3.4 OPEN DISASTER RISK REDUCTION DATA PLATFORM

June 2022

DRRPathways.ca

RESILIENCE PATHWAYS

CO-CREATING NEW KNOWLEDGE FOR UNDERSTANDING RISK AND RESILIENCE IN BC

This article is part of the Resilience Pathways Report. The report has the following objectives: a) to share knowledge about existing practices and recent advances in understanding and managing disaster and climate risk in BC, including some information on relevant federal programs, and b) to provide insights on gaps and recommendations that will help build pathways to resilience in BC.

This article belongs to *Chapter 3 Climate and Disaster Risk Management: Enabling Action*. To read all articles in the report, see DRRPathways.ca.

The Resilience Pathways Report is a project of Natural Resources Canada.

3.4 OPEN DISASTER RISK REDUCTION DATA PLATFORM

BY:

Joost van Ulden, Natural Resources Canada

William Chow, Natural Resources Canada

Drew Rotheram-Clarke, Natural Resources Canada

Damon Ulmi, Natural Resources Canada

Anthony Fok, Natural Resources Canada

Tiegan E. Hobbs, Natural Resources Canada and UBC

CONTRIBUTORS:

Phil LeSueur, Natural Resources Canada

EDITORS:

Sahar Safaie, Sage On Earth Consulting

Shana Johnstone, Uncover Editorial + Design

ABOUT THE OPENDRR PLATFORM

The Open Disaster Risk Reduction Platform (OpenDRR Platform) aims to provide tools to share hazard and risk data such that users can investigate, assess, and mitigate earthquake disasters. It is specifically aimed at policy makers, risk analysts, private and public institutions, and citizens, to facilitate decision making prior to a crisis. The platform is under development and is intended to be launched in the fall of 2022.

Developed by Natural Resources Canada (NRCan) to support with delivery of a national assessment of earthquake risks, the OpenDRR Platform is middleware between hazard or risk modelling platforms like OpenQuake and end users who need to understand and evaluate risk to make investment and policy decisions. The end-user interface will operate as a web application using standard web browsers in desktop, tablet, or hand-held device environments.

Development and execution of hazard and risk assessment models is a separate process, outside of the OpenDRR system. OpenDRR receives output from these models as input, using one or more interfaces and interchange formats based on existing standards or on specifications developed by the implementation team if no standards meet the requirements.

The high-level goals for OpenDRR are:

- Support open and collaborative development, using science outputs
- Provide centralized access to the science outputs
- Provide tools and applications to engage users, transfer information, and support decision making with respect to mitigating risk
- Improve the efficiency of disseminating risk assessments results

RISK MANAGEMENT

OpenDRR provides access to datasets that help improve understanding of the earthquake risks in Canada through a variety of industry accepted standards and best practices for geospatial data dissemination. The multichannel approach, which includes application programming interfaces (APIs), web applications, and dashboards, serves to reduce barriers to the reuse of project assets. By reducing the barriers for all stakeholders to access, explore, and visualize earthquake risk information, the platform ensures that timely access to authoritative information about earthquake risk can be utilized to build disaster resilience.

Emergency response planners will have at their disposal a suite of data products, as well as supporting web applications, that can be readily used in emergency planning. Two key elements of Open DRR are purpose-built to serve the community: 1) a dashboard that is highly customizable and allows individuals and organizations to create public or private spaces where they can query and intuitively visualize all available data; and 2) a purpose-built application, called RiskProfiler, that seeks to communicate the key messages relating to earthquake risks without any technical capacity on the part of the user.

> [OpenDRR's] multichannel approach . . . [reduces] the barriers for all stakeholders to access, explore, and visualize earthquake risk information . . . to build disaster resilience.

These elements will allow an emergency manager to obtain information about potential impacts from earthquake scenarios, such as anticipated demands on the healthcare system, disruption to housing, or financial impacts.

This information can be used to develop emergency response plans and training exercises. Emergency managers, for example, will be able to use the platform to build resilience into response plans by working with healthcare planners to expand hospital surge capacity. As well, OpenDRR will provide decision makers and the general public with comprehensive dynamic map visualization (showing earthquake scenarios) for all regions in Canada. This will give information, for example, on which construction types are most at risk and may therefore be in need of seismic retrofitting so as to prevent building collapse in the event of an earthquake.

The platform will support more efficient delivery of NRCan risk assessments over time while providing a place for ongoing contribution to NRCan's risk models, and it will address the current paucity of mechanisms by which to access seismic risk information. While existing applications like GeoBC's Common Operating Picture[1] provide situational awareness after an event occurs, OpenDRR will provide comprehensive, public, nationwide information about seismic risk that can be used for preparedness and mitigation.

PLATFORM DEVELOPMENT

Early stages of OpenDRR data platform development were part of the DRR Pathways project of NRCan, contributing to two objectives of the project: "Enhancing understanding of disaster risk" and "Strengthening risk governance through knowledge exchange and community engagement."

The initial requirement focused on earthquake risk, but with a desire to include other natural hazard types (e.g., landslides, wildfires, flooding, tsunami). The decision was made to focus on earthquake risk and bring in additional hazards when the platform was more mature. Earthquake risk data and information in OpenDRR does not include secondary perils like aftershocks, landslide, liquefaction, or tsunami. Data is provided for all populated regions of Canada.

The scale and scope of the data involved in the project necessitated an approach that not only streamlined the production of science-based outputs, but also provided for a high degree of collaboration across many disciplines (e.g., policy, technology, and science) and stakeholders (e.g., provincial and municipal).

Decision-support requirements were well understood at the outset. A comprehensive set of requirements for a multitude of stakeholders was developed.[2] The diversity of the stakeholders and their specific needs necessitated a multichannel approach since a single application (ex: only API) was deemed to be insufficient to serve all use cases effectively.

With development led by the Government of Canada, the platform had to comply with requirements for publishing science outputs, including standardized metadata, open data, support for both official languages, accessibility, and compliance with scientific integrity and publication policies. The scientific integrity and publishing requirements were particularly problematic as they have traditionally pushed the science and development behind closed doors. To comply with policy while supporting the objectives of the project, a balanced approach that prioritizes openness and transparency is needed.

The generation of science outputs is becoming increasingly reliant on software to automate processing, quality control, and publishing. Considerable attention was paid to alignment with best practices for Open Science.[3] As such, OpenDRR adopted FAIR Principles[4] (findability, accessibility, interoperability, and reuse of digital assets), which emphasize machine-actionability of data, and R5 Principles[5] (re-runnable (R1), repeatable (R2), reproducible (R3), reusable (R4), replicable (R5)), which describe ideal characteristics of software code that is released as a scientific output.

ALIGNMENT WITH THE SENDAI FRAMEWORK

OpenDRR supports Canada to achieve the first priority for action: understanding risk. Risk indicators provided on the platform also align with targets established by Sendai to support end users to develop resilience strategies that are aligned with this global framework. It allows practitioners to understand the current seismic risk facing Canadians and to explore ways in which that risk could be lessened, and by how much.

PLATFORM DESIGN

SOURCE DATA

Source data for this project includes the National Human Settlement Layer (physical exposure and social fabric), the National Seismic Risk Model for Canada (CanadaSRM, probabilistic), Canada's National Earthquake Scenario Catalogue (deterministic), and boundary geometries adapted from 2020 Statistics Canada, 2016 Census – Boundary files. The National Human Settlement layer includes a social vulnerability component that addresses the challenges posed to disadvantaged groups and help end users understand how they can create more equitable strategies to benefit the most vulnerable members of society.

TECHNOLOGY STACK

The OpenDRR technology stack (Figure 1) is made up of four main components: 1) Data Processing Pipeline, 2) GitHub, 3) Applications, and 4) Federal Geospatial Platform.

DATA PROCESSING PIPELINE

The OpenDRR data processing pipeline (Figure 2) is responsible for extracting, transforming, and loading data. It consists of several open-source technologies, namely PostgreSQL with PostGIS extension, and Python.

OPENDRR GITHUB

GitHub[6] was chosen as the platform (Figure 3) to support the development of the science outputs and related software, documentation, and tools. While well known in the software development community, it is relatively lesser utilized in the science community. However, the core functions of GitHub (e.g., versioning, repositories) were recognized to be beneficial by key contributors.

Where possible, runnable code is available to ensure transparency in the science. For example, an interested party could clone a repository and replicate a particular output, such as a dataset or even the entirety of the OpenDRR infrastructure. GitHub makes heavy use of containerization and infrastructure as code technologies for rapid deployment on personal computing devices or the cloud.

Built-in features of GitHub, such as continuous integration and deployment, community building, websites, peer review, and secure workspaces, are tools for achieving an open and collaborative approach to science, one that seeks to build consensus and drive engagement throughout its lifecycle.

Fortunately, GitHub provides many of the statistics that feed key performance indicators, such as visitor count, number and type of downloads, and number of followers, to name a few. These will help contributors and stakeholders measure impact, sentiment, and reuse of project assets over time.

GitHub also provides robust auditing, allowing project leads to follow up on contributions as required—for example, to determine what changes were made and by whom. This functionality helps to support integrity in the science carried out on the platform and allows users from outside of the core project team to comment or contribute safely.

APPLICATIONS

Due to the diversity of use cases and user profiles for the information products (e.g., maps, visualizations), it was clear that a single solution would not be sufficient. It was determined

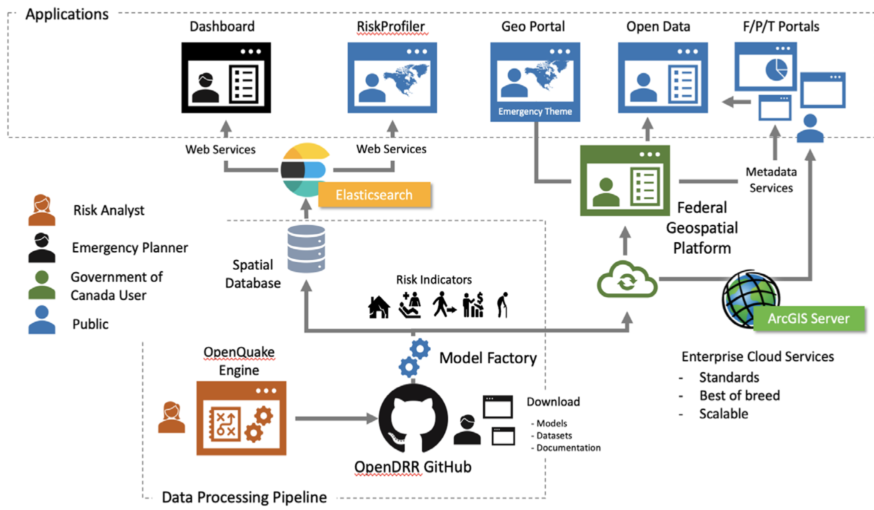

Figure 1: OpenDRR architecture overview.

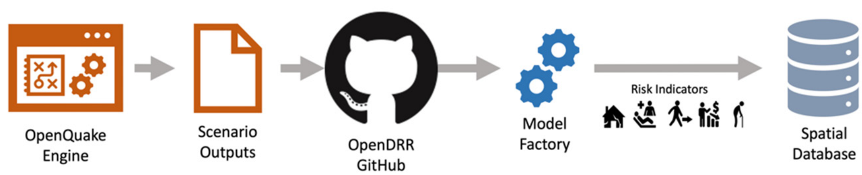

Figure 2: Data processing pipeline.

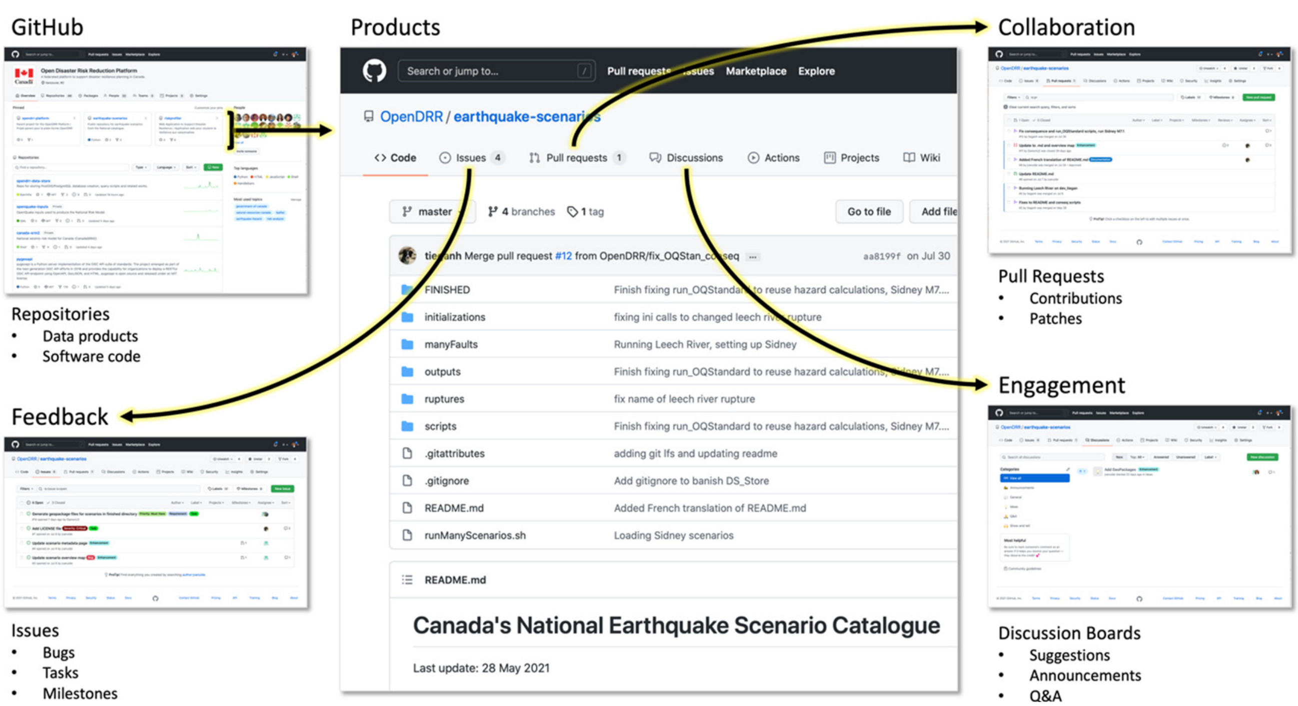

Figure 3: OpenDRR GitHub.

that a purpose-built web application (RiskProfiler) and a dashboard (Kibana) would be required to meet the needs of all users.

RiskProfiler Web Application

RiskProfiler (Figure 4) aims to provide planners and emergency managers with information on earthquake risks. This includes deterministic and probabilistic earthquake risk assessment results at a neighbourhood scale, across Canada. The scenarios are organized into a library that users can filter based on a variety of properties (e.g., location, magnitude). The library will grow over time to include more than one hundred deterministic scenarios covering the highest risk regions across Canada.

> RiskProfiler aims to provide planners and emergency managers with information on earthquake risks. This includes deterministic and probabilistic earthquake risk assessment results at a neighbourhood scale, across Canada. . . . Scenario properties can be adjusted to indicate how structural mitigation (retrofit) can affect loss.

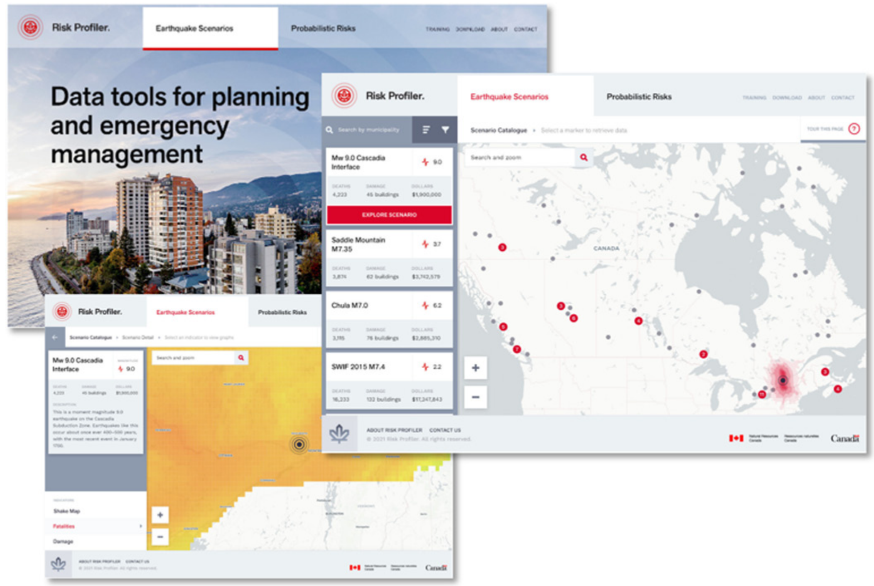

Figure 4: RiskProfiler.

The scenarios and associated risk information are intended to support emergency planning. Users are provided with a variety of visualizations that express the nature of the risk and the impacts that it may have on a community. These impacts are quantified, and in many cases, scenario properties can be adjusted to indicate how structural mitigation (retrofit) can affect loss. For example, the location and number of damaged buildings can be viewed for current conditions, or a toggle can be used to view the same metric if all buildings were brought up to modern design levels.

Kibana Dashboard Application

The Kibana Dashboard (Figure 5) is intended to support a more specific or sophisticated use case than that of RiskProfiler.

The dashboard allows users to customize workspaces where they can collaborate with others on data visualizations and reports. The entirety of available data in OpenDRR is available to dashboard users. The data is identical to what is available via OpenDRR's APIs. Users can add or link additional datasets to combine the DRR Pathway's project data with their own. For example, users could integrate risk assessment datasets provided through the OpenDRR Platform with their own linear infrastructure data to visualize the intersection between seismic risk to buildings and the existing road network.

Users can create reports, data visualizations, and maps for an area of interest. Visualizations can be exported and embedded in presentations and websites. Kibana can also be used as a platform to develop complex queries that can

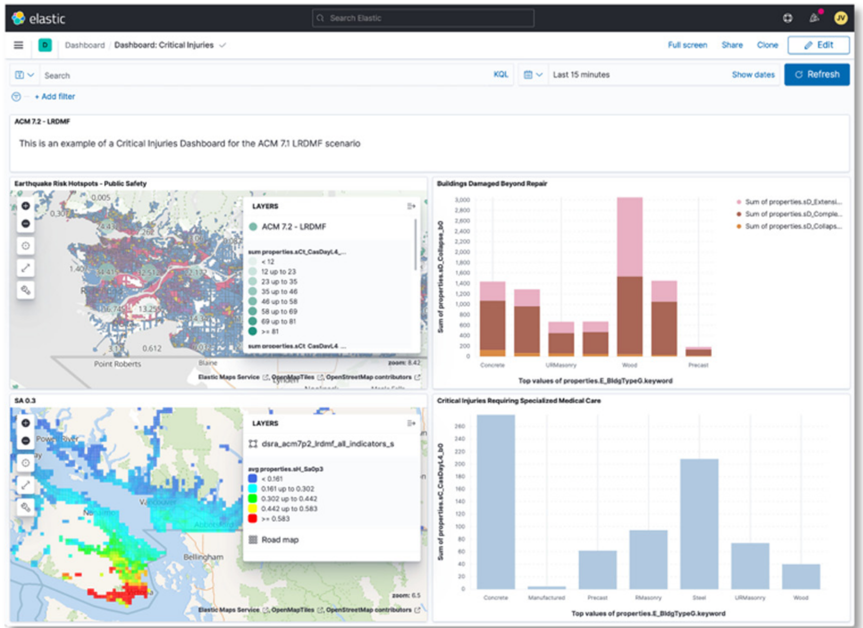

Figure 5: Kibana dashboard.

be sent directly to the Elasticsearch API. This is the first of such platforms to be made available to the public containing multi-scale risk data. It will be most useful to users who have one-off use cases, for example, those tasked with developing financial or insurance policy for a specific area.

FEDERAL GEOSPATIAL PLATFORM

The Federal Geospatial Platform (FGP) is a Government of Canada catalogue of geospatial data. The FGP provides enterprise-grade geospatial infrastructure and support services to facilitate the dissemination of data. OpenDRR datasets are hosted on the FGP and made available publicly via Esri REST services for organizations that utilize Esri-based tools and applications.

OPPORTUNITY

RECOMMENDATIONS

ADDITIONAL NATURAL HAZARD TYPES

Time will be needed to further build relationships with other federal partners who hold expertise in each of the other hazards. With this in mind, the decision was made to focus on earthquake risk and bring in additional hazards (e.g., landslides, wildfires, flooding, tsunami) when the platform is more mature.

There are several ways in which additional hazards can be added. They can be integrated fully or partially depending on the nature of the data and the capacity of the responsible party. For example, at the most basic level, a repository could be set up for each hazard type and the datasets, stored as release assets, could be easily propagated to the dissemination infrastructure by way of existing processes. A more advanced approach would be to generate the datasets from within the platform as it currently done for earthquakes; this would require a fair amount of scripting (e.g., Python, Shell) to achieve, but would not be impossible.

ENGAGEMENT

To engage the public more effectively and efficiently respond to queries about the science, the project will leverage the Discussions module in GitHub. The Discussions module can support FAQs, general discussions, feedback collection, or any other type of engagement. Outside of GitHub, the primary researcher would typically have to respond to queries on an ad hoc basis—a time-consuming but necessary task. The Discussions module could reduce the level of effort to support science hosted in repositories.

Other opportunities to engage with the user community will be explored as time and resources permit.

CHALLENGES

WORKING IN THE OPEN

Working in the open is standard practice in the software community, but this is not the case in the scientific community. Despite many science-based institutions promulgating FAIR and Open Science, they struggle to fully adapt to the very principles upon which these are based. Instead of working in the open from the outset, science continues to be carried out behind closed doors until such time that a final product is formally published. In part, this is a natural outcome of the way that the scientific community is structured—it is undesirable from a scientist's perspective to release to the public results which have not been reviewed. This is doubly true in the field of natural hazard and risk research, where outcomes may have direct tangible effects on the assets or safety of community members.

The science-based organization under which the OpenDRR platform is being developed decided to take a more liberal approach, open by default but closed where required. Internal policies regarding Open Science were not yet fully developed, therefore the approach was to do much of the science modelling and development in private repositories until peer review could be completed. It is expected that such work will be carried out in the open once Government of Canada policies and practices around Open Science are mature.

While not ideal, the OpenDRR did demonstrate that peer-review of science using GitHub was tractable. Transparency in the science that informs government policy is an important part of any democracy, and so the platform will continue to aspire to a future where policy-driven decisions are supported by data that is aligned with the principles of FAIR and Open Science.

CONTINUOUS INTEGRATION

It was readily apparent early in the development process that automation would be beneficial. In software development, continuous integration is a technology that integrates subsystems into larger systems on some predetermined event (e.g., tests have completed successfully for an update such as a bug fix). In the case of OpenDRR, continuous integration is used for software integration and data integration.

When new datasets are added and/or models are updated, automated tasks are run to deploy new services, downloadable assets, and metadata. This saves a significant amount of effort and reduces the amount of time it takes to make these assets available to the community.

In the case of OpenDRR software code, continuous integration scripts are used to prepare and publish containerized solutions, generate database scripts, generate configuration files, and run tests. Deploying systems into the cloud via continuous integration is in active development and is expected to further reduce the level of effort required to deploy the software stack.

SOFTWARE DEVELOPMENT LIFECYCLE

The intersection of science delivery and software development has traditionally been carried out independent of each other; that is, the science is completed and then the system is developed. With a very ambitious delivery schedule, it was decided to do both in parallel.

The Pathways OpenDRR development utilized the scrum process with two-week iterations. Tasks were assigned to each iteration and reported on every two weeks. The complexity of the products being developed necessitated a tight coupling of the raw science outputs with custom software and continuous integration processes.

This tight coupling presented several challenges. It was immediately apparent that a high degree of flexibility would need to be designed into the software to accommodate constantly changing data schemas. As well, the development of the science outputs moved at a much slower pace as it required collaboration with other scientists and peer reviews. The two-week iteration cycle resulted in too much overhead on the development of the science outputs, and therefore engagement with the science staff suffered.

GITHUB

The use of GitHub as a platform for the development of the science outputs was well received and uptake was high. The core concepts of GitHub were well understood, and the distributed nature of the platform proved to be a benefit during the COVID-19 pandemic, which saw most project participants working remotely, often disconnected from the enterprise network. Fine-grain control over access to repositories and associated assets was a critical factor in the success of the platform to support Open Science.

The 100-megabyte file size restriction imposed by the GitHub platform was an issue for some repositories. Thankfully, GitHub provides an alternative storage called "Git Large File Storage" (Git LFS) which was enabled on many repositories. Bandwidth quotas for Git LFS were exceeded. GitHub provides 1 GB of free storage and 1GB per month of free bandwidth. Additional costs were incurred to increase the quotas.

To mitigate the potential of increasing costs for managing large files, a strategy of including datasets and files in the release assets of a repository was adopted. GitHub allows for release assets (e.g., files, datasets) up to 2 GB to be stored and disseminated at no cost.

RESOURCES OR SIMILAR PROJECTS

As part of the requirements gathering exercise for the OpenDRR Platform a review of National and International risk portals, technologies and tools were reviewed and documented.[7] Descriptions and links are provided and can be found in Section 5.2 of *A Federated OpenDRR Platform to Support Disaster Resilience Planning in Canada: High Level Requirements – Risk Management Platforms*.

ENDNOTES

[1] Government of British Columbia "GeoBC's Common Operating Picture" last accessed March 18, 2022, https://bcgov03.maps.arcgis.com/apps/MapSeries/index.html?appid=11821451d60a49168d1f7602d379abf5 (requires password).

[2] Government of Canada, *A Federated OpenDRR Platform to Support Disaster Resilience Planning in Canada: High Level Requirements – Risk Management Platforms*, last modified December 19, 2019, https://opendrr.github.io/documentation/docs/opendrr-platform.html

[3] Government of Canada, "Open Science," last accessed March 18, 2022, https://science.gc.ca/eic/site/063.nsf/eng/h_98054.html

[4] Go Fair, "FAIR Principles," last accessed March 18, 2022, https://www.go-fair.org/fair-principles/

[5] Digitalization and Open Science, "R5 Principles," last accessed March 18, 2022, https://like-itn-digitalization.readthedocs.io/en/latest/4_R5/

[6] See GitHub at https://github.com/OpenDRR

[7] Government of Canada, *A Federated OpenDRR Platform to Support Disaster Resilience Planning in Canada: High Level Requirements – Risk Management Platforms*, last modified December 19, 2019, https://opendrr.github.io/documentation/docs/opendrr-platform.html#5-2-risk-management-platforms

Recommended citation
Van Ulden,J., Chow, W., Rotheram-Clarke D., Fok, A., Hobbs,T., Open Disaster Risk Reduction data; in Resilient Pathways Report: Co-creating new Knowledge for Understanding Risk and Resilience in BC; Safaie, S., Johnstone, S., Hastings, N.L., eds., Geological Survey of Canada, Open File 8910,2022 p. 290-298, https://doi.org/10.4095/330541