Gap Analysis on Open Data Interconnectivity for Disaster Risk Research

Final Version

A study report of the CODATA Task Group on FAIR data for Disaster Risk Research – Carol Song, Li Guoqing

Supported by

International Science Council
The global voice for science

CODATA
COMMITTEE ON DATA
INTERNATIONAL SCIENCE COUNCIL

IRDR
Integrated Research on Disaster Risk

Study Panel

Contributing Authors

Carol SONG (Co-Chair), Purdue University, USA
LI Guoqing (Co-Chair), Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, China

Bapon FAKHRUDDIN, Tokin and Taylor, New Zealand
Brenda K. JONES, USGS Earth Resources Observation and Science Center, USA
Edward T.-H. CHU, National Yunlin University of Science and Technology, Chinese Taipei
FAN Jinlong, National Satellite Meteorological Center, China
HUANG Shifeng, China Institute of Water Resources and Hydropower Research, China
HUANG Mingrui, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, China
LI Xiaotao, China Institute of Water Resources and Hydropower Research, China
Masaru YARIME, The University of Tokyo, Japan
Michael RAST, ESA-ESRIN (ESA's Centre for Earth Observation), Italy
Nuha ELTINAY, Public Health England and London South Bank University
QING Xiuling, National Science Library, Chinese Academy of Sciences, China
Rishma MAINI, Public Health England, United Kingdom
Siyka ZLATANOVA, Delft University of Technology, The Netherlands
Susan L. CUTTER, University of South Carolina, USA
Virginia MURRAY, Public Health England, United Kingdom
XIE Xinlu, Chinese Academy of Social Sciences, China
ZHANG Hongyue, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, China

Contact Information

LI Guoqing (ligq@radi.ac.cn)
Carol SONG (carolxsong@purdue.edu)
ZHAO Jing (zhaojing01@radi.ac.cn)

Acknowledgement

Table of Contents

01

Open and Linkable: New Strategies of International Science Data Management

[Please refer to the ISC international accord on open data for definition (Appendix A)]

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 or earthquake, but as the Sendai Framework notes, they can also be caused by man-made hazards and can be of slow onset. The Sendai Framework for Disaster Risk Reduction 2015-2030 was adopted at the Third UN World Conference in Sendai, Japan in March 2015 as a result of stakeholder consultations through the United Nations Office for Disaster Risk Reduction (UNDRR). Sendai Framework is a UN landmark agreement which applies "*to the risk of small-scale and large-scale, frequent and infrequent, sudden and slow-onset disasters caused by natural or man-made hazards, as well as related environmental, technological and biological hazards and risks. It aims to guide the multi-hazard management of disaster risk in development at all levels as well as within and across all sectors.*" (Paragraph 15). The framework therefore promotes an all hazards approach, and calls "*To strengthen technical and scientific capacity to capitalize on and consolidate existing knowledge and to develop and apply methodologies and models to assess disaster risks, vulnerabilities and exposure to all hazards.*" (Paragraph 24 i).

All disasters have damaging impacts on people's lives and properties, which can be aggravated by inappropriate risk management strategies. The Sendai Framework aims to achieve the outcome of a "substantial reduction of disaster risk and losses in lives, livelihoods and health and in the economic, physical, social, cultural and environmental assets of persons, businesses, communities and countries". In order to achieve this over the next 15 years, disaster research will help better understand the process and interaction of various natural phenomena among themselves and with human activities, leading to better quantification of vulnerability and risks. Broad dissemination of the new knowledge gained will aid decision-making in reducing risks and helping people cope with disasters and their aftermath.

More and more research domains are becoming data driven, especially in the face of vastly improved technology and infrastructure that can collect and make huge amounts of data available. Similarly, research on disasters 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. The need for open data and data interconnectivity, i.e., data which is accessible and usable by researchers, decision makers and the public, is most critical in the area of disaster research, management and mitigation.

Reports (Munich Re 2004) have indicated that more natural hazards have occurred in the past sixty years and that the economic and societal impact of disasters has climbed up by five times in the same period of time. In-depth analysis of the current state of disaster scientific data management and acquisition patterns indicates a greater need for interconnection of dispersed scientific data related to disaster risk assessment and reduction. Today, large amounts of disaster related scientific data exist, such as data from monitoring equipment, base maps, evaluation, progress, socio-economic statistics, and so on. They are typically dispersed geographically and owned by various government agencies, research centers, groups and, sometimes, individuals around the world. Researchers often find it difficult, if not impossible, to discover relevant data needed for their study. Even when they identify the data sources, they may not be able to obtain the data due to ownership issues, or the lack of tools to successfully select, transfer, interpret and use the data with their applications.

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 management. In this context, two related areas are being

studied: *open data*, and *data interconnectivity*. 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 stated in the Sendai Framework, disaster risk reduction requires a multidisciplinary approach to decision-making based on the open exchange and dissemination of disaggregated data. The work of the Linked Open Data for Global Disaster Risk Research (LODGD) is an increasingly important activity, endorsed by the Sendai Framework and the United Nations Office for Disaster Risk Reduction (UNDRR) Open-ended Intergovernmental Expert Working Group (OIEWG) on Indicators and Terminology Relating to Disaster Risk Reduction. On 2 February 2017, the UN General Assembly endorsed the 38 indicators proposed by OIEWG for monitoring the seven Global Targets of the Sendai Framework:

A. *Substantially reduce global disaster mortality by 2030, aiming to lower the average per 100,000 global mortality rate in the decade 2020–2030 compared to the period 2005–2015;*

B. *Substantially reduce the number of affected people globally by 2030, aiming to lower the average global figure per 100,000 in the decade 2020–2030 compared to the period 2005–2015;9*

C. *Reduce direct disaster economic loss in relation to global gross domestic product (GDP) by 2030;*

D. *Substantially reduce disaster damage to critical infrastructure and disruption of basic services, among them health and educational facilities, including through developing their resilience by 2030;*

E. *Substantially increase the number of countries with national and local disaster risk reduction strategies by 2020;*

F. *Substantially enhance international cooperation to developing countries through adequate and sustainable support to complement their national actions for implementation of the present Framework by 2030;*

G. *Substantially increase the availability of and access to multi-hazard early warning systems and disaster risk information and assessments to people by 2030.*

To summarize, Targets A-D are outcome targets, and will require disaster loss and damage data; E and G will require national self-assessment and Target F will relate to overseas development of system monitoring.

All the Sendai Framework indicators will be shared and used seamlessly by the Inter-Agency and Expert Group on Sustainable Development Goal Indicators (IAEG-SDGs) to reduce the reporting burden on countries implementing both the Sendai Framework and the SDGs. A similar collaboration by UNDRR will hopefully lead to collaboration for the sharing of a similar agreement for the Paris Agreement.

It is important to urgently enhance the scientific and technical work on disaster risk reduction and its mobilization through the coordination of existing networks and scientific research institutions at all levels and in all regions. In answering this call to the science community, it is of utmost importance to promote and practice the collection, management, opening up and sharing of scientific data related to disaster risk research, as well as the employment of relevant technologies and applications consistently and globally.

Aims

Endorsed by ICSU and CODATA communities, the overarching goal of this white paper is to systematically analyze the needs for the data infrastructure proposed in the Sendai Framework and provide the conceptual building blocks to help realize the Sendai imperative.

This paper aims to identify the gaps in technology and relevant policies that prevent effective interconnection of disaster related data and information for use in research, education and public engagement. It 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 data and data interconnectivity for disaster research.

Expected outcome

Our vision for the next generation disaster risk research data infrastructure is an interconnected, collective repository of observational and derived disaster-related data that is open, discoverable, and easily accessible and usable by all, enabled by the revolutionary digital technologies of today and an open access policy embraced by users and providers. The analysis from this white paper forms the basis for LODGD's next step – two additional white papers that focus on the next generation of data infrastructure and national polices for disaster risk reduction research and practice.

1.1 Background

The many benefits of open data, especially scientific data, both observational as well as simulation and analysis data, have long been recognized. Open data typically refers to data that is available to anyone who wishes to view, analyze and use it. In as early as 1957, the World Data Centre system was established to support open access to scientific data collected from the observational programs of the 1957–1958 International Geophysical Year. Originally established in the United States, Europe, Russia, and Japan, the World Data Centre system has since expanded to other countries and to new scientific disciplines. Its holdings include a wide variety of data that cover timescales ranging from seconds to millennia. These data provide baseline information for research in many disciplines, especially for monitoring changes in the geo-sphere and biosphere.

While the idea of open science data has been actively promoted, the rise of the Internet, web services, and the declining cost of computing and storage hardware has significantly lowered the barrier to publish or obtain data, bringing us much closer to a reality of broadly sharing scientific data to help improve society and people's lives.

In 1995,GCDIS (US) stated its position clearly in 'On the Full and Open Exchange of Science data' (National Research Council publication 1995): "*The Earth's atmosphere, oceans, and biosphere form an integrated system that transcends national boundaries. To understand the elements of the system, the way they interact, and how they have changed with time, it is necessary to collect and analyze environmental data from all parts of the world.*"

Studies of the global environment require international collaboration for many reasons: (a) to address global issues, it is essential to have global data sets and products derived from these data sets; (b) it is more efficient and cost-effective for each nation to share its data and information than to collect everything it needs independently; and (c) the implementation of effective policies addressing issues of the global environment requires the involvement, from the outset, of nearly all nations of the world. (d) International programs for global change research and environmental monitoring crucially depend on the principle of full and open data exchange (i.e., data and information are made available without restriction, on a non-discriminatory basis, for no more than the cost of reproduction and distribution) (Li, Liu and Jiang 2009).

International organizations advocating open data strategies include the Organisation for Economic Co-operation and Development (OECD), the United Nations Educational, Scientific, and Cultural Organisation (UNESCO), the International Council for Science (ICSU), the Interdisciplinary Committee on Data for Science and Technology (CODATA), the Inter-Academy Panel on International Issues (IAP), International Charter, the World Academy of Sciences (TWAS), Group on Earth Observation (GEO), Committee on Earth Observation Satellites (CEOS), Integrated Research on Disaster Risk (IRDR) and Future Earth. They have proposed their respective principles on how to access and share the data as in the following two examples:

Group on Earth Observations has accepted a set of high level Data Sharing Principles as the foundation for Global Earth Observation System of Systems (GEOSS 2015). Its 10-Year Implementation Plan GEOSS states that "The societal benefits of Earth observations cannot be achieved without data sharing" and sets out the GEOSS Data Sharing Principles:

- *There will be full and open exchange of data, metadata and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation;*

- *All shared data, metadata and products will be made available with minimum time delay and at minimum cost;*

- *All shared data, metadata and products being free of charge or no more than cost of reproduction will be encouraged for research and education.*

The International Council for Science (ICSU) has successively established the World Data Centre (WDC) and CODATA for information collection, exchange, service and sharing. ICSU is devoted to researching, observing and evaluating relevant data and information as well as their relation with decision-making, with open access of data being an important aspect; CODATA is committed to uplifting the quality, reliability, management and accessibility of data that holds great significance for the whole scientific community. The report on CODATA Strategy shows that data sharing holds a very important position in CODATA's strategic planning:

The mission of CODATA is to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use. Looking across and beyond specific scientific programs such as IRDR and Future Earth, CODATA is well positioned to promote coordination with key international initiatives and programs such as GEO, the European Union's Global Research Data Infrastructures (GRDI2020) project, the Eye on Earth initiative, the WSIS implementation and follow-up process, and the proposed Data Web Forum. For example, there are many potential synergies between the development of the GEOSS Disaster Societal Benefit Area and IRDR data needs, and between a range of GEOSS data and services and Future Earth activities. CODATA could play a lead role in harmonizing data policies, improving data access and interoperability, and developing long-term strategies for data stewardship. CODATA promotes direct cooperation among scientists and engineers by facilitating their participation in international data activities, establishing a worldwide exchange system of science data via Internet, organizing and supporting taskforces, work teams, committees and groups working on specific data issues to carry out international cooperation. The plan on implementation of data-sharing policy concerns both data strategies for International Science and Policy and Institutional Frameworks for Data.

1.2   Open Data Strategies for Disaster Research

Disasters result from interaction between different spheres of the earth and human activities. Ever since the time humans began to inhabit the earth, they have been subject to ceaseless confrontation against disasters and constantly draw lessons from their experience in coping with disasters. As people, countries, and the world become increasingly connected in many ways, understanding and responding to disasters has also become a global issue. The data-driven nature of the scientific research on disasters demands open access to data as it is impossible to fully understand the cause and impact of a disaster event without consulting multiple types of data. We submit that international cooperation based on open data is crucial to the advancement of scientific research on disasters.

Numerous technological advances in instrumentation and computation have enabled the observation and recording of various data before, during and after each disaster, creating a valuable database of information for research into disasters. The most recent improvement in satellite observation, aerial photogrammetry, in development of ground observation stations and advancement of various measurement instruments have significantly improved our ability to obtain information on disasters and increased our capacity to keep the exponentially growing amounts of data. This accelerated growth in data size is due both to the higher ability of data collection and to the increased number of disaster events in recent years.

The data challenge for disaster researchers comes from several aspects: the massive quantity of data, the distributed nature of these data, the heterogeneity and diversity of the data. Compounding these challenges is the lack of data sharing. Due to both policy and technology limitations, it is often difficult to share and access data across disciplines, organizations, and distant geographic locations. The reality of utilizing all the relevant data to better understand, respond to and mitigate disasters is closer today than ever before – progress in cyber infrastructure (e.g., computing, data management/federation/movement) as well as maturation of the Internet and web technologies have laid the groundwork for a framework of open access data to support disaster research and other relevant stakeholders. It should be recognized that open access and sharing of data is more than a concept; it needs to be implemented at both policy and technology levels, thus requiring collaboration among disciplines and cooperation among international groups and organizations.

Sharing of science data on disasters is realized through cooperation among states, institutions and organizations, which has received a lot of concern from international organizations and governments. International data organizations and disaster research entities show keen interest in open access of data; they have proposed relevant strategic plans and enhanced contact with policy makers and research institutes to implement the plans. Among them,

(1) The International Charters for Space and Major Disasters, an entity promoting the policy of open access of data, aims to help authorized users to access satellite imagery at no cost and in a rapid fashion in the event of natural or manmade hazards. Each member organization pledges to provide necessary resources in support for implementation of the rules set in the Charter, to help reduce losses of human lives and properties in the wake of disasters, and thus to save valuable time and provide an efficient means for emergency access to data on unexpected disasters.

(2) With regard to disaster research, ISC, ISSC, UNDRR and RADI are jointly engaging in a 10-year project, *Integrated Research on Disaster Risk (IRDR)*; their working group on disaster loss data is devoted to sharing of multidisciplinary data on disasters, including collection, storage and dissemination of data related to disaster loss.

(3) The *Integrated Risk Governance Project* (IRG) is a research plan of Future Earth, which aims to enhance global risk management capability through a ten-year effort. Achieving the goal necessitates application of data from multiple disciplines, such as weather, climate, ecology, hydrology, geophysics and environmental science.

Promoting open access and sharing of even more disaster related data has become part of the core contents in various joint international research plans and coordination mechanisms on dealing with disasters.

1.3   The Future: From Open Data to Data Interconnectivity

The future of data-driven science lies in the open and easy access of data as science is becoming increasingly more collaborative, both within and across disciplines and geographic boundaries of states and countries. The terms Open Data and Data Interconnectivity emphasize two different aspects of data sharing. The term "open data" refers to the availability of data, which should be freely available to all, while the term "data interconnectivity" emphasizes accessibility and usability, that open data should also be easily accessible by all and usable in modeling, further analysis, validation, etc. The value of science data manifests in its application. Compared with special access to exclusively held data, open access of data promotes and enables data reuse and repurposing across disciplines. With the campaign for open access of data going on, the ratio of exclusively held data will be in steady decline while open access data will increase, with the ultimate emergence of a worldwide open and accessible Internet space of science data. Currently, there is no unified definition of "open data" which is interpreted differently by various organizations and institutions, for example, the Open Data Centre Alliance regards open data as the Company's IT infrastructure, or an application mode and solution of cloud computing; Scholarly Publishing and Academic Resources Coalition (SPARC) regards open data as a new mode and notion of academic publishing of science data; the open data campaign advocated by W3C adopts the RDF data model to establish RDF links among data entities of different types and sources, so as to guide users through ordinary HTML web pages and structured data by specific semantic web explorers or search engines, and finally enable all to have free access to their desired data. The definition embraced by each group or domain serves its specific needs. However, we see "open data" as a guiding principle, both philosophically and practically, that scientific data should be freely accessible by anyone without restrictions such as copyright, patent and any other mechanism.

Open access of scientific data and linkage of open data represent a natural progression of data sharing. Open data access serves as the foundation of data sharing: the democratization of data – access by all without restrictions. Technology implementation is needed to make it easy for users, especially the non-domain experts, to obtain the data. Once open access is achieved, usability becomes the central issue: what does one need in order to make use of the data? Data users face a number of challenges because of complexity in data collection instrumentation and the physics represented by the acquired data, diversity in data formats and tools that deal with the variety of formats, difference in the spatial resolution of the data, and so on. On top of these, researchers face difficulties when dealing with data from a different discipline because of differences in vocabularies employed by different disciplines in describing data. Our concept of *data interconnectivity* focuses on the interconnection of datasets, bringing together data from diverse sources, disciplines and data providers to enable interoperability to support disaster research and applications.

In the area of disaster research, the international community has come to a consensus on the openness of relevant data to support disaster research, and, as a result, many data banks are available now. Due to reasons related to technology, policy and culture, various data are yet to be effectively connected, which has led to the low utilization rate of open data.

The Linked Open Data for Global Disaster Risk Research (LODGD) Working Group of CODATA aims to promote the interconnection and cross-domain utilization of disaster data.

The idea of LODGD was initiated from the experience of implementing the well-known GEO data sharing principle, and was further driven by the experience with technical implementation of distributed data linking. LODGD started from Earth Observation community and is now extending to cross-disciplinary communities.

There are two levels in LODGD concept model: data characterization (lower level) and data connection (higher level). At the lower level is the knowledge about disaster taxonomy and data dependency on disaster events. As a scientific study, this level aims to understand and present the correlation between specific disaster events and science data though integration of literature analysis and semantic knowledge discovery. The higher level concept deals with technical methods to connect the distributed data resources identified by the lower level knowledge given a specific disaster type.

The LODGD Task Group has been studying the mechanism for connecting dispersed disaster related science data to enable easier and faster discovery, access and to significantly reduce the barriers that researchers are facing today due to limited interconnection of various existing disaster-related data.

02

Status of Open Data for disaster research

It is necessary to clearly understand the status of open data that is relevant to disaster research before we analyze the gaps toward a comprehensive and interconnected data infrastructure. In this section, we describe the technical connotation of the concept of open data and examine the current state of open data in a number of scientific disciplines.

2.1 Issues related to open data

2.1.1 Data Accessibility

Accessibility is the driving force toward the landscape of open science data. Data sharing becomes a reality when there are few or no barriers to obtaining the data by experts and non-experts alike.

More than a decade ago, the issue of open access to research literature was discussed and debated. The work from initiatives such as Budapest Open Access Initiative (February 2002), Bethesda Statement on Open Access Publishing (June 2003), and Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (October 2003) directly prompted open access to scientific articles. The view held by this community is as follows:

There are many degrees and kinds of wider and easier access to research literature. By "open access" to this literature, it means its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited. Here's how the Bethesda and Berlin statements put it: For a work to be OA, the copyright holder must consent in advance to let users "copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship".

The "open access" movement initially focused on research literature, such as journal publications, and emphasized on barrier-free reuse of, reference to, revision, copying of and dissemination of scientific or research data, or taking some measures to simply ensure users' better access to data. This concept has since been broadened to include scientific data, observational or derived, especially data that has been obtained through publicly funded research.

In the past few years, ICSU began to discuss access to data and information, and in particular, science data and information, as reflected in ICSU's 2013-2017 Strategy Plan.

The long-term ICSU vision is for a world where science is used for the benefit of all, excellence in science is valued and scientific knowledge is effectively linked to policy-making. In such a world, universal and equitable access to high quality science data and information is a reality and all countries have the scientific capacity to use these and to contribute to generating the new knowledge that is necessary to establish their own development pathways in a sustainable manner. Science is increasingly dependent on access to, and integration of, large data-sets from multiple sources. Both the access to and the interoperability of these data sets present major challenges. Maintaining and expanding a quality-assessed science data bank is a multi-faceted challenge that must be addressed if we are to make scientific progress in areas such as earth system sustainability research.

Access to data and information is critical to the whole scientific enterprise. In particular, it is a rate-limiting step for scientific development in many less-developed countries. A World Data System is being established to ensure long-term stewardship and open worldwide access to essential data-sets and data products.

The first session of the United National Environment Assembly of the United Nations Environment Programme concluded that the state of environment should stick to the following core principles:1)enabling open access to data resources from the government, research projects, non-government organizations, communities and traditional knowledge; 2)sharing data for multiple purposes; 3) reliable data and information management; 4) archiving and tracking the use of data and information; 5)support the public's access to data and information by various means.

The requirement for data accessibility has evolved from the narrow connotation of making data available based on agreed-upon inter-organizational protocols a decade ago. Today, data accessibility almost always means timely access over computer networks. Any implementation of open data access will have to support online, on-demand access methods, as well as access through web services and Application Programming Interfaces (APIs).

2.1.2 Data Sharing

Data sharing is the practice of a data holder/owner making its data available to other users. The sharing aspect deals with issues on data ownership, restrictions (or the lack of) for others to access the data, the provisions of data use and acknowledgement. The broader scientific community recognizes that the value of data is in its utilization and demonstrated impact as a result of that use. Discussions at various forums have aimed at working out the principles and technicalities that would guide the broad sharing of scientific data.

In the 2010 Beijing Declaration, GEO members committed themselves to implement the GEOSS Data Sharing Principles by developing flexible policy frameworks that enable a more open data environment, which has influenced national and regional data policies including INSPIRE and Copernicus in Europe and Landsat in the United States.

The GEOSS Data Sharing Principles are as following:
- There will be full and open exchange of data, metadata and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation;
- All shared data, metadata and products will be made available with minimum time delay and at minimum cost;
- All shared data, metadata and products being free of charge or no more than cost of reproduction will be encouraged for research and education.

One of the first accomplishments of the Group on Earth Observations was the acceptance of a set of high level Data Sharing Principles as a foundation for GEOSS. Ensuring that these principles are implemented in an effective and flexible manner remains a major challenge. The 10-Year Implementation Plan says "The societal benefits of Earth observations cannot be achieved without data sharing". Further, the sharing of GEOSS data will in some cases be subject to important exceptions such as the protection of national security, privacy and confidentiality, indigenous rights, and threatened ecological and cultural resources.

The National Science Foundation (U.S.) requires a two-page data management plan for all proposals submitted. The agency stipulates that "all investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing." This requirement started in 2011, and in just three years, the impact is quite visible. Investigators not only are giving more thoughtful consideration to satisfy this requirement, many incorporate data sharing and accessibility into their proposals as a key component for dissemination of research outcome. This will accelerate inthe coming years as cyber infrastructure for data sharing becomes more mature and widely adopted in the U.S.

2.1.3   Data Interconnectivity

A key challenge in disaster research, whether analyzing causal relationship of various impacts or modeling to predict the impact of future disaster events, is to make use of multiple data sources, synthesize and discover the underlying relationships. At the very basic level, how does a researcher of a particular specialty find datasets that may be relevant to his/her study?

The academic community recognizes the importance of the interconnection of datasets, which often come from different scientific disciplines (e.g., hydrology, meteorology, climate, civil engineering, land use, and public health). The term *data interconnectivity* is about connecting data from diverse sources, disciplines and data providers, and enabling common understanding and interoperability through community-endorsed standards and methods for description and access. The realization of a data interconnectivity landscape will allow scientists to easily and efficiently discover, understand, access and utilize data across disciplines in disaster research, leading to new discoveries and solutions for mitigating impact of disaster and improving readiness in communities that are vulnerable to natural hazards.

Internationally, current efforts include those of the LODGD of CODATA, a research group dedicated to promote the linkage of disaster data. Experts in this group have formed a knowledge network and are researching the technology framework that will serve the disaster research community by connecting relevant open data from repositories at national, institutional and research group levels. Additionally, IRDR has established the Disaster Loss Data (DATA) Working Group. IRDR-Data focuses its efforts on issues related to the collection, storage and dissemination of disaster loss data. We envision a new kind of data infrastructure for disaster research that will connect disaster related datasets of observations, analyses, statistics, etc., from multiple scientific disciplines as well as through citizen participation.

2.2   The Challenges of Disaster Data

In this section, we examine the current state of open science data as related to the research of disaster risk reduction. While recognizing the existence of a large number of data sets, we have included a representative set of data in various disciplines ranging from earth observation, hydrology, meteorology, earthquake, geography to health, economic and disaster loss statistics. While all categories of data listed here present challenges for cross-domain utilization, each also presents unique challenges due to its characteristics.

2.2.1   Social and Economic Data

Societal factors intervene between nature (and the natural processes) and the built environment to redistribute the risk prior to an event, and to amplify or attenuate the losses after an event (Cutter 2010). Economic development, social structure, culture, ethics, legal system are important background knowledge to understand the interaction, to reduce vulnerability and build resilience to hazards. In assessing risks and vulnerability, the place-based (positioned) disaster loss data, as well as population, infrastructure, land use, building codes, socio-economic statistics, and disaster insurance data all contribute to the validity and accuracy of the assessment. In most cases, systematic information, such as the societal losses triggered by geo-hazards, is available (Cutter 2010).

The degree of open access to data varies among data owners/holders. Open access is also subject to the capability of the available data infrastructure, the legal system, as well as influence of the culture, public participation, and considerations for privacy, security or economic benefits. According to the *open data barometer: 2013 global report*[1], the availability of truly open data remains low, with less than 7% of the datasets surveyed in the Barometer published both in bulk machine-readable

---

[1]http://opendatabarometer.org/3rdEdition/report/

forms, and under open licenses. The developed countries provide more open available data than the developing countries do.

Governments collect a great deal of data, and make a portion of the data available as a public service. In some countries, datasets can be made open only when the public requests them. However, such requests may not be approved by the government. In other cases, processed data, rather than raw data, is open for public access, but such data often does not meet the needs of scientific research. Disaster loss data is typically treated as sensitive information, as it may be seen as a metric for government's performance or is highly relevant to disaster compensations. As a result, only select data may be made public to control public opinion. Data collected by research institutes or universities under nationally funded programs is open to public in developed countries where laws and regulations require public dissemination of research data. Disaster data collected by commercial entities, e.g., insurance companies, are typically not shared outside their own organizations as they are closely related to business strategies and practices.

Unstructured data relevant to disaster risk and loss assessment are scattered on web sites, media outlets and social media. Such data have a high degree of openness but require much more effort to utilize in research. Understanding how such data can be connected together or with other types of data to benefit disaster research and local communities is in itself a research topic.

Under the initiative of the US government in 2011, the Open Data Coalition is taking form in the world, involving more than 40 countries and regions. "Open Data Barometer: Global Report 2013" evaluates the openness of state governments' data from such perspectives as the completion of laws on information disclosure, formulation of policy on open access of data, demand of social organizations and professionals for open data. It points out that the developed countries outperform developing countries in opening the data, as reflected in the fact that America and Europe are in the first echelon, followed by the Asia-Pacific Region, Mid-east, Central Asia and then Africa. Across the world, the implementation of open government data policy remains at the state level, yet to reach the city level.

Challenge: Most of the social and economic data is not open. They tend to be unstructured and not easy to use.

2.2.2   Space based Data

When natural disaster happens, various kinds of observable objects are involved including weather information such as rainfall, snowfall, wind, etc., environment elements including topography, river and so on; exposure such as infrastructure, human population, livestock, etc. Earth observation has the advantage of monitoring all these objects unobtrusively. Based on the different electromagnetic radiation characteristics of disaster relevant objects, the different features can be detected by remote sensors and can be recorded in the form of images.

Earth observation data plays a crucial role of providing information throughout the chain of pre-warning, disaster response and relief, and post-disaster reconstruction. Satellite and other remote sensing instruments may provide accurate, near real-time Earth's surface information over the world. At present, the most commonly used Earth observation data acquisition platforms are satellites and aircrafts. Satellites platforms including optical satellites, Radar satellites, Geosynchronous satellites, small satellite constellations, et al. Aerial remote sensing has the advantage of mature technology, large-scale imaging, high-resolution, suitability for large scale terrain mapping and detailed information for small areas. However, there are limitations in flight altitude, endurance, altitude control, all-weather performance ability as well as a wide range of dynamic monitoring. Aerial and satellite remote sensing can play complementary roles in disaster monitoring.

As technologies in space exploration, photoelectric, microwave, and computing have advanced in the past decades, remote sensing technology has entered a new phase where it is possible to provide multi-resolution, multi-band, multi-polarization, multi-temporal earth observation data in time for

disaster mitigation. In order to fully utilize the vast amounts of earth observation data, coherent data infrastructure for storing, transmitting and processing needs to be developed to enable more efficient distribution, computation, and fast access for disaster mitigation.

The Earth observation community recognizes the importance of data sharing. The GEOSS 10-year Implementation Plan, endorsed by all GEO Members, states: "The societal benefits of Earth observations cannot be achieved without data sharing." The Earth observation community is at the forefront of open and shared data. At present, the kilometer level low-resolution data are almost fully open and available; the 10-30 meter level resolution data are becoming more open; the 1-meter high-resolution data are available through the commercial way. Including higher resolution commercial satellite data, some major space-based observation data for disaster application can be acquired through international organizations such as Disaster Charter, et al and made available free of cost.

However, in order to retrieve timely information from various satellite data, it is in urgent need of remote sensing data with the open and standardized format such as GSFC, ISO19000 series and so on, as well as the data processing tools.

Challenges: Higher resolution data is often necessary for disaster research and mitigation, but they are not widely available. Using such data, especially real time data, requires significant resources and expertise to process and analyze.

2.2.3   Hydrological Data

Hydrological data are an important part of national basic information resources. Hydrological data obtained through long-term and continuous observations are vital to disaster prevention and reduction, as well as management of water resources. Hydro-graphic stations mainly measure water level, discharge, sediment, precipitation, evaporation and water quality. Hydrological data are the basis for evaluation of water resources, hydrological calculation and forecast, water environment evaluation and climate change. Hydrological affairs in many countries are managed in different layers from the central government to local government.

Most of the countries in the world began continuous hydrological observation only in recent years except some which started decades ago, therefore long-term hydrological data series are very rare. Comparatively speaking, developed countries, such as the United States, Germany and Japan, have wide coverage of hydrologic station networks with high density distribution and a high degree of automation, while such network of hydrologic stations is new and sparse in African countries.

In some countries, hydrological data are widely shared. In the United States, USGS provides both real-time and historical hydrological data online and through web services (http://www.usgs.gov), including both time series stream flow and geospatial vector data. China has also built a website for sharing hydrology and water resources data (http://www.hydrodatanet.gov.cn) that provides various services, including online data query and download, off-line data product and yearbook printing.. The shared data include basic information of hydrology and water resource, real-time discharge, water level and rainfall information, as well as compilation and analysis results.

Under the framework of cooperation among countries and regions, sharing and exchange of hydrological data over a basin or large area is critically important for improving the ability of countries along the shoreline to prevent and mitigate disasters, and ultimately to reduce human injuries, and loss of lives and properties. However, due to the lack of a unified policy, standard and technical platforms for data sharing, cross-region and cross-sector hydrological data sharing has not been truly realized at present.

Current efforts by various countries have provided examples for an international mechanism for exchange and sharing of hydrological data. More than 140 countries have river agreements among them, most of which include detailed provisions on international river monitoring, hydrological data collection and sharing. The International Commission for the Protection of the Rhine River (ICPR),

established by several European countries, has technical and coordination working groups and more than 20 flood warning centers. A computer network connects these centers, hydrological stations and meteorological stations for sharing warning and other information. The 1995 cooperation agreement for Sustainable Development of Mekong River Basin requires that countries along the Mekong River regularly exchange the necessary data and information according to the Procedures for Data Exchange and Sharing among member countries.

Challenges: Hydrological data varies significantly in formats and data representation, hence making it difficult to use across regions and scientific domains.

2.2.4 Meteorological Data

Meteorological data refers to temperature, air pressure, moisture, wind, radiation, etc., that is measured in the environment with instruments at meteorological stations, radars and other methods.. Typically, the collection of meteorological data is conducted by government agencies or departments. This practice results in a well-developed (often with sustained funding) global network for collection, dissemination, exchange, storage and sharing of meteorological data, which also explains the high degree of availability and accessibility of the meteorological data in many countries.

A sound coordination mechanism underlies the open access of meteorological data around the world. The sustained efforts have promoted the open access of meteorological information across the world.

The World Meteorological Organization (WMO) promotes cooperation in the establishment of networks for making meteorological, climatological, hydrological and geophysical observations, as well as the exchange, processing and standardization of related data, and assists technology transfer, training and research. WMO facilitates the free and unrestricted exchange of data and information, products and services in real- or near-real time on matters relating to safety and security of society, economic welfare and the protection of the environment. WMO RESOLUTION 40 (Cg-XII) adopts the following policy: "WMO commits itself to broadening and enhancing the free and unrestricted exchange of meteorological and related data and products". Global Climate Observing System(GCOS) is a joint undertaking of the World Meteorological Organization (WMO), the Intergovernmental Oceanographic Commission (IOC) of the United Nations Educational Scientific and Cultural Organization (UNESCO), the United Nations Environment Programme (UNEP) and the International Council for Science (ICSU).   GCOS is to provide comprehensive information on the total climate system, involving a multidisciplinary range of physical, chemical and biological properties, and atmospheric, oceanic, hydrological, cryospheric and terrestrial processes. It includes both in situ and remote sensing components, with its space based components coordinated by the Committee on Earth Observation Satellites (CEOS) and the Coordination Group for Meteorological Satellites (CGMS). GCOS is intended to meet the full range of national and international requirements for climate and climate-related observations. As a system of climate-relevant observing systems, it constitutes, in aggregate, the climate observing component of the Global Earth Observation System of Systems (GEOSS).

At the national level, developed countries and major developing countries have established open data infrastructure with clear policies on data sharing, making available meteorological data ranging from ground-based observation statistics to satellite data. This is very important towards achieving the global target of early warning stated in the Sendai Framework.

For disaster research, the need is to meet the demand for disaster prevention and reduction by further opening meteorological data. The layout and distribution of ground-based meteorological stations remain to be critical issues in meeting the high demand for disaster prevention and reduction. For instance, higher density of common meteorological stations and faster observation frequency will improve data collection, which in turn will help regions, especially in the vast number of developing countries, to deal with emergency disasters and unexpected local disasters. Furthermore, although

globally shared meteorological data is within convenient reach thanks to efforts by World Meteorological Organization, meteorological information systems in different countries are not yet connected. Significant gaps exist in terms of integrated use and coordination of data from multiple sources.

Challenges: Weak or non-existent meteorological data services in many countries hamper data usage in early warning and other risk reduction efforts. For researchers, there is often inadequate data collection in developing countries. The data tends to be very large and have diverse formats, presenting barriers for multidisciplinary researchers and the public.

2.2.5   Seismic Data

Seismic data comes from two areas: seismology and earthquake engineering. The former is a comprehensive discipline about the occurrence patterns of earthquakes through the solid medium of Earth, the transmission patterns of seismological waves and the macroseismic survey of earthquakes. Engineering seismology studies the consequences and impact of earthquakes on urban areas and civil infrastructures using seismological theories and methods with the goal of improving the design and construction of structures to be earthquake resilient.

Opening of seismic data within the country borders has been adopted around the world, helping nations to build their national earthquake early warning and reduction system. Taking China's National Seismic Science Data Sharing Centre as an example, it is the main agency undertaking the seismic science data sharing around the country. Provincial-level data sources feed data into the national data sharing center. The dissemination and sharing of seismic science data are classified into four tiers: the first-tier data is open to the public, the second one is open to domestic and international users, the third is for domestic users, and the fourth is for exclusive use by authorized users. The Global Earthquake Model (GEM) develops and shares earthquake risk information, including datasets, models, methods and guidelines, and facilitates technology transfer and capacity development through broad participation of scientists and countries.

Sharing of seismic data is often a sensitive issue. There have been efforts to exchange seismic data internationally for decades. Although an international site-record sharing and exchanging mechanism is in operation, the number of sites and variables of exchanged data are very limited compared with national level sharing, especially, right after strong events, not all national seismological surveys are accessible by international experts to get the near real time data on event parameters. The primary reason is because of potential use of seismic data could be used to for sensitive purposes, such as monitoring the occurrence of nuclear tests.

Challenges: Earthquake risk information is not fully utilized. Certain countries and regions restrict the dissemination of seismic data because of sensitive nature of their potential usage.

2.2.6   Geographic Data

Geographic data is information related to location and space on land surface. A comprehensive collective of geographical information, referred to as the "fundamental geography data", describes the Earth's measuring points of control, drainage system, residents and facilities, transportation, pipeline system, the boundary and the administrative region, geomorphology, vegetation and soil, cadastral, toponym, etc., which is related to natural and social elements of the location, shape and attribute information. This data is not only critical for research in geo-sciences and other scientific domains, but also, perhaps more importantly, critical to the design, management and risk mitigation of urban infrastructure such as municipal facilities, transportation, etc. (Although most of these data sets are processed data using various technologies mentioned above, they represent a category of information that can stand on its own in this study.)

In most countries, this type of data is collected and managed by Government agencies. Access to such datasets is restricted with different authorization mechanisms in most countries. Certain less

sensitive data is accessible by the public. The definition of freely accessible datasets varies significantly among countries.

The past decade has seen progress in opening up of more fundamental geography datasets. The United States Geological Survey (USGS) provides free access to datasets of Digital Elevation Model (DEM), Digital Orthophoto Quadrangle (DOQ), Digital Line Graph (DLG), Digital Raster Graphic (DRG) data ranging from 1:24,000 to 1:250,000, as well as the LANDSAT 7 satellite remote sensing data, land use data, population densities data and soil surveys data (USGS Data Catalog).

Land Information New Zealand (LINZ, http://www.linz.govt.nz) established a topographic database (NZTopo) with scales ranging from 1: 50,000 to 1: 4,000,000. More than 40 datasets have been released on the web (LINZ Data Service), including LINZ topographic maps, New Zealand offshore islands and Pacific region topographic maps.

Geospatial Information Authority of Japan (GSI, http://www.gsi.go.jp/) has completed the national 1:25,000 topographic maps, and provides various maps, including the topographic maps, land use maps, vegetation and biological maps, and the volcano correlation maps, at different scales.

China has completed its national foundation geographic information system ranging from 1: 1,000,000, 1: 250,000, 1:50,000, and 12.5m grid interval digital elevation models of the key guard areas of the 7 main rivers, and 1m resolution numeral orthogonal projection database, which covers the 9,600,000 square kilometer national territory area. As of October 2014, China's Science and Technology Resource Sharing Network (escience.gov.cn), has shared 1789 datasets, ranging from geoscience, health, agriculture, forestry, weather, hydrology, to materials science data.

Challenges: The diversity of data format, the lack of standard format and necessary expertise in geospatial analysis and processing are barriers to users. Restricted access is often placed on these datasets.

2.2.7 Health data

The Sendai framework has a strong focus on improving health outcomes for people at risk of emergencies. It aims to achieve 'substantial reduction of disaster risk and losses in human lives, livelihoods and health, and in the economic, physical, social, cultural and environmental assets of persons, businesses, communities and countries' over the next 15 years. Health data can potentially be a useful source recognizing the Sendai framework Global Targets on mortality and morbidity. This type of data can provide important information about disease trends and risk factors, outcomes of treatment or public health interventions, functional abilities, patterns of care, and healthcare costs and use.

Given the wealth of health data in existence, the Global Burden of Diseases, Injuries and Risk Factors study has brought together the most recent epidemiological data according to year, age, and sex from 195 countries and territories, and is an excellent example of using big data to systematically and scientifically quantify the comparative magnitude of health loss due to diseases, injuries, and risk factors. As well as providing a worldwide update on disease burden, the study offers an independent analysis for monitoring progress towards the health-related Sustainable Development Goals which may also be used for reporting progress against the Sendai framework.

As approximately 70 to 90% of the causes of health disparities relate to environmental factors and the need to protect health from natural and man-made hazards, the field of ecosystem-human health research linking environmental and health datasets has emerged to be of increasing and critical importance. A review of environmental data resources for use with human disease datasets has established the diversity and potential of data sources. Juarez et al. characterized their public health exposome approach as using environmental data based around four broad domains; natural, built, social, and policy environments. Data sources ranged from remotely sensed meteorological data to census data on socio-demographic variables.

In the context of disaster risk reduction, 'data mashups' which explore the complex interactions between environment and health data, can assist in: identifying 'hot spots' rapidly for targeted prevention, interventions, and research; establishing surveillance and following trends in the interactions between different ecosystems and populations over time, for example tracking the global spread of the Zika and Ebola virus epidemics; and providing early warning systems to prevent and anticipate environmental impacts (such as heatwaves and floods) on health and wellbeing.

There are several projects employing big 'data mashup' applications to environment and human health research. For example, the Medical & Environmental Data Mash-up Infrastructure project hosts and links weather, environmental and human health data from different UK partner institutions in order to provide evidence-based decision support tools on the impacts of climate on health. In the United States, the Centers for Disease Control Wide-ranging Online Data for Epidemiologic Research database has integrated national spatial datasets on air pollution, heat or sunlight exposure obtained from remote-sensing with data related to cognitive decline and other health outcomes. The European Commission is also funding an 'Exposomics' project which aims to study the health effects of exposure to air pollution and water contamination in order to estimate the risk of disease in several population-based studies in Europe.

However, access to health data is not without its challenges. In particular, there may be data protection and confidentiality issues, which can add to the complexity of accessing data but may also contribute to the expense. In the UK, the National Health Service recognizes in its 'Five Year Forward View' that many of the IT systems are not interoperable, and that big data standards will need to be nationally specified in order to facilitate and maximize the use of available data.

With respect to environment and health 'data mashups', the myriad of disciplines and organizations involved can be complex to coordinate and manage. Such work also requires the identification of personnel skilled in the analysis and interpretation of diverse types of big data and able to collaborate effectively across a range of sectors. Another concern is the use of data by other agencies with potentially positive or negative, unintended consequences; the "Missing Map Project" uses crowd-sourced volunteers to digitize maps to aid in disaster response in developing nations, but this information could also be exploited by terrorists, corporations, and governments.

Challenges: Health disaster loss data is potentially a very rich source of information but the use of population health data raises important issues around data governance, ownership and privacy. The involvement of multiple disciplines and communities in disaster-health research can also be complex to coordinate, particularly as they may have different approaches to data synthesis, analysis and interpretation.

2.2.8 Disaster Loss Data

The disaster loss data landscape is complex, however, information on loss data is rapidly growing. Data on disaster losses refer to statistics of various costs related to disasters such as casualties, damaged buildings, GDP and other economic losses. When human, monetary or environmental losses occur as a result of a disaster, extensive loss data is often collected and stored by different organizations, but the thoroughness and accuracy of the data vary from country to country and even among local entities. While many methods exist, there is no standard which introduces a high level of variation into the resulting assessment results[2]. This results in gaps in the data, inconsistent overlaps and biases that ultimately affect the quality of research conducted and policies made based on the data. In 2015,three UN Landmark Agreements were adopted, i.e., the Sendai framework for Disaster Risk Reduction 2015-2030, the Sustainable Development Goals (SDGs), and the Paris Agreement (FCCC). All of these must be "*able to monitor and review implementation*" calling for "*a data revolution, rigorous accountability mechanisms and renewed global*

---

[2] De Groove T, Corbane C, Ehrlich D, Poljansek K (2014) Current status and best practices for disaster loss data recording in EU member states: a comprehensive overview of current practice in the EU member states. Scientific and technical research reports, report EUR 26879

partnerships"[3].   These frameworks all have the same goal – reduce damage and losses; understand vulnerability and the value of risk reduction measures and "build back better" – through sharing of loss data. The systematic follow-up after accidents, disturbances and disasters and the collection of disaster loss data enables learning and provides information that can be used to assess the costs and benefits of disaster risk management, and demonstrate the needs for public and private investment and for risk-sharing and social protection mechanisms.

Disaster loss data is key for measuring the implementation of the Sendai Framework. Paragraph 18 of the Framework states: "*To support the assessment of global progress in achieving the outcome and goal of the present Framework, seven global targets have been agreed. These targets will be measured at the global level and will be complemented by work to develop appropriate indicators. National targets and indicators will contribute to the achievement of the outcome and goal of the present Framework*". On 2 February 2017, the UN General Assembly adopted the resolution A/71/644 which defines a set of 38 indicators to monitor the following seven global targets of the Sendai Framework[4]:

A. *Substantially reduce global disaster mortality by 2030, aiming to lower the average per 100,000 global mortality rate in the decade 2020–2030 compared to the period 2005–2015;*

B. *Substantially reduce the number of affected people globally by 2030, aiming to lower the average global figure per 100,000 in the decade 2020–2030 compared to the period 2005–2015;*

C. *Reduce direct disaster economic loss in relation to global gross domestic product (GDP) by 2030;*

D. *Substantially reduce disaster damage to critical infrastructure and disruption of basic services, among them health and educational facilities, including through developing their resilience by 2030;*

E. *Substantially increase the number of countries with national and local disaster risk reduction strategies by 2020;*

F. *Substantially enhance international cooperation to developing countries through adequate and sustainable support to complement their national actions for implementation of the present Framework by 2030;*

G. *Substantially increase the availability of and access to multi-hazard early warning systems and disaster risk information and assessments to people by 2030. (UN 2015)*

Systematic disaster data collection and analysis ensures informed policy decisions and contributes to building the case for investments in disaster risk reduction. Therefore, an element in national and local level strategies is to establish and maintain a national disaster loss database. By building on UNDRR efforts in sharing methodology and supporting national development of data on disaster losses, currently Europe has developed the understanding of disaster data losses and has worked to align efforts (e.g. Joint Research Centre). The guidelines developed by the European Commission following consultations with the wider European region provide advice to countries on improving the coherence and completeness of the national disaster damage and loss data recording process, necessary for supporting evidence-based disaster risk management policies and actions. Other countries could follow such a systematic approach.

Furthermore, many of the proposed quantitative targets of the Sendai Framework will require the development of baselines and additional research: mortality rate (2005-2015), number of affected people (2005-2015), status of direct disaster economic loss to gross domestic product (2015). The development of disaster loss databases and case studies will be essential in order to review the

---

[3] United Nations (2015). The Road to Dignity by 2030: Ending Poverty, Transforming All Lives and Protecting the Planet
   Synthesis Report of the Secretary-General on the Post-2015 Agenda
[4] United Nations General Assembly, Resolution 68/261, "Fundamental Principles of Official Statistics", Mar. 3, 2014.
   Available from https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf [Accessed 22 February 2017]

framework implementation and to compare the achievements and progress made since 2015. For instance, research support required for loss expressed as monetary costs in Sendai Framework is not enough, because target D (critical infrastructure) indicators, in particular, require an assessment as description of physical damage and functional disruption (outages). At present the physical description of damage does not fully correspond to the costs expressed in monetary terms due to costs of amelioration (e.g., in order to safeguard against near or medium term failures) and/or costs that may be a translation of the physical damage but are more often the actual compensation received (e.g., from insurance or the State). The inconsistencies between physical description and costs must be overcome in the future by perhaps producing a double value: estimated and granted. Solving this problem is fundamental in improving risk modeling capacity.

There is currently no global standard for disaster loss data for measuring the progress toward global targets. The trend around the world, however, is toward open access and sharing of disaster loss data[5]. A number of initiatives are currently developing standards under the auspices of the EC Joint Research Centre (JRC) and Integrated Research on Disaster Reduction (IRDR) Data Group. An open-ended intergovernmental expert working group (OIEWG) comprised of stakeholders from Member States, NGOs, private sector, science and technology, etc. was formed to work on defining a set of indicators. Other groups, including EU Loss Data Guidance, CODATA Task Group LODGD, INSPIRE Directives, the Working Group on Disaster Data (WGDD) and many flagship projects (i.e. PPRD EAST, EU Floods Directive), are also working toward the same goal.

In the absence of an agreed-upon international standard, a set of minimum standards for disaster loss and damage data could be adopted that would contribute to the data quality and global comparability required to be able to measure progress against the Global Targets. An urgent priority is constructing a unified metadata repository and realizing open access to the metadata information. This step will significantly help improve disaster emergency management, response, and disaster scientific research on long term issues. Monitoring of disaster impacts is important but it alone is not sufficient. Recorded losses should be complemented by hazard simulations and model-based losses, improved exposure data, and better understanding of multiple vulnerabilities. Transparency of disaster loss data is a key and should be improved through broader cooperation, including the engagement of national statistical offices (NSOs) and national meteorological and hydrological services (NMHS) on data standardization, quality assurance, and accessibility.

Challenges: Disaster loss data is highly diverse in formats and access mechanism, lacking a global standard and unified metadata schema. Linking them with other data sources, including simulation and model losses, and exposure data, will contribute toward a comprehensive framework for assessment and enabling research on longer term issues. This type of data almost always has sensitive information about the region and country, and open access can be challenging in some countries.

2.3 Status of Open Data in Developing Countries

While almost all countries around the world suffer from natural hazards (earthquakes, volcanoes, floods, droughts, storm surges, etc.), the impact of such hazards on the developed and developing countries are quite different: heavier human losses in the developing countries while greater economic losses in the developed countries. The low-quality infrastructure, lack of effective mechanisms and polices to manage risks and a lower capacity for societal mobilization are often attributed to the greater loss of human lives and properties in the developing countries. International communities of disaster reduction research and practice recognize that the scientific use of disaster data can improve the timeliness, accuracy and effectiveness of disaster mitigation decision-making, and thus greatly reduce disaster losses.

---

[5] Fakhruddin et al. Disaster loss data policy Brief 2017. Global Platform on Disaster Risk Reduction. Cancun, Mexico.

In developing countries, the effect of technology advancement and joint international action may not be fully reflected. Developing countries should draw on the technological progress of mankind to effectively construct and increase their local disaster reduction capacities. Disaster data sharing capacity builds on a nation's economic development and scientific strength. As developing countries have no independent remote sensing satellites and cannot timely and effectively respond to disasters, equal access to data would not be guaranteed. UN ECA holds that timely acquisition of remote sensing data is a powerful instrument to promote regional sustainable development. Although remote sensing technology would provide the same benefits to all countries (i.e., quality and frequency of data), the high cost of the technology remains a barrier for developing countries. The international community may immediately start to facilitate transparent sharing and transferring advanced technology from developed countries to developing countries through "data democracy", as well as possible resources and funding to assist with data analysis.

Some developing countries are making progress toward this goal. As an example, China, one of the largest developing countries, has employed the following three mechanisms to improve the capability of opening and sharing disaster data with the international community:

National coordination mechanism to fully use the internal data resources

In 2008, Chinese government launched a project of Disaster Emergency Data Reservoir (DEDR). DEDR targets studies of emergency collaborative planning and scheduling technology, emergency data sharing technology and emergency cooperation mechanisms. In 2011, it was put into operation with a standardized and stable connection with China's metrology, ocean and land satellite data centers, as well as aerial remote sensing data centers and commercial UAV operators. As part of a national mitigation data support system, DEDR receives international support through cooperation with the International Disaster Charter, UNSPIDER and IRDR.

Regional cooperation on disaster data sharing

In cooperation with China and India, the United Nations Economic and Social Commission for Asia and the Pacific (UN-ESCAP) is promoting a Regional Cooperative Mechanism for Drought Monitoring and Early Warning in Asia and the Pacific (The Drought Mechanism), for development and operational provision of Earth Observation-based products and services to relevant countries in the region. Under the Mechanism, a pilot project for Mongolia and Pakistan is underway, and a project involving other countries is under consideration. It can help the less-capable countries to share the benefit of regional progress.

International assistance mechanism data for disaster

1On May 12, 2008, a Richter 8.0 Earthquake struck Wenchuan and its surrounding areas in Sichuan, China. The earthquake and subsequent disasters, like geological disasters, caused heavy casualties and huge property loss. China used 15 satellites for emergency observation, including Fengyun meteorological satellite, CBERS, Beijing-1 small satellite, and airborne instruments. Even so, they could not meet the extraordinary demand of coverage and timeliness required for disaster analysis, and data processing, and urgently needed assistance from the international community. Through CEOS, International Disaster Character, UNSPIDER and other coordination mechanisms, China established point-to-point connection with most of the world's leading earth observation organizations. International assistance to remote sensing data continued to be provided to the Chinese side from NASA, USGS, JAXA, ESA etc. Other international agencies also provided data assistance through other channels indirectly. On-Site Operations Coordination Centre (OSOCC) is a rapid response tool that provides a platform for the coordination of international relief activities with local

authorities in the immediate aftermath of a sudden-onset emergency or a rapid change in a complex emergency (https://vosocc.unocha.org/). OSOCC serves the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) to carry out its mandate of coordination and information management in emergency response, particularly at the field level.

This model of three-legged coordination and data sharing practice could be applied to other developing countries.

03

Gaps and Challenges in Linking Open Disaster Data

To enable broader use of open science data in the research of hazards and disaster prevention and reduction, it is critical to expand open access and data sharing from current practices and provide better linkage of data, especially datasets from different organizations and disciplines. Many barriers still exist today for majority of researchers, especially those in the developing countries, to find relevant datasets and access them easily. In this section, we discuss the gaps and challenges in three major areas: technological, policy and legal aspects, and governance and cultural.

3.1   Technology Gaps and Challenges

One of the priorities for action in the Sendai Framework is understanding disaster risk. To achieve this goal, it is vital to promote the collection, analysis, management and use of relevant data and practical information, and ensure broad dissemination of such data to meet the needs of different categories of users. At the same time, it is also crucial to promote real time access to reliable data and simulation data for expected impact, make use of space and in situ information, including geographic information systems (GIS), and use information and communications technology innovations to enhance measurement tools and the collection, analysis and dissemination of data.

Information technology is the backbone of the existing and certainly, future generation of disaster data infrastructure that will enable broad data sharing and interoperation locally, regionally and internationally. The technical elements comprising the data infrastructure include *data management, data discovery, data interoperability* and *data services*. These must work seamlessly together to manage and serve the diverse, distributed, multi-disciplinary disaster related data and information to various stakeholders (researchers, decision makers, and the public).

Today's data infrastructure that hosts and serves data to disaster studies has been developed by various groups, projects, institutions and agencies. Efforts have been directed at establishing coordination among different data repositories within a country and across country boundaries to share data. However, significant gaps remain, preventing the effective use and broader applications of these important data resources. As data acquisition techniques continue to improve, newer, higher resolution and more varieties of data will be generated, pushing the limit of today's data infrastructure.

The first major challenge is the management of the large amounts of diverse, heterogeneous datasets that are needed by the disaster research community and the public. This requires both hardware and software, such as storage systems (including storage media, interconnect fabrics and data structuring) and data management software, as well as people with expertise in running the facilities. In addition, analysis, processing and visualization capabilities (e.g., high-performance computing, big-data computing) are necessary to support of the use of the data. The challenge for various organizations to operate a large data facility lies not only in the initial investment and setup, but also the on-going maintenance and technology update necessary to support a reliable and on-demand data service. Although not unique to the disaster research community, this gap is exacerbated by the scale, time-criticalness and the broad involvement of multiple disciplines and sectors of the society in the domain. As more and more unstructured data, such as social media communications, is being considered and used in disaster studies and by local communities (public participation and crowd sourcing), data management systems need to adapt to support such data and provide corresponding data services on the unstructured data.

The second major challenge is that disaster data is difficult to discover and access across multiple repositories and disciplines. As pointed out in earlier sections, disaster related information resides in various geographically and organizationally distributed repositories, using different data formats, access protocols and polices, managed by institutions with very different governance models. The lack of interoperability standards makes it hard for individual researchers to access existing data

resources. Further compounding the issue is the multidisciplinary nature of disaster research. For example, when studying an earthquake, a researcher may also need to find data related to landslide, heavy rain, flood, or volcano eruption. It would be very difficult, if not impossible, to find all or most of the datasets relevant to a study topic – finding one dataset does not lead to other information that are directly and indirectly related (e.g., geographically and/or temporally). Even when one identifies a data repository that likely contains the datasets needed, the user typically has to spend significant amount of time to extract the right dataset (e.g., entering search criteria acceptable to the interface, which is non-trivial to researchers not familiar with that particular data resource and interface), to be able to obtain the data (e.g., having the network bandwidth and local storage to download the data). This segues naturally into the next major challenge related to the utilization of data.

Disaster data is often difficult to use across applications and disciplines. When researchers are able to access and obtain relevant data, they may not have adequate description of the data (lack of standards on data formats), and may not understand the data representation due to disciplinary background. The variety of data formats presents a significant challenge. As described in Section 2, disaster research demands many different types of data, such as data from remote sensing observation, meteorology, hydrology, geography, socio-economic statistics, and social media information. Some fields do not have established standard formats, making such data more difficult to discover and use because of the lack of broadly recognized vocabulary. Progress in the domains of taxonomy, ontology, data structuring and mapping between data structures, etc., is expected to help address some of the challenges.

The fourth major challenge is the poor interoperability of 'cross-disciplinary' datasets for disaster research. This issue stems largely from the multidisciplinary nature of disaster data. Data from both natural sciences and social sciences are needed for disaster research. For instance, in forecasting hazardous events due to weather, observational data such as satellite data, meteorological and hydrologic data will be considered; in assessing vulnerability, risks, and estimating potential damages (to aid local government in decision-making), various statistics such as urban infrastructure, population, and other socio-economic data will need to be considered together with hazard related data. These two types of data are often collected using very different approaches, governed by different policies, and managed by different types of organizations (research/technical groups vs. government/bureaucratic institutions). An even bigger issue is related to different spatial scales of the datasets – the socio-economic data are often at country or province level while observational data can be as high resolution as a meter. Downscaling of the coarse-scale data often uses unrealistic assumptions (e.g., distributions of infrastructures, economic activities and population at various times of the day), leading to inaccurate assessments and estimates.

In addition, increasingly large amounts of unstructured data, such as news reports, social media chatters, pictures, that certainly contains relevant information and may also contribute to vulnerability and potential loss analyses. Social science researchers are also using "storylines" to describe disaster events based on their observation and analysis "on the ground" where these events occur. Other researchers conduct quantitative analysis and collect data through interviews and questionnaires. The quality of such data may be debated as the method of collection and interpretation appears to be more subjective than objective, compared with observational data from physical sciences. The challenges of effectively utilizing unstructured data remain research topics today, including methods to collect, index, manage the data, and make them discoverable, accessible and usable by the broad research community.

Finally, data services will be a critical part of disaster research. The word "service" has at least several connotations such as reliability, persistence, and stable interfaces. Disaster data service

providers will need to address these aspects in addition to storing, managing and supporting access to the data itself. Disaster studies are highly data-driven and data-dependent, and the user community requires multiple interfaces for reaching at the data, e.g., database query, web service, and graphical user interface. Federation of data service providers is a major challenge due to lack of standards. Other areas of considerations include data validation, citation, and community feedback.

3.2 Policy and legal Gaps

While data and information sharing is understood to be critical to scientific progress, academic scientists do not necessarily share data or information with their colleagues (Thursby, Thursby, Haeussler and Jiang 2009). Often, incentives exist to encourage scientists not to disclose research. Their reluctance to share data and materials is increasingly considered as a major problem (Cohen and Walsh, 2007). Scientists' willingness to share data is also highly specific to the context in which they conduct research (Haeussler, Jiang, Thursby and Thursby 2014). Therefore, it is important to understand the benefits and costs of sharing data to scientists and consider carefully how policy can influence decisions on sharing and providing access to data.

Legal, political and institutional measures are part of the equation in achieving reduced risks and increased resilience at local, regional and global levels (Sendai Framework, II.17, pg. 12: *To attain the expected outcome, the following goal must be pursued: Prevent new and reduce existing disaster risk through the implementation of integrated and inclusive economic, structural, legal, social, health, cultural, educational, environmental, technological, political and institutional measures that prevent and reduce hazard exposure and vulnerability to disaster, increase preparedness for response and recovery, and thus strengthen resilience.*)

Policies on open access to scientific data would take the form of mandatory rules, infrastructure, or incentives (OECD 2015b). Mandatory rules could be implemented through requirements in research grant agreements or national strategies or institutional policy frameworks. Infrastructure, which could take the form of soft or hard one, would include initiatives undertaken to develop an open science culture. Since the Organisation for Economic Co-operation and Development (OECD) established principles and guidelines on access to public research data in 2007, the member countries have made efforts to adapt legal frameworks and implement policy initiatives to encourage greater openness in science (OECD 2015a). Development of the skills necessary for researchers to share and reuse the research data produced by others and data management guidelines for universities and public research institutes are also important infrastructure. Incentives would be provided through financial support to cover the cost of releasing data sets, proper acknowledgment of the efforts of researchers for data citations, and career advancement mechanisms based on metrics taking into account data-sharing efforts.

Compared with mandatory rules or infrastructure, however, incentives mechanisms for researchers involved in open data activities have not been widely introduced. Evaluations of universities and researchers are still mostly based on research bibliometric indicators, with little value attributed to the sharing of pre-publication inputs and post-publication outcomes including data. Also, data cleaning and curation,through the development of metadata, which requires a substantial amount of resources, is not acknowledged adequately in evaluation mechanisms or grant allocation procedures. This issue could be addressed to a certain extent by extending citation mechanisms to data sets. To scientists it is crucial that their work is recognized in their communities, and hence the reward to them would be scientific acclaim or a prize, such as the Nobel Prize. Scientists could consider reward in the form of financial returns, such as income from commercial application of the scientist's

solution. Hence a significant challenge in policy making is to incorporate these incentives to scientists into practical measures and instruments for encouraging sharing and providing access to data.

Tackling grand challenges such as disaster risk reduction requires close cooperation and collaboration on access to and sharing of data beyond sectoral or geographical boundaries. On the other hand, when research activities are conducted in partnership with the industrial sector, commercial interests would not be ignored, and consequently the mode of sharing research results could be different from the case in which only public actors are involved. Also, certain classes of data, such as medical records and national security, need to be treated with great care in their access and sharing, due to potential concern about sensitivity or confidentiality.

Open data access needs protection and guarantee from laws and policies, including information security and privacy. As stated in its 2013 Global Report by Open Data Barometer, absence of strong Right to Information Laws may prevent citizens from using open data to hold government accountable and weak or absent data protection laws may undermine citizen confidence in open government data initiatives. In the developed countries, laws of data access and data protection have been well established (e.g., in the U.S., the Freedom of Information Act provides that "that any person has a right, enforceable in court, to obtain access to federal agency records, except to the extent that such records (or portions of them) are protected from public disclosure by one of nine exemptions or by one of three special law enforcement record exclusions"), while they are lacking in most developing countries today. The lack of protection for public's right to information impedes work in the area of disaster research, for example, as it is difficult for non-governmental organizations, research institutions, and the public to obtain certain high resolution and important disaster information. In many countries, the increase in the demand for data sharing exceeds the pace at which the laws governing data access and sharing are established. Citizen's right to information (similar to the United States Freedom of Information Act) will ensure open access to government data for effective transparency and accountability (Open Data Barometer 2013). Policies and technical implementations are lacking in addressing the dynamic nature of data access. For instance, data that is normally restricted will need to be made available openly to aid in emergency situations.

While we advocate broader sharing of multidisciplinary data to support disaster research and practices, we also recognize that new issues may arise as a result. For example, a dataset that was acquired originally for a specific purpose or by one type of application could be used in a different domain and disseminated to a completely different audience. Examples of such data may include medical records and statistics, and social media communications. Data may be used in applications beyond the original intended use-- such expanded scope of use may have implications related to ownership, privacy and ethics. Laws and regulations to protect intellectual property and scope of use are needed to ensure appropriate use of data.

3.3   Governance and Cultural Gaps

Data may be considered as property of a particular organization in countries where management of data is distributed among different government departments, universities and research institutes without an authoritative coordination department (Fan 2006). A similar situation exists for disaster loss data. Disaster loss data at the provincial level is available in the statistical yearbooks, while the high-resolution data is not fully open. When high-resolution data is needed by other departments, administrative coordination is necessary in order to access the data. This process is time consuming and labor intensive. Many developing countries lack a basic foundational framework for managing governmental digital data, and need to build government data collection and management capacity (Open Data Barometer 2013). There is no systematic open-access inventory or accounting for

individual nations or aggregated to the global scale of hazard events and losses by location or by hazard agent (Cutter 2010).

Cultural and societal ethical standards also have a significant influence on the degree of data sharing and openness. Countries such as those in Europe have a high degree of open dissemination of governmental data as it is considered an integral part of civil rights, and having great benefit to governance, democracy and social development (Liu 2007). In some cultures, governments are reluctant to publish negative information, such as losses caused by disasters, for fear of the perception of incompetency on the government's part.

The level of awareness toward citizens' right to information varies among countries of different cultural backgrounds. In some parts of the world, people recognize that governmental data, especially those related to disasters, should be openly accessible by anyone and they demand data transparency from their governments. As a result, they often do have access to such data. In other countries, citizens are less active in public participation of disaster management. The lower levels of awareness, transparency, and public participation are also barriers to disaster data openness. Efforts on engaging the public through data sharing will empower the public to use data as well as contribute data via crowd sourcing, and ultimately impact risk reduction and resilience in local communities.

04

Scientific Issues behind Data Interconnectivity

There is a wide range of active research topics related to the use of scientific data. In this section, several scientific issues related to open disaster data have been described like: 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.

4.1   Data Dependency

Hazards of different types lead to disasters, such as flood, droughts, heavy rain, snowstorms, earthquakes, typhoons, landslides, wildfires and infestations, etc., causing great loss of life, damage and hardship. Research on each type of disasters involves (1) pre-warning, prevention and simulation prior to the disaster event; (2) the cause, process, and occurrence mechanism while the disaster is unfolding; (3) emergency responses, assessment of loss, post-disaster recovery and reconstruction. Research on long term strategies of vulnerability assessment and mitigation, as well as communication with the public, aims at making the society more resilient to hazards and reducing or eliminating loss of life and property. Disaster research is multi-disciplinary by nature and highly dependent on scientific data. Multiple processes and hazards may contribute to a disaster event, e.g., landslides and flooding may occur at the time of an earthquake. It would not be possible to study such phenomena thoroughly without access to relevant data sets.

We surveyed a number of peer-reviewed academic journal papers about earthquake and flood disaster events and examined the data used by researchers in studying different types of disasters (Zhang et al 2015). We found that the researchers of earthquake disasters relied on primarily eight types of data, namely, geological data, geophysical data, ground observation data, basic geographical data, earth observation data, space physics data, clinical medicine and socio-economic data. The researchers of flood studies relied on eight types of data including earth observation data, hydrology data, meteorology data, basic geographical data, ground observation data, biological genetics data, geophysical data and socio-economic data. Collection of disaster relevant data often depends on the condition of the natural environment, available technology, and economic strength of the affected area. Hence, data dependency may differ across regions, even for the same disaster event, due to temporal and regional characteristics. To consider the comprehensive impact from multiple systems, researchers rely on multi-source data. For example, Dankers and Feyen utilized data on climate, geography, plantation and land cover intheir simulation of flood disasters (Dankers & Feyen 2009; Dankers et al 2014); Ulbrich utilized socio-economic data, hydrologic data and observation station data to study the formation of precipitation and flood (Ulbrich et al 2013a, 2013b).

Researchers studying disaster events often face a significant technical challenge--finding and accessing the relevant data for the events being studied among a wide variety of data sources. The first challenge is the lack of consistent data representation. Disaster data comes from a wide range of disciplines. The methods of data collection vary significantly, from field monitoring, remote sensing observation, macro statistics, on-site investigation, to models and simulations. Data heterogeneity not only lies in the differences in measurement unit, recording method, description of data or metadata, and format, but also in the diversity of systems for data storage, discovery and access. Secondly, different disciplines classify data in different ways. The current classification system of disaster data comprises the four-category, five-category and seven-category classifications. The field needs to have a classification standard that is accepted by the disaster research community and implemented in data management systems. Such a standard shall not only consider the scientific classification of disaster data but also take into account the overlap among data from different disciplines, as well as effectiveness in practice in order to promote for wider adoption and usage. Developing an appropriate classification system holds the key to broader utilization of open disaster data and advance research to help people and governments in dealing with disasters.

4.2 Specialists and the Public toward Disaster Data

Disaster events have far-reaching impacts, often affecting lives of tens of thousands of people. As they bear the impact and participate in mitigation efforts, the public are both data collectors and consumers. Prior to industrialization, disaster reduction relied on societal mobilization. Due to the lack of necessary technical means, all members of the society needed to be mobilized and various techniques and skills were employed to observe as well as forecast disasters and conduct relief efforts. With industrial and technological development, especially the more recent progress in computation and information technology, disaster forecast and research to understand the various hazards have become more and more specialized and professional, leading to rather restricted access to disaster data. Data sharing has often been limited to expert communities that specialize in the relevant techniques and usage.

Advancements in information technology, especially the widespread use of the Internet, have made it easier to put data and information in the hands of the masses, creating new opportunities for both research and commercial interests. The open data movement in disaster data management is driven both by the need for increased transparency from the public and by the need for public's participation to improve disaster management.

One recent example of such impact is related to air pollution in Chinese cities. Public awareness of air pollution and its negative impact on health and daily life has significantly increased across China, resulting in hundreds of thousands of users accessing real-time air quality data via the Internet each day. Similar data consumption needs can be found in countries and regions prone to tornados where timely access to tornado related data would help prepare the citizen to avoid or reduce losses.

More and more, citizens are collecting and contributing disaster information in addition to consuming data. The participants in disaster data related activities now extend from scientists to the public. The broadening participation poses a major challenge in terms of both technological and policy aspects.The Tomnod project, sponsored by the Colorado-based satellite company Digital Globe, utilized crowdsourcing to identify objects and places in satellite images. After MH370 went missing, Tomnod swiftly pooled global remote sensing images and distributed them online so that anyone could examine and mark all the information potentially related to MH370. Prior to that, more than 10,000 people from various walks of life had participated in data processing for the website.

Making data openly accessible for the masses brings major challenges and incurs additional costs to address these challenges. While the data specialists have the tools and knowledge in dealing with complex and heterogeneous data, the data infrastructure to support public participation of data collection and data consumption remains a challenge. When the public participates in data collection and data consumption, policy issues related to privacy protection and proper use of data rise to be a top priority. We also recognize that disaster data may reveal different issues to different stakeholder groups. For example, data producers and users may have different considerations due to their different interests. Disaster data comes in all forms, shapes and places. The geographic distribution of the data resources may cause problems related to timeliness and accuracy. The lack of standards for integration of static and dynamic data and the co-existence of massive amounts of heterogeneous data present significant challenges in reconciling and utilizing the data. The full extent of issues, both scientific and technological, involving the public as a key component in disaster data collection and usage should be studied.

4.3 Autonomy of Disaster Data Resources

Today, disaster data resources typically reside within the confines of agencies or institutions, often with technology implementations and polices specific to the hosting institutions. In the long

term, it is necessary to grant autonomy to data resources. Autonomy of data recognizes the independence and transparency of data and promotes self-governance of data management and services. Progress in this direction will depend on technology advancement and, more importantly, on policy shifts towards open access.

From the technology perspective, the needs from specific applications and government mandates often drive the initial development of a data infrastructure. The resulting infrastructure is often constructed to meet the immediate demands by sacrificing portability and reusability of data resources. Compounding this deficiency is the policy requirement, e.g., restriction on the data fields to be exposed to open access, resulting in systems that are not broadly usable. This creates significant barriers to the utilization of the data resources, including:

- non-standard representation of data and meta data
- poor interoperability of data resources in different systems and applications
- inconsistent access methods, difficult for broad dissemination of data
- high cost of system development and customization to serve application-specific needs.

While we recognize the benefits of application-driven technology development, the multidisciplinary nature of disaster research can serve as an effective driver to increase usability and interoperability of data. By realizing autonomy of disaster data resources, the focus on data resource will now be shifted to establishing service-oriented data facilities to help the user community effectively use, reuse and share data resources. The neutrality of data service facilities will directly impact the open access and linking of data resources. Unlike the traditional data holdings, autonomous data facilities will be built with open access, interoperability, and policy management in the design at the beginning.

The autonomy of disaster data poses two main challenges: *technical* and *policies*. First, there needs to be an appropriate framework to define policies related to access, rights, credit, etc. Such policies will be integrated in the data infrastructure for presentation, enforcement and update. Second, the data resources must follow standards in order to support heterogeneous systems, including standards for metadata, access, sharing, etc. This construction of data standards need to be based on existing data standards and embraced by the community. To access data via the heterogeneous data sharing system, one can selectively get the meta-data expression and description corresponding to application needs, and thus meet the specific demand of the application system.

Management of disaster data in autonomy is the basis for long-term sustainable opening and sharing of data to the largest extent, thus a deeper understanding of autonomy will promote reforms of culture, policy, interest relations and technological conditions related to disaster data.

05

Cyberinfrastructure for disaster data interconnectivity

Technological barriers have been a major gap in our ability to link scientific data from multiple domains and types of instruments effectively for disaster researchers, as described in Chapter 3. However, the extraordinary innovations in information technology and their rapid adoption by researchers and public alike have reached the point where many of these barriers can now be addressed by a new and enhanced data infrastructure capable of supporting data producers and consumers of diverse sources and types of scientific data. In this section, we discuss a number of key enabling technologies that could contribute to such a cyberinfrastructure that will help realize our vision of an advanced data infrastructure of interconnected cross-domain disaster data for research and knowledge dissemination.

Although a new term a decade ago, *cyberinfrastructure* (*e-Science* in some countries) is widely used to refer to a collective of interoperable information systems, data and software that is fundamental to scientific discovery and collaboration. Analogy to the traditional physical infrastructure of roads, bridges, power grids, and telephone systems, cyberinfrastructure encompasses a set of complementary and interconnected areas including computing systems, data repositories and other information resources, networks, digitally enabled instruments and sensors, connected through interoperable software and tools, and services (Atkins 2003). Applications of specific communities can be built on top of and utilize resources in a cyberinfrastructure. The architecture of a data infrastructure would typically consist of various layers, from networking, computing, digital data, to interoperable services. In this section, we highlight several key areas of cyberinfrastructure relevant to an interconnected disaster data infrastructure.

5.1 Networking and Data Movement

While the desire to allow scientific data to be freely accessible has been around for a long time, the digital revolution in the past two decades that has put network access to almost any content into the hands of millions of people, promises to make the vision of open access to science data a reality. Network connectivity has become ubiquitous thanks to the increased connectivity among institutions, academic campuses, government agencies, and even homes, and to the proliferation of personal computation and communication devices in recent years. In the context of a data-centric cyberinfrastructure, the networking layer provides the foundation to data accessibility.

Computer networks are the most mature in terms of information standardization. They use well-defined formats, or protocols, for exchanging messages, although the underlying implementation may vary from device to device. A Wide-Area Network (WAN) spans regions, countries and even the world, connecting localized networks (e.g., LANs, MANs) and transmitting data over long distances and between different local networks.

Any initiative of open data and data interconnectivity relies on a well-connected network and adequate data transfer rate in order to support movement of data from data providers to data users and sharing of data among sites and users. As data size increases rapidly, data transfer rate and methods of transfer have important implications to success in linking and sharing of data. For example, it takes one day to transfer 1 terabytes of data over a 100 Mbps (megabits per second) network, in comparison to approximately 20 minutes over a 10 Gbps network (assuming a disk fast enough to store the incoming data).

Data transfer refers to the transmission of data from one physical location to another. File sharing is a primary example of transferring large amounts of data across the Internet. While emailing files as attachments or putting them up on web sites for download remain to be popular ways of sharing data, most datasets needed in disaster research are too big and complex to be shared effectively by these means. More recent technology and services, e.g., cloud storage such as Dropbox,

Google Drive, and Amazon S3 to name a few, have rapidly changed the way people share personal files, such as photos, videos, and other files, across computers and mobile devices. These services allow a user to put files in a remote storage server, and still have control over who can create, access and delete the files. Amazon S3 supports up to 5 terabytes in size. The cost and transmission rate remain in researchers' ability to share files of much larger sizes, from hundreds of gigabytes to terabytes. High-performance parallel file transfer methods (such as graft using parallel TCP streams and multi-node transfers to achieve high throughput) have been used by academics for many years. Services created based on these methods, such as Globus Connect (Foster et al 2011), iDrop Desktop (as part of iRODS (Moore et al, 2010)), Phedex used by the Large Hadron Collider (LHC) Experiment (Egeland et al 2010) to name a few, are being used by research institutions and individuals for big data transfers across the Internet. There is a well-established body of work that can be leveraged in building the next generation of disaster data infrastructure where supporting big data from multiple domains is one of the key requirements.

5.2 Advanced Computing

The need for computing power is everywhere, especially in the area of disaster research, from modeling and simulation, to data processing and visualization. The power of a single computer has steadily increased while the cost of the hardware has come down, making computing more affordable than ever. Today, the processing power in an iPhone, for example, rivals that of a supercomputer 30 years ago. A single multi-core server can meet the needs of simulation and data analysis for many researchers. However, significant computational power is critical to meet the needs of simulation and modeling at the regional and global scales, using high and super-high resolution data, and dealing with large amounts of data of different formats and sources.

There are several forms of computation: serial, parallel, and distributed computing, all of which are applicable for various problems in the disaster research domain.

Serial computing refers to software written as a serial stream of instructions executed on a CPU (central processing unit) on a single computer. The instructions are executed one by one in sequence. Software is easier to write, but for large problems, e.g., dealing with many points on a grid or a long simulated time period, the program may need to run a long time.

Parallel computing paradigm breaks up a problem into many smaller tasks and executes them concurrently on multiple CPUs. These CPUs may be on a single computer, as in multi-processor, multi-core machines, or on multiple computers. While executing the subtasks simultaneously may speed up the calculations, the amount of time needed for the subtasks to communicate with each other may affect the speed-up of the total processing time. Applications whose subtasks need to communicate with each other frequently are referred to as being *tightly coupled*. Modern supercomputers, such as Tianhe-2 (China), Titan (USA), and K-computer (Japan)[6], with high-speed (often custom designed) interconnects between the computing nodes are designed to support the tightly coupled parallel applications. Applications whose subtasks rarely or never have to communicate are called *loosely coupled*, or more often, *embarrassingly parallel* applications. These applications can run efficiently on systems connected with slower and less expensive interconnects (such as gigabit Ethernet). They lend themselves well on distributed computing systems (aka high-throughput, grid computing) where the computers may locate in geographically distributed locations.

[6]Top500 supercomputer sites http://www.top500.org

Load balancing, job and workflow management systems such as HTCondor[7], Pegasus[8] are widely used to harvest idle computer cycles, map tasks to distributed resources, and manage input and output data.

Affordability of today's cluster computing systems, along with the de facto standard operating system Linux and job management software on these systems, computing power is much more accessible to researchers in the disaster research domain than ever before.

5.3 Data-intensive computing

Computation in the domain of disaster research is typically data-centric. Whether researchers are modeling natural phenomena, or synthesizing and analyzing data from multiple sources and at multiple scales, they not only need computing power, but also need a system that can get large volumes of data into their application and store output data efficiently. Applications that deal with large volumes of data (TB or more) and spend most of their execution time on data input/output (I/O) and processing are considered data-intensive, in contrast to the computation-intensive applications that spend most of their execution time on calculations.

Analogous to compute-intensive applications, parallelization of data-intensive applications typically involves partitioning the data into multiple segments or regions which can be processed independently using the same executable application program in parallel on an appropriate computing platform, then reassembling the results to produce the completed output data. Such parallelization tends to be straightforward and can normally scale linearly according to the size of the data.

Major challenges for data-intensive computing from the domain of disaster research are managing and processing exponential growth in data volumes, significantly reducing associated data synthesis and analysis cycles to deliver results on-demand, e.g., assisting decision making, in a timely manner.

Innovations in system architecture and hardware are extending the capabilities of supercomputers to better support data-intensive sciences. The emerging new systems aim to better meet the demands of the rapidly growing data-intensive scientific applications, for example, by including massive high-performance storage for data storage, large-scale Flash memory capable of reading and writing files at ~1 terabytes (TB) per second, processor accelerators (such as Intel Phi and NVIDIA GPU) to support data parallelism, as well as software tools and databases to support big data analytics and transfers.

5.4 Cloud Computing

The term *cloud computing* may mean different things to different people. It is as much a computing paradigm as an IT service model. Cloud computing builds on a number of technologies that came before the new term was coined, including distributed systems (remote computers), utility computing (service provisioning), and virtualization (multiple virtual machines running on the same physical hardware). Many IT services are now "in the cloud", such as provisioning of computation time, data storage, and applications (e.g., Amazon EC2, Microsoft Azure, Gmail, Microsoft Office 365, and Apple iCloud apps).

Five characteristics are considered essential in cloud computing by NIST (National Institute of Standards and Technology): On-demand self-service, network accessibility, resource elasticity, resource pooling, and metered service. Traditionally, research computing and business applications have always been very complicated and expensive. The amount and variety of hardware and software

---

[7] HTCondor open source distributed computing software http://research.cs.wisc.edu/htcondor
[8] Pegasus open source scientific workflow management software http://pegasus.isi.edu

required to run them are daunting. A team of IT experts, such as system architects and administrators, is needed to design, install, configure, test, run, secure, fix and update the systems and software, a major barrier to many researchers and organizations of various sizes. In the world of cloud computing, service providers deploy tens of thousands of computer servers in distributed data centres. Users of all sectors can access computing cycles, storage and applications remotely, wherever and whenever needed, over the network, by themselves. Users may add or reduce resources instantly according to the demand of their applications, and only pay for what they consume. Organizations, large or small, no longer have to carry the burden and cost of maintaining and updating their own hardware and software.

Several service models are available to meet various needs, including:

Infrastructure as a Service (IaaS): IaaS clouds offer computers, physical or, more often, virtual machines, and can scale services up or down according to user requirements. Most of them also provide other value-added resources, such as VM disk images of various configurations, file or object storage, load balancers, firewalls, etc. Some research clouds allow researchers to select or create appropriate VM images of complex set-up for specific scientific applications (e.g., coupled climate models) that are time-consuming and need experts to configure.

Platform as a Service (PaaS): A PaaS cloud provides a computing platform, typically consisting of an operating system, programming language execution environment, often a web server and web service container, i.e., an environment for developers to create, host and manage applications without effort in managing the runtime system software, middleware, operating system, and other necessary tools required for development.

Software as a Service (SaaS): SaaS is a software delivery model in which software applications are hosted centrally and accessed by users remotely. The cost of using SaaS is often based on subscriptions, e.g., monthly or annually. Various software technologies exist for users to remotely access SaaS services either as a desktop tool or via a web browser, such as Citrix ICA client, Remote Desktop, ThinLinc, and VNC (virtual network computing) software. More recently, the lightweight virtual container technology, represented by the likes of *Docker*[9] and *Singularity*[10] is becoming widely used by the research community to deploy and share scientific software.

The scientific community has embraced the cloud service models. While the classic HPC system and software environment support a variety of scientific computations, such as simulations, that can benefit from tightly coupled parallelism and the batch processing paradigm, the computational needs in the broader scientific community and vast application domains emphasize a high level of interactivity, availability of resources when needed (on-demand, albeit may run a little slower), and ease of use, all of which are promised by the cloud computing model. Research and academic institutions are adopting the IaaS model in making computing more affordable and accessible to researchers. The PaaS model allows researchers and engineers to focus on developing scientific software, rather than investing time and energy managing system software and tools. The SaaS model can deliver scientific applications and tools directly to the user, allowing researchers to access sophisticated software and computational resources transparently without the burden of installation and maintenance, thus significantly lowering the barrier for the broader user community of researchers, educators, and public as well. This model will, in particular, enable the sharing of software in a more effective and usable manner, overcoming many barriers researchers face when attempting to use software created by others.

---

[9] https://www.docker.com
[10] http://singularity.lbl.gov

The computational needs of the disaster research and mitigation domain encompass a wide range of applications, from modeling, data processing, data synthesis, data analysis, to visualization and decision support. Most of these can be well served by cloud computing service models.

5.5   Service-Oriented Architecture and Data Services

Service-oriented Architecture (SOA) refers to the computer software architecture pattern in which computational functions are provided as interoperable services among software applications, typically from different computers over a network. A service encapsulates a unit of functionality in a service interface for other applications to use. These services may operate on different platforms and frameworks from various vendors and providers. By using a standard communication protocol to describe its function, interface and data presentation, SOA makes it easy to compose complex software applications from a variety of functional building blocks implemented as services. For example, a decision support application may obtain time series data from a data repository and a map from a map server, and call up a geo-computation service to aggregate data based on geographical regions. Web services supports SOA implementations over the Internet. They make functions, or services, accessible over standard Internet protocols across different systems, frameworks and programming languages. Since the early web service standards such as SOAP/WSDL and many proprietary protocols by various commercial vendors, today the scientific community has fully embraced the REST (Representational State Transfer) Web service technology as it has become one of the most important technologies for web applications. Web services built on the REST architectural style for networked applications are called RESTful services. A RESTful service focuses on access to resources across the network and does so through standard ways of exchanging information between a client and the service, i.e., uniform interface for resource representation, discovery, information exchange and linking to other resources.

Web services are being adopted by the science community. Traditionally, researchers focused on experimentation and exploration in their own labs and settings. Research output was primarily in the form of publications of journal papers. Data sharing is now more widespread, especially through publicly available data repositories from government agencies and government funded research projects. However, many barriers still remain, preventing a high degree of utilization of shared data. Top on the list of challenges are the scientific understanding of the data and adapting the data in a usable manner for the study at hand. While existing Web service technologies such as REST can be adopted to help make scientific data more accessible and usable, common protocols of ***data services*** are needed to address the specific challenges associated with scientific data, especially in domains such as disaster research where data from multiple disciplines are common and necessary. Data services aim at making data easier to understand and use by using common protocols to communicate the meaning and format of the data, in addition to access to the actual digital content of data. There are many on-going efforts in building data services, for example, the Data Observations Network for Earth (DataONE)[11] funded by the U.S. National Science Foundation, federates earth science data repositories at distributed member data nodes through three coordinating nodes, and provides services and toolkits to be used by scientists to search and utilize remote datasets. U.S. Geological Survey (USGS) provides data web services for access to its data catalogs. Data services are also available at U.S. NOAA regional climate centers' ACIS system, providing access to metadata and various forms of the raw and processed climate data through public web services. Common data service protocols will be a critical part of a data infrastructure for the disaster research community.

---

[11]DataONE project, https://www.dataone.org

5.6   Data science

Scientists have always conducted analysis on data, which they acquire through instruments, experiments and simulations, by employing mathematical and statistical methods. Today, the rapidly growing data size is changing the way we conduct scientific research, or simply, the way we approach the world around us. While the theoretical and fundamental understanding of natural phenomena remains important, progress in many fields is increasingly being driven by data analytics, visualization, mining and gaining insights from the data (e.g., genomics, biology, physics, cosmology, etc.) (Halevy et al 2009). The term "data science" is often used to focus on the theories and techniques applicable to large volumes of data, ranging from mathematics, statistics, to information science and computer science. Such methods include probability models, machine learning, data mining, relational and non-relationship database, predictive analytics, uncertainty modeling, data visualization, and many others. The focus on handling "big data" is rapidly increasing our ability to understand and gain insights from data at a scale that was impossible before.

The growth of data science will benefit many fields, especially in risk and disaster management. This field is driven by ever increasing volumes of data in various forms, acquired by different instruments, historical and in real-time, structured and unstructured, from diverse sources including social media platforms. More and more, researchers in the disaster risk field use their data and analytical ability to find and interpret rich data sources, combine and synthesize datasets, build models, incorporate uncertainty in data, visualize data to aid in understanding, share data across domains leading to new insights, and communicate their findings and new insights from the data to the science community and general public. The new data science focus and emergence of new tools and software will significantly improve the productivity and efficiency of the disaster research community. The next generation of the data infrastructure for disaster research will need to facilitate and support data science capabilities.

06

**Case Studies and Lessons learned on Linking
Opened Data for Disaster Mitigation
around the World**

In recent years, a number of international initiatives (listed below) have been established to make data available for humanitarian and emergency response from international partners.

The International Charter for Space and Major Disasters, which was established in 1999, proposed the objective of utilizing space-based assets to contribute to the response to natural or technological disasters. The international Charter provided the strategy foundation for disaster data sharing.

In October of 2005, the ICSU 28th General Assembly agreed on the launch of Integrated Research on Disaster Risk Program (IRDR). IRDR aims to address the challenge of natural and human-induced environmental hazards. The DATA group of IRDR is committed to establishing disaster loss data infrastructure for stakeholders.

On December 16, 2006, the United Nations General Assembly agreed to establish the "United Nations Platform for Space-based Information for Disaster Management and Emergency Response (UN-SPIDER)". UN-SPIDER is the first to focus on the need for ensuring access to and use of Space-based Information throughout the disaster management cycle.

During China-US Roundtable Meeting of CODATA held on March 30, 2009, Prof. Li Guoqing proposed the idea of establishment of the Historical Disaster Data Grid (HDDG) to facilitate event-oriented data management and sharing for disaster scientific communities.

After the release of the report USGS Natural Hazards Response in 2012, USGS launched a project of Hazards Data Distribution System (HDDS). The HDDS site provides event-based download access to remotely sensed imagery and other datasets acquired for emergency response.

In November 2012, European Space Agency made its contribution to GEO by releasing GEO Geohazards Supersites and Natural Laboratories (GSNL). It committed itself to data sharing of Geohazards, and formed the idea of building the Supersite (Geohazard Supersite references; Lengert, Popp and Gleyzes 2012; Rowan, Baker and Wier 2013; Amelung, Lengert and Puglisi 2010).

China also published its National Spatial Data Acquisition and Emergency Data Sharing Platform in 2013, which aims to enable timely data coordination and sharing for emergency assistance (mainly for flood and earthquake).

In middle of 2014, CEOS (Committee on Earth Observation Satellite) proposed the idea of Recovery Observatory Infrastructure to further promote disaster data sharing and initiated 3 pilots (Flood, Seismic hazards and Volcano oriented, respectively) to demonstrate how satellite EO data can be used to improve monitoring and response of hazardous events, when accessing data at no-cost.

The On-Site Operation Coordination Centre (OSOCC) originally developed by United Nations' Office for the Coordination of Humanitarian Affairs (OCHA) aimed at improving the international assistance in support of the Governments of affected areas. In the case of emergency, OSOCC makes available a great number of data useful for making decisions about response and humanitarian assistance.

Non-profit and citizen driven projects have also been created to aid relief efforts using open data and technology, such as the Humanitarian OpenStreetMap Team (HOT), making free, up-to-date maps available, providing a critical resource when relief organizations are responding to disasters or political crises.

In the sections below, we describe four examples of these initiatives in more detail to highlight the efforts and showcase the specific actions in making them successful.

In the United States, many different federal and state agencies collect hazard data. For example, the U.S. Geological Survey collects geophysical and hydrological data, and the U.S. National Ocean and Atmospheric Administration focuses on atmospheric and hydro meteorological data. Only some parameters of a hazard event are monitored consistently (e.g., magnitude, date of occurrence, location), while the documentation of others (e.g., deaths, injuries, or economic losses) is incomplete and inconsistent. In the absence of a national inventory of hazard losses, the Hazards and Vulnerability Research Institute at the University of South Carolina developed the Spatial Hazard Events and Losses Database for the United States (SHELDUS®, www.sheldus.org). In the early 2000s, this database was originally supported by grants from the National Science Foundation (Grant No. 99053252 and 0220712) and the University of South Carolina's Office of the Vice President for Research. Periodic support for the database has been provided by the South Carolina Emergency Management Division for South Carolina updates. Given the lack of consistent federal, or private sponsorship in subsequent years, SHELDUS® switched to a user-fee publicly accessible hazard loss database in 2014.

SHELDUS® (http://www.sheldus.org/) is a U.S. county-level hazard data set covering 18 different hazard types and a time span from 1960 to present (Table 1). The data originates from several existing national data sources such as the National Climatic Data Centre's monthly storm data publications and the National Centers for Environmental Information's (formerly National Geophysical Data Centre) Tsunami and Earthquake event databases. Event-related information includes location (state and county), deaths, injuries, property losses, crop losses, and beginning/ending dates.

Originally, SHELDUS® contained only those events that generated more than $50,000 in damage or at least one death. Since the release of SHELDUS® Version 13.1 these thresholds have been removed and SHELDUS® now includes every loss-causing hazard event since 1960. The annual database updates include new data releases (i.e., losses that occurred in the most recent calendar year), data correction, data additions using supplementary data sources, and/or new data download and analytical features.

In addition to providing georeferenced loss information, SHELDUS® resolves historic boundary changes related to geography (e.g., new counties) as well as climatological forecast zones, the latter of which changes numerous times during a calendar year. The database also matches hazard event records with other identifiers (e.g., Presidential Disaster Declaration Numbers (PDDs), the Global Identifier Number (GLIDE) and the Billion Dollar Disasters product produced by NOAA's National Climatic Data Center. The GLIDE number (http://glidenumber.net) is an important feature because it is a globally-accepted identifier linking the data in SHELDUS® to international databases.

Table 1. Overview of loss information in SHELDUS®

SHELDUS® version 13 (1960-2013)	
Number of Records:	831,182
Direct Property Losses (in $2013)	$655.7 Billion
Direct Crop Losses (in $2013)	$148.5 Billion
Fatalities:	31,495

Injuries:	235,739
Costliest Year:	2005 ($120 Billion)
Costliest Hazard Type:	Tropical Cyclones ($260 Billion)
State with Highest Losses:	Florida ($105 Billion)
State with Most Fatallities:	Texas (2,202,)

Functionality

The SHELDUS® website provides online data download functions along with maps, reports, references to SHELDUS® applications, and FAQs. There are four querying options: location, PDDs, GLIDE, or named major disasters. Under the *Location* option, users can download data either in aggregated form (e.g., by location, time, and/or hazard type) or individual, event-specific records (Figure 2). Furthermore, the *Location* option offers additional query options (e.g., time period, hazard type) to create user-defined data sets. *Named major disasters* allows users to query based on more readily identifiable events such as Hurricane Katrina, Superstorm Sandy, or the 1993 Mississippi River Floods without knowing the precise location of the event and its attributed losses. All the counties affected by the particular event are included in the query.

At its finest resolution (i.e., without any data aggregation), SHELDUS® outputs include the beginning and end dates of the event, the hazard type, state, county, injuries, fatalities, property damage, and crop damage. Additionally, SHELDUS® offers the choice to retrieve economic loss data either in current U.S. dollar and/or as inflation-adjusted loss data.
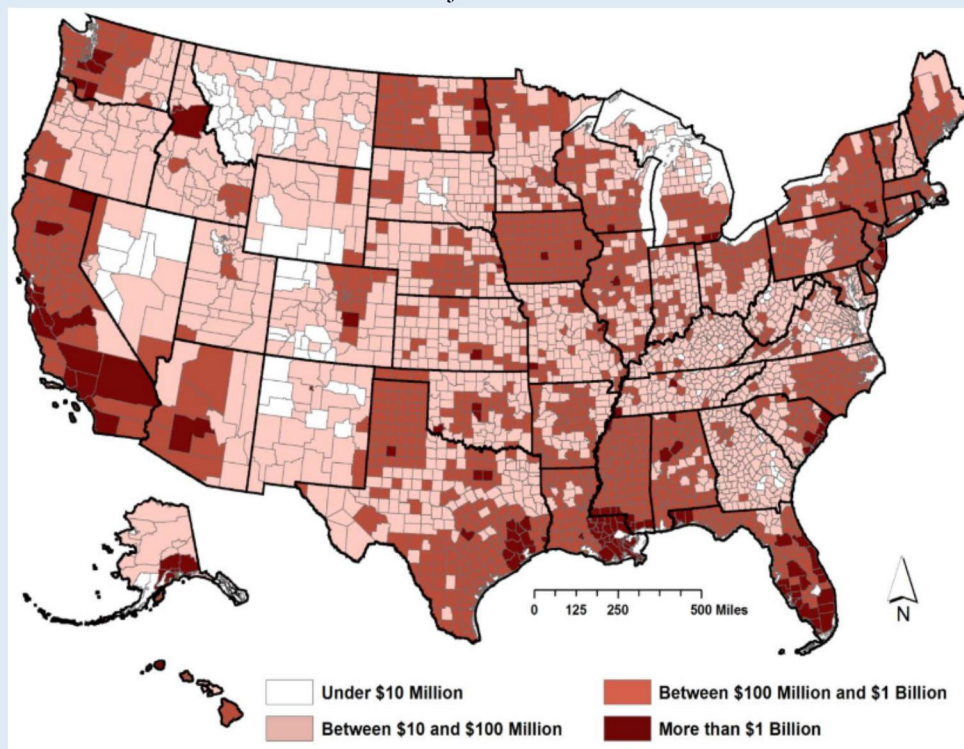
Figure 1. Spatial distribution of direct, economic losses (property and crop) from natural hazards between 1960 and 2013 (in billion, US$ 2013)

Figure 2. Location-based search interface of SHELDUS®

Applications

The SHELDUS® database is widely used, particularly among researchers and planners (Figure 3). As of May 2015, there are more than 85 peer-reviewed publications and 50 thesis/dissertations that heavily utilize SHELDUS® data. The topics of research range from resilience models for the tourism and hospitality industry to statistical analyses on fat-tailed distributions. The most frequent users of SHELDUS® come from disciplines such as economics, geography, environmental studies, engineering, climatology, urban planning and public policy.

In the realm of planning, SHELDUS® data is extensively integrated into local and state hazard mitigation plans. Since 2000, local and state governments must maintain such plans in order to remain eligible for federal disaster dollars. The goal of these plans is to identify mitigation actions that reduce the impacts of natural hazards. A central component of a hazard mitigation plan is a risk assessment, which details the type of hazards affecting a community as well as historic events and their losses. This is where SHELDUS® data plays an important role because it allows planners to download historic event and loss information for the locality of their choice without any additional post-processing. Nearly half of all U.S. states and countless numbers of counties use SHELDUS® in their risk assessment portion of hazard mitigation plans. There is also an increasing use of SHELDUS® in climate change adaptation and resilience plans.

While research and planning represent the main usage areas of SHELDUS®, there are creative uses of the data elsewhere. For example, college instructors have incorporated SHELDUS® into their lesson plans and even museums' displays have drawn on SHELDUS® to communicate the socio-economic impacts from natural hazards on the U.S.
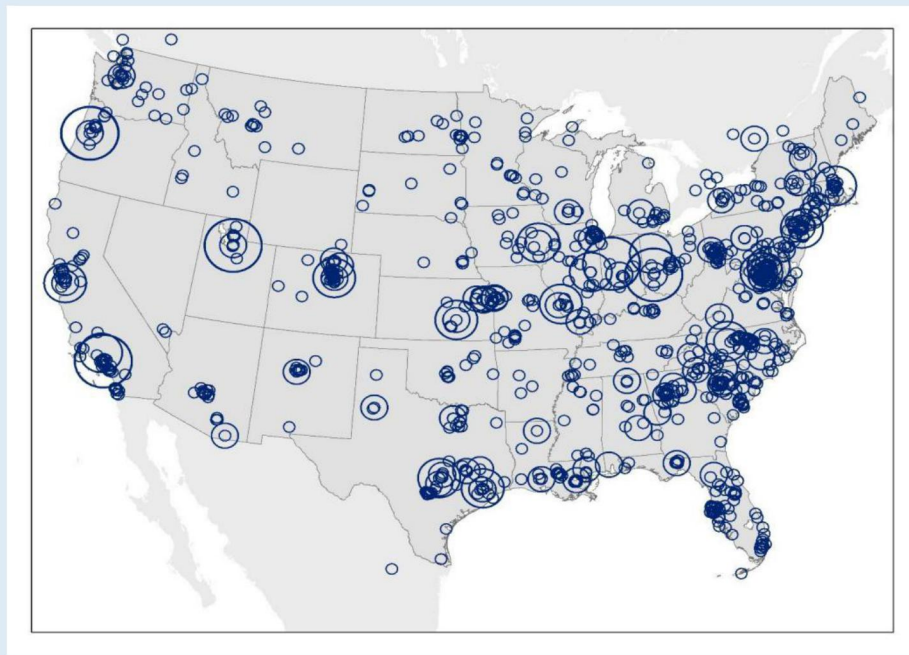
Figure 3. Locations of SHELDUS downloads between July and December 2012. The size of the circle represents the number of downloaded SHELDUS records.

6.2   The USGS Hazards Data Distribution System (HDDS)

Remotely sensed datasets such as satellite imagery and aerial photography can be invaluable resources to support the response to and recovery from many types of emergency events such as floods, earthquakes, landslides, wildfires, and other natural or human-induced disasters. When disasters strike, there is often an urgent need and high demand for rapid acquisition and coordinated distribution of pre- and post-event geospatial products and remotely sensed imagery. These products and images are necessary to record change, analyze impacts of and facilitate response to the rapidly changing conditions on the ground.

The primary goal of U.S. Geological Survey (USGS) Emergency Response project is to ensure that the disaster response community has access to timely, accurate, and relevant geospatial products, imagery, and services during and after an emergency event. The USGS Hazards Data Distribution System (HDDS) provides quick and easy access to the remotely sensed imagery and geospatial datasets that are essential for emergency response and recovery operations (Lamb 2012; Jones & Risty; Jones & Bewley 2010; Bewley; Jones & Lamb).

The USGS HDDS serves as a single, consolidated point-of-access for relevant satellite and aerial image datasets during an emergency event response. The coordinated and timely provision of relevant imagery and other datasets is an important component of the USGS support for domestic and international emergency response activities.

The disaster response liaison provides access for satellite tasking, imagery acquisition and distribution, image registration, creation and distribution of disaster-extent maps, and Web-based mapping services for products, and for pre– and post-disaster data storage and distribution. The USGS Earth Resources Observation and Science Center (EROS) staff also works on disaster preparedness,

including providing basic pre-event data such as satellite images, vector data layers, and other pre-disaster data layers.

The Hazards Data Distribution System (http://hddsexplorer.usgs.gov/), as a disaster response system, incorporates satellite tasking and data acquisition, product development, Web applications, and data storage.

Data will be acquired for each event that meets the specified criteria. The data will be obtained from the providers whose imagery meets the criteria needed for responding to the event. The data will be made available around the clock and acquired in a timely fashion by USGS staff. The data will be processed up to a standard level, preferably precision terrain corrected. If that level of processing is not possible, then the next highest level will be sought. The USGS EROS will also take on the responsibility of invoking the International Charter for Space and Major Disasters. The data provided by the Charter is free of charge but may have provider distribution restrictions that need to be enforced.

In the product development process, any value-added processing that is done can be made available via a web-based delivery system.

The USGS EROS can provide data sets from Landsat 1-5, 7, and 8, digital orthophoto quadrangles, digital raster graphics, digital line graph, and digital elevation models in accordance with applicable distribution policies and agreements. Other data sets that can be supplied include Advanced Spaceborne Thermal Emission and Reflection radiometer (ASTER), Moderate Resolution Imaging Spectroradiometer (MODIS), Advanced Very High Resolution Radiometer (AVHRR), Hyperion, Advanced Land Imager (ALI), IKONOS, Quickbird, WorldView, SPOT, Radarsat, aerial photography, and Disaster Monitoring Constellation (DMC).
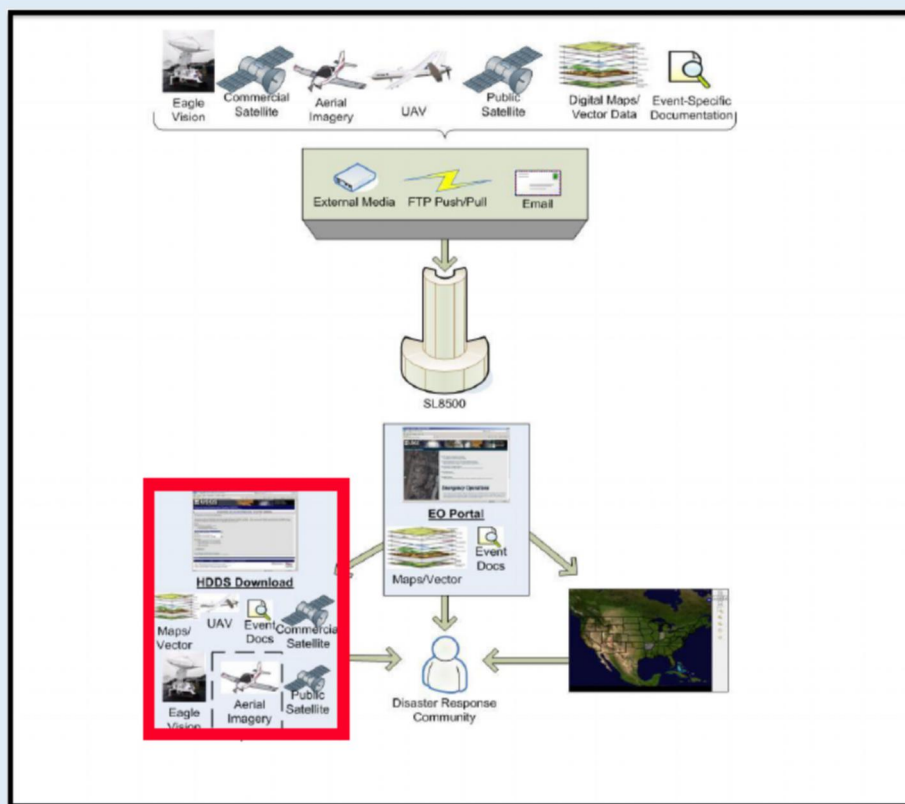
Figure 4. Hazards Data Distribution System (HDDS)

The data will be delivered in standard formats and map projections via the web, ftp, and media. The media deliveries will occur only if the web and ftp are not functioning as viable alternatives to a speedy response. All participants will have their own unique user code/password to access the HDDS. Access to the restricted/licensed imagery will be requested and granted on an event basis. All new data acquisitions will be distributed as licensed to promote the sharing of information by as many participating agencies as possible. USGS EROS recognizes the importance of agencies working together to respond effectively to a major terrorist incident or natural hazards. Therefore, participating agencies are encouraged to submit their value-added products or other data layers for distribution on the HDDS.

Currently, HDDS holds over 354 Tb of data from over 8.8 million files, specifically, corresponding to over 1300 baseline and disaster events. Public access data comprises about 248TB (7,761,504 files), which provides general public with data of unrestricted access; restricted data holdings account for approximately 106 TB (1,116,571 files), which provides data for designated emergency response agencies with password-protected access.

The functions of interactive HDDS interface include:
- Immediate download access to event-related imagery.
- Geographical data visualization with browse image and footprint area overlay.
- Extensive metadata available in multiple formats (for example, CSV, KML, SHP, XML, FGDC).
- XML-based Really Simple Syndication (RSS) feeds for newly ingested data.
- Registration service and log-in access for restricted datasets.
- Standing request services
- Custom web-mapping services
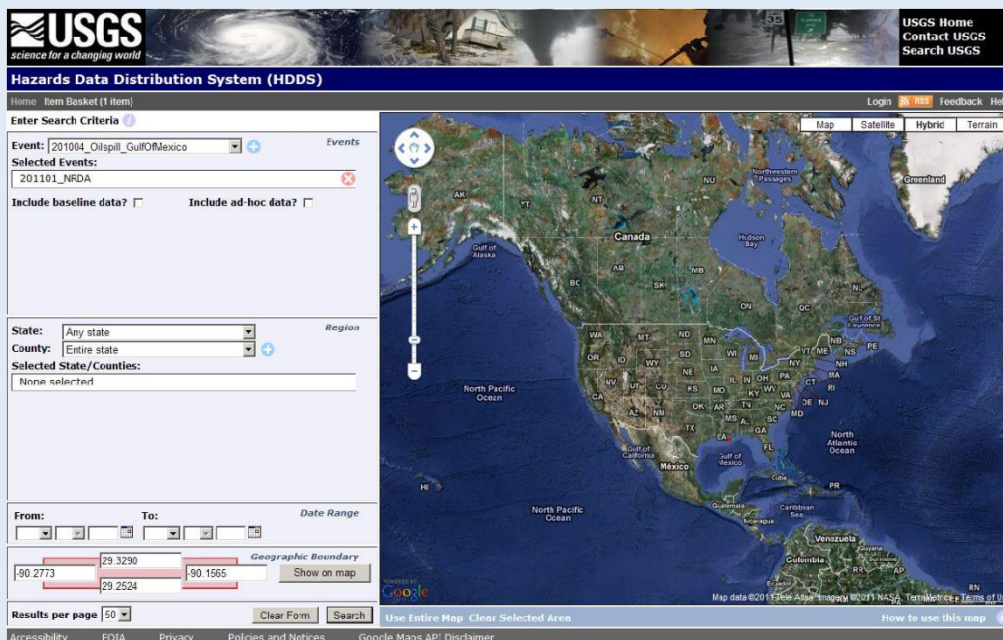- Bulk data download capability.

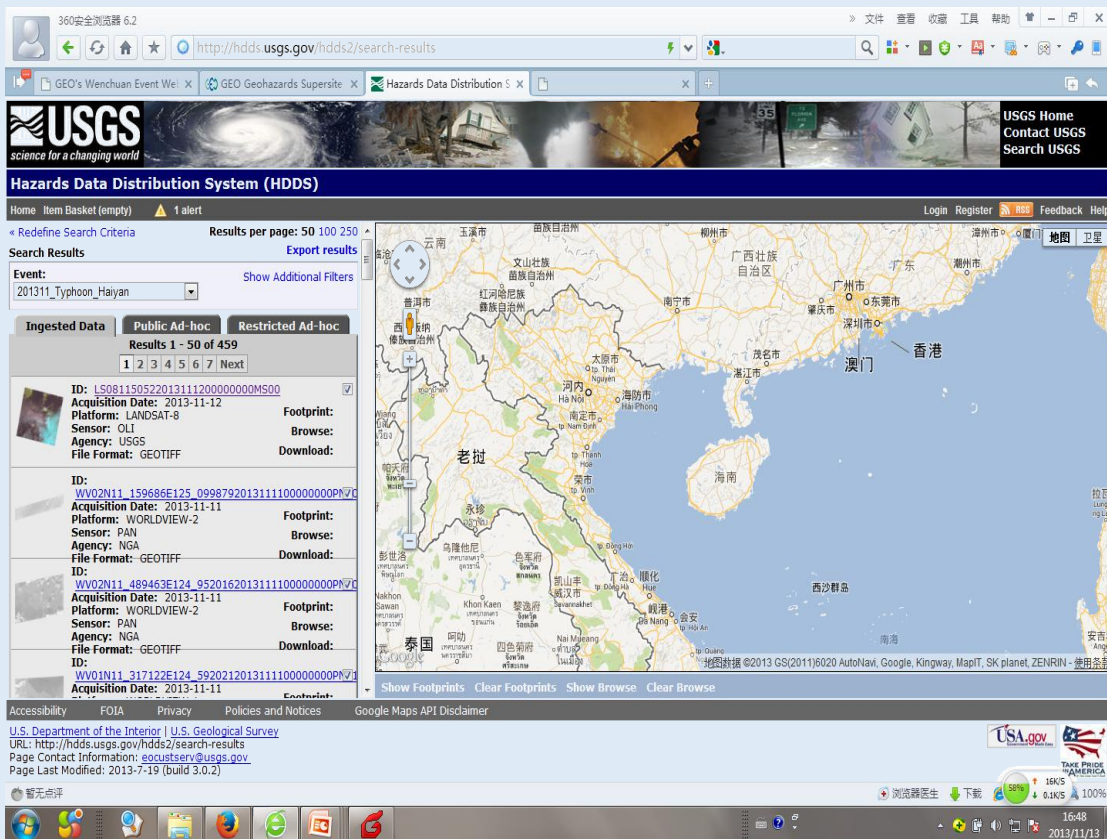**Figure 5. HDDS2 - Graphical Interface of data retrieval**

Figure 6. Home Page of the USGS Hazards Data Distribution System (HDDS) The HDDS site provides event-based download access to remotely sensed imagery and other datasets acquired for emergency response.

The data can be easily accessed in standard formats, and the response community will have the opportunity to share their value-added products with one another.

- Data sharing among agencies ensures that the same images are being used. This will allow for cooperation and sharing of value-added products, because standards are in place.

- Near-line and off-line archiving and retrieval ensure that the data is preserved for historical evaluation and reuse. This also saves dollars for future studies and for response to events that occur in the same location.

6.3   ESA Geohazards Exploitation Platform (GEP)

This section summarizes the capabilities of the Geohazards Exploitation Platform developed by the European Space Agency (ESA). For a more complete description, please refer to the Appendix B.

As far as the value-added products are concerned, ESA focuses effort on hosted processing capabilities offered by its Geohazards Exploitation Platform (GEP), providing users with a range of data processing and dissemination services. In general, the processing capabilities are made available to users as-a-Service, where the users can define a set of input data and processing parameters, and trigger the execution of the algorithms.

In order to support the development of new algorithms, the ESA GEP also provides a Platform-as-a-Service capability (PaaS). In particular, Cloud Sandboxes enable developers and integrators to easily implement new algorithms. This solution utilizes the Virtual Machine technology and employs a middleware to provide a transparent interface to Cloud services. The cloud services are used to scale up the processing when the dimensions of the input dataset increase. GEP offers direct access to dedicated Cloud Sandboxes to researchers in MARsite (marsite.eu) who are interested in experimenting with their own algorithms on ESA data.

The use of GEP removes the need of transferring huge amounts of input product data from the ESA archives and from other agencies' archives synchronized with ESA's virtual archive to the users' machines, resulting in significant savings for the users. With GEP, the agency is providing partners with tools and infrastructure aimed at supporting geohazards researchers and practitioners with easy and open access to Earth Observation satellite data (e.g. ESA, JAXA, DLR, ASI missions), community knowledge and expertise, and collaborative research.

In the geohazard domain, science users require satellite EO to support mitigation activities designed to reduce risk. These activities are carried out before the earthquake (or other geological peril) occurs, and they are presently the only effective way to reduce the impact of earthquakes on society. Short-term earthquake prediction today offers little promise of concrete results. The assessment of seismic hazard requires gathering geo-information for several aspects: the parameterization of the seismic sources, knowledge of historical and instrumental rates of seismicity, the measurement of present deformation rates, the partitioning of strain among different faults, paleo-seismological data from faults, and the improvement of tectonic models in seismogenic areas. Operational users in charge of seismic risk management have needs for geo-information to support mitigation. Satellite EO can contribute by providing geo-information concerning crustal block boundaries to better map active faults, maps of strain to assess how rapidly faults are deforming, and geo-information concerning soil vulnerability to help estimate how the soil is behaving in reaction to seismic phenomena (read more from https://geohazards-tep.eo.esa.int/#!pages/initiative).

An "Exploitation Platform" refers to a virtual ICT environment, often cloud-based, providing users with very fast access to: (i) a large volume of data (EO/non-space data), (ii) computing resources (e.g. hybrid cloud/grid), and (iii) processing software (toolboxes, RTMs, retrieval schemes and visualization routines).

The idea underpinning the exploitation platforms is to enable users to perform effective data-intensive research by providing them with a virtual machine running dedicated processing software close to the data, thereby avoiding moving large volumes of data through the network and spending non-research time on developing ICT tools. The GEP platform is accessible online for users (including public access level) at: http://geohazards-tep.eo.esa.int(Figure 7).
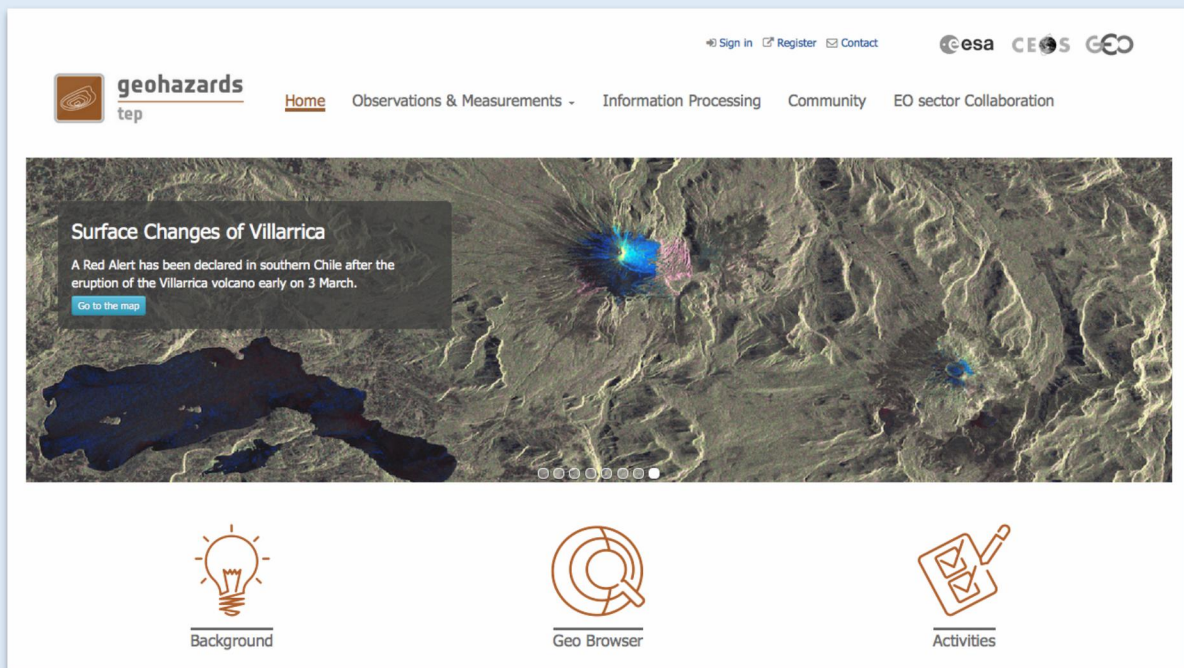
Figure 7 - The Geohazards Exploitation Platform

It is providing the following set of capabilities, which are made available to the MARsite partners.

EO data discovery service through a single point of access to visualize data collections in terms of acquisition footprints and sensor parameters, with resources available from ESA missions (especially the SAR missions from ENVISAT, ERS and SENTINEL-1) and third party missions (currently DLR TerraSAR-X and, upcoming for, ASI Cosmo-Skymed and CNES).

EO data access service over distributed repositories supporting the dissemination of imagery either stored in the GEP cloud platform environment or accessed through the GEP portal in other remote data repositories from the pool of contributing agencies. Data access is based on the authentication of registered users and the granting of data dissemination according to the user profile. For instance, EO data constrained by license terms and distribution restrictions can be accessed from the platform's geographic interface via active links to the repository of the data provider.

Accounting service for EO data consumption allowing the monitoring of the volumes of data use per EO source and according to the activity associated to the user profile. The accounting service can be used to support reporting concerning the exploitation of EO data, either from the Platform (e.g. hosted processing) or in the framework of application projects (e.g. CEOS pilots).

EO processing services for on-demand processing, exploiting software to transform EO data into measurements; the user may run an EO processor provided on the platform (ready to use software-as-a-service, SaaS), or integrate an application he/she has developed (platform-as-a-service capabilities, or PaaS).

The SaaS Processing can be invoked either interactively though a web browser, or through scripting using the OGC Web Processing Service (WPS) interface; The PaaS provides software development and integration tools, and enables users to perform their data exploitation activities with

large flexibility and autonomy, by using one or several virtual hosts directly provisioned on the cloud platform and deployable on demand.

Access to Value-Added products generated on the GEP, or products contributed by third parties. The platform allows cataloguing and dissemination of products relevant to the geohazards community. It can be used to provide access to elaborated products in support of thematic exploitation goals.

As of today, the GEP already allows public data search over a large archive of EO data from ESA sensors and from third-party missions (CEOS partners) like DLR's TerraSAR-X.

Practitioners can access the GEP infrastructure through the Geobrowser service, which is now made available with a set of baseline functionalities to all users (unregistered users, general public) for data search, data selection and data processing at: http://geohazards-tep.eo.esa.int/geobrowser (Figure 8).

The user documentation for GEP is also available online. It features an overview of the Platform concepts, a Community Portal User Guide, a Cloud Operations Administrator Guide and a growing set of data processing tutorials (SAR processing with ADORE DORIS, GMTSAR, ROI_PAC, and a set of G-POD services such as GAMMA-L0, SBAS). This user documentation will continue to evolve in the coming months (currently available at: http://terradue.github.io/doc-tep-geohazards/).
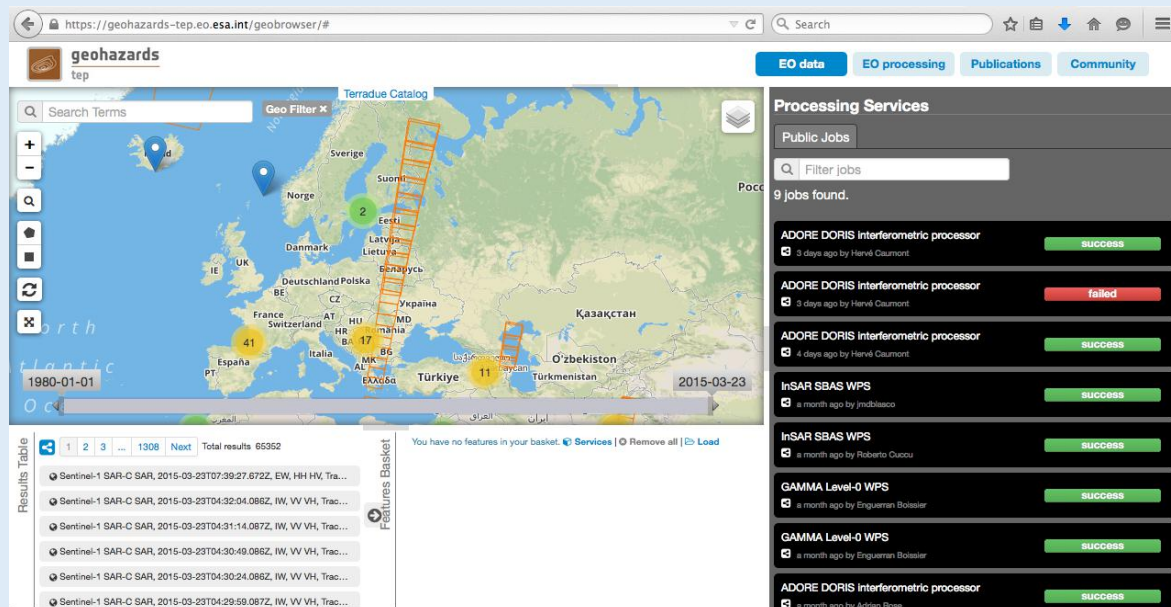

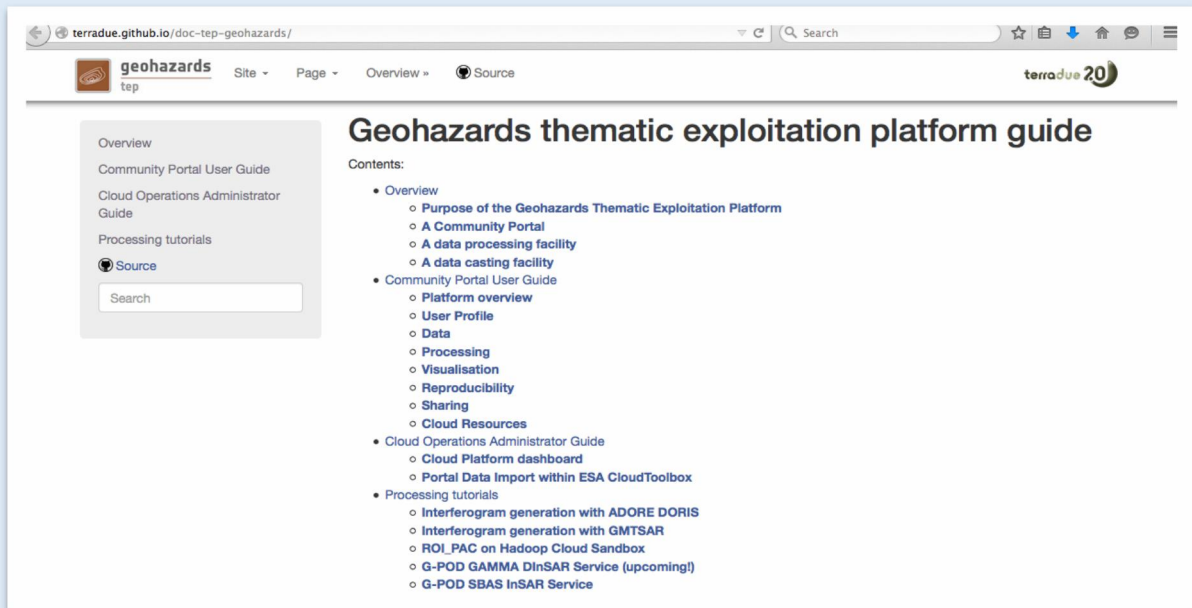**Figure 8 - User access to the Geohazards Exploitation Platform**

Figure 9. User documentation for the Geohazards Exploitation Platform

6.4 The "Disaster Reservoir" Project in China

The Disaster Emergency Data Reservoir (DEDR) project illustrates the importance of integrating technological solutions with governmental coordination and resource sharing.

When the Wenchuan earthquake occurred on May 12, 2008, transportation and communication infrastructure collapsed and the area was devastated. In the relief efforts shortly after the earthquake, timely and quality aerial remote sensing data was urgently needed. In the post-quake phrase, various departments needed to assess losses in the quake-hit areas in a timely fashion to appropriately guide the reconstruction. Images from military and civilian satellites and other aerial remote sensing resources were assembled in real-time through a massive campaign. Nonetheless, it was apparent that China lacks the technical system to support the coordinated multi-agency emergency response to disasters, and the inadequate capability in sharing the relevant spatial data severely impedes the effort at such critical times. In July 2008, the National Sci-tech Support Plan launched the project of *National Emergency Coordination System of Spatial Data Acquisition and Application and Data Sharing Service Platform*, also referred to as *Disaster Emergency Data Reservoir* (DEDR), to research spatial data acquisition and application, emergency coordination plan and dispatch technique, emergency data sharing technique, acquisition technology of aeronautical and spatial data, and spatial data application emergency cooperation mechanism. Central to the project is to build a nationally coordinated platform for spatial data acquisition, sharing, distribution and application for use in nationwide disaster prevention and relief efforts.

Aside from the technical aspects of the data sharing platform, the system also assists with coordination of various agencies within the government tasked with spatial data acquisition, processing and distribution. For example, the platform is designed to plan data acquisition tasks and dispatch these tasks to the designated aeronautical and spatial data acquisition agencies during national Level I or II emergency response modes. These organizations will acquire and process data

according to task requirements, and send the assembled spatial data products to an emergency data sharing service center that archives, manages and distributes relevant data to users. The emergency service system activates various units within the government to perform pre-planned configurations, such as activating emergency communications systems, deploying military satellites and aerial remote sensing planes for imagery of the disaster-hit areas, to aid disaster response efforts. In addition to domestic remote sensing capabilities, the Chinese government has agreements with approximately ten satellites operated by organizations in different countries, including UNESCAP, France, and United States, who will share their information acquired during the two most recent days of an earthquake for the relevant areas. Communication and coordination were set up by the Ministry of Science and Technology to disseminate the data from international partners to relevant domestic units in a timely manner.

This data reservoir platform was subsequently tested in the relief effort of Yushu Earthquake on April 15, 2010, confirming the benefit of utilizing the coordination and data sharing system in large scale emergency situations. For instance, on the next day after the earthquake, the aerial remote sensing planes flew to quake-hit zones for 0.4-meter high resolution remote sensing images, the first-hand materials for seismic resistance and disaster mitigation in Yushu. Meanwhile, the emergency coordination platform acquired, according to plan, archived historical data and programming data from China's satellites, including data from meteorology satellites, resource satellites, a series of remote-sensing satellites, Beijing-1 Satellite, foreign commercial satellites. These data sets provided substantial support for monitoring and evaluation of Yushu Earthquake, and in the post-quake reconstruction. The data reservoir platform was used again to aid the emergency response to Lushan Earthquake (April 20, 2013), providing satellite and aerial imagery of pre-quake and post-quake data within 12 hours after the quake hit the area. These real-life applications informed the developers of user needs and coordination requirements, and helped government units and commercial entities improve their capabilities in data acquisition, business coordination and resource sharing.

This project has demonstrated the importance of governmental role in leading and building up data exchange technology infrastructure to overcome the boundaries and self-interests of organizations and the barriers to timely decision-making. Standardization of technical interfaces for data acquisition, access and sharing has enabled the interoperation of multiple data sources, leading to timely dissemination of relevant data in emergency responses.

07

Conclusions and Recommendations

Natural hazards pose a great threat and challenge to all human societies. It is the inevitable phenomenon resulting from interactions between natural processes of the earth and human activities, which make disaster recognition and disaster mitigation a complex task. The diversity in hazards requires a multi-disciplinary approach, integrating scientific research and application of the findings with pre-disaster prediction, decision making and assessment based on complete, scientific and reliable data.

As our societies have entered the information age, the collection, storage and services of disaster related data have also entered the digital phase, enabling large scale analysis and processing of such data by the scientific community. Open data strategies have been adopted in more and more disaster related data service infrastructure, especially with the trend of global open data. In all phases of disaster management including forecast, emergency response and post-disaster reconstruction, the need for interconnected multi-disciplinary open data for collaborative study and analysis is apparent, in order to recognize and discover the rules and discipline of disaster completely, scientifically and in time.

The extraordinary progress in computing and information technology in the past decade, such as broad local and wide-area network connectivity (e.g., Internet), high performance computing, service and cloud computing, and big data methods and mobile devices, provides the technical foundation for connecting open data to support disaster research. Organizations, especially in the earth observation community, have launched efforts to connect global disaster related data resources and achieved great successes in studies and real-life cases. A new generation disaster data infrastructure based on the interconnected open data begins to form (this topic will be discussed in a separate white paper from the LODGD Task Group). In the science community, interconnected open data for disaster is beginning to influence how disaster data is shared and will need to extend coverage of data and provide better ways of utilizing data across domains where innovation and integration are very much needed.

One way to heed the call of the Sendai Framework for greater use of science is to build strong links between such National Platforms and leading networks of scientists, researchers and other academics. The Integrated Research on Disaster Risk Programme (IRDR), funded by the Chinese Academy of Science and co-sponsored by the International Council for Science (ISC), the United Nations Office for Disaster Risk Reduction (UNDRR), and the International Social Science Council (ISSC), aims to serve as that link to bring more science and data-driven approaches to disaster risk management.

What can the scientific community and governmental organizations do to advance the state of disaster relevant data for a better understanding of disasters and more effective ways of mitigating and reducing impact of disasters on lives and properties?

We hereby make the following recommendations:
1. Academia, disaster management agencies and international organizations should strengthen global collaboration on disaster data by coordinating the utilization of disaster data from multiple data repositories, and promoting the interconnection and use of multi-domain data as a high-priority scientific activity in disaster research.
2. Organizations like the United Nations (UNDRR), national governments and relevant agencies should mobilize resources and accelerate the effort of establishing common definitions and data standards for disaster data to ensure effective implementation of data interconnectivity at both technical and policy levels. Scientists and research institutions are urged to actively participate in such efforts to ensure that these standards meet the needs of disaster research.

3. Disaster data acquisition, preservation and service organizations should improve the accessibility and usability of disaster related data and realize the CODATA principle of data sharing. Suggestions include:
    a. Help connect data users (researchers, disaster management agencies, policy makers, corporate managers, citizens) to data providers through IRDR/CODATA workshops and other professional meetings
    b. Provide best practice guidelines on implementing CODATA open data strategies and showcase example implementations
    c. Moving data repositories from research to service, reducing barriers for data access and use.
    d. Expand open data beyond the domain of earth observations to include other types of data, especially the economic, population, public health, infrastructure, and social media data, etc.
4. Consult with relevant agencies and communities, and establish disaster data copyright protection and acceptable use policy to ensure the legality and appropriate use of data during disaster mitigation.
5. Study, design and ultimately create the next-generation disaster data infrastructure to enable the discovery of and easy access to highly usable, distributed, multi-disciplinary datasets for disaster mitigation stakeholders and applications on a global scale. Developed countries should take the lead in enabling global data service capabilities.
6. The international community should focus on the urgent needs of developing countries in disaster management and help them by establishing appropriate mechanisms of global and regional cooperation and basic data infrastructure to utilize the open data resources from the international community over Internet.
7. Innovative ideas are needed to encourage the private sector to join this effort. The private sector and the public also have incentives to support efforts to open and link data for disaster research and reduction.
8. Establish one or more pilot projects on applications of cross-disciplinary approaches and use of data for studying disaster risk reduction. This could involve academic experiments, datasets, infrastructure, test beds, and institutions who are at the forefront of supplying and using data to assist with disaster management. The scale could be institutional, national, and regional.

Abbreviations

ADPC: Asian Disaster Reduction Centre (ADRC)

ALI: Advanced Land Imager (ALI)

ALOS: Advanced Land Observation Satellite

ASI: Agenzia Spaziale Italiana

ASTER: Advanced Spaceborne Thermal Emission and Reflection radiometer (ASTER),

AVHRR: Advanced Very High Resolution Radiometer (AVHRR), Hyperion,

BJ-1: Beijing-1 satellite

CBERS: China-Brazil Earth Resources Satellite

CEODE: Centre for Earth Observation and Digital Earth

CEOS: Committee on Earth Observation Satellites

CGMS: Coordination Group for Meteorological Satellites (CGMS)

CNES: Centre National d'Etudes Spatiales (CNES)

CODATA: Committee on Data for Science and Technology (CODATA)

CRED: Centre for Research on the Epidemiology of Disasters (CRED)

CRESDA: China Centre for Resources Satellite Data and Application (CRESDA)

DDI: Disaster Data Infrastructure (DDI)

DEDR: Disaster Emergency Data Reservoir (DEDR)

DLR: Deutsches Zentrum für Luft- und Raumfahrt

DMC: Disaster Monitoring Constellation (DMC)

EDC: Earth Resources Observation Systems (EROS) Data Center (EDC)

EPOS: European Plate Observatory System

ESA: European Space Agency

FEMA: Federal Emergency Management Agency (FEMA)

FY: Fengyun satellite

GCOS: Global Climate Observing System(GCOS)

GEO: Group on Earth Observations

Geonet: A geological hazards monitoring service in New Zealand run by GNS Science

GEOSS: Global Earth Observation System of Systems (GEOSS)

GPS: Global Positioning System

GRIP: Global Risk Identification Programme (GRIP)

GSNL: Geohazards Supersites and Natural Laboratories (GSNL)

HDDG: Historical Disaster Data Grid (HDDG)

HDDS: Hazards Data Distribution System (HDDS)

ICSU: The International Council for Science (ICSU)

INGV: Istituto Nazionale di Geofisica e Vulcanologia

InSAR: Interferometric Synthetic Aperture Radar (InSAR)

IOC: Intergovernmental Oceanographic Commission (IOC)

IRDR: Integrated Research on Disaster Risk

IRG: Integrated Risk Governance Project (IRG)

IRIS: Interface Region Imaging Spectrograph (IRIS)

JAXA: Japan Aerospace Exploration Agency

LA RED: Red de Estudios Sociales en Prevención de Desastres en América Latina (Network of Social Studies in the prevention of Disasters in Latin America - LA RED)

LODGD: Linked Open Data for Global Disaster Risk Research.

MODIS: Moderate Resolution Imaging Spectroradiometer (MODIS),

NASA: National Aeronautics and Space Administration

NCDC: National Climatic Data Centre

NMA: National Meteorological Agency

NOAA: The National Oceanic and Atmospheric Administration

NSF: National Science Foundation

NWS: The U.S. National Weather Service (NWS)

OECD: the Organisation for Economic Co-operation and Development (OECD)

PRC: The People's Republic of China

SCI: Science Citation Index

SCU: University of South Carolina

SHELDUS: The Spatial Hazard Events and Losses Database (www.sheldus.org)

SJ-9A: Shijian-9A satellite

SPARC: Scholarly Publishing and Academic Resources Coalition (SPARC)

SPOT: Systeme Probatoired'Observation dela Tarre

TWAS: The World Academy of Sciences

UAV: Inmanned Aerial Vehicle (UAV)

UN ECA: United Nations Economic Commission for Africa

UN-ESCAP: United Nations Economic and Social Commission for Asia and the Pacific (UN-ESCAP)

UN-ESCAP: United Nations Economic and Social Commission for Asia and the Pacific (UN-ESCAP)

UNAVCO: A Non-Profit University-Governed Consortium, Facilitates Geoscience Research And Education Using Geodesy.

UNDP: United Nations Development Programme (UNDP)

UNEP: United Nations Environment Programme (UNEP)

UNESCAP: United Nations Economic and Social Commission for Asia and the Pacific (ESCAP)

UNSPIDER: United Nations Platform for Space-based Information for Disaster Management and Emergency Response

USGS: United States Geological Survey

WDC: World Data Centre (WDC)

WDS: World Data System

WGDD: Working Group on Disaster Data (WGDD)

WMO: World Meteorological Organization (WMO)

References

Amelung F, Lengert W, Puglisi G. Supersites: An Initiative Towards Open Access for InSAR Data[C]//EGU General Assembly Conference Abstracts. 2010, 12: 11682.

Atkins, Daniel. "Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure." (2003).

Bewley R D. USGS Emergency Response Resources[J].

Cohen, W.M. and Walsh, J.P., 2007. Real impediments to academic biomedical research. *Innovation policy and the economy*, *8*, pp.1-30.

Committee on Scientific Accomplishments of Earth Observations from Space, National Research Council (2008). Earth Observations from Space: The First 50 Years of Scientific Achievements. The National Academies Press. p. 6. ISBN 0-309-11095-5.

Committee on Scientific Accomplishments of Earth Observations from Space, National Research Council (2008). Earth Observations from Space: The First 50 Years of Scientific Achievements. The National Academies Press. p. 6. ISBN 0-309-11095-5. Retrieved 2010-11-24.

Cutter, S.L. Social Science Perspectives on Hazards and Vulnerability Science. T. Beer (ed.), *Geophysical Hazards,* International Year of Planet Earth, 17. DOI 10.1007/978-90-481-3236-2_2, © Springer Science+Business Media B.V. 2010

Dankers, R. and Feyen, L., 2009. Flood hazard in Europe in an ensemble of regional climate scenarios. *Journal of Geophysical Research: Atmospheres*, *114*(D16).

Dankers, R., Arnell, N.W., Clark, D.B., Falloon, P.D., Fekete, B.M., Gosling, S.N., Heinke, J., Kim, H., Masaki, Y., Satoh, Y. and Stacke, T., 2014. First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. *Proceedings of the National Academy of Sciences*, *111*(9), pp.3257-3261.

Egeland, Ricky, Tony Wildish, and Chih-Hao Huang. "PhEDEx data service." Journal of Physics: Conference Series. Vol. 219. No. 6. IOP Publishing, 2010.

Fan, Xu'er, Li, Xinxin and Zhang, Yingjie. Public Health Data Sharing Policy Study [J]. Chinese Journal of Public Health Engineering, Vol. 5, No. 2, April 2006.

Foster, Ian. "Accelerating and democratizing science through cloud-based services." IEEE Internet Comput. 15.ANL/MCS/JA-69753 (2011).

Geohazard Supersite: The Geohazard Supersites Partnership White Paper and Implementation Plan, 11 October 2011; Selection Process for GEO Geohazard Supersites, A Proposal of the Committee on Earth Observing Satellites. September 18, 2012; Geohazard Supersites and Natural Laboratories (GSNL) Initiative "Supersites Definitions" 2012-2015 GEO Work Plan – Disasters (DI-01) Task, September 6 2012

Haeussler, C., Jiang, L., Thursby, J. and Thursby, M., 2014. Specific and general information sharing among competing academic researchers. *Research policy*, *43*(3), pp.465-475.

Halevy, A., Norvig, P. and Pereira, F., 2009. The unreasonable effectiveness of data. *IEEE Intelligent Systems*, *24*(2), pp.8-12.

http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData#FAQ.

Jones B K, Bewley R D. The Hazards Data Distribution System is Updated[M]. US Department of the Interior, US Geological Survey, 2010.

Jones B K, Lamb R. USGS Emergency Response and the Hazards Data Distribution System (HDDS)[C]//AGU Fall Meeting Abstracts. 2013, 1: 1601

Jones B K, Risty R R. The National Map Hazards Data Distribution System[J].

Lamb R M, Jones B K. United States Geological Survey (USGS) Natural Hazards Response[M]. 2012.

Lengert W, Popp H J, Gleyzes J P. GEO Supersites Data Exploitation Platform[C]//EGU General Assembly Conference Abstracts. 2012, 14: 11933.

Li, J., Liu, D. and Jiang, H. Research on International Current Status on Sharing Science Data [J], *Library Development*, 2009.

Liu, Kejing. Information Sharing Thoughts in Developed Countries and the Relevant Legal Systems [J]. Library and Information Service, Vol. 5, No. 3, 2007.

Moore, Reagan W., and ArcotRajsekar. "Irods: Data sharing technology integrating communities of practice." Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International. IEEE, 2010.

Munich Re Annual Report 2004.

On the Full and Open Exchange of Science data (1995) National Research Council, Washington, DC

Open Data Barometer, 2013 Global Report

Open Data Centre Alliance. Defining a New Class of Data Centre and Cloud Infrastructure Solutions. http://www.opendatacentrealliance.org/

Organisation for Economic Co-operation and Development, *Inquiries into Intellectual Property's Economic Impact*, OECD Publishing, Paris (2015a).

Organisation for Economic Co-operation and Development, *Making Open Science a Reality*, OECD Publishing, Paris (2015b).

Rowan L, Baker S, Wier S, et al. Integrated Data Search and Access to Geophysical Data for Geohazards Supersites and Natural Laboratories[C] AGU Fall Meeting Abstracts. 2013, 1: 1621.

Scholarly Publishing and Academic Resources Coalition (SPARC), https://sparcopen.org/open-data/

The Global Earth Observation System of Systems (GEOSS) 10-Year Implementation Plan, (As adopted 16 February 2005). https://www.earthobservations.org/documents/10-Year%20Implementation%20Plan.pdf

Thursby, M., Thursby, J.G., Haeussler, C. and Jiang, L. 2009. "Do academic scientists share information with their colleagues? Not necessarily," VOX – CEPR's Policy Portal, November 25. http://www. voxeu. org/article/why-don-t-academic-scientists-share-information-their-colleagues.

Ulbrich, U., Brücher, T., Fink, A.H., Leckebusch, G.C., Krüger, A. and Pinto, J.G., 2003. The central European floods of August 2002: Part 1–Rainfall periods and flood development. *Weather*, *58*(10), pp.371-377. (2013a)

Ulbrich, U., Brücher, T., Fink, A.H., Leckebusch, G.C., Krüger, A. and Pinto, J.G., 2003. The central European floods of August 2002: Part 2–Synoptic causes and considerations with respect to climatic change. *Weather*, *58*(11), pp.434-442. (2013b)

United Nations Office for Disaster Risk Reduction, *Sendai Framework for Disaster Risk Reduction 2015 – 2030*, Geneva (2015).

USGS Data Catalog. https://data.usgs.gov/datacatalog/#fq=dataType%3A(collection%20OR%20non-collection%20OR%20supercollection)&q=*%3A*

W3C. Linking Open Data［EB/OL］ 2011-03-15.

Zhang, H.Y., Qing, X.L., Huang, M.R. and Li, G.Q. 2015.A Correlation Analysis Model for Multidisciplinary Data in Disaster Research. Data Science Journal, 14(19):1-12. http://datascience.codata.org/articles/10.5334/dsj-2015-019/

Integrated Research on Disaster Risk (IRDR) Community of Practice (CoP) for IRDR's Disaster Loss Data (DATA) project http://www.irdrinternational.org/projects/data/

UN General Assembly (UNGA). Open-ended intergovernmental expert working group on indicators and terminology relating to disaster risk reduction Geneva, 29-30 September 2015, 10-11 February 2016 and 15-18 November 2016 Report of the Chair of the Open-ended Intergovernmental Expert Working Group on Indicators and Terminology relating to Disaster Risk Reduction on Results of informal consultations from February to November 2016 on indicators and terminology relating to Disaster Risk Reduction A/AC.285/CRP.1. Dated 12 November 2016. Available at http://www.preventionweb.net/files/50683_reportofthechaironresultsofinformal.pdf Accessed on 30 November 2016

Rappaport, S.M. (2011) Implications of the exposome for exposure science. *Journal of Exposure Science and Environmental Epidemiology*, *21*(1), pp.5-9.

Goldstein, B.D., Osofsky, H.J. and Lichtveld, M.Y. (2011) The Gulf oil spill. *New England Journal of Medicine*, *364*(14), pp.1334-1348.

Juarez, P.D., Matthews-Juarez, P., Hood, D.B., Im, W., Levine, R.S., Kilbourne, B.J., Langston, M.A., Al-Hamdan, M.Z., Crosson, W.L., Estes, M.G. and Estes, S.M. (2014) The public health exposome: a population-based, exposure science approach to health disparities research. *International journal of environmental research and public health*, *11*(12), pp.12866-12895.

Nichols, G.L., Yvonne A, Lindgren E, Devaux I, Semenza J.C. (2014): European monitoring systems and data for assessing environmental and climate impacts on human infectious diseases. *International journal of environmental research and public health* 11, no. 4; 3894-3936.

Al-Hamdan, M.Z., Crosson, W.L., Economou, S.A., Estes Jr, M.G., Estes, S.M., Hemmings, S.N., Kent, S.T., Puckett, M., Quattrochi, D.A., Rickman, D.L. and Wade, G.M. (2014) Environmental public health applications using remotely sensed data. *Geocarto international*, *29*(1), pp.85-98.

Boyd, D., Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, 15:5, 662-679, DOI: 10.1080/1369118X.2012.678878

Fleming, L.E., Haines, A., Golding, B., Kessel, A., Cichowska, A., Sabel, C.E., Depledge, M.H., Sarran, C., Osborne, N., Whitmore, C., Bloomfield, D. (2014). Data Mashups: Potential Contribution to Decision Support on Climate Change and Health. *International Journal of Environmental Research and Public Health,* 11:1725-1746; doi:10.3390/ijerph110201725

Granella, C., Belmonte Fernández, O., Díaz, L. (2014). Geospatial information infrastructures to address spatial needs in health: Collaboration, challenges and opportunities. *Future Generation Computer Systems*, 31: 213–222.

McMichael, A.J., Haines, A. (1997). Global climate change: the potential effects on health. *BMJ*, 27; 315(7111): 805–809. PMCID: PMC2127530.

Roche, D.G., Jennions, M.D., Binning, S.A. (2013). Data deposition: Fees could damage public data archives. *Nature*, 502(7470):171. doi: 10.1038/502171a.

Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014). A World that Counts: Mobilising the Data Revolution for Sustainable Development. Report prepared at the request of the United Nations Secretary-General.

Fleming LE, Tempini N, Gordon-Brown H, Nichols G, Sarran C, Vineis P, Leonardi G, Golding B, Haines A, Kessel A, Murray V, Depledge M, Leonelli S. (2017) Big Data in Environment and Human Health: Challenges and Opportunities. In: Oxford Encyclopedia of Environment and Human Health [Unpublished book].

Lim, S.S., Allen, K., Bhutta, Z.A., Dandona, L., Forouzanfar, M.H., Fullman, N., Gething, P.W., Goldberg, E.M., Hay, S.I., Holmberg, M. and Kinfu, Y. (2016) Measuring the health-related Sustainable Development Goals in 188 countries: a baseline analysis from the Global Burden of Disease Study 2015. *The Lancet*, *388*(10053), pp.1813-1850.

Nass, S.J., Levit, L.A. and Gostin, L.O. (2009) The value, importance, and oversight of health research. In: Institute of Medicine (US) Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule. National Academies Press US.

Wang, H., Naghavi, M., Allen, C., Barber, R.M., Bhutta, Z.A., Carter, A., Casey, D.C., Charlson, F.J., Chen, A.Z., Coates, M.M. and Coggeshall, M. (2016) Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet*, *388*(10053), pp.1459-1544.

Five year forward view. (2014) *London: NHS England*. Available at: Stevens, S., 2014. Five year forward view. *London: NHS England*. [Accessed 6th January 2017].

Revision History

Date	Editors	Sections changed	Description
May 20, 2016	Carol Song, LI Guoqing	All sections	Revision 1.0 released for external review
Jan-Feb 2017		All sections	Revision 2.0 - Revision after external reviews. Comments and suggestions from the external review have been incorporated.
March 7, 2017	Carol Song, LI Guoqing	All sections	Revision 3.0 – Revision to incorporate additional comments from the LODGD Task Group.
April 12, 2017	Carol Song, LI Guoqing	Sections 1, 2.2.7, 2.2.8, 2.3	Version 1.0. Release for publication.