Gap Analysis on Open Data Interconnectivity for Global Disaster Risk Research

Summary

Disasters are sudden calamitous events that bring great damage, loss or destruction to large populations and regions. They are most often caused by natural hazards such as flood, hurricane, fire, earthquake, etc., and their damaging impacts on people's lives and properties are often aggravated due to inappropriately managed risks. Disaster risk research aims to better understand the process and interaction of various natural phenomena among themselves and with human activities, to better quantify vulnerability and risks, and to gain and disseminate knowledge to aid decision-making in reducing risks and helping people cope with disasters and their aftermath.

This field of research relies heavily on scientific data, including both observations, analysis and simulation data, that are multidisciplinary, heterogeneous, and dispersed across institutional and country boundaries. Increasingly diverse sources of data, including unstructured data such as information from communications, social media, etc., are also beginning to play an important role in disaster studies.

However, significant barriers exist today, preventing the effective use and broader applications of relevant data in disaster research. Gaps in data infrastructure, data sharing policies and data use governance must be addressed to unleash the potential of disaster research in helping regions, especially the developing countries, to improve risk assessment, reduction and mitigation. Two related areas are being studied: *open data*, and *linked data*. The data-driven nature of disaster research demands open access to scientific data, as it is impossible to fully understand the cause and impact of a disaster event without consulting multiple types of data. In addition to open data, disaster researchers face perhaps a greater challenge – to find relevant data sets in a "sea" of distributed and disparate data resources. The next generation data infrastructure must provide linkage of data, helping researchers to find relevant data across distributed data holdings.

As people, countries, and the world become increasingly connected in many ways, understanding and responding to disasters has become a global issue. International cooperation based on open data is crucial to the advancement of scientific research on disasters. A number of international organizations are working toward the goal of global data sharing. Among them, GEO (Group on earth observation), CODATA (The Committee on Data for Science and Technology), ICSU WDS (World Data System), and UN-SPIDER (United Nations Platform for Space-based Information for Disaster Management and Emergency Response) represent such efforts. The earth observation community has practiced and gained valuable experience in sharing their data, while data in many other domains, such as economic and disaster loss data, urban infrastructure data, etc., are still difficult to share, due to the lack of metadata and/or standard structure/format.

The LODGD (linked open data for global disaster risk research) task group originated from the Earth Observation community and is now expanding to include communities of other disciplines. The task group aims to study the mechanism to connect dispersed disaster related scientific data to enable easier and faster discovery, search, and access to data, reducing the barriers faced by researchers today due to limited interconnection of various existing disaster-related data. *Our vision is an interconnected, collective resource of observational and derived disaster-related data that is open, discoverable, and easily accessible by all, enabled by the revolutionary digital technologies today and open access policy embraced by users and providers.*

Organization of the paper

This white paper aims at identifying the gaps and challenges in technology and relevant policies that prevent effective interconnection of disaster related data and information for use in research, education and public engagement. The paper examines the current state of information technology for data management and sharing, as well as policies regarding data availability at various levels, and discusses potential solutions and examples toward open and linked data for disaster research. A second white paper to focus on data infrastructure design and implementation issues is being drafted by the LODGD task group, in collaboration with other CODATA scientists.

This paper is organized as follows:

Chapter 1 begins with an introduction and background on global efforts in the sharing of scientific data.

Chapter 2 examines the current state of disaster data, discussing issues related to open access and linking of the data in various disciplines.

Chapter 3 describes the gaps and challenges of open access and linking of disaster data in the areas of technology, policy and law, and governance and culture.

Chapter 4 describes the scientific issues related to open disaster data: autonomy of disaster data resources, data classification based on dependency relationship of disaster event-supporting data, and the conflict between specialists and the masses toward open access of disaster data.

Chapter 5 summaries the potential technology approaches toward building the next generation data infrastructure to support the community of disaster research and the public.

Chapter 6 discusses lessons learned from several examples of best practices in developing disaster data infrastructures, including Spatial Hazard Events and Losses Database from the University of South Carolina, Hazards Data Distribution System from USGS, Geohazard Supersites from European Space Agency, and Disaster Reservoir from China Ministry of Science and Technology.

Chapter 7 outlines our thoughts on the next-generation disaster data infrastructure. Three potential cyber platforms for linking disaster data are discussed: Disaster Emergency Data Infrastructure, Historical Archived Data Infrastructure, and Disaster Loss Database Infrastructure.

The final chapter documents a number of recommendations from the task group on what the science community and governmental organizations need to do to advance the state of data-driven disaster risk research to improve our understanding of risks and eventually lead to more effective ways of mitigating and reducing impact of disasters on lives and properties.