

Special Topics

9 - Data Management throughout the National Risk Assessment Process

Key words: data management, data license, open data, metadata.

Data management is an important part of a national risk assessment and can help ensure that the process is both effective and impactful. Risk assessment is an extremely data-intensive process and conducting a national risk assessment may involve accessing information from a wide range of stakeholders including mapping agencies, scientific and technical ministries from across government, universities and other research institutions, and the private sector. In addition, valuable new data and analysis is created during risk assessments. It is therefore necessary to develop a strategy to efficiently organize and manage the data as it comes in as well as to distribute the results to participants and key stakeholders. Data management plans govern the process by which data is gathered from participating entities, the technical and quality standards to which new data will be produced, how data will be maintained during the risk assessment, and the means by which the output data will be shared or secured. This section provides a general introduction for non-specialists to some of the more important concepts involved in data management for national risk assessments.

Why Invest in Data Management for National Risk Assessments?

The datasets necessary for conducting risk assessment are valuable resources. They can be expensive to create and, when managed properly, will be used by a diverse set of users for multiple purposes beyond those for which they were initially produced. Conversely, if data is managed poorly, the investment made in creating it will not yield a full return. In addition, improper management or limiting access to data can lead to duplication of efforts, as other organizations are forced to recreate data that already exists. A well-crafted data management plan can help encourage stakeholders to share their data, and ensure that the processes for sharing data are effective and transparent. This will both increase the value of the investment in the data, as it will encourage use, but also build trust in the results of the risk assessment as more stakeholders have access to the raw data that underlie it.

Stakeholder Involvement and Accountability

Successful development and implementation of a data management strategy for national risk assessment requires involvement of stakeholders, comprising data producers and users, early on in the planning. This involvement will both help ensure that the data management activities meet the needs of participants as well as increase their sense of ownership of the process -- which is vital for successful implementation. Many government entities involved in national risk assessment will already be subject to legal regulation that control the conditions under which they produce, maintain, and share data. Identifying these constraints at the start of the process is also important. The various obligations that stakeholders have related to creating or sharing data for risk assessment purposes can be documented in

Memorandums of Understanding (MOUs), signed by each participating organization, in order to formalize the agreement.

Standard Data Formats

Data management plans should also specify the preferred formats in which datasets should be created, maintained, and shared. Generally speaking, these should always be standard data formats agreed upon by groups such as the Open Geospatial Consortium¹. This will help alleviate compatibility challenges that have in the past made it difficult for data created in one software to be used in another. Some examples include .csv for tabular data or .shp, .geojson, and .kml for spatial data. Other data standards, the resolution at which spatial data is recorded or the attributes associated with records in an asset database for example, should be considered during risk assessments where significant amounts of new data will be created.

Data Licenses

A central element of any data management strategy is clear articulation of the conditions under which data is shared. These conditions are specified by a data license or terms of use that should accompany each dataset. A well-written data license should cover, at minimum, attribution, modification, and redistribution. Attribution refers to citation of the owner of the data on products in which it is used. Modification governs the conditions that users of the data must comply with when altering the dataset or combining it with other data. Redistribution refers to the permissions that users have to redistribute the data or any derived works once they have accessed it, and whether it may be used for commercial purposes. In recent years, there has been increasing advocacy for adopting open data policies across government and academic research. Open data advocates argue that liberal, “open”, data licensing supports transparency, efficiency and participation in government, peer review of science, and more widespread and effective data use for decision-making overall. If the country has concerns over sensitive asset data, it is important not to lose sight of the potential value of releasing this information in aggregate form and making other components of the risk assessment, such as hazard data, openly available for further use by different public and private sector and academia. The Global Facility for Disaster Reduction and Recovery (GFDRR)’s Open Data for Resilience Initiative (OpenDRI) has been working on these issues related to disaster and climate risk assessment since 2011 [Box 1].

Metadata

The creation and maintenance of metadata is a necessary component of data management. Metadata, literally means “data about data” and provides information about how and when datasets were created, what their attributes signify, who the initial authors and owners were, and the terms of the data license. There are several well recognized standards for metadata, including those put out by the International Standards Organization² and United States Federal Geospatial Data Committee³. Many GIS software also includes tools for authoring and sharing metadata. Data management strategies should also include plans for storing, sharing, and updating metadata when necessary for every dataset they cover.

BOX 1 - Open Data: The Malawi Spatial Data Working Group

Since 2012, the Malawi Spatial Data Working Group has worked to share spatial data using the MASDAP platform - www.masdap.mw - in 2012. The group began as a partnership between government ministries and other organizations working on flood risk assessment in the Shire River Basin. Participants formed the Malawi Spatial Data Working Group to manage the activity and share necessary data during the project. The working group, which meets monthly, continued their efforts to gather and share data following the conclusion of the risk assessment and MASDAP is now a valuable source of risk information for the whole country. MASDAP received support from the Global Facility for Disaster Reduction and Recovery (GFDRR)'s Open Data for Resilience Initiative (OpenDRI). OpenDRI has partnered with national governments, universities, and community-based organizations to launch data sharing platforms like the Sri Lanka Disaster Risk Information Platform (<http://riskinfo.lk>), to support community mapping projects for disaster risk assessment (<http://www.opencitiesproject.org>), and to build tools to communicate complex risk information to diverse stakeholders (<http://inasafe.org>). More information about data available through OpenDRI projects can be found at: <https://opendri.org>

BOX 2 - Free & Open Source Tools for Data Management

There are a number of free and open source tools that have been used to support management and sharing of spatial and tabular data.

GeoNode - <http://www.geonode.org> - GeoNode is tool that allows users to share and visualize geospatial and tabular data on the internet. The software is free but it requires installation and customization. Metadata authoring tools are also included.

CKan - <http://ckan.org/> - CKan is another tool that acts as a full featured web-based data and metadata sharing platform.

QGIS - <http://www.qgis.org/> - QGIS is a desktop-based, GIS software that provides features for data editing, manipulation, and conversion. Free extensions can be used to automate some parts of metadata creation.

Recommendations

- Incorporate stakeholders from both potential contributors to and users of risk assessment data early in the planning process. Provide stakeholders an understanding of the importance and value of their data for quality of the risk assessment results. Give them an opportunity to make substantive contributions to the data management plan.
- Agree upon the data licensing, metadata standards, acceptable formats, and other protocols as early as possible.
- Whenever possible, release data under open licenses that encourage wide use for many purposes.
- Develop a common repository for data during the risk assessment that can also be used to share the results and outputs when the assessment is completed.

- Document the data sharing plan into an MoU or other formal agreement that can clarify the expectations and responsibilities of participating stakeholders

Resources for further information

- Field Guide to the Open Data for Resilience Initiative: https://www.gfdr.org/sites/gfdr/files/publication/opensdri_fg_web_20140629b_0.pdf
- Future Trends in Geospatial Information Management: The Five to Ten Year Vision: http://ggim.un.org/docs/UN-GGIM-Future-trends_Second%20edition.pdf
- A Guide to the Role of Standards in Geospatial Information Management: <http://ggim.un.org/docs/Standards%20Guide%20for%20UNGGIM%20-%20Final.pdf>
- Why Information Matters: A Foundation for Resilience: https://www.internews.org/sites/default/files/resources/150513-Internews_WhyInformationMatters.pdf

Authors: Robert Soden (GFDRR)

Contributors and Peer Reviewers: Peer Reviewer: Paul Henshaw (Global Earthquake Model Foundation), Keiko Saito (World Bank)

¹ Open Geospatial Consortium. 2017. Retrieved from site:

<http://www.opengeospatial.org/>

² ISO 19115-1 Geographic Information Metadata Part 1: Fundamentals. 2014. International Organization for Standardization

³ The Federal Geographic Data Committee. 2017. Content Standard for Digital Geospatial Metadata. Retrieved from site: <https://www.fgdc.gov/metadata/csdgm-standard>