

# TP3: Multi-disciplinary Community Datasets

**Acting Leader:** Brendon Bradley

**Deputy:** -

This Technology Platform fosters the contribution to, and utilisation of, existing community databases, as well as enabling the development of new disciplinary databases, and the extension of regional databases to a nation-wide scale. Research funded via QuakeCoRE will also contain the requirement that data contributed to a relevant community database, or the open-access QuakeCoRE web repository (where a relevant community database is not currently established), be available to the entire QuakeCoRE community as well as national and international end-users, within the constraints of privacy considerations.

## Key research thrust areas:

1. Facilitate opportunities for linking of research and stakeholder community datasets relevant to earthquake resilience.
2. Integration of multi-disciplinary datasets into existing open-source community databases to facilitate transformative research.
3. Development or prototyping of database tools for understanding multidisciplinary links between datasets and facilitating database development and usage.
4. Collection and/or integration of multi-disciplinary datasets into existing open-source community databases

## Overarching Objectives

The key top level objectives of TP3 are:

**'Increase accessibility to existing open-source datasets':** Provide both a centralized location by which datasets relevant to a wide range of researchers in earthquake resilience can be accessed (this does not imply that datasets will all be hosted by QuakeCoRE, often there may be links to third-party data), and also facilitate the access to datasets through the use of interfaces/APIs to download subsets of data from complex multi-faceted datasets in different formats.

**'Focus on transformative datasets':** Data is critical for all aspects of QuakeCoRE research, and is collected and archived by groups of researchers already. Open-source 'community' datasets should be datasets that require centralized hosting (and possibly curation) because they are mission-critical to achieving strategic QuakeCoRE goals.

## Principles

The underlying principles for TP3 to attain the overarching objectives are:

**Centralized vs decentralized:** It is not efficient for QuakeCoRE to host/curate all datasets that exist. There is a natural progression that all datasets start as the focus of a subset of researchers which have a specific focus on the data and utilize it in a decentralized manner (relative to QuakeCoRE as a community of researchers). As the widespread value of some specific datasets becomes apparent (because of their size, comprehensiveness, or multidisciplinary nature) then it makes sense for a more coordinated approach to enable their access in a centralized manner. Furthermore, to handle datasets that are both dynamic (i.e. continually growing at a rapid pace) and multi-faceted it may be necessary to have interfaces/APIs that enable the users to interact with the dataset before identifying a specific subset of the data to download and manipulate remotely. Thus datasets can be seen to sit in one of three categories: (1) decentralized, QuakeCoRE does nothing; (2) centralized, QuakeCoRE either hosts the data in a 'raw' format, or provides a link to a third-party location; (3) centralized data with appropriate interfaces/APIs to enable interrogation of the dataset prior to download to a remote location.

**Open Source:** The data must be open source to maintain flexibility, enhance collaboration, and not be excessively dependent on individual organisations. This also recognises that NZ researchers represent a small portion of the global human resource in this field and the use of OS datasets enables a greater leverage of international initiatives.

**Flexible:** Any curation and hosting that QuakeCoRE provides to external datasets must be flexible to enable the datasets to be used in a multitude of ways by different researchers from different disciplines

**Recognise QuakeCoRE's position in the data ecosystem:** Data curation and hosting is extremely expensive, and several research entities that we have interacted with (PEER/SCEC) has said that they have avoided this topic because it is a black hole for resources. We must be extremely careful not to fall into this trap. Staying true to the centralized vs decentralized mantra is one important concept, as well as the idea that in order for QuakeCoRE to continue to innovate we must continually shift resources onto new datasets. In most cases there should be an emphasis on transferring QuakeCoRE community datasets to external third parties that can provide a long-term 'home' for them (particularly if curator of the dataset costs more than simply server space).

## Personnel

The QuakeCoRE (and aligned) staff for Technology Platform 3 are:

Sung Bae, IT Architect (0.15 FTE)

Sharmila Savarimuthu, Software Engineer (1.0 FTE)

## Key performance indicators:

| Thrust Areas                      | Key tasks/Deliverables  | Start     | Finish     |  |
|-----------------------------------|---|-----------|------------|--|
| TP3.1 Link datasets               | 1. Undertake a stocktake of metadata from Canterbury Earthquake Sequence  | 1/01/2016 | 1/7/2017   |  |
|                                   | 2. Develop multi-disciplinary datasets from New Zealand earthquakes   | 1/7/2017  | 31/12/2020 |  |
| TP3.2 Multi-disciplinary datasets | 1. Develop a framework for the DIVE Platform  | 1/1/2016  | 1/1/2017   |  |
|                                   | 2. Assess and contribute to the development of at least 1 additional multi-disciplinary dataset initiative  | 1/1/2017  | 31/12/2020 |  |
| TP3.3 Database tools              | 1. Select data platform for use by QuakeCoRE researchers  | 1/01/2017 | 1/1/2019   |  |
|                                   | 2. Host workshops with QuakeCoRE stakeholders to understand available relevant datasets and opportunities for integrations with research datasets   | 1/01/2018 | 31/12/2018 |  |
|                                   | 3. Contribute to the development of a national strategy for post-earthquake data management and integration with other data sources (eg: RiskScape) | 1/01/2018 | 31/12/2020 |  |

#### 2019 RfP Information

No applications or Expressions of Interest (EOI) are being sought via the RfP. The Technology Platform Leaders are responsible for delivering a long-term coordinated Technology Platform Programme; investigators are encouraged to engage with the Technology Platform Leaders to see where they may contribute to the Technology Platform Programme.

#### Periodic Updates:

---