RESILIENCE & DATA IN NEW ZEALAND: THE DATA INTEGRATION AND VISUALISATION EN MASSE (DIVE) PLATFORM 2016 SUMMARY

Joanne R. Stevenson, Ph.D. & John Vargo, Ph.D.[1]

Chris Thompson, Ph.D.[2] & Lucy-Jane Walsh[2]

Resilient Organisations Research Report 2017/01

April 2017

Resilient Organisations Ltd.[1]

University of Canterbury Digital Lab[2]

Summary

In 2016, the NZ Centre for Earthquake Resilience (QuakeCoRE) and the Resilience to Nature's Challenges (RNC) – National Science Challenge, funded a small team of researchers from Resilient Organisations Ltd. in collaboration with UC CEISMIC (the Canterbury Earthquake Digital Archive) to investigate how to best enable teams of researchers to address complex social problems that will make New Zealand more resilient. The focus of this programme was to identify the types of data QuakeCoRE and RNC research teams would be using, how they planned to analyse and share that data, and how data management practices could enhance the impact of these research programmes.

Between March and November 2016, the team initiated a consultation process involving a series of workshops, surveys, interviews, and software prototype design and testing.

The outcomes of this consultation process resulted in several key outcomes:

1. The identification and classification of data types that researchers will be using.
2. The identification of critical data needs for researchers, including:
    a. Systems for knowing about ongoing research (before publication).
    b. Enhanced searchability of data across institutions.
    c. Systems that make sharing research data safe, easy, and desirable.
    d. Establishing standards and guidance for transdisciplinary data management in a way that facilitates data integration, analysis, and visualisation.
    e. Enhancing access to public, proprietary, and sensitive data sources.
    f. Streamlining and clarifying data sharing agreements for datasets that have significant reuse value or to which researchers will add value.
    g. The ability to track data reuse.

The consultation process also involved:

3. Evaluating pre-existing systems that can meet some of the immediate needs of resilience researchers in this space including DesignSafe, the New Zealand Geotechnical Database, and EERI Clearinghouse System.
4. Fostering relationships between key data providing organisations and researchers.
5. Identifying human and institutional factors that inhibit the success of such boundary pushing, transdisciplinary, and cross-institutional research programmes.

Two final outcomes moved the consultation process into the design phase for a system that can begin to meet the needs of resilience researchers and practitioners:

6. The development of several software use cases to guide the development of future data sharing systems.
7. The creation of a working prototype data federation portal system, which we are calling the Data Integration and Visualisation En Masse (DIVE) Platform.

The consultation process made it clear there are systems that can meet some of the needs of those working to improve resilience in New Zealand. There are, however, still significant unmet needs that will hinder the progress of truly trans-disciplinary and transformative research.

Outcomes would be enhanced by a system that is problem-focused, rather than divided by funding or disciplinary boundaries. Such a problem-focused system will enhance the visibility of the work going on to improve the resilience of New Zealand. It will be a place where communities of researchers, decision makers, data holders, private industry, and citizen scientists can view, upload, and download data. Such a system should facilitate the creative collision of secondary and primary data, local narratives, real-time hazard monitoring, Mātauranga Māori knowledge, and multi-media information.

We propose the continued development of DIVE into an interactive online space for researchers and practitioners to organise and communicate information relevant to their *ongoing research* and *information gathered from disaster events as they unfold*. Capturing this data in a federation portal that is curated, properly archived, and strategically shared will facilitate future research, aid response and recovery actions and decision making, and may become a resilience building tool as broader communities are able to contribute data on the hazards they are experiencing or the trends they are seeing in their communities.

Acknowledgements

This work was funded by the Tertiary Education Commission through the Centre of Earthquake Resilience (QuakeCoRE) and by the Ministry of Business Innovation and Employment through the Resilience to Nature's Challenges National Science Challenge. The authors would like to extend sincere thanks to everyone who gave up their time to attend workshops, fill out surveys, and share their expertise throughout the development of the DIVE Platform. We would also like to thank Toni Wi and Louise Home-Dewar for their help in the preparation of this report.

Table of Contents

Summary ii

1. Introduction 1

2. DIVE Development Process 2

3. Findings and Outputs 5

3.1 Processes 6
3.2 Institutional structures 6
3.3 Software 7
3. 4 Stakeholders 7
3.5 Observations about the Data Development Context 7

4. Conclusions 9

References 10

Appendix A: The DIVE Platform Interface 11

Appendix B: Use Case Summaries from Workshop 3 13

Appendix C: Affiliated researchers round table presentations 18

From Workshop 3 18

1. Introduction

New Zealand is exposed to a wide range of natural disasters, in no small part because New Zealand straddles the boundary between two tectonic plates with its attendant risk of earthquakes and volcanic eruptions. The New Zealand Government has committed across several platforms to building the resilience of its people, places, and economy to ensure safety, stability, and prosperity in the face of significant exposure to disruption.

Understanding the current state of the nation's resilience and achieving systemic improvements requires cross-institutional and transdisciplinary collaboration and research innovation. Such models, however, present challenges to the status quo of data sharing and management. Information inefficiencies and gaps hinder the progress of those tackling New Zealand's most complex and important issues. The success of these research programmes requires a new knowledge development ethos. Meaningful collaboration across institutions and disciplines requires effective information management. This means creating spaces where data can be captured, safely shared, and managed to ensure quality, appropriate use, and ongoing development.

In 2016, the NZ Centre for Earthquake Resilience (QuakeCoRE) and the Resilience to Nature's Challenges – National Science Challenge, funded a small team of researchers from Resilient Organisations Ltd. in collaboration with UC CEISMIC (the Canterbury Earthquake Digital Archive) to investigate how to best enable teams of researchers to address complex social problems that will make New Zealand more resilient.

Using a design-thinking approach the research team developed the Data Integration and Visualisation En mass (DIVE) platform. Although it is still in the early phases of development, the DIVE platform encompasses data management processes, institutional structures, web-based software, and stakeholders (i.e., the people and organisations that will use DIVE). The aim of the DIVE platform is to enable teams of researchers, decision makers, and practitioners to address complex problems by:

1. Making data sharing safe, easy, and desirable;
2. Establishing standards and guidance for transdisciplinary data management;
3. Enhancing access to public, proprietary, and sensitive data sources; and
4. Facilitating data integration, analysis, and visualisation.

This report details the work conducted throughout 2016 to develop the DIVE platform. We begin by describing the methods used to develop the DIVE platform, followed by a summary of findings and outputs of the stakeholder consultation and initial development processes. This section features the current capabilities of the DIVE platform in its initial prototype form. The report concludes with an overview of the benefits and opportunities provided by the platform, and an outlook for future development.

2. DIVE Development Process

The DIVE development team used a design-thinking approach to scope and design the first DIVE prototype. Design thinking is a solution-focused process which incorporates the in-depth insights of end-users into iterative prototype development (Brown & Wyatt, 2010). Descriptions of design thinking refer to it as a system of overlapping spaces, as opposed to sequential steps. Brown and Wyatt (2010) label these spaces: inspiration, ideation, and implementation, where *inspiration* is, "the problem or opportunity that motivates the search for solutions" (p.33); *ideation* is "the process of generating, developing, and testing ideas" (p.33); and *implementation* is, "the path that leads from the project stage into people's lives" (p.33).

The inspiration phase typically begins with a 'brief', which provides a general framework of constraints and goals of the design process and benchmarks against which progress can be measured. For DIVE the brief was to create data management systems that enable teams of researchers to address complex social problems that make New Zealand more resilient to hazards and disasters. The standard against which this system was benchmarked is whether it is useful, usable, and used. These principles were loosely defined in the early phase of the project and can be guided by a series of questions based on an evolving understanding of the system (See Box 1).

Box 1. Principles for evaluating the success of DIVE.

Useful: Is the data up-to-date? Is the quality of the data being managed? Can others understand and use the data that is being uploaded?

Useable: Is data searchable and accessible for a wide range of users? For example, can it facilitate 'citizen science' or council data collection efforts if that is what the users need?

Used: Are communities of practice being established and self-sustaining?

Once the brief is set, the 'inspiration' phase is focused on exploring the needs of stakeholders through direct consultation and observation. QuakeCoRE and RNC researchers, along with key data providers such as government ministries, local councils, and the Earthquake Commission, comprised the initial stakeholder group whose needs we assessed. We began this process with a workshop and an assessment of the way stakeholders are interacting with other data management systems (Figure 1). These processes were supplemented by an online survey (the results of which are summarised in Stevenson, Brown, & Vargo 2016) and several informal interviews with subject matter experts to gather more in-depth information about user-needs and processes that may be useful.

The second space of design thinking is 'ideation', where insights generated in the inspiration phase are synthesised. The most salient observations are identified and translated into visions and choices that guide the design of the system (Brown & Wyatt, 2010). In the first iteration of the ideation space we synthesised the discussion at Workshop 1, the stakeholder survey, and expert interviews, along with observations of how researchers are interacting with extant systems (i.e., DesignSafe, the New Zealand Geotechnical Database, and the Australian Urban Research Infrastructure Network) into a report and a series of use cases.

Operational use cases describe a flow of operations for those interacting with a system, and can be used to identify the functions, operating systems, boundaries, and constraints that are relevant to potential users (Summers, 2012). The use cases focused on important data-related challenges facing QuakeCoRE and Resilience to Nature's Challenges National Science Challenge researchers.

The cases synthesize concepts that emerged from the initial inspiration phase, including:

- How sharing primary datasets (i.e., generated by original research for a given purpose) can inspire novel secondary uses in a way that enhances the value of the original work.
- The advantages of a federated search engine and the ability to view data that is not stored locally.
- The complexity of datasets with confidential elements and the need for streamlined processes for sharing and accessing sensitive datasets.
- The processes and systems needed to standardise and ease the collection of building inventory data, enhance building data integration, and facilitate data sharing & visibility.
- The need to capture ongoing and planned research to reduce overlap, reduce the burden on data providers and the subjects of social research, and to improve opportunities for collaboration.

More on these use cases can be found in Appendix B.

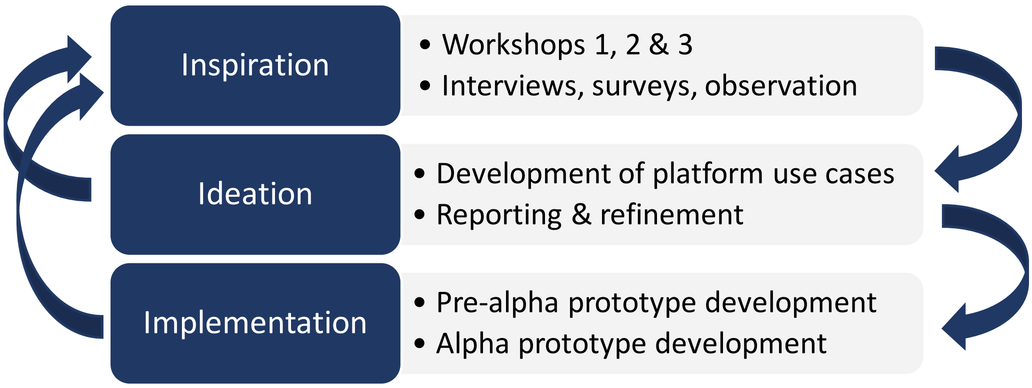


Figure 1. Design Thinking Methodology for the DIVE Platform showing the progression of phases from inspiration to ideation to implementation, and the associated stages of development.

In Workshop 2 we presented some of these synthesized findings and examined workflow and assimilation processes and problem solving in extant systems seeking feedback from workshop participants. The outcomes from this workshop were again synthesized in a report. The subjects discussed and outcomes resulting from Workshops 1 and 2 can be found in Stevenson et al. (2016) and Stevenson and Vargo (2016), respectively.

After Workshop 2 we entered the 'implementation' phase – developing the pre-alpha prototype[1] of the DIVE software system. This represents the first iteration in a multi-prototype roll-out. The proposed DIVE versioning programme (Figure 2) shows the current status of the DIVE Platform and estimates development progress over the next few years.

The development team re-entered the inspiration phase of the design process in Workshop 3. This workshop was held in November 2016 at the University of Canterbury, and included an overview

[1] 'Pre-alpha' refers to all software development activities before formal testing.

and reflection on the DIVE Platform development process over the previous year, a demonstration of the DIVE prototype's current features, and a discussion about how the DIVE Platform may be used to address key issues in resilience research. This was followed by an Affiliated Researchers Round Table, which included a series of 9 short presentations (summaries of these presentations can be found in Appendix C), and a robust group discussion.

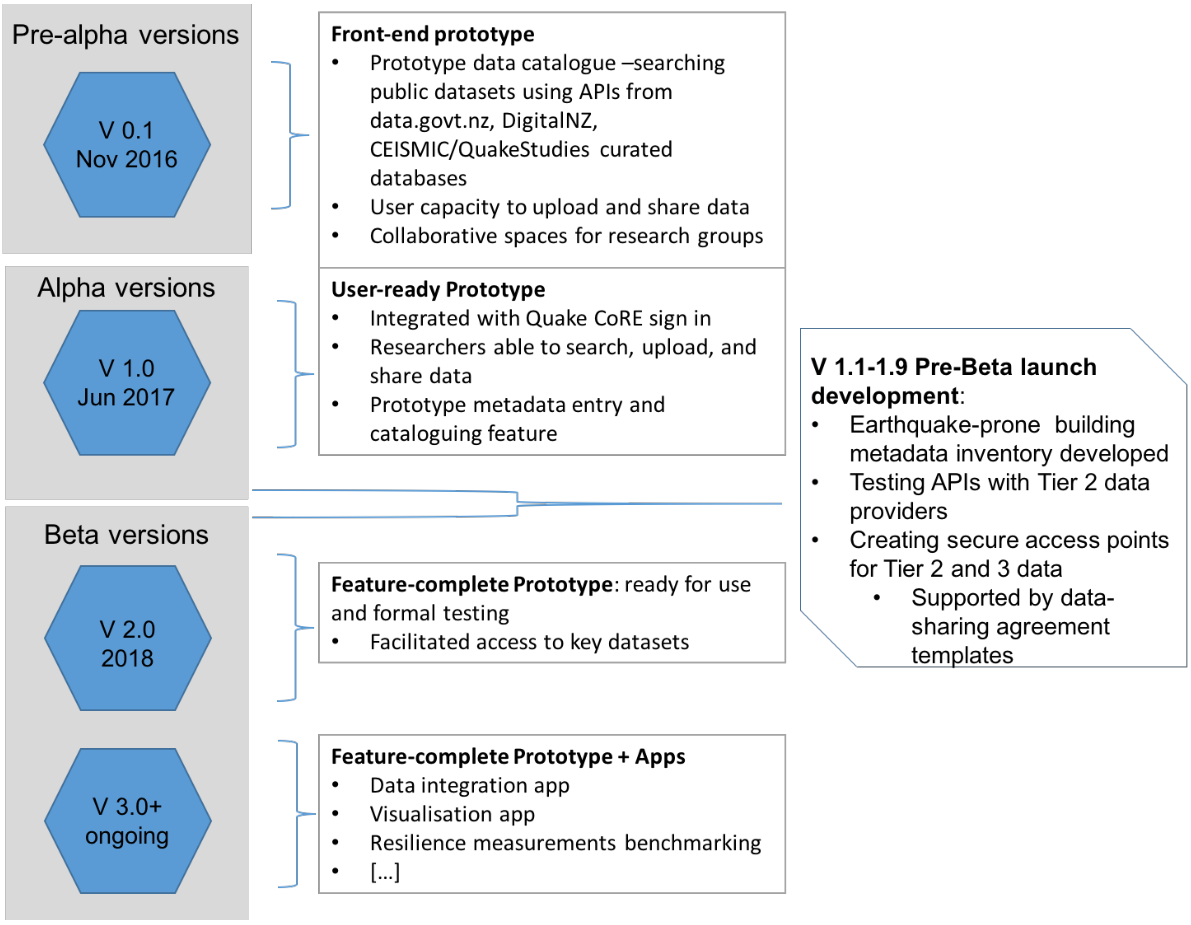


Figure 2. DIVE Versioning Programme as of November 2016.

Throughout the initial development phase, approximately 65 stakeholders and subject matter experts contributed to the DIVE development process, engaging in workshops (Figure 3), surveys, and interviews. The workshops, surveys, and interviews helped the development team outline critical processes for successful initiation of the DIVE Platform and provided important input to the initial design of the DIVE software pre-alpha prototype. The workshops were also a helpful starting point for forming the relationships that will be central to the successful implementation of a transdisciplinary and cross-institutional collaborative platform.

We will continue this iterative design-thinking process to ensure that DIVE becomes a system that is useful, useable, and used by the those working to solve New Zealand's resilience challenges.

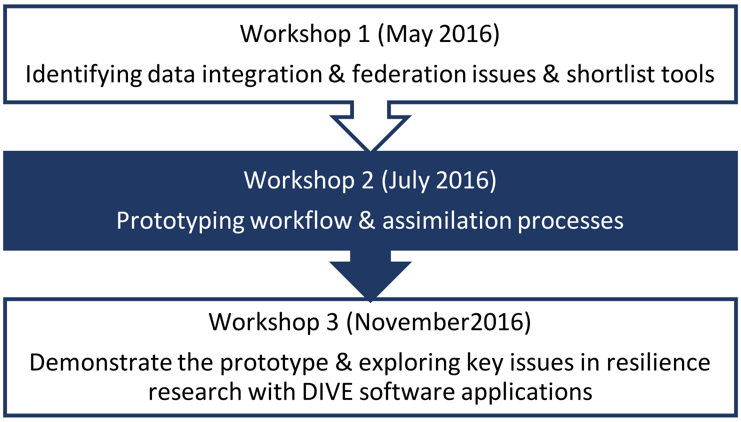


Figure 3. Titles of DIVE workshops held in 2016.

3. Findings and Outputs

The DIVE programme is constructed around a brief to develop a data management system for researchers working toward enhancing resilience in New Zealand. It became clear early in the consultation phase that there is justified scepticism among researchers across the National Science Challenges and QuakeCoRE about the potential efficacy of a new data management platform. Resistance is associated with three primary issues:

1) There is already a significant reporting burden on researchers. Learning a new system or meeting data sharing expectations represent additional time costs to researchers.
2) Building a system around a relatively short-lived research programme (which includes both the National Science Challenges and QuakeCoRE) presents the risk of future data loss or a system that is not maintained and becomes obsolete.
3) As there are already large-scale data integration schemes and data repositories in New Zealand and abroad, there is a risk of creating redundancy, inefficiencies, and multiple copies of data.

While these issues do not obviate the need for data management systems, they serve as important guides for ongoing development. It was clear from the design and consultation process that existing management systems are not currently meeting the needs of researchers engaged in the Resilience to Nature's Challenges and QuakeCoRE research programs. Problems with the way data is currently being managed include:

- Researchers and stakeholders not being aware of ongoing work or datasets that have not yet been published;
- Researchers having no central portal for federated data sharing;
- Researchers having different understandings of key concepts, and therefore different systems for recording and communicating data;
- Having no agreed system for managing and communicating the quality or completeness of datasets;
- Poor or inconsistent management of geographic data; and

- Lack of clarity around sharing data for restricted access datasets (i.e., Tier 2 and Tier 3 data as in Table 2).

Table 2. Tiers of access for key disaster resilience related datasets.

Tier 1	Tier 2	Tier 3
Some key datasets (i.e. tax statistics from the Inland Revenue Department and Census data) are publicly available and will be federated by DIVE through the data.govt.nz and DigitalNZ API's.	Other key datasets have restricted access and will require further negotiation to set up streamlined access through DIVE – for example, items held in the NZ Geotechnical Database, EQC claims data, and the ACC earthquake related-injury database. Also, included in this category are files created by QuakeCoRE and RNC researchers such as the Economics of Resilient Infrastructure business survey data.	The most difficult key datasets include proprietary data held by commercial providers, including Paymark and Marketview retail transactions, cell phone usage, and private insurance claims. Currently this data has been obtained ad hoc by some researchers – a possible solution to explore is negotiating more inclusive data sharing agreements through the DIVE platform with holders of proprietary data.

With these findings in mind we are taking a multifaceted approach to the design of DIVE going forward. The DIVE Platform includes the processes, institutional structures, software, and stakeholders using the platform.

3.1 Processes

Processes are the foundation of a useful data management platform. Processes include the standards that developers and users espouse to ensure that data is captured and shared in ways that allow it to be understood and used by others. Processes also refer to the way data and data management systems are maintained. Such processes include:

- Identifying potential users and assessing their needs (which we have done as part of the DIVE 2016 workshop series);
- Establishing ongoing monitoring of user needs and prioritising areas for investment;
- Building and curating relationships with data providers (e.g., EQC, LINZ, Crown Research Institutes);
- Evaluating and creating procedures for addressing research ethics and security issues;
- Creating and maintaining metadata standards, data dictionaries, data catalogues, and other related systems that enhance data searchability and integration.

These and other processes will be instrumental to ensuring that the DIVE Platform achieves its goal of *being useful, usable, and used*.

3.2 Institutional structures

Institutional structures refer to the long-term governance and resourcing of DIVE. Ensuring the ongoing development and maintenance of data required to improve New Zealand's resilience –

including unique datasets created by QuakeCoRE, Resilience to Nature's Challenges, and others – needs to be part of the Platform development. Identifying appropriate institutional structure and plans is a development priority going forward.

3.3 Software

The centrepiece of the DIVE Platform is the web-based software that provides a digital portal for federating data, search engine, and data management and analysis applications. The DIVE (pre-Alpha) prototype has been developed using CKAN, an open source data platform which provides tools to streamline publishing, sharing, finding, and using data. At its base install, CKAN acts as a data federation portal and catalogue, as well as having several built-in tools for exploring and understanding data.

Currently, the platform allows users to upload and explore data with plans to add to the *integrate* and *analyse* functionalities later. The platform also comes with many inbuilt ways of visualising data, such as spreadsheets which can be viewed as raw data, as a graph, or a map if coordinates are included. Images can be displayed and HTML and PDF pages can be viewed and interacted with on the site.

The DIVE prototype also includes an API which allows data to be pulled out or pushed into the platform. The API will allow researchers to integrate some of the tools they are already using or tools created specifically for the DIVE website.

For further examples of the DIVE platform interface including search capabilities and data visualisations see Appendix A.

3. 4 Stakeholders

The drivers behind the DIVE Platform are the stakeholders. Stakeholders refer to data providers and data users, though many stakeholders will be both. Stakeholders guide the design and development of DIVE. They will test and refine prototype iterations, and through their investment of time and information will determine the success of the platform.

The stakeholders for the initial prototype of the DIVE Platform are people involved in QuakeCoRE, the Resilience to Nature's Challenges – National Science Challenge, CEISMIC, and several government and private research organisations that were represented at DIVE workshops throughout 2016. This body of stakeholders will continue to grow as the DIVE web interface becomes more widely available.

3.5 Observations about the Data Development Context

The design thinking process associated with the development of DIVE lends itself to an iterative creative process. The current DIVE software prototype provides a basic design for meeting researchers' immediate needs and indicates pathways for more advanced processes going forward.

Figure 4 shows the hierarchy of data management needs that we identified through stakeholder consultation. The most basic and foundational need for stakeholders working to enhance resilience in New Zealand is ensuring that data is captured and preserved. This requires a portal for hosting and federating data and, more importantly, appropriate integration of research data management and ethics considerations into technical systems. The New Zealand Government Open Access Licensing (NZGOAL) Framework provides a foundation for achieving this, but work is still required within the research community to make robust mechanisms for the re-use of research data part of mainstream practice.

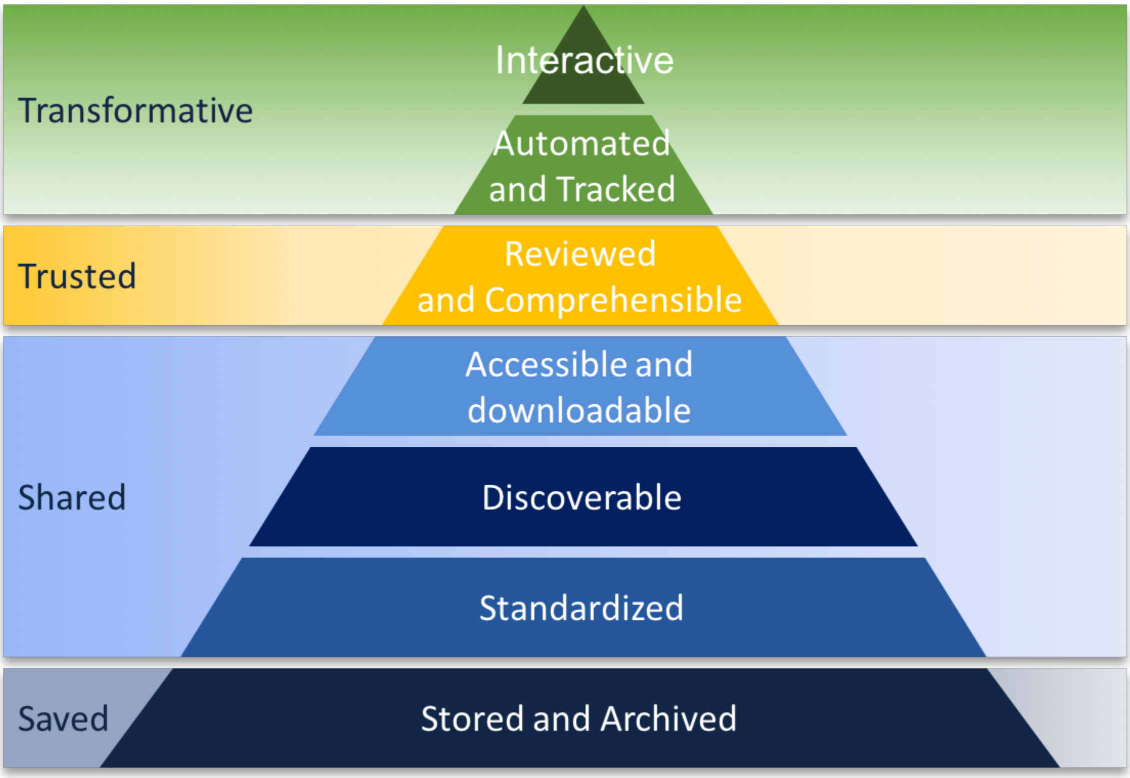


Figure 4. Hierarchy of data management needs as identified through stakeholder consultation

Some datasets will only need short-term storage. In most instances, this will be managed by researchers who intend to terminate all data after the project is complete and the results are published.

Datasets with re-use value, however, should be archived, and where possible made available for future use. Researchers will design their own ad hoc systems for short-term storage (e.g., internal or external hard drives). Storage that extends beyond the life of the research programme (i.e., archiving) needs to be saved differently, ideally in a trusted repository managed by a stable institution.

It will be valuable for the DIVE Platform processes to include guidelines on the design of research projects so that researchers can make up-front decisions about how their research data might be reused, and understand the trade-offs involved in either waiving confidentiality or removing personal or identifying information from the shared dataset. The data that is shared for such projects would likely be subject to stricter controls to ensure no personal information is shared in error.

The current DIVE interface can assist in applying relevant metadata to records and can enhance and transform pre-existing information when it is ingested via the CKAN API. As part of the data ingestion process, the DIVE interface can use validation and required fields to ensure data quality.

The next layers of data management on the hierarchy are systems that make data shareable. When data is standardised it means that it is consistently described and recorded using a set of accepted rules (i.e., metadata standards) to systematise the format as well as the meaning (U.S Geological

Survey, 2017). Metadata standards can also be developed to improve data discovery, including guiding researchers to:

- Include terms found in published taxonomies in their metadata,
- Write descriptive titles and abstracts, and
- Provide the direct URL to online resources in metadata.

The next layer, 'Trusted' refers to the way others perceive the quality, accuracy, and precision of the data being shared in a repository. Metadata standards and tools that facilitate consistent record keeping (e.g., electronic lab notebooks) can enhance the quality of data, and make it more likely that it is properly structured and adequately annotated, and therefore comprehensible.

Trust is built in the quality of data when it is accessed through a portal where information is reviewed and curated (i.e., organized, described, cleaned, enhanced, and preserved) for re-use. A data manager for such a system may review frequencies or produce summary statistics for missing values in datasets, review data for confidentiality issues, and perform other checks that increase the consistency, completeness, and usability of the data.

The last level, transformative data management, refers to data management systems that become self-sustaining, lead to improved research, and enhance outcomes for communities. For example, systems can allow published data to be tracked to understand how it's being used to answer new questions. Certain aspects of data capture can be automated and systems can allow 'citizen scientists' to input their own data.

4. Conclusions

This report details the development of the Data Integration and Visualization En mass (DIVE) platform. The aim of the platform is to enable stakeholders to address complex problems by streamlining data sharing, enhancing access to sensitive data sources, and facilitating data integration, analysis, and visualization. The DIVE platform offers many potential opportunities for researchers, such as a greater opportunity to collaborate with an open-access data source and easier access to data for faster problem solving.

Using a design thinking approach with stakeholder consultation and workshops an initial prototype was created allowing data to be uploaded and explored on a single platform. However, there are a number of challenges associated with data sharing such as standardizing data from diverse sources, data governance, data interoperability, and creating organizational frameworks that must be tackled alongside future iterations of the prototype as the platform matures (Medyckyj-Scott et al., 2016).

Rather than dealing with data issues reactively and in an ad hoc fashion, an integrated platform with challenges acknowledged up-front and as part of the design can help to deal with issues as they arise. Ultimately, the value of the DIVE platform is evident as a proof-of-concept, whereby a software-based platform was created with the purpose of facilitating data processes to support a problem-based field of enquiry. Enhancing resilience in New Zealand is critical to investing in the wellbeing of communities, organizations, and individuals. Using data more effectively is a key step in achieving this goal.

References

Medyckyj-Scott, D., Stock, K., Gibb, R., Gahegan, M., Dzierzon, H., Schmidt, J., & Collins, A. (2016). *Our Land and Water National Science Challenge: A Data Ecosystem for Land and Water Data to Achieve the Challenge Mission.* Landcare Research Contract Report.

Stevenson, J. R., & Vargo, J. (2016a). *Data Integration and Visualisation Platform: Using the DIVE Prototype to Answer Some Questions and Raise Others*. [Workshop Presentation]. Data Integration and Visualisation En masse Workshop: 29 November 2016. Christchurch, New Zealand. Retrieved from: https://wiki.canterbury.ac.nz/pages/viewpage.action?pageId=52692301

Stevenson, J. R., Vargo, J., & Brown, C. (2016). *Data Integration and Visualisation: Prototyping the QuakeCoRE Data Platform for Diverse Needs.* Resilient Organisations/QuakeCoRE Research Report. Retrieved from: https://wiki.canterbury.ac.nz/pages/viewpage.action?pageId=52692301

Stevenson, J. R., & Vargo, J. (2016b). *Data Integration and Visualisation En Masse (DIVE) Platform Report: Programme Update September 2016.* Resilient Organisations/QuakeCoRE Research Report. Retrieved from: https://wiki.canterbury.ac.nz/pages/viewpage.action?pageId=52692301

Summers, B. L. (2012). *Effective methods for software and systems integration*. CRC Press.

U.S Geological Survey. (2017). *USGS Data Management*. United States Geological Survey, U.S. Department of the Interior. Retrieved February 15th, 2017, from: https://www2.usgs.gov/datamanagement/plan/datastandards.php

Watt, J., & Brown, T. (2010). Design thinking for social innovation. *Stanford Social Innovation Review.* Retrieved June 20th, 2013.

Appendix A: The DIVE Platform Interface

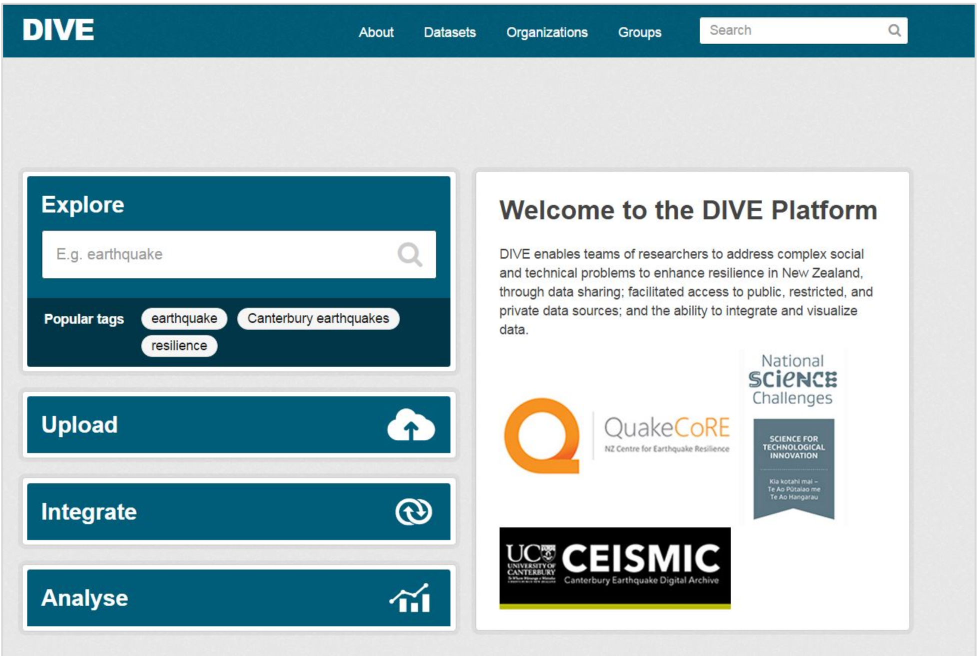


Figure 1. Home page of the DIVE pre-alpha prototype built on a CKAN interface.

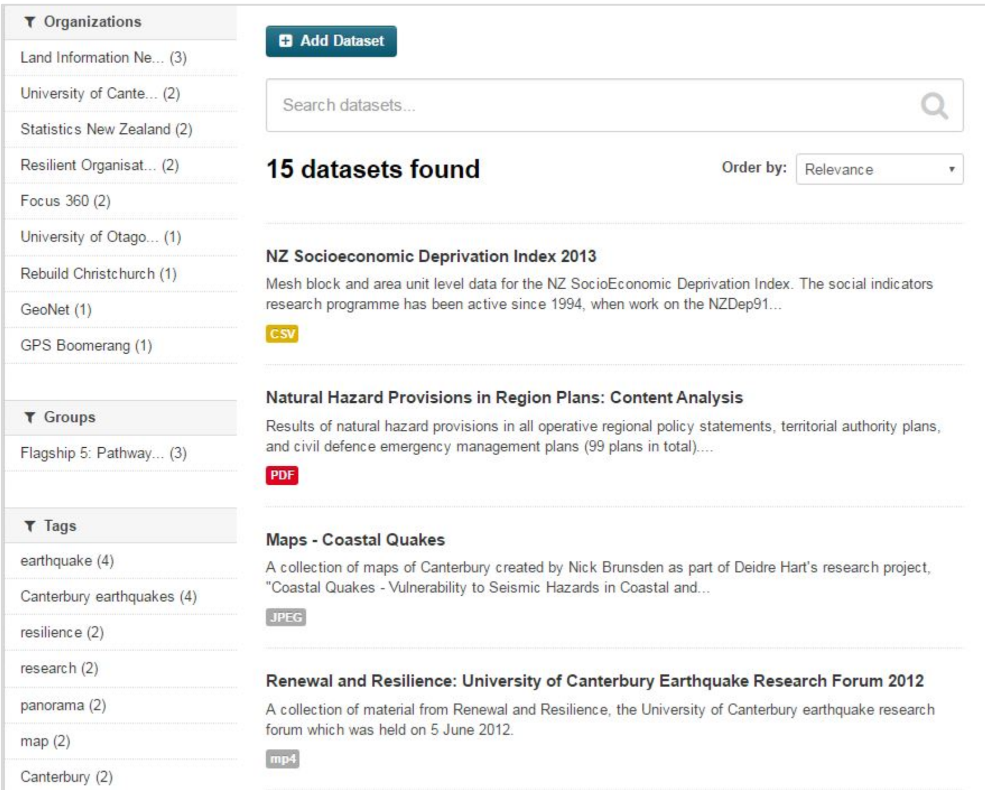


Figure 2. A search with facets in the DIVE platform.

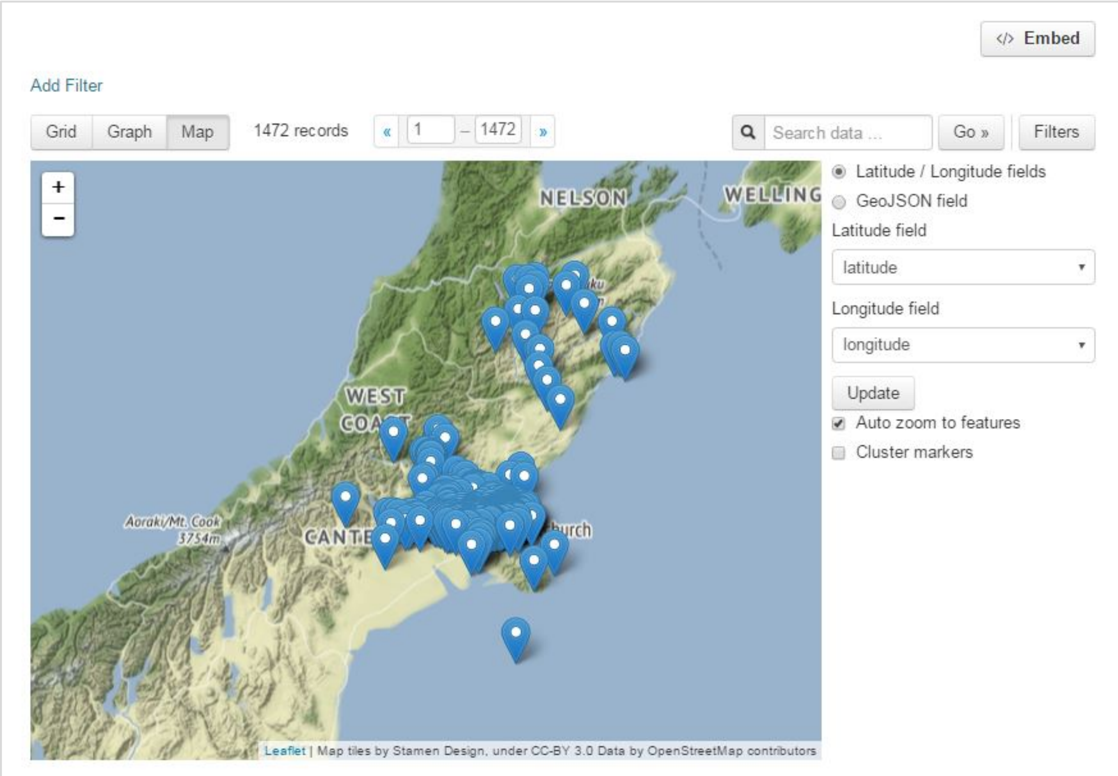


Figure 3. Map view in DIVE search interface of Geonet earthquake data from February 2011.

Figure 4. Graph view of Geonet earthquake data from February 2011.

Appendix B: Use Case Summaries from Workshop 3

The following slides summarise the DIVE Software use cases presented in Workshop 3 in November 2016. The full presentation can be found on the DIVE Wiki Page.[2]

Use Cases

Use Case 1: Federation & analysis

Use Case 2: Addressing obstacles to data sharing & data integration

Use Case 3: Creating a safe home for unique data & metadata standardisation

Use Case 4: Enhancing searchability

Use Case 1: New Questions from Old Data

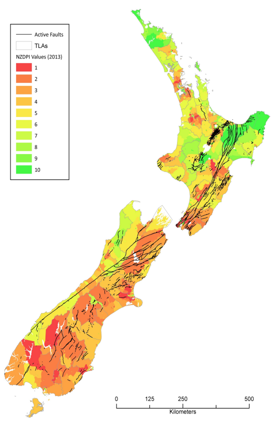
Is hazard planning adequate in socially deprived communities?

DIVE Platform Hosts & Harvests Existing Data

- Plan Content Analysis 2014 Uploaded by QC researcher
- NZDep Data, University of Otago - Wellington
- Active faults database, GNS

Integrates & Displays

- Joins the relevant data in a relational database and displays preview. This may also occur after download or by integrating all of the data on an open source GIS Web App like Koordinates.



[2] DIVE Wiki: https://wiki.canterbury.ac.nz/pages/viewpage.action?pageId=52692301

Use Case 1: Conclusions

- Researchers are creating datasets that have value far-beyond their initial research questions.
- These can be uploaded to DIVE and inspire secondary uses that the original researchers had not even considered.
- Federated search and viewing data that is not stored locally will enhance exploration and innovation

Use Case 2: Improving Data Access & Adding Value to Data

Linking Building Properties to Earthquake Induced Damage and Business Downtime using FEMA P-58 and REDI Assessments

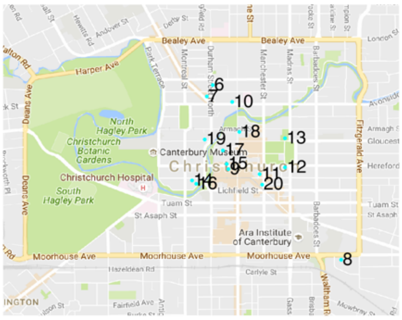
Gemma Cremen, Jack W. Baker, Sonia Giovinazzi, Erica Seville[1]

Building Properties	Damage Benchmarks
CEBA Database	CEBA Database
+	+
CCC Building Footprint	ERI Survey Database

- 3 Non-disclosure agreements (NDAs)
- 3 ways of identifying the same buildings
  - Prupi*, Unique ID, Lat/Long

* Prupi is the ID used by the Christchurch City Council before the Christchurch earthquake. They switched the unique ID system after the earthquake. The EQC uses a different set of IDs for buildings.

Use Case 2: Improving Data Access & Adding Value to Data



"Preliminary results [show] from an effort to use the FEMA P-58 and REDi seismic assessment procedures to predict damage and downtime for a sample of buildings… [that] both procedures appear to overestimate earthquake impacts...Future evaluations of a larger set of buildings will provide more definitive results." (Cremen et al., 2016).

Use Case 2: Conclusions

- Gemma Cremen and colleagues will create a valuable unique dataset with the combined databases, but there is no clear way to share this value-added dataset without all future researchers signing non-disclosure agreements with all of the original data holders.

- The DIVE platform has the potential to develop processes that ease administrative burden.
  - Building on existing expertise and experience with data sharing agreements, permissions, ethics, and copyrights.

Use Case 3: Hosting Unique Data & Metadata Standardisation

<u>Collection Fields for Building Inventories</u>
Building Name
Buildings Address
Year Built
Construction type
Description of concrete extraction location
Testing standards used
Specimen dimensions
Concrete compression strength
Concrete elastic modulus
Concrete density
Description of steel extraction location
Testing standards used
Specimen dimensions
Steel tensile strength
Steel elastic modulus
Capital Value (and date of value) …

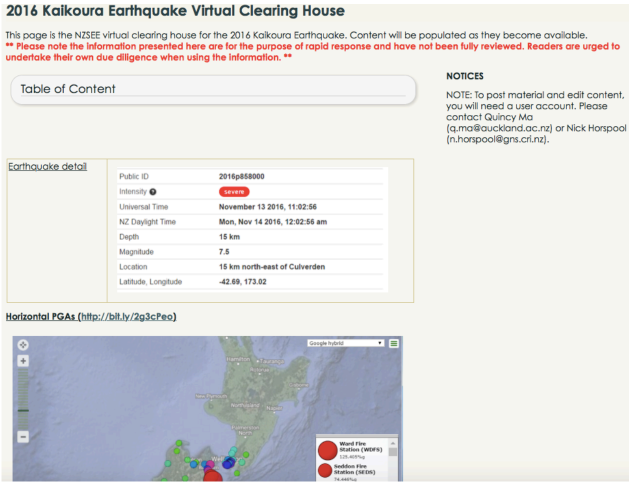
[2] Scattered masonry that has fallen from St John the Baptist Church in Latimer Square

Use Case 3: Conclusions

- QuakeCoRE Researchers and others will be collecting rich databases about New Zealand's building inventories. There is currently no 'home' for this data.
- We will work toward developing a Building Inventory Module for DIVE that will standardise and ease the collection of building inventory data, enhance building data integration, and facilitate data sharing & visibility.
- Work toward better live updating, archiving, and version control processes for collaboratively produced datasets.

Use Case 4: Kaikoura Earthquakes

- Capturing planned and ongoing research/ data collection
- Parallel databases
- Sourcing, uploading, and managing data quality is a lot of work



Use Case 4: Conclusions

- There is a great need to capture ongoing and planned research
- Sourcing, uploading, and capturing high quality metadata data is a lot of work
  - How can better automate and/or distribute this burden
- Visualising metadata can make it more searchable

The full presentation can be found on the DIVE Wiki Page:

https://wiki.canterbury.ac.nz/pages/viewpage.action?pageId=52692301

Appendix C: Affiliated researchers round table presentations From Workshop 3

Slides are available for the presentations described in Table 1 in a separate document.

Table 1. Affiliated researchers round table presentations from Workshop 3, 2016.

Presenter	Selected Affiliations	Presentation Title	Thesis or Highlights
Matthew Hughes	University of Canterbury & QuakeCoRE	*Finding and sharing data in a post-disaster environment*	Relationship building and establishing strong degrees of trust with data providing agencies is critical for successful and meaningful co-creation.
David Johnston	Natural Hazards Research Platform, Joint Centre for Disaster Research (Massey University) & United Nations Office for Disaster Risk Reduction	*Largescale cross-institution collaboration*	The International Council for Science's CODATA, the IRDR's Disaster Loss Data (DATA) project, and the RHISE group (researching the health implications of seismic events) provide useful international and national guidance for the development of DIVE.
Austen Ganley	University of Auckland & the Biological Heritage National Science Challenge	*A virtual hub for nationwide sharing of environmental DNA sequence data*	As part of the BH – NSC, developing a system where DNA data can be shared and used by those with all levels of expertise. Critical areas include visualization, database creation, standardizing metadata collection, and potentially creating an app so people collect and contribute metadata easily (like GeoNet).
Byron Cochrane	Land Information New Zealand	*International standards and best practices*	LINZ is drawing on international standards, best practices, and existing tools and expertise (e.g., spatial data on the web working group) to develop their spatial data infrastructure.
Sheng-Lin Lin	GNS	*Principles for developing ONE standardized, extensible, updatable building dataset*	There are significant challenges with managing and sharing earthquake affected and earthquake prone building data, including no nation-wide unique ID, no standardized data capture process, and issues with confidentiality.

Chris Bowie	Opus International Consulting	*Natural Hazards Research Platform – Neighbourhood Recovery Trajectories*	Researchers need to budget to assess the quality of and to 'clean' large datasets – it is a resource and time consuming process. Researchers would benefit from a collaborative web-space to share data.
Kyle Dow	Christchurch City Council	*'Addresses – who's do you use' & 'Knowing who is using your data'*	There are data quality issues to be aware of when using addresses to spatially locate data. A better option is to use building IDs rather than addresses where possible. Data providers (e.g., the CCC) have concerns about managing use and monitoring how people use and change the data.
Alistair Ritchie	Landcare	*Pervasive issues affecting our operational and commercial systems*	Data management presents technical, social, and financial problems. A key challenge for Landcare: how to integrate and disseminate data describing many aspects of the environment in a way that supports monitoring, analysis and reporting.
Simon Kingham	University of Canterbury & the Better Homes, Towns, and Cities National Science Challenge	*Next generation information for better outcomes*	We need to align the way we talk about spatial data. If we want data that can 'work together' focusing on spatial standards will help significantly.

Appendix D: Existing Hazards Databases & DesignSafe-CI

In New Zealand, there are several information aggregation portals for hazards data – notably Tonkin & Taylor's Geospatial Portal, the NZ Geotechnical Database, the CEISMIC digital archive, and the post-earthquake data clearinghouse system hosted by The Earthquake Engineering Research Institute (EERI).

These are useful repositories, but are heavily focused on earthquake and geotechnical data. NZGD focuses exclusively on data relevant to engineers, and CEISMIC provides mainly documentary resources, with an emphasis on images, news media, video and audio files, and associated cultural heritage collections.

The Earthquake Engineering Research Institute (EERI) supports an international virtual clearinghouse system, which has been employed to capture valuable data gathered by researchers and practitioners working across many institutions in the aftermath of the 2010 and 2011 Canterbury earthquakes and the 2016 Kaikoura earthquake. The Kaikoura Earthquake Virtual Clearinghouse website was established for those wishing to publish information relevant to the public and international researchers. This effort was largely driven by the earthquake engineering community, but managed to capture media articles, photographs, maps and social-economic information that are relevant across a broad range of applications. However, the EERI data clearinghouses are geared towards initial data capture after an earthquake, and are not updated with recovery data.

https://www.nzgd.org.nz/HelpSupport/AboutNZGD.pdf?AspxAutoDetectCookieSupport=1

http://www.ceismic.org.nz

http://www.eqclearinghouse.org

http://www.eqclearinghouse.org/2016-11-13-kaikoura/

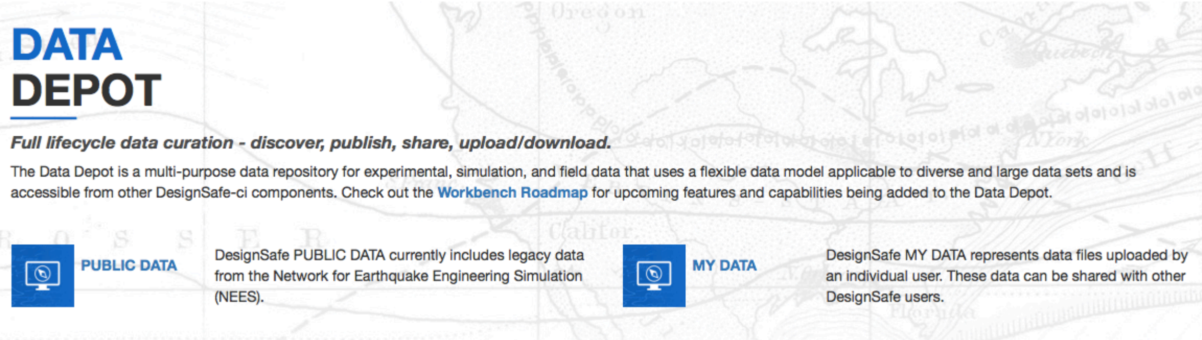
Other aggregation portals for hazards data in New Zealand include NIWA's Historic Weather Events Catalogue, GeoNet's Geohazards applications and data, and places where citizens can engage in the collection of critical environmental data, such as GeoNet's Felt Reports and NIWA's community air quality observation network.

There is not yet a unified space where independent researchers and citizen scientists can share the information they are collecting across multiple hazards in a way that is curated and archived. Mountains of information are collected following these events. Some of the data collected by crown researchers is published on their institutions' websites, and lessons learned are published in the months and years following the event, but most data are stored in an ad hoc manner and easily lost or buried.

https://felt.geonet.org.nz

https://www.niwa.co.nz/atmosphere/researchprojects/CommunityObservationNetworksforAir/rangiorasurvey

DesignSafe-CI is a flexible, extensible, community-driven cyber-infrastructure for the natural hazards engineering research community, based in the US.



Features include:

- Excellent tools for analysing data
- Interactive forums
- Regular workshops for users

DeseignSafe-CI facilitates collaboration and data sharing with other researchers, and supports the full lifecycle of data needed to address civil infrastructure threats posed by natural hazards.

Like the NZGD it is geared toward engineers, and although they are very interested in engaging with a global research community, most of the data on the platform is currently from researchers in the US.

https://www.designsafe-ci.org