 2, p. 267–280, 2008. Cited on page 77.

IMRAN, M.; CASTILHO, C.; DIAZ, F.; VIEWEG, S. Processing social media messages in mass emergency: A survey. **ACM Computing Surveys**, v. 47, n. 4, p. 1–37, 2014. Citations on pages 42, 64, 130, and 140.

JANSEN, J. M.; LIJNSE, B.; PLASMEIJER, R. Towards dynamic workflow support for crisis management. In: **Proceedings of the 7th International Conference on Information Systems for Crisis Response and Management (ISCRAM)**. Seattle, USA: [s.n.], 2010. p. 1–5. Cited on page 68.

JANSSENS, L.; SMEDT, J. D.; VANTHIENEN, J. Modeling and enacting enterprise decisions. In: **Proceedings of the 28th International Conference on Advanced Information Systems Engineering (CaISE)**. Ljubljana, Slovenia: [s.n.], 2016. p. 1–12. Citations on pages 69 and 129.

JHA, A. K. Safer homes, stronger communities: A handbook for reconstructing after natural disasters. In: ____. [S.l.]: The World Bank, 2010. cap. Disaster Types and Impacts, p. 339–344. Citations on pages 39 and 73.

JHA, A. K.; BLOCH, R.; LAMOND, J. **Cities and Flooding: A Guide to Integrated Urban Flood Risk Management for the 21st Century**. Washington, DC, USA: The World Bank, 2012. 638 p. Cited on page 91.

JIRKA, S.; BORING, A.; KJELD, P.; MAIDENS, J.; WYTZISK, A. A lightweight approach for the sensor observation service to share environmental data across europe. **Transactions in GIS**, v. 16, n. 3, p. 293–312, 2012. Citations on pages 94 and 98.

JONKMAN, S.; PENNING-ROWSELL, E. Human instability in flood flows. **Journal of the American Water Resources Association**, v. 44, n. 4, p. 1–11, 2008. Cited on page 98.

KAEWKITIPONG, L.; CHEN, C.; RACTHAM, P. Lessons learnd from the use of social media in combating a crisis: A case study of 2011 Thailand flooding disaster. In: **Proceedings of the 33th International Conference on Information Systems (ICIS)**. Orlando, USA: [s.n.], 2012. p. 1–17. Citations on pages 65, 77, 78, 86, and 91.

KEENAN, P. B. Decision-making support systems: Achievements and challenges for the new decade. In: ____. **Decision Making Support Systems: Achievements and challenges for the New Decade**. [S.l.]: IGI Global, 2003. cap. Spatial Decision Support Systems, p. 28–39. Cited on page 60.

KELLY, G. C.; TANNER, M.; VALLELY, A.; CLEMENTS, A. Malaria elimination: moving forward with spatial decision support systems. **Trends in Parasitology**, v. 28, n. 7, p. 297–304, 2012. Cited on page 60.

KELMAN, I.; GAILLARD, J. C.; MERCER, J. Climate change's role in disaster risk reduction's future: Beyond vulnerability and resilience. **International Journal of Disaster Risk Science**, v. 6, n. 1, p. 21–27, 2015. Cited on page 42.

KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. [S.l.], 2007. Technical Report. Citations on pages 71, 72, 80, 81, and 89.

KLEINDIENST, D.; PFLEGER, R.; SCHOCH, M. The business alignment of social media analytics. In: **Proceedings of the 23rd European Conference on Information Systems (ECIS)**. [S.l.: s.n.], 2015. p. 1–14. Citations on pages 42, 65, 128, 130, 134, 151, and 168.

KOGA, I.; MEDEIROS, C. B. Integrating and processing events from heterogeneous data sources. In: **Proceedings of the 32th Brazilian Computer Society Conference (CSBC)**. [S.l.: s.n.], 2012. p. 1–8. Cited on page 103.

KOŚCIELNIAK, H.; PUTO, A. Big data in decision making processes of enterprises. **Procedia Computer Science**, v. 65, p. 1052–1058, 2015. Citations on pages 105, 130, and 133.

KOTHARI, C. R.; GARG, G. **Research Methodology: Methods and Techniques**. [S.l.]: New Age, 2004. 470 p. Cited on page 46.

KUMAR, S.; HAVEY, T. Before and after disaster strikes: A relief supply chain decision support framework. **International Journal Production Economics**, v. 145, n. 2, p. 613–629, 2013. Citations on pages 86 and 87.

KURNIAWATI, K.; SHANKS, G. G.; BEKMAMEDOVA, N. The business impact of social media analytics. In: **Proceedings of the 21st European Conference on Information Systems (ECIS)**. Utrecht, Netherlands: [s.n.], 2013. p. 1–12. Citations on pages 42, 63, 128, 133, 134, and 151.

LAU, F. A review on the use of action research in information systems studies. In: ____. **Proceedings of the International Conference on Information Systems and Qualitative Research**. Philadelphia, USA: Springer US, 1997. p. 31–68. Cited on page 134.

LEE, J.; KIM, J.; KIM, D.; CHONG, P. K.; KIM, J.; JANG, P. RFMS: Real-time flood monitoring system with wireless sensor networks. In: **Proceedings of the 5th IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS)**. Atlanta, USA: [s.n.], 2008. p. 527–528. Citations on pages 40 and 91.

LEVY, J. K.; HARTMANN, J.; LI, K. W.; AN, Y.; ASGARY, A. Multi-Criteria Decision Support Systems for Flood Hazard Mitigation and Emergency Response in Urban Watersheds. **JAWRA Journal of the American Water Resources Association**, v. 43, n. 2, p. 346–358, 2007. Cited on page 60.

LEYH, W.; FAVA, M. C.; ABE, N.; RESTREPO-ESTRADA, C.; HORITA, F. E. A.; MENDIONDO, E. M.; ALBUQUERQUE, J. P. de. Sdi-node to interlink information, essential for disaster preparedness and management, with other linked open data. In: **Proceedings of the 13th International Conference on Information Systems for Crisis Response and Management (ISCRAM)**. Rio de Janeiro, Brazil: [s.n.], 2016. p. 1–1. Cited on page 20.

LIANG, L. Y.; MIRANDA, R. Dashboards and scorecards: Executive information systems for the public sector. **Government Finance Review**, v. 17, n. 6, p. 14–19, 2001. Citations on pages 60, 61, 95, and 148.

LINK, D.; HELLINGRATH, B.; LING, J. A human-is-the-loop approach for semi-automated content moderation. In: **Proceedings of the 13th International Conference on Information Systems for Crisis Response and Management (ISCRAM)**. [S.l.: s.n.], 2016. p. 1–13. Cited on page 40.

LINK, D.; HORITA, F. E. A.; ALBUQUERQUE, J. P.; HELLINGRATH, B.; GHASEMIVAND-HONARYAR, S. A method for extracting task-related information from social media based on structured domain knowledge. **Proceedings of the 21st Americas Conference on Information Systems (AMCIS)**, Rio Grande, Puerto Rico, p. 1–16, 2015. Citations on pages 21, 38, 42, and 57.

LONGUEVILLE, B. D.; ANNONI, A.; SCHADE, S.; OSTLAENDER, N.; WHITMORE, C. Digital Earth's Nervous System for crisis events: real-time Sensor Web Enablement of

Volunteered Geographic Information. **International Journal of Digital Earth**, v. 3, n. 3, p. 242–259, 2010. Citations on pages 77 and 91.

LONGUEVILLE, B. D.; LURASCHI, G.; SMITS, P.; PEEDELL, S.; GROEVE, T. D. Citizens as sensors for natural hazards: A VGI integration workflow. **Geomatica**, v. 64, p. 41–59, 2010. Citations on pages 55 and 65.

MACEACHREN, A. M.; JAISWAL, A.; ROBINSON, A. C.; PEZANOWSKI, S.; SAVELYEV, A.; MITRA, P.; ZHANG, X.; BLANFORD, J. SensePlace2: GeoTwitter Analytics Support for Situational Awareness. In: **Proceedings of the 2011 IEEE Symposium Visual Analytics Science and Technology (VAST)**. Providence, Rhode Island, USA: [s.n.], 2011. p. 181–190. Cited on page 86.

MAHENDRAWATHI, E. R.; PRANANTHA, D.; UTOMO, J. Development of dashboard for hospital logistics management. In: **Proceedings of the 2010 IEEE Conference on Open Systems (ICOS)**. Kuala Lumpur, Malaysia: [s.n.], 2010. p. 86–90. Cited on page 61.

MALSBENDER, A.; BEVERUNGEN, D.; VOIGT, M. *et al.* Capitalizing on social media analysis–insights from an online review on business models. In: **Proceedings of the 19th Americas Conference on Information Systems (AMCIS)**. Chicago, Illinois, USA: [s.n.], 2013. p. 1–9. Citations on pages 42, 63, 128, 133, 134, and 151.

MANDVIWALLA, M.; WATSON, R. Generating capital from social media. **MIS Quarterly**, v. 13, n. 2, p. 97–113, 2014. Citations on pages 38, 64, and 128.

MANSOURIAN, A.; RAJABIFARD, A.; ZOEJ, M. V.; WILLIAMSON, I. Using SDI and web-based system to facilitate disaster management. **Computers & Geosciences**, v. 32, n. 3, p. 303–315, 2006. Citations on pages 40, 41, and 168.

MANYIKA, J.; CHUI, M.; BROWN, B.; BUGHIN, J.; DOBBS, R.; ROXBURGH, C.; BYERS, A. H. **Big data: The next frontier for innovation, competition, and productivity**. [S.l.], 2011. Technical Report. Cited on page 63.

MARCELINO, E. V. **Desastres naturais e geotecnologias: conceitos básicos**. [S.l.], 2007. Technical Report. Cited on page 39.

MARKOVIC, N.; STANIMIROVIC, A.; STOIMENOV, L. Sensor web for river water pollution monitoring and alert system. In: **Proceedings of the 12th International Conference on Geographic Information Science (AGILE)**. Hannover, Germany: [s.n.], 2009. p. 1–9. Citations on pages 40, 41, 103, and 168.

MARKUS, M. L.; MAJCHRZAK, A.; GASSER, L. A design theory for systems that support emergent knowledge processes. **MIS Quarterly**, v. 26, n. 3, p. 179–212, 2002. Cited on page 135.

MARSH, J. A.; PANE, J. F.; HAMILTON, L. S. Making sense of data-driven decision making in education. **Evidence from Recent RAND Research**, p. 1–15, 2006. Citations on pages 38 and 42.

MCAFEE, A.; BRYNJOLFSSON, E. Big data: the management rrevolution. **Harvard Business Review**, v. 10, p. 59–68, 2012. <https://hbr.org/2012/10/big-data-the-management-revolution>. Citations on pages 38 and 64.

MEIER, P. **Digital Humanitarians: How Big Data Is Changing the Face of Humanitarian Response**. [S.l.]: Crc Press, 2014. Cited on page 40.

MENDES, H. C.; MENDIONDO, E. M. Histórico da expansão urbana e incidência de inundações: O caso da bacia do Gregório, São Carlos–Brazil. **Revista Brasileira de Recursos Hídricos**, v. 12, n. 1, p. 17–27, 2007. Cited on page 96.

MENDIONDO, E. M. Reducing vulnerability ti water-related disasters in urban areas of the humid tropics. In: **Proceedings of the Integrated Urban Water Management**. Foz de Iguaçu, Paraná, Brazil: [s.n.], 2010. p. 109–127. Citations on pages 42 and 56.

MERZ, B.; ELMER, F.; KUNZ, M.; MÜHR, B.; SCHRÖTER, K.; UHLEMANN-ELMER, S. The extreme flood in june 2013 in Germany. **La Houille Blanche**, v. 2014, n. 1, p. 5–10, 2012. Cited on page 91.

MILLER, H. G.; MORK, P. From data to decisions: A value chain for big data. **IT Professional**, v. 15, n. 1, p. 57–59, 2013. Citations on pages 38, 42, and 130.

MISRA, A.; SHARMA, A.; GULIA, P.; BANA, A. Big data: Challenges and opportunities. **International Journal of Innovative Technology and Exploring Engineering (IJITEE)**, Citeseer, v. 4, n. 2, p. 41–42, 2014. Cited on page 63.

MITCHEL, J. J.; RYDER, A. J. Developing and using dashboard indicators in student affairs assessment. **New Directions for Student Services**, v. 2013, n. 142, p. 71–81, 2013. Cited on page 61.

MITTEN, P. T.; PARSONS, B. M. Geospatial portfolios for analysis/decision making (GeoPAD). In: **Proceedings of the IEEE/MTS Oceans Conference**. Waikoloa, USA: [s.n.], 2011. p. 1–5. Cited on page 59.

MOLINA, M.; BAYARRI, S. A multinational SDI-based system to facilitate disaster risk management in the andean community. **Computers & Geosciences**, v. 37, n. 9, p. 1501–1510, 2011. Citations on pages 41 and 103.

MORA, M.; FORGIONNE, G.; GUPTA, J.; CERVANTES, F.; GELMAN, O. A Framework to Assess Intelligent Decision–Making Support Systems. In: **Proceedings of the 7^{th} International Conference Knoweledge and Engineering Systems (KES)**. Oxford, UK: [s.n.], 2003. p. 59–65. Cited on page 58.

MOREIRA, R. B.; DEGROSSI, L. C.; ALBUQUERQUE, J. P. An experimental evaluation of a crowdsourcing-based approach for flood risk management. In: **Proceedings of the 12^{th} Workshop on Experimental Software Engineering (ESELAW)**. Lima, Peru: [s.n.], 2015. p. 1–11. Cited on page 104.

MUMFORD, E. The story of socio-technical design: reflections on its successes, failures and potential. **Information Systems Journal**, v. 16, n. 4, p. 317–342, 2006. Cited on page 166.

MUNICHRE. **Münchener Rückversicherungs-Gesellschaft, Geo Risks Research, NatCat-SERVICE**. http://www.munichre.com/en/reinsurance/business/non-life/natcatservice/annual-statistics/index.html: [s.n.], 2015. Cited on page 39.

MUNRO, R. Crowdsourcing and the crisis-affected community. **Information Retrieval**, v. 16, n. 2, p. 1–57, 2012. Cited on page 77.

NEVILLE, K. M.; DOYLE, C.; MUELLER, J.; SUGRUE, A. Supporting cross border emergency management decision-making. In: **Proceedings of the 21st European Conference on Information Systems (ECIS)**. Utrecht, The Netherlands: [s.n.], 2013. p. 1–7. Cited on page 59.

NIKO, D. L.; HWANG, H.; LEE, Y.; KIM, C. Integrating User-generated Content and Spatial Data into Web GIS for Disaster History. **Computers, Networks, Systems, and Industrial Engineering 2011**, v. 365, p. 245–255, 2011. Citations on pages 65, 66, and 77.

NORRIS, F. H.; STEVENS, S. P.; PFEFFERBAUM, B.; WYCHE, K. F.; PFEFFERBAUM, R. L. Community resilience as a metaphor, theory, set of capacities and strategy for disaster. **American Journal of Community Psychology**, v. 41, n. 1-2, p. 127–150, 2008. Citations on pages 40, 42, and 56.

NOVAK, D. D.; PAULOS, A.; CLAIR, G. S. Data–driven budget reductions: a case study. **The Bottom Line**, v. 24, n. 1, p. 24–34, 2011. Citations on pages 38 and 151.

OGC. **Sensor Observation Service Interface Standard**. [S.l.], 2012. <http://www.opengeospatial.org/standards/sos>. Citations on pages 93, 94, 99, and 100.

____. **Observations and Measurements (O&M)**. [S.l.], 2013. <http://www.opengeospatial.org/standards/om>. Citations on pages 93, 94, 110, 140, and 141.

____. **OGC Sensor Observation Service 2.0 Hydrology Profile**. [S.l.], 2014. <https://portal.opengeospatial.org/files/?artifact_id=57327>. Cited on page 94.

OH, O.; KWON, K. H.; RAO, H. R. An exploration of Social Media in extreme events: Rumor theory and twitter during the Haiti Earthquake 2010. In: **Proceedings of the 32th International Conference on Information Systems (ICIS)**. Saint Louis, Missouri, USA: [s.n.], 2010. p. 1–14. Cited on page 86.

O'LEARY, D. E. Artificial intelligence and big data. **IEEE Intelligent Systems**, Piscataway, NJ, USA, v. 28, n. 2, p. 96–99, 2013. Cited on page 64.

OMG. **Business Process Model and Notation (BPMN), Version 2.0**. [S.l.], 2013. <http://www.omg.org/spec/BPMN/2.0/>. Citations on pages 67 and 128.

____. **Decision Model and Notation (DMN)**. [S.l.], 2014. <http://www.omg.org/spec/DMN/>. Citations on pages 69, 128, and 140.

OSTERMANN, F. O.; SPINSANTI, L. A conceptual workflow for automatically assessing the quality of volunteered geographic information for crisis management. In: **Proceedings of the 14th International Conference on Geographic Information Science (AGILE)**. Utrecht, Netherlands: [s.n.], 2011. p. 1–6. Cited on page 65.

OTHMAN, S. H.; BEYDOUN, G. Model-driven disaster management. **Information & Management**, v. 50, n. 5, p. 218–228, 2013. Cited on page 129.

PATEL, R.; KAUSHIK, B. RIFMAS: River flow management system using wireless sensing agents. In: **Proceedings of the 2009 IEEE International Advance Computing Conference (IACC)**. Patiala, Punjab, India: [s.n.], 2009. p. 853–857. Cited on page 91.

PEARCE, L. The Value Of Public Participation During a Hazard, Impact, Risk And Vulnerability (HIRV) Analysis. **Mitigation and Adaptation Strategies for Global Change**, v. 10, n. 3, p. 411–441, 2005. Cited on page 77.

PERHOVAZ, R. P. **Estudo experimental e teórico da qualidade de água da drenagem urbana com base ecohidrológica**. Dissertação (Mestrado) — Engineering School of São Carlos, University of São Paulo, 2010. Cited on page 96.

PETERSEN, K.; FELDT, R.; MUJTABA, S.; MATTSSON, M. Systematic mapping studies in software engineering. In: **Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (EASE)**. Bari, Italy: [s.n.], 2008. p. 68–77. Citations on pages 48, 71, and 89.

PETERSON, R. **Constructing Effective Questionnaires**. [S.l.]: SAGE Publications, 2000. Cited on page 109.

POHL, D.; BOUCHACHIA, A.; HELLWAGNER, H. Supporting Crisis Management via Sub-Event Detection in Social Networks. In: **Proceedings of the 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)**. Toulouse, France: [s.n.], 2012. p. 373–378. Citations on pages 77 and 86.

POIANI, T. H.; HORITA, F. E. A.; ALBUQUERQUE, J. P. de. Análise geográfica entre mensagens georreferenciadas de redes sociais e dados oficiais para suporte à tomada de decisões de agências de emergência. In: **Proceedings of the XVI Brazilian Symposium on GeoInformatics (Geoinfo)**. Campos do Jordão, Brazil: [s.n.], 2015. p. 38–43. Cited on page 20.

POSER, K.; DRANSCH, D. Volunteered geographic information for disaster management with application to rapid flood damage estimation. **Geomatica**, v. 64, n. 1, p. 89–98, 2010. Citations on pages 56, 57, 66, 73, and 77.

POWER, D. Decision Support Systems: From the Past to the Future. In: **Proceedings of the 10th Americas Conference on Information Systems (AMCIS)**. New York, USA: [s.n.], 2004. p. 2025–2031. Cited on page 58.

POYNTON, T. A.; CAREY, J. C. An integrative model of data-based decision making for school counseling. **Professional School Counseling**, v. 10, n. 2, p. 121–130, 2006. Cited on page 38.

PROVOST, F.; FAWCETT, T. Data science and its relationship to big data and data-driven decision making. **Big Data**, v. 1, n. 1, p. 51–59, 2013. Citations on pages 38, 42, 63, 133, and 134.

RIJCKEN, T.; STIJNEN, J.; ; SLOOTJES, N. “SimDelta” - inquiry into an internet-based interactive model for water infrastructure development in the netherlands. **Water**, v. 4, n. 2, p. 295–320, 2012. Citations on pages 86 and 87.

ROBSON, C. **Real world research**. [S.l.]: John Wiley & Sons, 2011. 608 p. Cited on page 46.

ROCHA, R. S.; DEGROSSI, L. C.; HORITA, F. E. A.; PEREIRA, J. P. A. AGORA-PL: uma proposta para desenvolvimento de famílias de sistemas colaborativos baseados em vgi para a gestão do risco de inundação. In: **Proceedings of the 11th Brazilian Symposium on Collaborative Systems (SBSC)**. Curitiba, Brazil: [s.n.], 2014. p. 118–125. Cited on page 21.

ROCHE, S.; PROPECK-ZIMMERMANN, E.; MERICSKAY, B. Geoweb and crisis management: issues and perspectives of volunteered geographic information. **GeoJournal**, v. 78, n. 1, p. 21–40, 2011. Citations on pages 77 and 91.

ROSEMANN, M.; EGGERT, M.; VOIGT, M.; BEVERUNGEN, D. Leveraging social network data for analytical crm strategies: the introduction of social bi. In: **Proceedings of the 20th European Conference on Information Systems (ECIS)**. Barcelona, Spain: [s.n.], 2012. p. 1–12. Citations on pages 42 and 64.

ROTAVA, J.; MENDIONDO, E. M.; SOUZA, V. C. B. Simulação de instabilidade humana em inundações: primeiras considerações. In: **Proceedings of the 20th Simpósio Brasileiro de Recursos Hídricos**. Bento Gonçalves, Brazil: [s.n.], 2013. p. 1–8. Cited on page 98.

RUIZ, M. C.; ROMERO, E.; PÉREZ, M. A.; FERNÁNDEZ, I. Development and application of a multi-criteria spatial decision support system for planning sustainable industrial areas in Northern Spain. **Automation in Construction**, v. 22, p. 320–333, 2012. Cited on page 60.

RUNESON, P.; HÖST, M. Guidelines for conducting and reporting case study research in software engineering. **Empirical Software Engineering**, v. 14, n. 2, p. 131–164, 2008. Citations on pages 109, 118, and 137.

RUSHTON, G. Spatial Decision Support Systems. In: SMELSER, E.-i.-C. N. J.; BALTES, P. B. (Ed.). **International Encyclopedia of the Social & Behavioral Sciences**. Oxford: Pergamon, 2001. p. 14785–14788. ISBN 978-0-08-043076-8. Citations on pages 59, 60, 61, and 151.

SAADATSERESHT, M.; MANSOURIAN, A.; TALEAI, M. Evacuation planning using multiobjective evolutionary optimization approach. **European Journal of Operational Research**, v. 198, n. 1, p. 305–314, 2009. Cited on page 128.

SAATY, T. L. Decision making with the analytic hierarchy process. **International journal of services sciences**, v. 1, n. 1, p. 83–98, 2008. Cited on page 37.

SACKMANN, S.; HOFMANN, M.; BETKE, H. Towards a model-based analysis of place-related information in disaster response workflows. In: **Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management (ISCRAM)**. Baden Baden, Germany: [s.n.], 2013. p. 78–83. Citations on pages 68 and 129.

SAEEDI, K.; ZHAO, L.; SAMPAIO, P. R. F. Extending bpmn for supporting customer-facing service quality requirements. In: **Proceedings of the 2010 IEEE International Conference on Web Services (ICWS)**. [S.l.: s.n.], 2010. p. 616–623. Cited on page 68.

SALDAÑA, J. **The coding manual for qualitative researchers**. Ed. 2. London: Sage Publications Ltda, 2015. Citations on pages 49, 109, and 114.

SAMBAMURTHY, V.; DESANCTIS, G. An experimental evaluation of GDSS effects on group performance during stakeholder analysis. In: **Proceedings of the 23th Annual Hawaii International Conference on Systems Sciences (HICSS)**. Grand Wailea, Maui, Hawaii: [s.n.], 1990. p. 79–88. Cited on page 59.

SANTORO, F. M.; BORGES, M. R. S.; PINO, J. A. Acquiring knowledge on business processes from stakeholders' stories. **Advanced Engineering Informatics**, v. 24, n. 2, p. 138–148, 2010. Cited on page 129.

SANTOS, L.; COUTINHO-RODRIGUES, J.; ANTUNES, C. H. A web spatial decision support system for vehicle routing using Google Maps. **Decision Support Systems**, v. 51, n. 1, p. 1–9, 2011. Cited on page 60.

SAVELYEV, A.; XU, S.; JANOWICZ, K.; MüLLIGANN, C.; THATCHER, J.; LUO, W. Volunteered geographic services: developing a linked data driven location-based service. In: **Proceedings of the 1st ACM SIGSPATIAL International Workshop on Spatial Semantics and Ontologies (SSO)**. Chicago, USA: [s.n.], 2011. p. 25–31. Citations on pages 77 and 78.

SCHADE, S.; DIAZ, L.; OSTERMANN, F.; SPINSANTI, L.; LURASCHI, G.; COX, S.; N, M.; LONGUEVILLE, B. Citizen-based sensing of crisis events: sensor web enablement for volunteered geographic information. **Applied Geomatics**, v. 5, n. 1, p. 1–16, 2011. Cited on page 77.

SCHAFER, W.; GANOE, C.; CARROLL, J. Supporting community emergency management planning through a geocollaboration software architecture. **Computer Supported Cooperative Work (CSCW)**, v. 16, n. 4-5, p. 501–537, 2007. Citations on pages 77 and 78.

SCHNEBELE, E.; CERVONE, G.; WATERS, N. Road assessment after flood events using non-authoritative data. **Natural Hazards and Earth System Sciences**, v. 14, n. 4, p. 1007–1015, 2014. Citations on pages 41, 103, and 133.

SCHOOLEY, B.; HILTON, B.; LEE, Y.; MCCLINTOCK, R.; OLUSOLA, S.; HORAN, T.; SCHOOLEY, B.; MCCLINTOCK, R.; HORAN, T. CrashHelp : A GIS Tool for Managing Emergency Medical Responses to Motor Vehicle Crashes. In: **Proceedings of the 7th International Conference on Information Systems for Crisis Response and Management (ISCRAM)**. Baden-Baden, Germany: [s.n.], 2010. p. 1–11. Cited on page 60.

SEAL, V.; RAHA, A.; MAITY, S.; MITRA, S. K.; MUKHERJEE, A.; NASKAR, M. K. A simple flood forecasting scheme using wireless sensor networks. **International Journal of Ad hoc, Sensor & Ubiquitous Computing**, v. 3, n. 1, p. 45–60, 2012. Cited on page 91.

SEAMAN, C. B. Qualitative methods in empirical studies of software engineering. **IEEE Transactions on Software Engineering**, v. 25, n. 4, p. 557–572, 1999. Cited on page 49.

SEIN, M. K.; HENFRIDSSON, O.; PURAO, S.; ROSSI, M.; LINDGREN, R. Action design research. **MIS Quarterly**, v. 35, n. 1, p. 37–56, 2011. Citations on pages 28, 49, 134, 135, 136, 147, and 158.

SHIM, K.; FONTANE, D.; LABADIE, J. Spatial Decision Support System for Integrated River Basin Flood Control. **Journal of Water Resources Planning and Management**, v. 128, n. 3, p. 190–201, 2002. Citations on pages 57 and 58.

SHUKLA, S.; PANDEY, G. To design an architectural model for flood monitoring using wireless sensor network system. **International Journal of Computer Science & Information Technologies**, v. 5, n. 1, p. 502–507, 2014. Cited on page 91.

SIMON, H. A. **The New Science of Management Decision**. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1977. Cited on page 37.

SIMONOFF, J. S.; RESTREPO, C. E.; ZIMMERMAN, R.; NAPHTALI, Z. S.; WILLIS, H. H. Resource allocation, emergency response capability, and infrastructure concentration around vulnerable sites. **Journal of Risk Research**, v. 14, n. 5, p. 597–613, 2011. Cited on page 128.

SIMONOVIC, S. P. Decision Support System for Flood Management in the Red River Basin. **Canadian Water Resources Journal**, v. 24, n. 3, p. 203–223, 1999. Cited on page 57.

SINNAPPAN, S.; FARRELL, C.; STEWART, E. Priceless Tweets! A Study on Twitter Messages Posted During Crisis: Black Saturday. In: **Proceedings of the Australasian Conferences on Information Systems (ACIS)**. Brisbane, Australia: [s.n.], 2010. p. 1–11. Cited on page 77.

SOBREPEREZ, P. Using plenary focus groups in information systems research: More than a collection of interviews. **Electronic Journal of Business Research Methods**, v. 6, n. 2, p. 181–188, 2008. Cited on page 49.

SOLTANI, S. R.; MONAVARI, S. M.; MAHINY, A. S. Urban land use management, based on GIS and multicriteria assessment (Case study: Tehran Province, Iran). In: **Proceedings of the 2011 International Conference on Multimedia Technology (ICMT)**. Hangzhou, China: [s.n.], 2011. p. 6644–6647. Cited on page 60.

SPAGNOLETTI, P.; RESCA, A.; SÆBØ, Ø. Design for social media engagement: Insights from elderly care assistance. **The Journal of Strategic Information Systems**, v. 24, n. 2, p. 128–145, 2015. Cited on page 134.

SPRAGUE, R. H.; CARLSON, E. **Building Effective Decision Support Systems**. [S.l.]: Prentice-Hall: Englewood Cliffs, New Jersey., 1982. Cited on page 58.

STOLLBERG, B.; ZIPF, A. Geoprocessing Services for Spatial Decision Support in the Domain of Housing Market Analyses Experiences from Applying the OGC Web Processing Service Interface in Practice. In: **Proceedings of the 11th International Conference on Geographic Information Science (AGILE)**. Girona, Spain: [s.n.], 2008. p. 1–10. Cited on page 60.

SUGUMARAN, V.; SUGUMARAN, R. Web-based Spatial Decision Support Systems (WebS-DSS): Evolution, Architecture, Examples and Challenges. **Communications of the Association for Information Systems (AIS)**, v. 19, n. 1, p. 1–23, 2007. Cited on page 151.

SULAIMAN, N. A.; HUSAIN, F.; HASHIM, K. A.; SAMAD, A. M. A study on flood risk assessment for Bandar Segamat sustainability using remote sensing and GIS approach. In: **Proceedings of the 2012 IEEE Control and System Graduate Research Colloquium (ICSGRC)**. Shah Alam, Malasya: [s.n.], 2012. p. 386–391. Cited on page 60.

SUSMAN, G. I.; EVERED, R. D. An assessment of the scientific merits of action research. **Administrative Science Quarterly**, v. 23, n. 4, p. 582–603, 1978. Cited on page 134.

TAYLOR, K.; GRIFFITH, C.; LEFORT, L.; GAIRE, R.; COMPTON, M.; WARK, T.; LAMB, D.; FALZON, G.; TROTTER, M. Farming the web of things. **IEEE Intelligent Systems**, v. 28, n. 6, p. 12–19, 2013. Citations on pages 42, 63, 130, and 163.

TSAI, M. H.; SUNG, E. X.; KANG, S. C. Data-driven flood analysis and decision support. **Natural Hazards and Earth System Sciences Discussions**, v. 2016, p. 1–14, 2016. Cited on page 38.

TVERSKY, A.; KAHNEMAN, D. Judgment under uncertainty: Heuristics and biases. **Science**, v. 185, n. 4157, p. 1124–1131, 1974. Cited on page 37.

UN. **Sendai Framework for Disaster Risk Reduction 2015-2030**. [S.l.], 2015. Technical Report. Citations on pages 42 and 168.

UNISDR. **2009 UNISDR Terminology on Disaster Risk Reduction**. [S.l.], 2009. Cited on page 55.

VASILIU, A. **Dashboards and Scorecards: Linking Management Reporting to Execution,**. 2006. <http://dssresources.com/papers/features/vasiliu/vasiliu04302006.html>. Cited on page 61.

VERA-BAQUERO, A.; COLOMO-PALACIOS, R.; MOLLOY, O. Business process analytics using a big data approach. **IT Professional**, v. 15, n. 6, p. 29–35, 2013. Citations on pages 42 and 63.

VIEWEG, S.; CASTILLO, C.; IMRAN, M. Integrating social media communications into the rapid assessment of sudden onset disasters. In: **Proceedings of the 2014 International Conference on Social Informatics (SocInfo)**. Barcelona, Spain: [s.n.], 2014. p. 444–461. Citations on pages 38, 41, 42, 64, 105, 128, 130, 133, 140, 163, 165, and 168.

VIVACQUA, A. S.; BORGES, M. R. Taking advantage of collective knowledge in emergency response systems. **Journal of Network and Computer Applications**, v. 35, n. 1, p. 189–198, 2012. Citations on pages 56, 57, and 77.

WALLS, J. G.; WIDMEYER, G. R.; SAWY, O. A. E. Building an information system design theory for vigilant eis. **Information Systems Research**, v. 3, n. 1, p. 36–59, 1992. Cited on page 135.

WAMBA, S. F.; AKTER, S.; EDWARDS, A.; CHOPIN, G.; GNANZOU, D. How "big data" can make big impact: Findings from a systematic review and a longitudinal case study. **International Journal of Production Economics**, v. 165, p. 234–246, 2015. Citations on pages 42, 63, 105, 130, 133, 134, 140, 151, 163, and 168.

WAN, Z.; HONG, Y.; KHAN, S.; GOURLEY, J.; FLAMIG, Z.; KIRSCHBAUM, D.; TANG, G. A cloud-based global flood disaster community cyber-infrastructure: development and demonstration. **Environmental Modelling & Software**, v. 58, p. 86–94, 2014. Citations on pages 41, 103, and 168.

WANG, S.; LI, M. Analysis of Optimized Logistics Service of 3C Agents in Taiwan Based on ABC-KMDSS. **Applied Mathematics & Information Sciences**, v. 7, n. 1L, p. 307–312, 2013. Cited on page 59.

WEAVER, A.; BOYLE, J.; BESALEVA, L. Applications and Trust Issues When Crowdsourcing a Crisis. In: **Proceedings of the 21st International Conference on Computer Communications and Networks (ICCCN)**. Munich, Germany: [s.n.], 2012. p. 1–5. Cited on page 77.

WIDENER, M. J.; HORNER, M. W.; METCALF, S. S. Simulating the effects of social networks on a population's hurricane evacuation participation. **Journal of Geography Systtems**, v. 15, n. 2, p. 193–209, 2013. Cited on page 86.

YATES, D.; PAQUETTE, S. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. **International Journal of Information Management**, v. 31, n. 1, p. 6–13, 2011. Citations on pages 77 and 86.

YIN, R. K. **Case study research: Design and methods**. [S.l.]: Sage publications, 2013. Cited on page 193.

YONGSONG, Z.; SIUMING, L.; KWOKKIT, Y. GIS Based Urban Emergency Decision Support Model. In: **Proceedings of the 4^{th} International Conference on Management and Service Science (MASS)**. Wuahn, China: [s.n.], 2010. p. 1–4. Cited on page 60.

YOU, Z.; SI, Y.-W.; ZHANG, D.; ZENG, X.; LEUNG, S. C.; LI, T. A decision-making framework for precision marketing. **Expert Systems with Applications**, v. 42, n. 7, p. 3357–3367, 2015. Citations on pages 105 and 133.

ZHANG, C.; LI, W. The roles of web feature and web map services in real-time geospatial data sharing for time-critical applications. **Cartography and Geographic Information Science**, v. 32, n. 4, p. 269–283, 2005. Citations on pages 41, 94, 103, and 147.

ZHANG, H.; BABAR, M. A.; TELL, P. Identifying relevant studies in software engineering. **Information & Software Technology**, v. 53, n. 6, p. 625–637, 2011. Cited on page 73.

ZHENG, L.; SHEN, C.; TANG, L.; LI, T.; LUIS, S.; CHEN, S.; HRISTIDIS, V. Using data mining techniques to address critical information exchange needs in disaster affected public-private networks. In: **Proceedings of the 16^{th} ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)**. Washington, USA: [s.n.], 2010. p. 125–134. Cited on page 61.

ZHOU, K.; FU, C.; YANG, S. Big data driven smart energy management: From big data to big insights. **Renewable and Sustainable Energy Reviews**, v. 56, p. 215–225, 2016. Citations on pages 130 and 165.

ZOOK, M.; GRAHAM, M.; SHELTON, T.; GORMAN, S. Volunteered geographic information and crowdsourcing disaster relief: A case study of the haitian earthquake. **World Medical & Health Policy**, v. 2, n. 2, p. 7–33, 2010. Citations on pages 40 and 66.

APPENDIX

A

CASE STUDY PROTOCOL AT CEMADEN

The case study protocol is an essential document when a researcher is performing a case study (YIN, 2013). This is because it defines all the elements of the research design of the study, from the research question to the rules for analyzing and publishing the data. As a consequence, it increases the credibility and reliability of the study by pointing out its threats to validity, and providing a research guide not only for guiding the researcher but also for allowing the reproducing the case study.

Therefore, this appendix presents the case study protocol that was used as a basis for conducting the case study at the Brazilian National Center for Monitoring and Early Warning of Natural Disasters (Cemaden).

1. Introduction

This document is structured as follows: the following sub-sections outlines the objective of this case study, as well as its research question and research context. Section 2 outlines the research design, namely the object of study, the variables, and criteria for the case selection. Section 3 details the research instruments while Section 4 describes the process of data analysis. Section 5 presents the threads to validity. Eventually, Section 6 details the research reporting.

1.1. Objective and Research Question

This case study aims to gather data from an emergency agency in Brazil for analyzing its processes of managing an alert of natural disaster. By managing an alert, we mean all the processes from the monitoring of environment variables through sensors (e.g. metereological stations) to giving and closing an alert. Therefore, this objective leads to the following research question:

> *RQ) How are the warnings about a natural disaster managed by the decision-makers of a Brazilian emergency agency?*

Furthermore, this study follows the activities defined by the case study life-cycle proposed by Yin (2009) and Runeson & Höst (2009).

1.2. Research Context: The National Early Warning and Monitoring Centre of Natural Disasters (CEMADEN)

The National Center for Monitoring and Early Warning of Natural Disasters (CEMADEN), which is a branch of the Brazilian Ministry of Science, Technology and Innovation (MCTI), was created in 2011, after the disaster of the Serrana Mountain Region of Rio de Janeiro State. Among CEMADEN's activities, the following tasks play a crucial role in supporting disaster management in Brazil: (1) the development of warning for impending disasters that can support prevention and response measures; (2) the development and implementation of monitoring systems for natural disasters; (3) the operation of these monitoring systems; and (4) the issuing of warnings of imminent natural disasters to the National Center of Disaster Risk Management (CENAD, in Portuguese).

Although CEMADEN has its own monitoring equipment (including hydrology stations, meteorological stations, and automated rainfall gauges), the center works in collaboration with several institutions such as the National Water Agency (ANA), the Brazilian Geological Research Company (CPRM), and the National Institute of Meteorology (INMET). These provide further data about weather conditions, risk maps, and environmental variables, which thus supplement the existing data of the center. All these data can thus be made available to support the monitoring activities and warning services carried out inside a Monitoring Room (Figure 1).

Figure 1. Monitoring room at the CEMADEN

The room is equipped with a video wall, which can display different kinds of information, e.g. the number of open warnings and data from the hydrology stations, or flooded areas. This is used by a monitoring team that works in six-hour shifts, starting at midnight. The size of these teams ranges from five to seven members, and includes at least one specialist in each area (hydrology, meteorology, geology, and disaster management). The monitoring team has three channels of communication: (1) between the members, (2) with the CENAD, and (3) for urgent notifications. Since all of the members are inside the monitoring room, the communication can take place on an individual basis, or a meeting of the team might be held in the middle of the room. In turn, the communication with CENAD, which is a branch of National Secretariat of Civil Defense, is conducted through e-mail by the disaster management specialist. If it is an emergency - e.g. either a serious disaster is imminent or a disaster is already occurring - the warning can be issued to CENAD by phone or through video conferencing so that the contingency plans of the National System of Civil Defense can be activated.

Finally, the early warning levels are defined by a matrix (Table 1), which intersects two variables - the potential impact and the possibility of an occurrence of a natural disaster - and this can divided into three levels - moderate, high, and very high.

		Potential Impact		
		Moderate	High	Very High
Possibility	Very	Moderate	High	Very

of Occurrence	High			High
	High	Moderate	High	Very High
	Moderate	Moderate	Moderate	Moderate

Table 1. A matrix of warnings showing different levels of disaster

Here, there is an assessment of the potential impacts of a disaster and the estimated damage in terms of the people in the communities and buildings affected; while the risk factor is represented by the analysis of variables such as the vulnerability of communities, weather conditions, land use, and inhabitants exposed to risk.

1.3. Theoretical Background

Early warning has been defined as "the provision of timely and effective information, through identified institutions, that allows individuals exposed to a hazard to take action to avoid or reduce their risk and prepare for effective response" (UNISDR, 2004). Early warning systems define the technological infrastructure that can assist in carrying out these tasks. A technological infrastructure has been installed as a means of supporting data processing and forecasting natural disasters; it is based on expert models in early warning systems. However, these systems need to go beyond this infrastructure, by taking account of how risks are understood and providing information for an early warning system. This is because these factors might be required for triggering actions that can prevent or mitigate a disaster.

Early warning systems are widespread within the field of disaster management and act as an important alternative to supporting disaster preparedness and response. Picozzi et al. (2015) devised an early warning system for earthquakes which provides alert messages within about 5 to 10 s for seismic hazard areas, while Alfieri et al. (2012) analyzed a European operational warning system for water-related disasters. Another line of inquiry has been to use information from crowdsourcing platforms - e.g. Twitter,, and OpenStreetMap - to provide updated information for early warning systems. In their work, Chatfield and Brajawidagda (2013) have demonstrated that social media messages could act as a supplementary source of information in disaster detection. The use of crowdsourcing was also explored in the work of Meissen and Fuchs-Kittowski (2014). This work employs crowdsourcing either as input data for further model processing or as input data for checking the plausibility of prediction model outputs or to augment the overall picture of the hazardous situation. There has also been some work on disaster management that is aimed at modeling the tasks of the decision-makers. Within this group, McEntire and Myers (2004) have outlined the tasks and procedures that should be carried out to prepare communities for disasters. Blecken (2010) introduced a reference task model, which supports humanitarian organizations in modeling and optimizing their supply chain management.

The works cited above make clear that there are a considerable number of studies on the use of early warning systems for disaster management but most of these only lay emphasis on the technological infrastructure of the systems. Similarly, although there are a very few works that examine the question of decision-making in disaster management, none of them discusses the early warning process. It is important to analyze this because it might provide evidence of the information bottleneck or the misuse of information systems by the players, In this way, improvements could be made with the aim of increasing efficiency, reducing the number of false alarms, and shortening the response time required for warnings.

2. Design

This section presents the case study design. First, it describes the object of study (or: unit of analysis), i.e. the phenomenon that has been observed and analyzed. This definition is essential for limiting the scope of the study, as well as linking the case study to its objective and research question. Finally, this section describes the criteria for selecting a case.

2.1. Object of Study

Since the objective of this case study is to analyze the management of warnings of a natural disaster from the CEMADEN, its object of study are all the processes performed by its members, mainly those that are related to this task. A multiple-case unit of analysis aims at getting a clear understanding about these processes. Here, the processes performed by a member of the CEMADEN are considered an unit of analysis. Figure 2 displays these definitions of context, case, and unit of analysis.

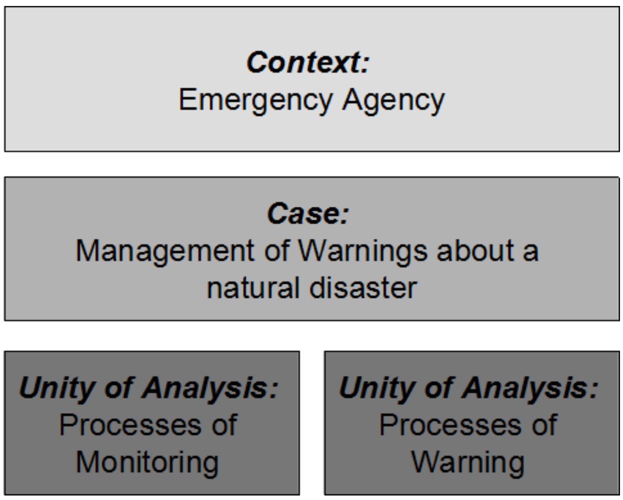


Figure 2. Definition of the case study

The unit of analysis then can be analyzed by the following variables, which can be used for analyzing the processes of an institution. These variables were raised from the a literature review, as well as well-known standards of modeling and notation (Blecken, 2010; OMG, 2013; OMG, 2014; OGC, 2013; Horita et al. 2015; NOAA, 2010; Léon et al., 2006; Red Cross, 2009).

- ***Activity:*** this variable relates to a generic term for the work which a institution performs.
- ***Business Rule (or: Business Knowledge):*** this variable is associated with the functions or rules, which can be used for supporting decision-making. This could be a decision table, or an analytic model (e.g. hydrology model).
- ***Decision:*** this represents the task of choosing an alternative among different options by giving a specific input data and business rules.
- ***Information Requirement:*** this variable represents all the information which is required for supporting decision-making, e.g. the input data for a decision. It can be represent by either a simple data items or complex data structures but rather than a data source.
- ***Actor:*** the stakeholders related to an activity or a decision by interacting directly (e.g. making a decision) or indirectly (e.g. processing data). They can be a single actor (e.g. hydrologist or disaster manager) or an institution (e.g. National Water Agency).
- ***Data Source:*** this is associated with all resources that could used for providing the required information to an activity or a decision, e.g. a report, an institution, or an information system. In contrast to an information requirement, this represents the source of information rather than the information itself.
- ***Relationship:*** this variable denotes the relationship between other two elements, e.g. a disaster manager (element actor) decides whether give or not an alert (element decision) or an institution report (element data source) may inform that an airport is closed (element information requirement). This also indicates which element is the consumer - when requires the output of other element - and which is the provider - when provides an input for other element.
- ***Sequence Flow:*** this variable defines the order, which activities will be performed.
- ***Warning Levels:*** the warnings are normally separated in three levels - e.g. moderate, high, and very high - that indicate which activities should be performed. Moreover, some of the early warning systems define colors for these level as a way of making easier the communication and dissemination of a warning. This variable includes the elements associated with warning levels.
- ***Warning Elements:*** this variable defines elements about the way of how the warning is written and spoken, e.g. speed, accuracy, and clarity.
- ***Risk Knowledge:*** this variable indicates the elements related to the understanding of risk, and thus supporting risk assessment, e.g. measurements of risk, concept of risk, definition of risk, areas vulnerable, risk maps, etc.
- ***Communication***: this variable comprehends the elements, which are related to the dissemination of information.

These variables were used in turn for defining a set of case study topics, which supported the development of the questionnaire (see Table 1).

2.3. *Case Selection*

Since this case study is looking to the management of a warning about a natural disaster into an emergency agency in Brazil, the selected case needs to fulfill the following case criteria:

CC1: the case of study works for an emergency agency.

CC2: the case of study manages (or had managed) warnings in the emergency agency.

CC3: the case of study has attended to at least one emergency situation before.

CC4: the case of study is a hydrologist, meteorologist, geologist, or a disaster manager.

3. Research Instruments

This section details the information about the process of data collection. In this case study, we will collect data by means of semi-structured interviews with decision-makers, and participatory observation at the CEMADEN. The semi-structured interviews were chosen due to the flexibility given to the interviewer, which can improve the interview by asking further questions beyond to those defined beforehand (Yin, 2009). Moreover, these interviews are expected to last a maximum of an hour, which needs to fit at the participants availability of schedule. Finally, the interview will be recorded only if the participant allows it.

For guiding these interviews, the following sub-section presents an interviews plan, which is composed by all the information and questions that need to be gathered with the interviewees.

3.1. Interview Plan and Questionnaire

The phases of the interview plan are threefold: (a) the initial brainstorming about the general information of the CEMADEN; (b) the semi-structured questions; (c) the final considerations. In the beginning of the interview, the participant is asked for fulfill the consent form, which details the scope of the study, and procedures used.

The initial brainstorming aims to gather the following data:

- Academic and Professional Background

- Current position at the CEMADEN
- Department
- Previous experiences in the field of disaster, and on the current position
- Size of the team, e.g. number of members, and background of the members
- The interviewee has started as a temporary staff or permanent staff

This first phase is expected to take no more than 15 minutes (or 1/4 of the interview time). After this, we are ready to start the second phase, the semi-structured questions. This will last more time like 30 minutes (or 2/4 of the interview time) because it aims to gather all the data that could be used for answering the preliminary questions. Table 1 displays the case study topic, the variables, questions, and the question to the interviewee.

Table 1. Questionnaire

<table>
<tr><th>Topics</th><th>Variables</th><th>Questions</th><th>Questions to the interviewee</th></tr>
<tr><td rowspan="8">Processes (OMG, 2013; OMG, 2014)</td><td rowspan="2">Activity</td><td>1. Which are the activities that you perform?</td><td rowspan="6">1 What are the activities (or: tasks) that you perform within the CEMADEN?</td></tr>
<tr><td>2. Do you think that a specific activity happens in different events?</td></tr>
<tr><td>Decision</td><td>3. Which of the activities could be defined as a decision?</td></tr>
<tr><td>Sequence Flow</td><td>4. Is there any clear sequence of activities?</td></tr>
<tr><td>Information Requirement</td><td>5. What are the information that you need to make the decision?</td></tr>
<tr><td rowspan="2">Data Source</td><td>6. What are the data sources considered for gathering the required information?</td></tr>
<tr><td>7. Is there any artifact, which guides the activities of CEMADEN?</td><td rowspan="2">2. Is there any artifact, which guides the activities of CEMADEN?</td></tr>
<tr><td>Business Knowledge</td><td>8 Is there any decision logic (or: algorithm) that you use for supporting the decision?</td></tr>
<tr><td></td><td>Risk Knowledge</td><td>9. What are the measurements used for indicating an upcoming</td><td>3. What do you understand by risk management? How</td></tr>
</table>

<table>
<tr><td rowspan="9">Warning (Léon et al., 2006; Red Cross, 2009; NOAA, 2010)</td><td rowspan="3"></td><td>event?</td><td rowspan="3">can we measure risk?</td></tr>
<tr><td>10. Which are the elements that indicate an upcoming event?</td></tr>
<tr><td>11. What do you understand by risk and warning?</td></tr>
<tr><td>Warning Levels</td><td>12. In which case the warning level can be changed? e.g. from moderate to high or high to very high.</td><td>4. How is the warning managed? e.g. increasing the warning level, canceling it, or improving its quality.</td></tr>
<tr><td rowspan="3">Warning Elements</td><td>13. What is the efficiency of the warnings?</td><td rowspan="3">5. How is the efficiency and accuracy of a warning assessed?</td></tr>
<tr><td>14. How is the accuracy of a warning assessed?</td></tr>
<tr><td>15. Are the information provided by the warning useful?</td></tr>
<tr><td>Actor</td><td>16. What are the stakeholders involved on the activity of monitoring and warning?</td><td rowspan="4">6. What are the stakeholders involved with your work? How is the communication among them?</td></tr>
<tr><td>Relationship</td><td>17. What is your relationship with the other actors?</td></tr>
<tr><td></td><td rowspan="2">Communication</td><td>18. How is the communication between actors performed? e.g. via document, phone, information system.</td></tr>
<tr><td></td><td>19. Are the risks and the warnings well understood?</td></tr>
</table>

Once we are done with the application of the questionnaire, the third phase - the final considerations - can be started. Within this final phase, the interviewee is asked for making any final consideration about the topics discussed during the interview. The interviewer can also come up with some comment or observation. When both the interviewee and the interviewer have no further consideration, the interview is then closed. This final phase should take no more than 10 minutes.

3.2. Observations

In addition to the interviews, this case study also expects to perform observations within the monitoring room at the CEMADEN. Observations are used for case studies because they may provide a deep understanding of the phenomenon that is studied from its real scenario (Runeson & Höst, 2009). Moreover, the analysis of the daily activities into the monitoring room might provide new insights, and even consolidate the outcomes of the interviews.

The approach adopted for the data collection is to monitor a group of decision-makers in their daily activities at the monitoring room, take notes of the events of interest, and later on analyze them in combination with the interviews outcomes. We have chosen to perform simple observations with *low degree of interaction by the researcher* (Runeson & Höst, 2009). This is because the access to the decision-makers into to the monitoring room is limited, and deeper data collection techniques (e.g. think aloud) are difficult to be executed.

The script defined for collecting data at the monitoring room is composed by two phases: (1) the understanding of the monitoring room, and (2) the observation of the day-to-day activities. The first phase should be performed with a member of the monitoring room with the aim of collect the following data:

1. The objectives of the monitoring room;
2. The tasks, which are performed into the room;
3. The organization of the members within the room;
4. The role of each member;
5. The workspace of each member;
6. The available information systems;
7. The information systems, which are used;
8. The other available resources, e.g. whiteboards and videowall;
9. The role of each available resource;
10. The protocols inside the room.

This first phase is expected to take no more than 20 minutes. Into the second phase, we expect to collect data for analyzing the processes (or: activities) of managing warning of natural disaster. The variables defined in Section 2.1 - *Activity, Business Knowledge, Decision, Information Requirements, Actor, Data Source, Sequence Flow, Warning Levels, Warning Elements, Risk Knowledge, Communication, and Relationship* - were used as points of interest for structuring the data collection, e.g during the field work, the observer will consider the variable "*Communication*" for analyzing how the communication among the teams is done, or the analysis about the order of when the activities are performed is part of the variable "*Sequence Flow*". This second phase probably will take more time like four or six hours.

References

Adger, N. (2000) Social and ecological resilience: are they related?. *Progress in Human Geography*, v. 4, No. 3, pp. 347–64.

Alfieri, L.; Salamon, P.; Pappenberger, F.; Wetterhall, F., and Thielen, J. (2012) Operational early warning systems for water-related hazards in Europe. *Environmental Science & Policy*, v. 21, pp. 35-49. DOI:10.1016/j.envsci.2012.01.008.

Blaikie, P., Cannon, T., Davis, I., and Wisner, B. (1994) At Risk: Natural Hazards, People, Vulnerability, and Disasters. London and New York: Taylor & Francis Group.

Blecken, A. (2010) Humanitarian Logistics: Modeling Supply Chain Processes of Humanitarian Organisations, Haupt. [Online] https://goo.gl/pSOzYH.

Chatfield, A. T. and Brajawidagda, U. (2013) Twitter Early Tsunami Warning System: A Case Study in Indonesia's Natural Disaster Management. *Proceedings of the 46th Hawaii International Conference on System Sciences (HICSS)*, pp. 2050-2060, Wailea, Hawaii. DOI: 10.1109/HICSS.2013.579.

De León, J. C. V.; Borgardi, J.; Dannenmann, S. and Basher, R. (2006) Early Warning Systems in the context of Disaster Risk Management. *Entwicklung and Ländlicher Raum 2*, pp. 23-25.

Haworth, B. and Bruce, E. (2015) A Review of Volunteered Geographic Information for Disaster Management. Geography Compass, pp.237-250. DOI: 10.1111/gec3.12213

Horita, F. E. A., Link, D., Albuquerque, J.P., and Hellingrath, B. (2016) oDMN: An Integrated Model to Connect Decision-Making Needs to Emerging Data Sources in Disaster Management. In the *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*, pp.2882-2891, Kauai, Hawaii. DOI: 10.1109/HICSS.2016.361

Horita, F. E. A., de Albuquerque, J. P., Degrossi, L. C., Mendiondo, E. M., and Ueyama, J. (2015). Development of a spatial decision support system for flood risk management in brazil that combines volunteered geographic information with wireless sensor networks. *Computers & Geosciences*, v. 80, pp. 84-94. DOI:10.1016/j.cageo.2015.04.001

Horita, F. E. A.; Degrossi, L. C.; Assis, L. F. F. G., Zipf, A., and Albuquerque, J. P. (2013) The use of Volunteered Geographic Information (VGI) and Crowdsourcing in Disaster Management: a Systematic Literature Review". Proceedings of the 19th Americas Conference on Information Systems (AMCIS), Chicago, USA, pp. 1–10. [Online] http://goo.gl/ILdjwG.

Lindell, M., Prater, C. and Perry, R.W. (2007) Introduction to Emergency Management. New York: John Wiley.

Link, D., Hellingrath, B., and Bültemann, C. (2015) Information Categories for Infrastructure and Logistic Resource Assessments in Humanitarian Logistics. In Dethloff, J., Haasis, H., Kopfer, H., Kotzab, H., & Schönberger, J. (Eds.), Logistics Management (pp. 445–453). Lecture Notes in Logistics. Springer International Publishing. DOI: 10.1007/978-3-319-13177-1_34

Londe, L; Coutinho, M. P.; Di Gregório, L. T.; Santos, L. B. L.; Soriano, S. (2014) Water-related disasters in Brazil: perspectives and recommendations, *Ambient. & Soc.*, v. 17, pp. 133-152. DOI: 10.1590/1809-4422ASOC1082V1742014

McEntire, D. A. and Myers, A. (2004) Preparing communities for disasters: issues and processes for government readiness. *Disaster Prevention and Management: An International Journal*, v. 13, pp. 140 - 152. DOI: 10.1108/09653560410534289

The United Nations Office for Disaster Risk Reduction (UNISDR). (2004) 2004 Terminology: basic terms of disaster risk reduction, *International Strategy for Disaster Reduction Secretariat*, Geneva. [Online] http://goo.gl/UT0P5W.

Brazilian Institute of Geography and Statistics (IBGE). (2014) Perfil dos Municípios Brasileiros 2013.

Meissen, U. and Fuchs-Kittowski, F. (2014) Towards a reference architecture of crowdsourcing integration in early warning systems. *Proceedings of the 11th International Conference on Information Systems for Crisis Response and Management (ISCRAM)*, pp.334-338, University Park, USA. [Online] http://goo.gl/39QYVv.

Object Management Group (OMG). (2013) Business Process Model and Notation (BPMN), Version 2.0 (formal/2013-12-09). [Online] http://goo.gl/ev6tUR.

Object Management Group (OMG), Decision Model and Notation (DMN), (dtc/2014-02-01), 2014. [Online] http://goo.gl/VDV4tW.

Open Geospatial Consortium (OGC), Observations and Measurements (O&M), (OGC 10-004r3), 2013. [Online] http://goo.gl/udPX9T.

Picozzi, M.; Zollo, A.; Brondi, P.; Colombelli, S.; Elia, L., and C. Martino. (2015) Exploring the feasibility of a nationwide earthquake early warning system in Italy. *Journal of Geophysical Research*, v. 120, pp. 2446-2465. DOI:10.1002/2014JB011669

Runeson, P. and Höst, M. (2009) Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, pp. 131-164.

United Nations (UN). Sendai Framework for Disaster Risk Reduction 2015-2030. [Online] http://goo.gl/E6lM74.

Yin, R. K. (2009) Case study research: design and methods. 4. ed. Thousand Oaks: SAGE Publications, Inc.

APPENDIX

B

FOCUS GROUP PROTOCOL AT CEMADEN

Since the generated model diagrams were evaluated together with members of Cemaden, we decided by conducting focus group sessions with the aim of gathering feedback from practice on the effectiveness of the framework, as well as at identifying recommendations for improvements and/or gaining new insights. These focus group sessions would also provide evidence on whether the modeled business process, decision, and data sources were implemented into the context of Brazilian National Center for Monitoring and Early Warning of Natural Disasters (Cemaden). Therefore, this protocol was adopted as a guideline for conducting the those sessions at the Cemaden. It consists of two parts: (a) the sequence of activities and (b) the questionnaire, and should last no more than 30 minutes.

Study Protocol

- **Welcome.** Start with a short presentation of the moderator, acknowledge the attendance of the participants, and provide an explanation of the overall concept of the focus group.

- **Objective.** Introduces briefly the objective of the study, i.e., end-user assessment in regarding the effectiveness of the modeled documents for supporting the activities in the monitoring control room of Cemaden; and mention the research context, i.e., Pró—alerta project, the previous collaborations (e.g., ISCRAM '16 and DSS), and future potential collaborations (e.g., Natural Hazard).

- **Further information about the session.** Mention that the whole session with being audio-recorded but all the participants will be anonymized. Due to the audio recording, recommend that the participants speak in an appropriate volume with the aim of ensuring the quality of the recording. The session is expected to last no more than 30 minutes.

- **Introduction of the modeled documents.** Introduce the following documents: (a) the modeled business process by describing the stakeholders, activities, sequence flow, and lanes (3 min); (b) the bridge of an activity with a decision and detail the elements of the decision, the information requirements and business rules (3min); and (c) the way of required information connects to data sources (3min).

- **Ask the questions.** Ask the questions for the participants. Do not forget to turn on the audio recorder.

- **Closure.** Mention the possibility of future collaborations, as well as the ongoing works.

Questionnaire

- Do you believe that the presented business process reflects the reality in the monitoring control room?
  - If not, why? What would you change?
  - For you, which are the strengths and weakness of the business process?
- In regarding the presented business process, do you believe that it would be useful if adopted as a guideline for the monitoring teams, e.g., as a protocol describing the sequence of activities that all the teams should follow?
  - If not, why? What would you change?
  - If yes, why?
  - For you, which are the strengths and weakness?
- In regarding the visualization format, do you believe that it makes easy the understanding of the business process as a whole?
- In the presented business process, what are the most critical activities and how they should be prioritized?
  - Why?
- In regarding the presented decision, do you believe that the visualization format makes easy the understanding of the decision as a whole?
- Based on the presented decision, do you have a clear understanding of the other associated decisions, as well as the information required for decision-making?
  - Which are the strengths and weakness?
- How useful is the presented decision for supporting decision-making?

  - If not, why?

- Do you believe that this presented decision reflects the reality in the monitoring control room?

  - If not, why? What would you change?

- Apart of this presented decision, what are the other decisions that you can identify in the presented business process?

  - Which are the most critical and how they should be prioritized?
  - Why?

- Finally, there is a connection of decision-makers' tasks with data sources, how useful is this for understanding the decision-making as a whole?

  - Why?
  - Which are the strengths and weakness?

- Do you consider that these presented documents can be applied for modeling another decision-making, e.g., monitoring of landslides?

  - Why?

APPENDIX

C

PROTOCOL OF DESIGN PRINCIPLES EVALUATION AT CEMADEN

This study protocol was defined for collecting data to investigate the acceptance of the features in supporting (or improving) decision-making in the monitoring control room of Brazilian National Center for Monitoring and Early Warning of Natural Disasters (Cemaden). Since it was basically focused on the features developed for operationalizing the design principles proposed in Chapter 6, the results provided evidence about the proposed principles for supporting decision-making in dynamic environments with heterogeneous spatial data sources. It consists of two parts: (a) the sequence of activities and demonstration session, and (b) the questionnaire, and should last no more than 15 minutes.

Study Protocol

- **Welcome.** Provide a short explanation of research context and previous works.
- **Objective.** Introduces briefly the objective of the study, i.e., to gather the end-users' feedback about the features of the presented SDSS. Mention that the SDSS prototype should be understood as a supplementary component to the existing system (SALVAR), which adds new features for analyzing geospatial data and configuring the system.
- **Introduction of the developed features.** Introduce the following features: (a) the configuration of decisions; (b) the configuration of data sources; (c) the configuration of the location of interest; and (d) the multiscale visualization of data.
- **Ask the questions.** Ask the questions for the participants. Do not forget to turn on the audio recorder.
- **Closure.** Acknowledge the support and time of the participant.

Questionnaire

- What is your opinion in regarding the multiscale visualization?
  - Which are the strengths and weakness?
  - If it is not interesting, why?
- How important are the geospatial data for your analysis?
  - Why?
  - For you, which are the strengths and weakness?
- Do you believe that the connection of decisions and information requirements is important for your decision-making?
  - If yes, how this supports (or improves) your decision-making?
- Is the configuration of data sources important for your decision-making?
- Do you have any other comment, recommendation, or suggestion for improving the system?
- Finally, how important are the use of geospatial data and the configuration of a decision for your decision-making?