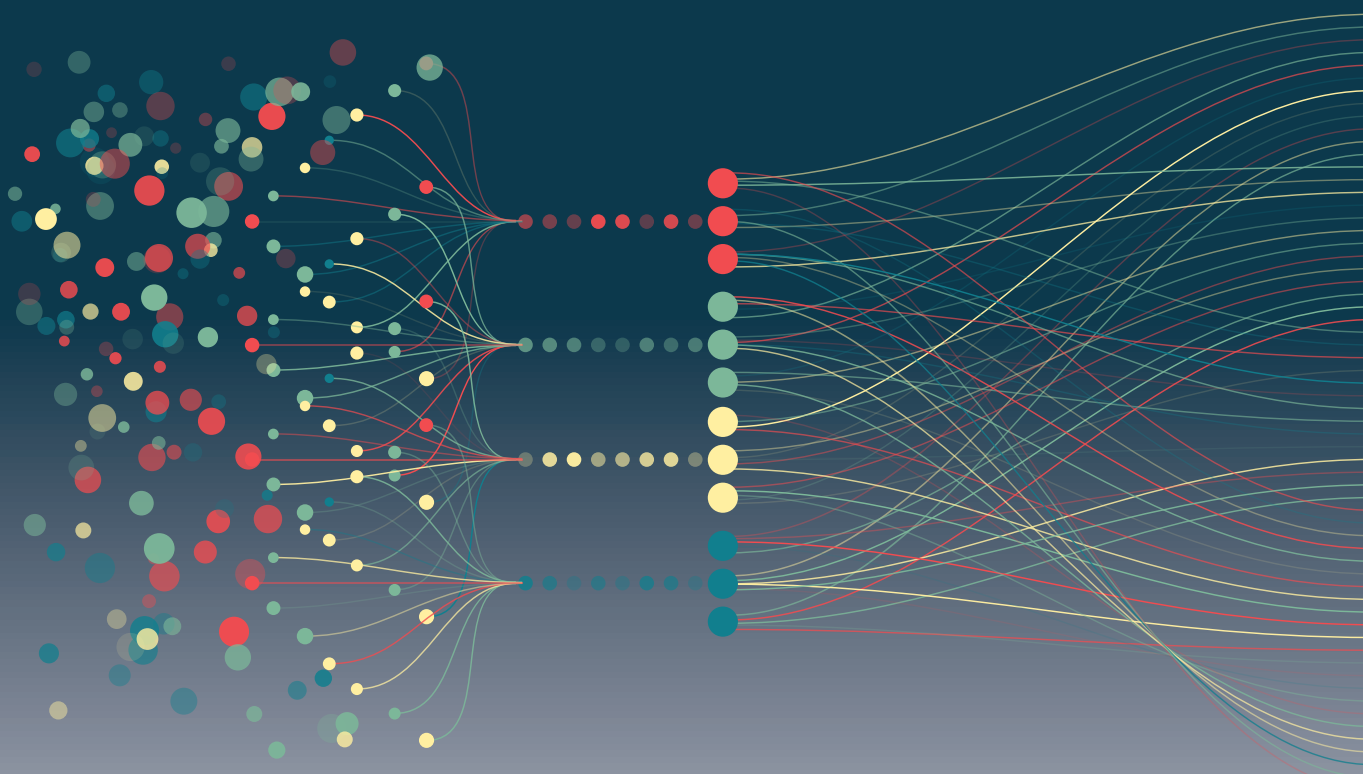
Data challenges to understanding construction sector capability and capacity in Aotearoa New Zealand – defining the minimum data set



CanConstructNZ
Balance in the sector

TE KUNENGA KI PŪREHUROA
MASSEY UNIVERSITY
UNIVERSITY OF NEW ZEALAND
TE WĀHANGA PŪTAIAO
COLLEGE OF SCIENCES

BRANZ

Key insights:

- The construction sector faces many data challenges in understanding construction capability and capacity.
- Data quality across the construction sector is variable.
- Data standardisation is a critical challenge that needs to be addressed in Aotearoa New Zealand.
- A proposed minimum data set (MDS) for understanding construction sector capability and capacity is provided to enable standardisation and coordination. The MDS focuses on four key categories: project preamble, project characteristics, project budget and time, and project threats.

Introduction

Data provides a powerful tool for understanding the capability and capacity of the construction sector to deliver its pipeline of work. This white paper discusses the data challenges to understanding capability and capacity in the construction sector and seeks to define a minimum data set for the sector. . The availability of data on previously executed construction projects could be valuable in determining future trends, generalising lessons learnt, and making valid assumptions following historical trends. Data about expected construction projects could help industry stakeholders to forecast and plan their future workloads and workforce requirements. . A better understanding of construction sector capability and capability data would provide greater clarity for the sector, allowing enhanced planning, more integrated cross-sectoral coordination, and better visibility for informed scheduling of investments and required future skills.

Judging the quality of high-level data can be a challenge. It is usual to define data quality in terms of its fitness for use and purpose (Watts et al., 2009). However, the availability of several high-level data providers in the construction sector that can provide appropriate and "fit-for-purpose" data sets cannot be assumed.

And furthermore, despite increasing volumes of data being available, challenges arise because much of this data is poor quality and/or difficult to gather and analyse (Watts et al., 2009). Poor quality data could result in lower client satisfaction, higher operational costs, lower decision-making effectiveness, and impaired business strategies (Ge et al., 2011; Samitsch, 2014).

What is quality data?

The founding principles for assessing the quality of data were laid by the works of Wang and Strong (1996), who proposed four primary data quality categories:

1. intrinsic
2. contextual
3. representational, and
4. accessibility.

Within these four data quality categories are 18 different data quality dimensions. Data quality dimensions are "attributes of data quality that can, when measured correctly, indicate the overall quality level of data" (Cichy & Rass, 2019, p. 4). Figure 1 illustrates four primary data quality categories and dimensions proposed by Wang and Strong (1996).

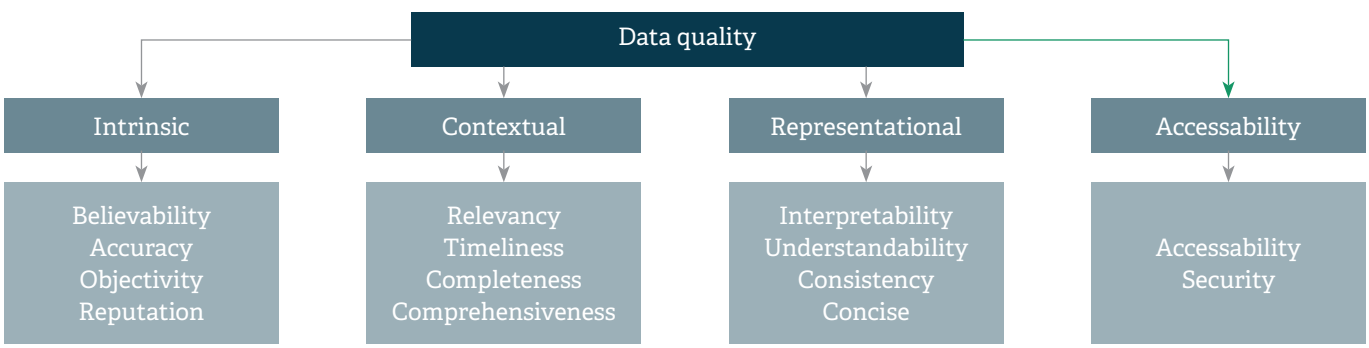
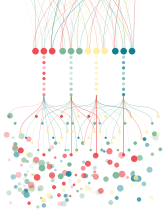


Figure 1. Data quality dimensions. Source: Wang and Strong (1996).



The Ministry of Business, Innovation, and Employment (MBIE) estimated the size of Aotearoa New Zealand's construction sector to be NZ$60.2 billion in 2022 (MBIE, 2023). The availability of adequate and fit-for-purpose construction capacity and capability data is crucial for government policymakers, developers/owners, service providers and consumers. However, secondary data from CanConstructNZ research activities has highlighted several challenges with construction data in Aotearoa New Zealand. This white paper outlines the contextual data quality of available construction data sets in Aotearoa New Zealand. The paper identifies the high-level challenges and shortfalls of the available data sets. It also provides recommendations to enhance the contextual quality and content of the construction data in New Zealand. A minimum data set is proposed as an initial step towards standardising the contextual quality of the construction capability and capacity data in Aotearoa New Zealand.

Construction capability and capacity data

Aotearoa New Zealand's construction capability and capacity data is currently being compiled, analysed and presented by an array of different enterprises that could be categorised as follows:

- **Category 1:** Specialised government agencies
- **Category 2:** Private enterprises
- **Category 3:** Government agencies undertaking the role of developers or procuring buildings.

Each construction activity data collector aims to achieve a different strategic objective. For example, Te Waihanga | the Infrastructure Commission is a Category 1 enterprise, which collects construction capability and capacity data to advise the Government on infrastructure needs and challenges and to enable better planning and policy formulation. It also aims to provide data insights and projections for the construction sector to aid it in planning, coordination and future preparedness. Category 2 enterprises may be characterised as data collection businesses that market their commodity (construction activity data) to their clients, mainly the construction sector stakeholders, such as Pacifecon. Category 3 enterprises could be classified as government agencies that are active players in the construction sector, often through the procurement and construction of their own buildings. These government enterprises strive to partner with and provide data to the construction sector. For example, the Ministry of Education is one of the largest construction project providers in Aotearoa New Zealand. The Ministry of Education relies on the construction sector for building, maintenance and redevelopment projects.

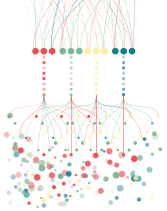
Key construction data challenges

Drawing on the experiences, insights and secondary data from interactions with construction sector and policymakers, several factors were perceived to affect the availability and quality of the construction pipeline data in Aotearoa New Zealand. The following articulates the extent of the challenges and issues with this data.

Construction sector capability and capacity data is being collected through different enterprises with different aims, perspectives and methods. These differences have led to variations in the availability of the data, its quality and its content. A leading national construction company official said, "There is no construction [activity] data [in Aotearoa New Zealand]. It's information. You must look quite hard for the data."

Moreover, the construction project life cycle is long and complex, and the different data collectors may focus on different project phases according to their inclinations and strategic objectives. Thus, the amount of data collected across the construction sector differs in terms of comprehensiveness and depth. Private commercial data collectors (Category 2) identify projects in the planning stage to enable construction suppliers to plan appropriately. However, their interest in tracking projects after the planning stages may vary, and consequently, the comprehensiveness of the data is uncertain. Similarly, the amount of data collected and presented may differ in the level of detail and granularity as each provider views 'fitness for purpose' differently.

There is concern about the accuracy of the construction sector capability and capacity data within the sector. The construction sector has no standardised data collection methodology, as it varies considerably across the data providers. Some data providers use surveys and try to communicate with the project stakeholders for updates. Other data providers run systems that are updated by the developers or owners. Differences in data compilation methodologies yield different levels of availability and accuracy across the available data sets.



To add to the challenge, some developers or owners from the same category or field may use different platforms to publish their project data. For example, Te Waihanga | the Infrastructure Commission publishes data for public (government-focused) projects, rather than private sector projects. Construction data inaccuracy is further exacerbated by external factors, such as projects being postponed, so projects could be seen in the pipeline for years without knowing when they will be commissioned, or even projects being removed from the pipeline altogether.

The difference in objectives and strategic inclination within the construction sector is perceived to influence the presentation and format of the collected data. Consequently, the adopted jargon and presented format may not be suitable for universal use and may contribute to the problematic interpretability of the data. A reputable construction company official stated that while Stats NZ (a Category 1 data collector) collects and provides construction capability and capacity data, they tend to use economic terminology and standard economic measurement methods. The official added that other collectors or platforms use Stats NZ data as a source but are more successful in presenting it in formats that are easier to use and understand.

Evidence from data sets

To further articulate the problem and assess the extent of the challenges, a sample of four construction pipelines from different construction data providers was considered. The sample was selected to represent the three categories of construction capability and capacity data presented earlier. The selected data providers were:

- **Te Waihanga | the Infrastructure Commission** (Category 1). Te Waihanga | the Infrastructure Commission is a government agency that compiles pipeline project data from major public and a limited number of private organisations involved in construction projects. Its objective is to offer advice to the Government for improved planning and policymaking. Additionally, the agency provides information and future trajectories to assist the planning of the major public construction projects. Currently, Te Waihanga | the Infrastructure Commission publishes pipeline information from approximately 73 entities, such as selected local authorities and the New Zealand Transport Agency | Waka Kotahi, which contribute substantial volumes of construction projects.
- **Pacifecon** (Category 2). Pacifecon is a private business that provides research information on planned commercial, residential and infrastructure projects. Pacifecon has been operating for around 30 years, and its main stakeholders are central government, local government, private developers and the wider construction sector.
- **The Ministry of Education (MoE)** (Category 3). The Ministry of Education is a significant actor and procurer of construction projects. The Ministry is highly reliant on the construction sector for building, maintenance and redevelopment projects across Aotearoa New Zealand. MoE publishes its quarterly pipeline of work to provide the construction sector with the needed information to plan accordingly.
- **Local government (Category 1)**. Local government comprises city, district and regional councils, as well as unitary authorities. Local government's issue long-term plans (LTPs) that signal the high-level description of infrastructure projects planned for execution within the 10-year life of the plan within their respective boundaries. The LTP is reviewed every three years and is modified on a rolling basis. The LTP provides only high-level data, including the forecasted annual budget for each project or sector. Some local authorities publish detailed construction project data on other platforms, such as the Te Waihanga Insights Platform.

A comparison of the latest versions of construction pipeline data published by these providers was conducted. The contextual category was compared, focusing on the comprehensiveness dimension to provide details on the types of data being collected and the level of detail and granularity adopted by each construction data provider. To complement the focus on comprehensiveness, the level of completeness of the data sets and the number of missing data points across the column were reviewed to evaluate if the data domains to be collected were achievable; that is, are the construction data providers able to collect all the claimed data points across the different projects?

A summary of the comparison is given in Table 1.

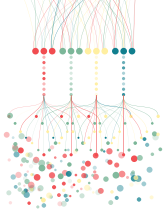
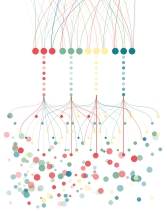


Table 1. High-level contextual comparison of construction capability and capacity data from selected providers

Parameters	Data Attributes	Infrastructure Commission	Pacifecon	MoE	Local Government LTP
Comprehensiveness	Project identifiers	Name, ID, short description	Name, ID, short description	Name	Name
	Procuring agency/ developer	Yes	Yes	Default (MoE)	Default (Local Government)
	Project location	Region (city, suburb, coordinates)	Region (suburb and street)	By Region	By suburb/area
	Project work type	Fifteen categories: commercial, communication, community facilities, defence, education, energy, health, housing, justice, science, transport, waste, water	Nine categories: civil, commercial, education, health, energy, industrial residential, sport, multi-category	Two categories: new school, and existing school redevelopment	Four–sixteen categories, depending on the issuing authority. Common categories include: transport, roads, water supply, wastewater treatment, stormwater, parks and community, waste management, environmental services, engineering and technical services
	Project life cycle reported	Planning, procurement, construction	Planning, procurement, construction	Procurement	None, only
	Consents approvals	No	Partial /Inconsistent	No	No
	Procurement type	Yes (direct, limited, open, selective)	No	No	No
	Procurement scope	Yes (alliance, construction only, D&B, ECI, PPP)	No	Yes (traditional, D&B, ECI)	No
	Supplier information for awarded projects	No	Yes	No	No
	Funding	Yes (funded, source confirmed, source to be confirmed)	No	Only approved projects are published	High-level annual capex
	Preliminary in-house value estimate	No	Yes	No	Yes
	Value bands	Nine bands: < 1 M, 1–5 M, 5 –25 M, 25–50 M, 50–100 M, 100–250 M, 250–500 M, 500 M–1 Bn, > 1 Bn	Five bands: <1 M, 1–5 M, 5–10 M, 10–25 M, < 25 M)	Seven bands: VB1 < 1.5 M, VB2 1.5–.5 M, VB3 3.5–7 M, VB4 7–14 M, VB5 14–21 M, VB6 21– 30 M, VB7 > 30 M	No
	Reported schedules	Business case, procurement, construction (start and end)	No	Tender, contract (start date)	No
	Date formats	Nearest quarters	Month/Year	Month/Year	Year
Completeness	Total columns	29	14	9	N/A
	Blank columns	4	0	0	N/A
	Columns with missing data	10	3	0	N/A



Variances and differences of construction data

A significant variance detected across the data sets is the difference in reported projects' life cycles. Some data sets report all the primary cycles, such as the planning, procurement and construction phases, while others concentrated on single phases, such as the MoE, which publishes its projected pipeline information focusing on the tentative timings through the procurement phase. Information provided by the local governments' 10-year LTPs was only high level and does not provide details. However, some local authorities publish the appropriate details using other platforms, such as the Te Waihanga | the Infrastructure Commission's Pipeline report, after completing internal planning processes and confirming project commissioning data. Another point of variance is the reporting of building consents and resource management consents across the data sets. Only one data set reports on the consent status of each project, meaning there is lower certainty around projects' status, progress and expected completion time frames in the other three sampled data sets.

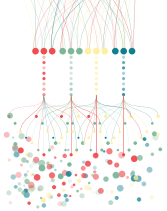
How a project is defined across the sampled data sets also differed. While all the data sets provide a project name, not all assign unique identification numbers to the projects. Furthermore, the assigned project names and identification numbers differ across the data sets, making identifying and cross-checking project information from different sources difficult. Similar challenges could be identified when considering how a project location is presented – while some data sets report only high-level locations, such as regions or cities, others provide variable detail levels, ranging from the broad regional level to exact location and coordinates.

The level of detail and sub-categorisation of the projects' construction type varies significantly between the four sampled data sets. Some data sets adopt a detailed approach with as many as 16 project work types reported, while others use general work scope classifications, such as "new build" or "redevelopment". While some of the data sets describe the projects with the generally used construction categories, such as residential, non-residential, commercial or civil work, others use (also include?) categorisation phrases related to the developer or owner's sector, such as "defence" or "justice". Despite giving more detail, these differences in construction classification may create confusion and present challenges for the ease of use, understandability, interpretability and manoeuvrability of the data or information.

The sampled data sets also adopt different formats in presenting the procurement criteria for the listed projects. For example, not all data sets report the expected procurement type and the level of detail relating to procurement varies. Some explicitly mention how a project will be procured, such as direct award or open tendering, while others do not present such information. Similarly, the procurement scope is not consistently reported across the sampled data sets. Some data sets detail the procurement scope – for example, alliance, public-private partnership, design and build, early contractor involvement, or construction only – while other data sets do not report procurement scope at all. Likewise, supplier information is only published in one of the four sampled data sets, for awarded projects that have progressed beyond the procurement phase. Another data set only mentions if a preferred supplier was selected for the project. but give no further details or information.

Reporting on project funding status and expected costs varies considerably across the four sampled data sets. Some data sets report if a project is funded or if its funding source has been confirmed or is not yet confirmed. It is understood that some of the sampled data sets would only publish projects if funding was already approved or secured. High-level data sets such as the local government 10-year LTPs only project the annual expected capital expenditure of the project, with no detail on funding certainty. Three of the sampled data sets categorise the projects according to pre-defined value bands. However, the number of bands and the band intervals also varies considerably. One data set categorises projects under nine value bands with tighter intervals, enhancing confidence levels, while other data sets use fewer value bands with greater intervals. For example, a project with an expected value of NZ$1.1 billion would be classified under the value band "> 25 million" in one data set, while it would be (more accurately) represented in another data set as "> 1 billion". Additionally, only one data set provides a lump-sum estimate of the expected costs.

Time frames and reporting of schedules also differs across the four data sets. Generally, each data set has adopted its own time-reporting format: nearest quarter, month and year, or year only. In addition, some data sets report the expected schedule for each project's life cycle or phases, others focus on specific phases, such as tendering and



contract dates, and some provide only the year with no detail on the projected phases. The high variability in time-reporting formats reduces the certainty and confidence in any aggregated construction activity data.

The number of data columns within each data set was considered to provide insight into the level of comprehensiveness of the data being collected. Significant variation was identified between the four data sets: one data set has 29 data columns, another 14 columns and a third 9 columns, while the fourth data set does not have a data column format. However, the data set that has 29 data columns had four columns that were completely blank with no data points recorded, and ten columns had variable degrees of missing data points. Neither of the other two data sets that have data columns contained blank columns, although one had three columns with missing data points.

Towards a minimum construction capability and capacity data set

The previous sections illustrated some challenges around construction capability and capacity data in Aotearoa New Zealand. It may prove challenging for the construction sector to gather all the data requirements from a single source, forcing the data seekers to triangulate and articulate through multiple platforms. Accessibility issues further complicate the availability of adequate and meaningful construction capability and capacity data. The four sampled data sets exhibited various levels of variance and differences which can be mainly summarised as:

- different data collection aims, objectives, scopes and perspectives
- differences in data columns and hence the types of data being collected and reported
- lack of a standardised data reporting format across provides
- variability in the level of detail and granularity of the data
- lack of compatibility between the different data file formats or platforms, and
- lack of information on projected risks, shocks and stresses.

The identified challenges surrounding construction capability and capacity data and the variance across the available data sets pose a real threat to stakeholders across the construction sector. Indeed, these challenges around the collection and analysis of construction capability and capacity data may endanger the health and resilience of the construction sector in the future.

A path forward?

Following a series of research workshops and several interviews and interactions with construction sector stakeholders, a standardised set of data covering all the required domains and with the with appropriate detail and granularity of data is proposed. The data set would have four main data categories, with various levels of subcategorisation:

Category 1: **Project preamble** presents the project's general information in terms of name and location.

Category 2: **Project characteristics** provides more information about the main project features and type.

Category 3: **Project budget and time** provides project-specific data and information.

Category 4: **Project threats** provides specific information about the projected risks and shocks the project could undergo during its life cycle.

The proposed minimum data set (MDS) for construction capability and capacity data fields is shown in Table 2 and illustrated in Figure 2.

It is hoped by outlining a MDS for construction capability and capacity data, we might facilitate dialogue and coordination within the construction sector to start addressing the data challenges the sector has, with the end goal to help the sector to be more efficient, productive and resilient, and so able to enhance Aotearoa New Zealand's built environment.

Table 2. Proposed a minimum data set for construction capability and capacity data n New Zealand

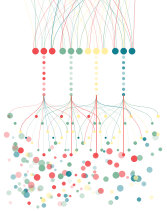


Table 2. Proposed a minimum data set for construction capability and capacity data n New Zealand

Category	Subcategory 1	Subcategory 2	Definition
Preambles	Identity	Project name	This field should assign a unique project name to the project throughout its life cycle.
		Unique ID	
	Location	Region	This field should detail the exact location of the project works.
		City	
		GPS (lat. & long)	
	Owner	Owner name	This field should state the identity of the procuring organisation.
Characteristics	Type	Civil	This field should identify the specialisation of sector organisations.
		Commercial	
		Residential	
		Environmental	
		Industrial	
		Institutional	
		Utilities	
	Phase	Initial concept	This field reflects the current phase of the project.
		Concept	
		Approvals	
		Design	
		Procurement	
		Construction	
		Completion	
		Maintenance	
	Priority	Asap	This field indicates the urgency and priority of the project.
		Scheduled	
		Flexible	
	Funding	Funded	This field should express the funding status of the project.
		Unfunded	
	Procurement	Undecided	This field should state the procurement method adopted.
		Open tender	
		Negotiated	
		Early contractor involvement	
		Collaborative	
Budget and time	Value	Estimated value	This field should provide an estimate of the construction-only or contracted value (expressed in New Zealand dollars).
		Awarded value	
	Time	Start date	This field is generally related to the construction phase but may report other phases (expressed by year and quarter).
		Projected duration	
Threats	Shocks	Political	This field highlights risk factors selected from a matrix of start date and delay factors based on project type, funding, status and value.
		Economic	
		Social	
		Technological	
		Environmental	
		Legal	
	Stresses	Stresses	
	Other risks	Risks	

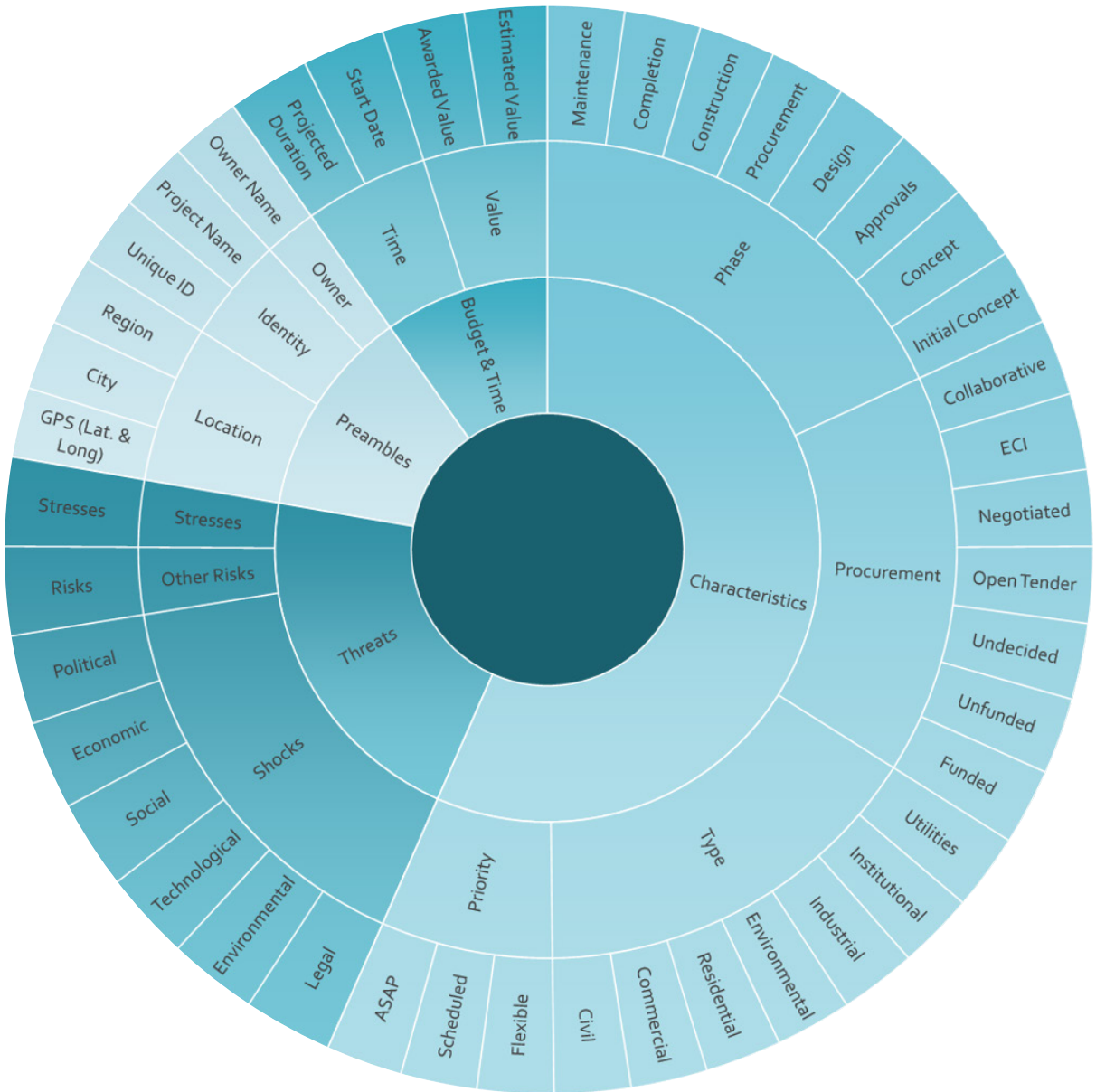


Figure 2. Proposed construction activities pipeline data to be collected for the MDS

Acknowledgement

This paper is part of the ongoing work-in-progress of CanConstructNZ.

The authors acknowledge the funding of the New Zealand Ministry of Business, Innovation and Employment's Endeavour Programme Research Grant. Contract Number: MAUX2005.

To find out more visit: https://canconstructnz.org.nz/