Chapter 4: Opportunities and enablers of change

4.1

Changes in technology and data sharing

Knowing where people and things are, and their relationship to each other, is essential for informed decision-making. Real-time information is useful to prepare for and respond to disasters. Location-based services are helping governments to develop strategic priorities, make decisions, and measure and monitor outcomes.

As identified by the Global Facility for Disaster Reduction and Recovery (GFDRR),[204] for communities and governments to build resilience to hazards, they must have access to information about disaster risk that is understandable and actionable. Advances in science, technology and innovation can further the understanding of disaster risk and help achieve this goal. Especially when a wide variety of stakeholders across the public, private, academic and NGO sectors form partnerships and work together.

Improvements in technology have been exponential since the publication of GAR15. This, coupled with the increased awareness and willingness to share data, information and data processing capabilities, has enabled a greater understanding of global change and the ability to forecast how natural systems will respond to human activity and political decisions.

Ongoing efforts to engage the science and technology community in developing, implementing and providing data and services to the risk management community are being strengthened. This ensures that the DRR community benefits from the best possible scientific and technological advances and advice. One of the greatest areas of technological enhancement has been in the availability of, and access to, computational processing power. This can be seen through the greater availability of supercomputers and virtual servers, which have increased the overall availability of cloud-based computing capabilities for hazard modelling. In turn, the data available has also improved. As an example, the ESA Copernicus satellite marks a significant improvement in globally available, open, high-resolution satellite imagery.

4.1.1

Hazard knowledge

Data collected on the Earth systems (climate, oceans, land and weather), as well as the societal systems (population location, density and vulnerability), is a fundamental input for many of the calculations to permit a better understanding of the nature and drivers of risk.

The science and technology community have an essential role in the continual advancement of the understanding of hazards, exposure and vulnerability and its effect on reducing the risks to people, infrastructure and society. Satellites have a unique vantage point for monitoring many kinds of large-scale processes, from forest fires to overflowing rivers, to earthquake-prone zones as well as patterns of human settlement, herd migration trends and degradation of coral reefs. Remotely sensed data can be provided in near real time. This can include maps, optical images or radar images that accurately measure the affected areas.

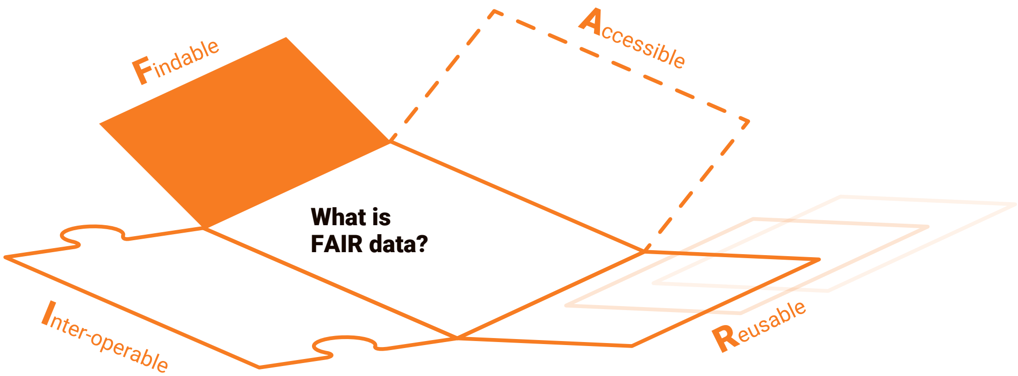
4.1.2

Open data

Open data can have many different interpretations and meanings. Here, open data is described as "data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirements to attribute and share alike."[205]

Open data policies have been shown to be an economic force enhancer for nations, with value created many times over and providing greater returns on investment through increased tax revenues on the products and services created with the data. Open data also meets society's needs for ethical principles for accessing and using public data. Within the research and innovation sectors, open data can facilitate interdisciplinary, inter-institutional and international research. It also enables data mining for automated knowledge discovery among the growing amount of big data available to researchers and policymakers. Finally, open public data supports improved decision-making and enhances transparency in government and society.

Figure 4.1. FAIR data is findable, accessible, interoperable and reusable



(Source: UNDRR 2019: https://www.nature.com/articles/sdata201618)

An open science approach, complementing open data principles, is often followed by research and academic institutions. This works on the basis that data is as open as possible but recognizes that it can be closed if necessary. The findable, accessible, interoperable and reusable (FAIR)

204 (GFDRR 2018a)
205 (Open Data Handbook 2019)

data principles are also a core facet of open and exchangeable knowledge.

For data created with public funds or where there is a strong demonstrable public interest, open data should be the default. There are, of course, several reasons for keeping some data more proprietary or secret, and these need to be balanced against the benefits of openness outlined above. For example, proportionate exceptions include restrictions based on national security, law enforcement, personal privacy and commercial proprietary concerns. Less well known and sometimes more relevant include the protection of indigenous people's rights, and the exact location of cultural artefacts or endangered species.[206]

There are movements advocating for open data. For example the Open Data for Resilience Initiative is designed to support teams of regional risk management specialists to build capacity and long-term ownership of open data projects. The creation of the global open data index also helps by ranking, at the State level, the various degrees to which data is openly available with a view to encouraging use of data from more open jurisdictions.

Some countries have open data policies, while others may have open policies but derive their funding through consulting, which places limitations on how open they can be. Protectionism remains a barrier to open sharing of tools, data and knowledge as people are naturally concerned for the long-term viability of their livelihoods and believe their competitive advantage is rooted in their access to the exclusivity of their knowledge.

There are some cases where the best available data is produced and owned by private companies. Private risk modelling in the private sector is also not open and is dominated by a few big companies that supply "black-box" models. These are models that – whether they are available for public use or not – do not divulge the nature of the calculation used in the model. When data is made publicly available, it is often at least one version behind the most current; in some cases, it is simply not available for free. This can then lead to the challenge of clear data accountability. If data is being used for risk and hazard modelling, it needs to be accurate, trusted and reliable, leading to important questions about the provenance and refresh rates of data. Without clear information about the provenance, history and processing of a given data set, it is difficult to determine how reliable it might be.

Advancements in open data provided from satellites have made more advanced models possible. Landsat and Copernicus are the two contemporary examples by the United States Geological Survey/NASA and ESA, respectively. Landsat provides the longest temporal records of moderate resolution multispectral data of the Earth's surface, while Copernicus is providing the highest-resolution imagery available openly and globally. In 2014, the Sentinel-1 mission provided a polar-orbiting, all-weather, day and night radar imaging mission for land and ocean services. In 2015, Sentinel-2A was launched followed by Sentinel-2B in 2017, providing spatial resolutions of 10, 20 and 60 m. This has improved the resolution previously available and provides high-resolution imagery to be used in various hazard models. The fact that the data is open has resulted in a boom in scientific research based on satellite data.

The initial two Sentinel missions have since been joined by Sentinel-3 (which measures sea-surface topography, sea- and land-surface temperature, ocean colour and land colour) with high-end reliability that helps inform ocean forecasting systems and environmental and climate monitoring. Sentinel-5P was launched in 2017 and provides data on air quality and climate. The variety of data available from Copernicus through the Sentinel missions has revolutionized the scale of open source data available.

It is recognized that while open and available data is useful for many applications within disaster risk management, during an extreme event there is often the need for higher-resolution imagery. In this regard and with the sharing of relevant data enshrined in the International Charter on Space and Major Disasters, private sector providers can work with space agencies to provide timely and accurate data for disaster recovery.

4.1.3

Open source software

Open source software can be described as the provision of source code that is available at no cost and for use by anyone for any purpose. The opposite of open source software is proprietary software, where a user normally must pay to access the software and abide by several restrictions in its use and distribution.

Open source software was rare 10 years ago, but it is now commonplace. Perhaps the greatest benefit of open source tools is their flexibility and evolving capacity that develops as more people use and adapt the software for their specific needs. Shared software helps promote greater levels of understanding of hazards rooted in the same methodology.

Community-driven open source software is increasingly being used in government organizations, and there is a growing number of private sector companies focused on providing technical support to open source software. This movement by governments to use open source software has gone a long way in overcoming barriers to adoption. As with any technology, significant assessments need to be made on the total cost of ownership of open source software. While there may be an initial economic benefit from using open source software, it can be expensive to customize and maintain, as this is dependent on the community developing the software, and the knowledge of the user.

Future-proofing is also a consideration. With open source software, the software itself is less likely to be affected if the company behind its design closes. As other developers can simply pick up where the original ones left off, its sustainability is better ensured. The vision of future-proofing underpins this philosophy. If the base information is available and comprehensible broadly, the likelihood of continued interest in and research about the topic is more likely to continue. These systems emphasize testing and continuous integration where every change in the engine is reviewed by someone else and can include a scientific review and publication. When a change goes into the system, all tests are re-run. Having the whole processes visible and transparent ensures that if a bug is fixed, it will often result in improvements to the tests.

Open software and tools are becoming the software of choice within research institutions. In the early stages, open source implied a free but often primitive version of the commercial software. However, in the last few years, open source software has progressed exponentially and often represents best-in-class versions of scientific modelling tools. With the science rooted in open source tools, more users have access to them, enabling greater contributions and allowing their knowledge and research to feed back into improved development of the tool itself.

Not all software is open source, and their remains reliance on some proprietary software. Proprietary software can have its benefits for organizations using their own data and information for risk modelling, especially if it has been produced by a commercial enterprise and is for commercial use.

One area where open data and open source cross paths is in crowdsourcing. Growing interest in the use of crowdsourced data to solve certain kinds of data problems has led to the development of a number of layers in use within risk science. A notable example is the use of OpenStreetMap, which is foundational to almost all risk sciences. The Humanitarian OpenStreetMap Team has worked on several projects that use community volunteers to produce locally sourced context information. It is training volunteers to collect and code messages in a quality-controlled manner, feeding data to centres that can use it for better understanding of a multitude of hazards. Because there is still some reluctance to rely on crowds to answer important contextual information about risk, exposure and vulnerability, these systems are supplemented in some cases with "expert opinion" to reinforce the pedigree of the data.

---

206 (GEO 2015)

4.1.4

Interoperability

Interoperability may be defined as "the ability of a computer system of software to work with other systems or products without special effort on the part of the user."[207] The interoperability of data has technical, semantic and legal dimensions. From a technical standpoint, the data needs to have compatible formats and well-known quantities that make diverse data possible to integrate to form new data and products.[208]

From the semantic point of view, one of the main challenges to interoperability is contained within the metadata used to describe any given data set. When trying to combine data, challenges can be as simple as the native language of the data creator being different from the data user, meaning that it can be difficult to combine. Another semantic challenge can be with the naming conventions and descriptive terms used in different disciplines (or even subdisciplines). These issues of nomenclature are very important, especially for identifying and measuring risks and hazards.

Legal interoperability can be described as having occurred when multiple data sets from different sources have been merged, and users are able to access and use each of the different data sets without having to seek explicit authorization from each creator of the data.

It is not only the interoperability of data and systems that is important for disaster risk management. DRR is inherently interdisciplinary, and this is reflected in the discussions around cascading risks and hazards. Researchers and professionals often work in silos within their own disciplines. Improving the availability of knowledge and data can encourage practitioners to think about the wider implications of risk-informed decisions.

In terms of interoperability of model components, one suggestion is to bridge the gap between different hazards models using machine learning, leading to a harmonized model across hazards creating a whole simulation model that produces global Earth simulations systems. This is a goal in the future, and could be a very useful policy and advocacy tool. However, it cannot be done at a scale that would make any sense beyond the global level at this stage. For models to inform risk reduction, preparedness and response efforts, they need to be at the local level. Machine learning may be able to assist in this, but it requires a lot of effort to ensure data is fed into the system in the right way. This is an area that is likely to expand in the future as multi-hazard risk continues to be considered.

For data to be used for disaster risk management, it must be discoverable, available, accessible and usable.[209] Initiatives such as the United Nations Committee of Experts on Global Geospatial Information Management work on Geospatial Information and Services for Disasters highlight that during a crisis, the sharing of data about citizens and infrastructure among international organization, NGOs and governments can be critically important.

In recent years, the impacts from natural hazards such as typhoons and hurricanes, as well as epidemics such as the Ebola outbreak in West Africa, have heightened the gaps in availability and access of data. The increasing need for data to be used in DRR and management has also highlighted challenges in coordination and collaboration among stakeholders. This led the United Nations Committee of Experts on Global Geospatial Information Management to create a Strategic Framework for Geospatial Information and Services for Disasters.

Successful implementation of the strategic framework will lead to an outcome where "the human, socioeconomic and environmental risks and impacts of disasters are prevented or reduced through the use of geospatial information and services."[210]

The strategic framework builds on key documents such as the Sendai Framework and United Nations General Assembly Resolution 59/12, and calls for all Member States and other stakeholders to institutionalize good governance practices and science-based policies, supported by improved capacities on human resource, infrastructure and geospatial data management. By supporting countries in addressing the challenges and social, economic and environmental impacts of disasters, it contributes to sustainable development efforts.

4.1.5

Data science

The ability to create data is still ahead of the ability to solve complex problems by using the data. There is no doubt that there is a huge amount of value yet to be gained from the information contained within the data generated. The growth in the amount of data collected brings with it a growing requirement to be able to find the right information at the right time, and challenges of how to store, maintain and use the data collected.

The concept of using computer science and computational processing in science and technology is not new. For nearly two decades, there have been evolving practices and processes in the use of data science. What is becoming more mainstream is the shift to a context where there is no longer a reliance on costly supercomputers to host and process data. The growth of cloud computing, using a distributed network of computing where processes can run parallel on many machines, is lowering the cost of entry for many users. This means that there is now greater uptake and use of cloud computing for risk management. Coupling this with the developments in machine learning and artificial intelligence allows greater interactions within disparate data sets and enables more granular modelling of the drivers of risk.

The cloud computing model is becoming the prevailing mode of work for most medium- and large-scale global data sets, including Earth observation (EO) applications. This is due to the ability of cloud services to archive large satellite-generated data sets and provide the computing facilities to process them.

As cloud computing services are being more widely used, the technology is maturing rapidly. Taking the example of EO analysis as a use case, there are many different platforms and applications available for the risk community to use. These include the Open Data Cube,[211] Copernicus Data and Information Access Services,[212] Earth on Amazon Web Services,[213] Google Earth Engine,[214] the JRC Earth Observation Data and Processing Platform,[215] NASA Earth Exchange,[216] and the European Centre for Medium-Range Weather Forecasts Climate Data Store.[217]

Each of these cloud computing services has different benefits. These range from the way the data is ingested (some include pre-loaded data, which reduces the effort on the part of the user) to scripting language (which is used for the processing). One of the main disadvantages of using cloud services is their lack of interoperability. This means that for users, there must be a trade-off between flexibility and ease of use. For example, Amazon Web Services are flexible, but they require users to be capable of developing applications using basic content libraries. This flexibility comes at the cost of needing to have a steep learning curve. By contrast, Google Earth Engine provides immediate access to functions and data, reducing the barrier to entry.

Set against the benefits of cloud computing, there are some issues that need to be considered in its

207 (Belmont Forum 2015)
208 (GEO 2015)
209 (Murnane et al. 2019)
210 (United Nations Committee of Experts on Global Geospatial Information Management 2017)
211 (Open Data Cube 2019)
212 (EU 2019)
213 (Amazon 2019)
214 (Google 2019)
215 (Soille et al. 2018)
216 (NASA 2019a)
217 (EU 2019)

use. These include the recognition that the distribution of available technology is rarely even, and that many areas still have challenges meeting the needs of basic electricity let alone the high-speed Internet connectivity required for accessing, sharing and processing large quantities of data. For this reason, it is often necessary for software developers to factor in the ability to function offline along with the capacity for downloading the required data sets, so models can be run locally. Access to electricity is a particular concern in an active disaster scenario, so the capacity to work offline is essential. Some models can take multiple days to run, and if power is cut or technology fails during that period, the model must be re-run, which costs valuable time and computing resources.

Large amounts of data (from traditional in situ sources as well as satellite sensors) are now being exchanged rapidly and across the globe by researchers and practitioners in many different fields. The growing interdependence among traditional scientific disciplines leads to the practice that data collected in one discipline is likely to be used in other disciplines. This leads to the greater need of sharing of data for the advancement of science.[218]

One of the main benefits from the large amount of data that has been created from EO sensors and many other sources has been developments in automated knowledge discovery. The ease of access to computational processing power, as well as better access to data, has led to the development of machine learning techniques. As identified by GFDRR, with any new and emerging technologies, there are many ambiguous and overlapping terminologies such as artificial intelligence, machine learning, big data and deep learning.[219] For this purpose, it is accepted that the terms are interchangeable.

Risk management is no exception to the use of machine learning, and there are new applications and uses continually being developed. Many of the uses of machine learning within disaster risk management focus on the improvement of the different components of risk modelling, such as exposure, vulnerability, hazard and risk.

Machine learning is moving beyond hard-coded algorithms to algorithms that continually learn and update themselves. This is facilitated by the development of methods where a machine may be instructed to seek information within large quantities of apparently unstructured data.[220] Although recent developments are delivering very powerful machine learning algorithms, it is important to remember that a model is only as good as the data used within it.

4.2

Conclusions

It is clear from recent developments that open data and analysis, shared and interoperable software, computing power and other technology, are the technical enablers of improved data science, risk assessment and risk modelling. For their success, they also rely on the willingness of people to work with other disciplines, across cultural, language and political boundaries, and to create the right regulatory environment for new and urgent work to proceed.

218 (Kunisawa 2006)
219 (GFDRR 2018b)
220 (UN-GGIM 2015)