

Geospatial Data Management

Strategic Review

Version 1.0

Prepared by: The GDMR Team

Location Information

Released under the Official Information Act 1982

Executive Summary

The Geospatial Data Management Strategic Review forms the apex of the Geospatial Data Management Review by providing the vision, business drivers, goals, and guidance to lead Location Information (LI) towards having a highly effective data practice. The observations and recommendations in this document are in context of LI, however, we believe that these could also be applied to LINZ at enterprise level.

In its current form, LI data management is hindered in the use of modern standards, technologies, and practices. Toitū Te Whenua Land Information New Zealand's (LINZ) vision is to be the New Zealand geospatial leader through customer-centric innovation in data collection, management, and distribution. Although significant effort has been made in the past to align data management with modern best practices, obstacles described in this report preclude full commitment and adoption of these solutions. As a result, our position as a government leader in geospatial data management is at risk.

Business Drivers for Improved Data Management

1. **Customer Value** - Data that Toitū Te Whenua LINZ collects and distributes must be fit for customer's needs, meet quality expectations, and be easy to find, use and understand, including focusing on value for Māori as a key customer.
2. **Geospatial Leadership** - LINZ will provide strategic leadership in the NZ geospatial community in areas of data, standards, governance, and collaboration.
3. **Risk Reduction** - LINZ must implement systems and processes that properly ensure the integrity, availability, and security of data.
4. **Productivity** - Producing and maintaining our data in a way that is efficient and effective.
5. **Agility** - Using agile approaches to respond faster to customer demands, technological shifts, and government requirements.
6. **Strong Data Foundations** - Building strong data foundations to ensure ongoing data maturity and to set LINZ up for future growth.

Strategic Themes

1. **Creating Customer Value** – Our latest customer surveys and interviews highlight that our channels by and large meet their needs, but there are some challenges especially access around large datasets. Customers feedback also indicates that some data is not fit for purpose. We must make informed decisions and prioritise on customer value. To get there we need good product ownership, good customer feedback loops and improved data publishing capability.
2. **Improving the Way we Work** – Our way of working is ineffective, being a mixture of agile and legacy practices. We must be responsive to change, focussing on continually delivering value to our customers, making our work open and transparent and building a learning culture. We need to move away from siloed data practices and take a portfolio view of our data and products. This will enable better prioritisation and drive the most effective value. Change is required in our organisational structure. We need lean-agile leadership, formally established communities of practice, and the people and capability to execute on our strategy. These requirements align with LINZ's current organisational change programme *Organising Ourselves*.
3. **Improving Data Management** – Data management is inconsistent and loosely applied. We want data of high quality, easily accessible, available and easy to maintain. We need better data

governance, better control of our master data, centrally accessible and managed metadata, convergence on data management tooling and improved data supply processes.

4. **Designing Systems for Agility** – Our data products rely on many point-to-point integrations and cannot easily be changed due to the strong interdependency between systems. To work as independent autonomous teams and to respond to change, we need to change the way we design systems and adopt self-service technology platform. We must evolve architecture, design thinking, integration and platform capability to enable and align with agile. We need to capture our business and data domains to design and create independent, loosely coupled systems that enable autonomous teams and responsive delivery to customer value. Automating manual processes will further improve efficiency, predictability and quality while supporting frequent and safe software and data delivery.
5. **Improving Business Intelligence** – We don't have accessible data to enable business intelligence. Our priorities, insights and drivers will benefit from having data to inform our decision making. We must establish a business analytics capability and converge our effort and resources to establish a data warehouse.

Next Steps

To deliver benefits to the organisation, our customers and our data management capability, we recommend changing the way we work, creating a central data catalogue and accelerating centralisation of our data storage in Amazon Web Services (AWS).

1. **Improving the Way We Work – Fully adopt agile at all levels**

- Establish a lean-agile leadership competency: enablement, principles, value driven, customer driven, frameworks
- Create a single portfolio view starting with the themes and epics in this document, adding in the overall group value streams and epics
- Re-align our roles, responsibilities and decision making with agile practices
- Shift from projects to products with ownership from dedicated resources, responsible for product and data maturity
- Formally establish communities of practice to improve collaboration, knowledge sharing and standardisation
- Empower teams to make decisions, deliver and adapt
- Establish a product enablement team to help provide self-service technology capabilities and enable product teams

2. **Improving Data Management**

- **Metadata Management – Create a Central Data Catalogue**
Centralise metadata management and discoverability through the development of a central data catalogue. This will support data discovery, centralised data access, data ownership, metadata management, standardisation, interoperability, data management processes, data supply and significantly improve our data foundations.
- **Master Data Management - Centralise Data Access by Collocating our Data in the Cloud**
We must rationalise our existing data to improve our master data management. We must continue the drive to move our data to cloud storage for centralised data access. Initiatives around this has already started with the Geostore. We recommend a strategic and accelerated approach away from legacy infrastructure to migrate data to a common cloud provider.

Contents

Executive Summary	2
1 Foreword	6
2 Creating Customer Value	7
2.1 Customer Insights	7
2.2 Data Publishing.....	8
3 Improving the Way We Work	9
3.1 Organisational Structure	10
3.2 Community of Practice.....	12
3.3 People and Capability.....	13
4 Improving Data Management	14
4.1 Data Governance	15
4.2 Master Data Management.....	15
4.3 Metadata Management.....	16
4.4 Common Data Management Tools.....	17
4.5 Data Supply	18
4.6 Data Quality.....	18
5 Designing Systems for Agility	19
5.1 Cloud Enablement	20
5.2 Cloud Migration.....	21
5.3 System Integration	22
5.4 Automation.....	23
5.5 Platform Services.....	23
5.6 Domain Modelling.....	24
6 Improving Business Intelligence	25
6.1 Business Analytics	25
6.2 Data Warehousing	26
7 Next Steps	27
7.1 Improving the Way of Work	27
7.2 Improving Data Management.....	28
8 Appendix – Backlog	29
9 Appendix – Data Management Principles	31

9.1	Background	31
9.2	Data is a Strategic Asset	31
9.3	Data Stewardship	32
9.4	Data Ethics	34
9.5	Data Collection.....	35
9.6	Data Access and Availability	36
9.7	Data is Fit for Purpose	37
9.8	Design for Data Compliance.....	38
10	Appendix – Review Team	39

Released under the Official Information Act 1982

1 Foreword

The purpose of Toitū Te Whenua LINZ' (LINZ) is to maintain and build confidence in property rights, geographic and hydrographic information and encourage land information markets to develop and mature. The Location Information business group (LI) provides much of this geographic and hydrographic information, supporting essential services such as national security and emergency service responses and enabling commercial applications. LI is committed to providing the most authoritative, complete, accurate, and up-to-date information to enable this.

LINZ recognises that high-quality data is critical to both LINZ and LINZ customers. Our vision is to be the New Zealand geospatial leader through customer-centric innovation in data collection, management, and distribution. To drive this vision effectively, LINZ needs to ensure that the data we deliver is of high value to customers. In addition, we need to employ agile approaches to respond faster to new customer demands, technological shifts, and government requirements - no longer should LINZ find itself in a position of widespread technology legacy and technical debt. Finally, while delivering to these customer needs, our systems should be safe, robust, and efficient using modern industry approaches. Through this execution, LINZ can walk the talk, become a data management exemplar for the rest of NZ Public Sector, and ultimately lay strong foundations to ensure ongoing data maturity and set LINZ up for future growth.

Over the past 20 years, LINZ's systems that manage data have organically grown without much consideration for re-use, interoperability, or in some cases, robustness. In most cases, siloed projects within individual domains have delivered these systems due to a lack of a strategic view, capability, or cost and time pressures. Unfortunately, this approach has resulted in many ad-hoc solutions, outdated technologies, lack of ownership, poor support models, and risky processes. At the same time, some data management improvements have occurred within business groups, but more extensive enterprise-wide changes were not established to lift the level of data management maturity or provide re-usable solutions. As a result of LINZ's past delivery approach, our position as a government leader in geospatial data management is at risk.

In June 2020, the LI Leadership Team commissioned the Geospatial Data Management Review to understand the investments to transform LINZ into a modern data enterprise. This project brought together a team of domain data leads, key product owners, and architects from across LI. This team reviewed our practices, technology, data systems, distribution channels, structures, people capability, governance, and processes to provide a clear current state analysis of our data management. We also engaged our customers on our data publishing channels to ensure we understand any shift in their needs and expectations. LI teams resolved some high-risk data management issues throughout the project, including master data stored on hard drives and metadata standardisation.

This is an open and honest view of where we are and offers guidance on the changes required to transform us into a high performing data organisation. It is structured to create a backlog¹ of work for LI. Sections represents agile Themes, while subsections represent agile Epics. This report takes a critical and reflective view of our data, people, practices, processes, platforms and systems. It also considers how to collect, process, manage, store and distribute our data.

¹ See Appendix 1 for a Consolidated table of Themes and Backlog items

2 Creating Customer Value

THEME

Improve customer experience. Respond to the changing needs of customers. Prioritise on customer value

LINZ has always made our customers a high priority. Our external facing systems are generally easy to use and support our customer requirements reasonably well. However, our offering has stagnated to some degree, while our customer's needs have evolved. Our current rate of delivery is not keeping up with changing requirements.

Some of the data products we provide are not fit for purpose. A lot of data comes from legacy sources, e.g. traditional Hydro charts, Topo maps where the data is a subset of the original dataset and of limited use to customers beyond its designated purpose. Continued engagement with our customers will help us understand what fit for purpose means from their point of view.

Fit-for-purpose data is quality data that is readily discoverable, easily interpreted and meets the needs of intended customer use. Data is curated and published by considering the breadth of customer needs, in order to generate maximum value. Ensuring that data is fit-for-purpose should include careful consideration of any ethical concerns in data collection, sharing, use, representation of the information intended, rapid data integration, and minimisation of any sources of unintended bias.

Understanding our customers, the data and systems they use is critical to prioritising our effort. This, supported by other recommendations in this document, will enable us to respond effectively and naturally to change and help keep our datasets easy to locate and intuitive to use.

The way we deliver our data to customers is also changing. Fixing the Foundations included an in-depth review of our data publication channels (Geospatial Data Management - Channels Review Report - Objective ID: A4547639). This document also builds on the findings presented in the Channels Review Report.

2.1 Customer Insights

EPIC

Measure customer engagement and collect data distribution metrics to improve understanding of customers and the services they use to inform our decisions.

We don't have comprehensive methods and metrics to understand how our customers use our systems and data which limits our ability to make improvements from a customer-centric perspective. We must improve our understanding of customer requirements and incorporate these insights into our product decisions. Currently, we don't have the expertise or processes to make well-informed investments across LINZ products and services.

We need to establish infrastructure and capability in customer insights to support the following:

- Improved customer satisfaction
- Better understanding of customer requirements
- Better understanding of who our customers are

- Understanding our customer types
- Strong feedback loops
- Customer guidance and education
- Get metrics to use for business intelligence

2.2 Data Publishing

EPIC

Improve data delivery and discovery to align with changing customer needs and strategic programs of work.

LINZ customers need access to the most up to date and authoritative data. Therefore, we need to consistently look for ways to improve our publication channels with a focus on the following:

- Publication channels and data aligned with customer needs
- Easy and timely access to LINZ data
- Increased re-use and value of LINZ data
- Innovation opportunities for us and our customers
- Reduced handling of customer requests

Prioritisation should be ongoing and driven by customer value. Our agile guidelines must support the practice, design, implementation and operational support of our publication channels. For example, a portfolio of themes and epics across LI needs to constantly focus our publishing effort and inform autonomous data publishing product teams with a vision, goals, principles and guidelines to help us cooperate and collaborate while remaining agile.

In consideration of the findings of the Channel Review Report, we suggest future work includes:

- **Consolidation** – Data channels are fragmented. We need to consolidate similar channels which would reduce costs and improve the customer experience.
- **Onboarding** – Customers struggle to set up. We must improve onboarding of new users and businesses with clear direction and documentation.
- **Customer Experience** – User experience is failing to keep track with modern expectations. We need to improve user experience on the channels and API's continuously.
- **Customer Segmentation** – We don't have a good view of our customer market segments. We must gather better intelligence to help support prioritisation and understanding.
- **Move away from OpenTopography** – Provide an improved consolidated customer experience and improve data discovery, reduce management costs, provide better insights into customer usage, and provide future capability for future point cloud dataset. e.g. bathymetry.
- **Invest in Improvements to Large Data Access** – Provide better online access to our large datasets via improvements to our existing established LDS channels and new modern cloud storage services for our emerging cloud-ready customers.
- **Discovery** – Customers are often not aware of all our data channels. We need to improve discovery of data channels and the key data held within those channels.

3 Improving the Way We Work

THEME

Aligning to agile management, data and development practices to respond quickly to change, focussing on continually delivering value to our customers, making our work open and transparent and building a learning culture.

LI has a diverse set of business domains covering hydrography, topography, addressing, property data, positioning, and resilience, each with multiple teams that provide business functions and services to customers. The teams within these domains are responsible for many IT systems and datasets, with meagre resources to draw upon, despite data being essential to our customers. In many cases, active ownership roles are not in place to ensure the systems and data are well managed. We have many ongoing and overlapping projects, and switch people between them to provide coverage resulting in lower productivity, quality, and business agility. Collaboration occurs across teams in pockets, and there are common platforms for data publishing and storage. However, there is still siloed collaboration and approaches to delivery, resulting in a lack of reuse or the implementation of good practice. We often miss the value of making the right decisions at the right level. Some enterprise decisions are implemented at team level, while some team decisions are frustrated by bureaucracy and impediments.

LINZ leadership is committed to adopting agile practices to respond quickly to change, focus on continually delivering value to our customers, making our work open and transparent, and building a learning culture. This commitment is part of the Organising Ourselves change programme. We need to adopt these practices at scale to enable effective delivery of our data products and services. Moving away from siloed data management practices and taking a portfolio view of our data enables better prioritisation, creates opportunities, drives the most effective value, and manages critical risks.

Change is required at multiple levels to reap the benefit of being an agile organisation. While some teams have implemented agile practises, internal roles, leadership structures, budgeting and development practices are mostly reminiscent of pre-agile methods. Agile must be enacted at all levels: leadership, project management, people, technology, governance, platforms and solution design. We need bold actions to make this change quick and effective or risk falling into limbo, where we are neither committed to agile nor invested in traditional approaches.

The culture, mindset and decision making of leaders and teams will need to shift to support team autonomy, with lean-agile leadership providing strategic direction and enablement. As an organisation, we must avoid making large monolithic decisions. Our focus should be on small iterative and incremental timely decisions that deliver value. Decisions need to be made by the people closest to the customer. As an agile practice, we must become comfortable with deferring commitment, waiting until the last responsible moment to decide. Leadership must focus on empowering teams, driving change, and organising around essential value. People need to organise into teams that deliver products and services and focus on upskilling skillsets to support incremental delivery.

LINZ data is a valuable asset. Over the past ten years, we have gradually shifted our focus from data as a by-product of our business function to data as product. We must align our practice, structures and resourcing to manage our datasets with the same rigour as our other products.

3.1 Organisational Structure

EPIC

Align organisational structure and ownership with agile roles and responsibilities. Lay the foundation to prioritise with purpose, communicate better, enable product teams, focus on value and deliver efficiently.

We need to maximise our existing resources to put roles and support structures in place to deliver value, enable collective prioritisation, and reduce key business risks. To solve current structural problems, we must shift to agile ways of working by the following:

- **Move towards a 'scaled agile' framework** - Re-organise around customer value streams to focus on value, deliver efficiently, and govern our data products' maturity. This change is a significant undertaking, and it is recommended that we acquire support from an agile coach to help us through this change and embed the agile principles and values into our work.
- **Products over projects** – Driven by the customer value streams, shift from delivery by projects to continuous delivery within an established product lifecycle. Re-organise our people around these products and make them responsible for the development and day-to-day operation of these products.
- **Enablement** – A data enablement team is needed to provide self-service data infrastructure to support product team delivery. Currently, we have the Data Infrastructure Team providing a similar function to this. However, the team is not entirely focused on platform services and will need optimisation.
- **Customer Insights** – Establish capability and process to make well-informed investments across LINZ's products and to help drive roadmaps and prioritisation.
- **In-house resources** – Preference for in-house capability rather than outsourced managed services for domain specific data management activities. We have a unique domain of technical work, and it is hard to procure adequate short-term resources or work with managed IT suppliers with a high turnover in staff.

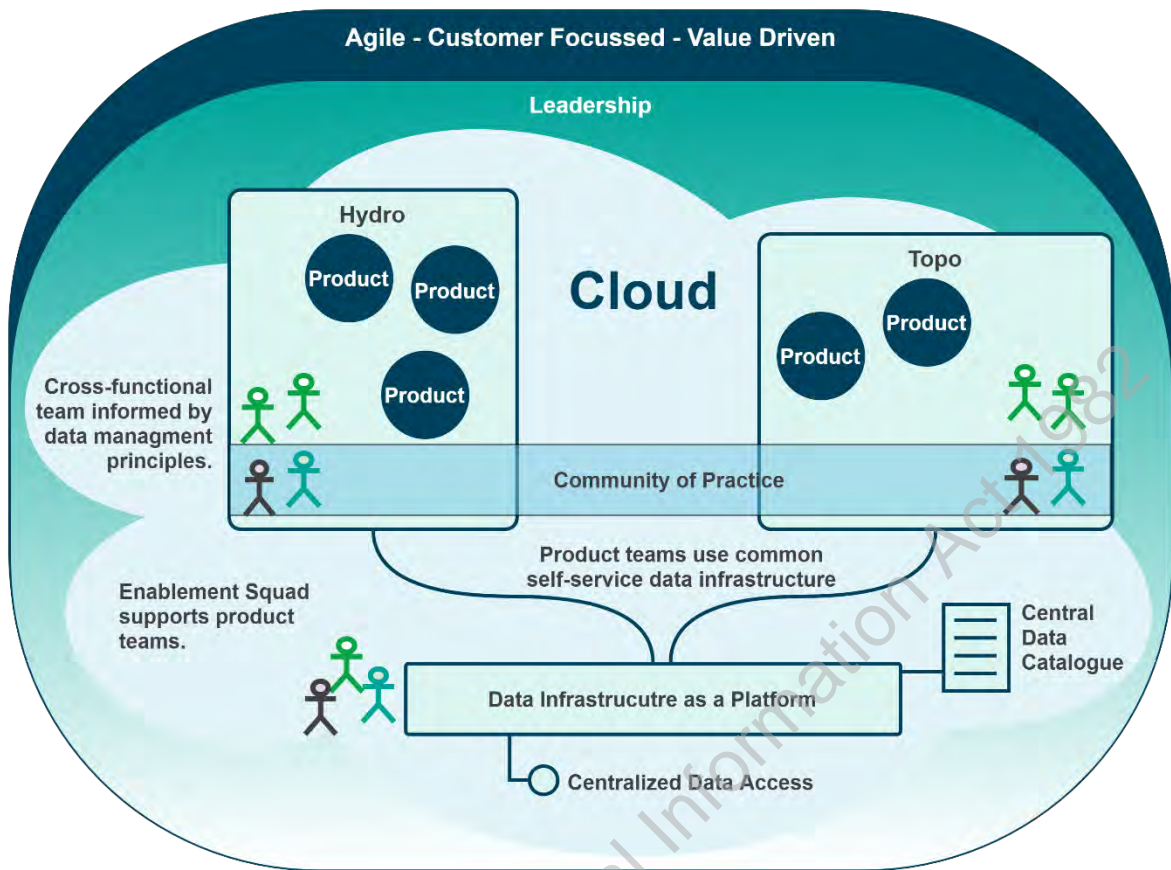


Figure 1 Conceptual view of the ways of working to support data delivery

To effectively implement these new ways of working and support best practise data management, the following guiding design principles should be employed:

- **Data as a Product** - Shift from 'managing data as a by-product' to 'delivering data products'
- **Customer Centricity** - Customer value should be at the centre of all decisions
- **Scalability** – Our product teams and system design strategy must promote new system features to be created as new products and services so that new products can be supported independently by new teams, and retirement of product will translate to reduced teams
- **Data-Driven Decisions** - Remove biases, assumptions and making decisions based on analytics and metrics to make decisions
- **Principles of Data Management** - Embed the principles and responsibilities of the LINZ's data and information policy into the business
- **Continuous Learning and Improvements** – A learning organisation is committed to relentless improvement and promotes a culture of innovation
- **Self-Service Data Infrastructure Platform** - Provide domain agnostic data platform that hides all the underlying complexity and provides the data infrastructure components self-service for autonomous team re-use
- **Cross-Functional Teams** - High performing teams are cross-functional and have all the resources they need to deliver value. They have the authority to self-organise and make their own decisions

- **Product over Projects** – Products need to be owned by teams. Projects have a start and end. Products are developed and managed by the same team for the life of the product. We cannot move central capability from product to product.

Cross functional teams will work on one or more products, and typically do this in two weekly Scrums (Figure 2).

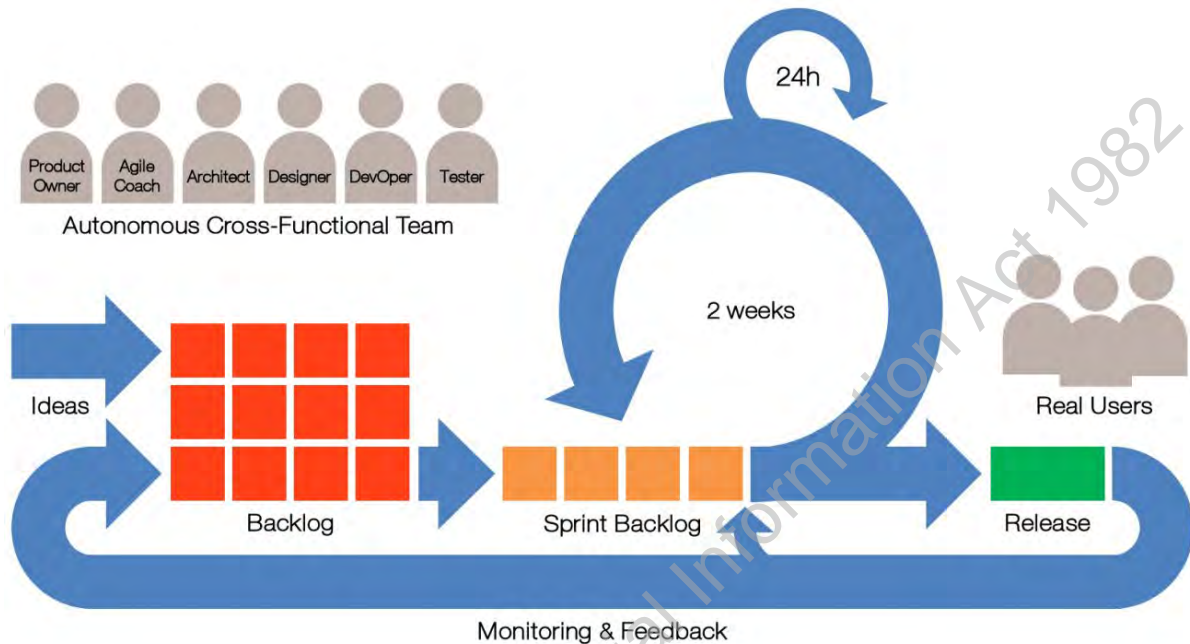


Figure 2 High level Product Team Development Process

3.2 Community of Practice

EPIC

Establish, shape and drive our communities of practice to raise our level of maturity by becoming tools for collaboration, standardisation and knowledge sharing.

To support the proposed agile ways of working, we must invest in a formal community of practice for data. These communities will primarily be set up for people who are part of product teams. They will be responsible for delivering products that involve data but would be open to anyone with an interest in the topic. We have already established an informal data community of practice. It is proposed that this group is enhanced to increase our data management maturity. The primary outputs of the community are:

- Improve data management maturity across all teams
- Create a collaborative forum for knowledge sharing, testing ideas, gaining feedback and educating to improve the capability
- Define best practice, and make it well understood and agreed by the community
- Standardisation, including identifying and implemented common tools, data specifications, processes and approaches that improve delivery.
- Provide a forum for testing ideas and gaining feedback.

A practice leader needs to be appointed to drive the community of practice successfully. The leader would be responsible for the practice policies and guidelines and ensure collaboration is effective. The role requires a technically capable individual.

It is proposed that this data community of practise would replace the existing TLG S&I subgroup.

3.3 People and Capability

EPIC

Create a shared common capability around platforms, people and communities of practice to support product teams. Upskill our existing capability in product teams to work with our modern platforms.

Moving to new ways of working and adopting modern industry best practice in product and data management requires new capabilities and augmenting existing capabilities. We already have a wealth of specialists and data analyst roles who understand the domain's requirements, processes, standards, and best practises. However, these roles are generally not equipped to develop the IT processes required to improve our data management systems. In addition, we need capabilities to drive effective product management, implement agile methods and streamline teamwork. Within LI, we have already invested in building this support capability within the Topographic and Data Services teams.

Across LI, there is a need to improve data management maturity by increasing data capability. This capability can be acquired by upskilling and recruiting into our in-house product teams. As a result, the following key roles have been identified:

- **Product Owners** - Owns the product, its backlog of work, understands the customer and prioritises the value
- **Customer Engagement** – Engages customers and communities to understand their needs and support them in using our data and products
- **Data Engineers** - Develops automated data management processes and analysis using the cloud.

Detailed analysis will be required to determine the exact capability and resourcing required within each product team. Furthermore, we need to add additional development capacity for infrastructure platforms, including data, cloud, and networking, to enable product teams to accelerate our data management work. We should establish dedicated teams with individuals that build knowledge and maintain ownership of domains and products. The alternative, of using support squads to support our business teams, is not recommended. This is because product ownership is key to the success of agile delivery and operations.

Because system design maturity needs improvement, we should undertake further analysis work with Digital Delivery to determine a model to support architecture decisions for business domains and product teams.

4 Improving Data Management

THEME Improving how we manage our data

The heart of our core business is managing data to support our work. This work is expensive due to the effort required to collect and maintain it using suppliers, people and IT systems. Our data management is generally in good shape, however, some fundamental changes are needed to support business agility, reduce risk, and set ourselves up for the future.

Managing geospatial and property data can be complex because of its structure, volume and integrations. Therefore, data management principles, frameworks, and capabilities are required to ensure data maturity that delivers to our business goals. Having the right people, resources and technology get us partway there, but without established practices for specific aspects of data management, we will not achieve higher levels of reuse, standardisation, safeguards, and overall effectiveness. Lastly, having the data management capabilities foundations in place also provides an easier pathway for new products and services to enter the enterprise without high establishment costs.

We need to establish data management capability in the following key areas, as highlighted in Figure 3. The establishment and running of these capabilities will need to be supported by leadership and our communities of practice. The following epics highlight the most critical capabilities that we need to focus on to provide operational excellence.

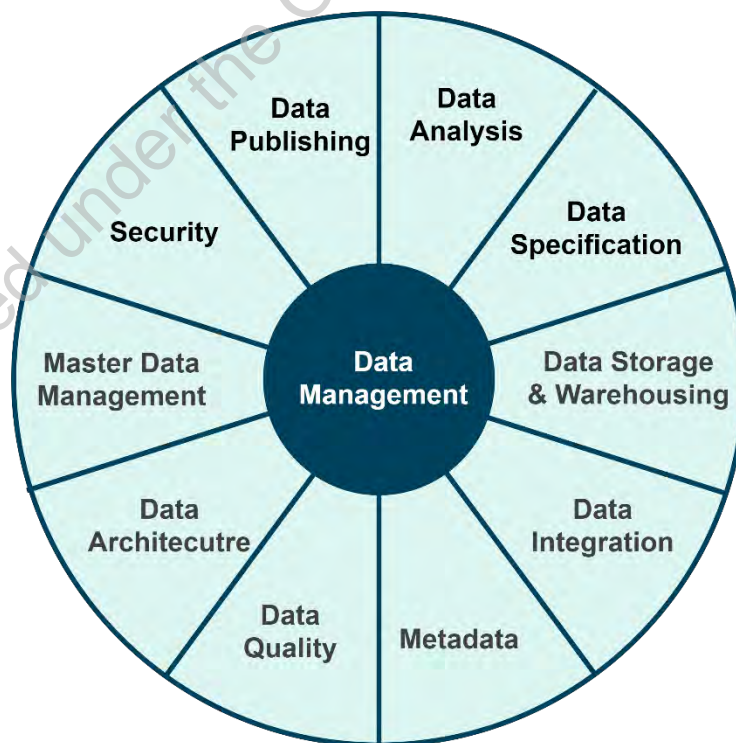


Figure 3 Data management capabilities

4.1 Data Governance

EPIC

Create a framework to assess our maturity and discover gaps in data management to support data owners with consistent data management principles and practices.

We need an operational data governance framework to define, measure and report our expected levels of data management maturity. Our leadership and the data community of practice must implement principles that help create a maturity framework that people understand and can efficiently implement. This ensures that new product teams have successful recipes for building and operating high-quality processes and systems. While some of our data processes and systems have been made to industry standards and are well-supported, many of our data systems have significant issues that expose us to risk, inefficiencies or limitations in use. In some cases, teams are not aware of issues and their impacts. Having frameworks to measure against also helps raise awareness and provide established pathways for implementation.

Surrounding the data governance framework, should be an ongoing process and support network that sets people up for success at all stages. Multiple levels of the organisation, from leaders, product owners, to product team members, should be aware of what is required. We should report compliance transparently, ideally in real-time, via automation processes that are baked into how we work and use the systems. Finally, guided by our lean-agile leadership, our principles and framework should be further established and evolved by communities of practice. LI data leaders have created the first draft of our principles and goals in Appendix 2.

4.2 Master Data Management

EPIC

Improve master data management. Ensure access, standardisation, completeness, accuracy and ownership of LINZ's master data.

Our master data is not managed consistently, and we have developed a strong dependency on external partners for our core business. Some data is stored on legacy systems and is not easily accessible. In some cases the location of the latest authoritative copy is unknown. We need to consolidate our data in our legacy systems to create consistent, integrated and enhanced master data.

Bring master data in-house

LINZ must take control of its data. Koordinates currently holds and manages some of our authoritative data and metadata due to past and current data management decisions. While working with external groups is good and necessary to provide the best possible service to our customers, we should not outsource our key capabilities.

Relying on an external party for fundamental data management capabilities is exposing us to risk and hampering innovation. Currently, our ability to respond to change is limited by the ability of Koordinates to respond, which has previously been slow, and in some cases, not a priority for them. Koordinates effectively has a monopoly on LINZ's data publishing. This reduces our bargaining power and limits our options. Having externally hosted data, where we cannot fully access the original data, makes it difficult to execute our Master Data Management principles and goals.

It is important to design our future state so that dependencies on external parties for master data management are significantly or totally removed.

Co-locate data on a single cloud platform

Data is stored in multiple locations with vast variations in the accessibility and reliability of the storage. This environment creates challenges when working with our data, including:

- Some data is not easily accessible
- It is inefficient to work with our data
- Data management process automation is challenging to achieve and maintain
- Cost overheads are high for data shipping, support, storage, etc
- Data is not well preserved from loss
- Capability and knowledge of modern platforms and tooling cannot be applied

We must identify all data sources, clean and store the data, and make the master and appropriate derived copies of data available. In addition, we must also improve access and operations to migrate from legacy solutions to centrally accessible platforms for cloud-based data processing.

4.3 Metadata Management

EPIC

Implement modern metadata standards. Create a data catalogue and manage all metadata centrally.

Metadata is a foundational enabler for many aspects of data management. Our metadata is supplied, edited and managed in various shapes and forms depending on the data type and system that it is managed in. However, the publication is primarily based on legacy closed standards with poor technology integration options. We need to standardise the use of metadata within our organisation to drive better productivity, discovery, access, use, governance, and interoperability.

As part of the metadata review, work has already been done to identify key fields and modern open standards. For example, metadata stored in Basemaps and Geostore is stored in JSON format alongside the data files in compliance with the STAC (SpatioTemporal Asset Catalog) standard. As an open and extendable standard, this metadata is already adopting the data fields identified and formalised as the output of the metadata review.

Create a Central Data Catalogue

Tooling and resources are required to enable the adoption and management of standardised metadata across LINZ. Providing metadata editing capability as part of a central data catalogue will allow strong metadata management practices.

The proposed data catalogue will keep a central index of metadata about all data inside of LINZ. This includes the structure, quality, definitions, lineage and usage of the data. By indexing all LINZ's metadata in a central catalogue, we can obtain data collection, management, and distribution metrics.

As the number of datasets in LINZ grows the catalogue will become a key enabler ensuring:

- **Productivity** – Consumers of data can quickly find and get access to the data they need, while also identifying what additional data is available to inform their decision-making
- **Agility** – Consistency and collaboration across all LINZ's datasets and their metadata
- **Strong data foundations** – A central location facilitates data governance
- **Consistency** – By using the same tool to manage all metadata, keeping information consistent becomes a natural side-effect.
- **Discoverability** – A central catalogue will enable the searching of descriptive information across all datasets

4.4 Common Data Management Tools

EPIC

Establish common tooling for data management to reduce the complexity of supporting multiple technology sets and support convergence in data management processes and standards.

There are some cases where different tools are used by various groups to fulfil similar purposes. For example, Addressing (e.g. addresses) and Topo (e.g. roads, tracks) data editing processes being comparable and highly aligned, both teams could benefit from using the same tools.

We must consider all areas and identify opportunities to simplify and converge tools to edit and manage our data. Any implemented rationalisation will lower the support burden and variance associated with multiple tools that mostly do the same job.

Benefits include:

- **Simplicity** – Reduced system landscape complexity
- **Supportability** – Less tools mean that we can create a common knowledge base and support base for the tools we use. This is especially valuable for more advanced tools.
- **Consistency** – Natural convergence around data standards
- **Process Improvements** - Promotion of standardised data management processes

4.5 Data Supply

EPIC

Extend our data catalogue to allow suppliers to upload data and capture metadata to reduce delays and double handling of data by LINZ.

We still receive data from our partners and suppliers using legacy methods such as hard drives, often with metadata about the data being manually captured upon receipt. These legacy methods were established when transferring large data using the internet or using cloud storage options were not possible. In some cases, internal validation of received data causes us to request a resupply of data. We need to shift our supply processes for more automated online processes that allow external suppliers to submit data online via the cloud, including metadata, and support automated validation before submitting to LINZ for further QA or processing.

If we support these processes using our existing catalogue and infrastructure, it will simplify our workflows and maximise our investments. Using cloud infrastructure will also open the process to efficiencies for suppliers. Those who already have the data in their cloud storage will have access to streamline processing, will be able to shorten the feedback loop and reduce data transmission cost and times.

4.6 Data Quality

EPIC

Implement and automate data quality management. Review and establish quality best practices and data quality metrics.

The levels of our data quality management and insights are widely varied. We don't have standardised data quality tracking in place, and in some cases, data quality tracking does not exist.

We need to create consistency around how we manage and measure data quality. For example, most of our datasets do not track the six standard data quality metrics. Tracking these metrics would instil confidence that data is fit-for-purpose and enable reporting required for good prioritisation for product management and leadership. The standard quality metrics that we must track are:

1. **Accuracy** – How well does our data depict reality
2. **Completeness** – Is our data as comprehensive as you need it to be
3. **Consistency** – Do disparate data stores have the same matching data records
4. **Validity** – Does our data exist in the right format and data type
5. **Timeliness** – Is our data acceptably up to date
6. **Uniqueness** – Is our data free of duplicate records

Most datasets have some data validation as part of their publishing, but this is generally not monitored. To track this, LINZ must have a standardised way of storing data quality metadata, ideally as part of automated processing pipelines, and exposing this in ways that support action. We need to explore options of using the proposed data catalogue and standard processing design patterns as an enabler for this. Managing both our general metadata, which already may include some quality information, with additional quality information as needed, may get us a long way to improve data quality management.

5 Designing Systems for Agility

THEME Evolve architecture, design thinking, integration and platform capability to enable and align with autonomous teams and agile practices.

System architecture, development practices, platforms and company structure are all co-dependent. Therefore, the success of an agile development practice will primarily rely on the way systems are developed. If systems are tightly coupled, teams not enabled, infrastructure not self-managed, and regular releases hamstrung by interdependencies between systems and databases, then the promise of agile will remain a pipedream.

To reap the rewards of agile practices as an organisation, we must change how we design, integrate, manage and govern our systems. Development teams need to converge around products and services that they own, operate and take responsibility for. They need to be able to make the right decisions and use the technology that suits their requirements. If our products are centrally operated, this becomes a challenge, and there will be a natural driver to converge around technology. However, if teams own their products, then technology becomes an enabler, not a constraint, with value to the customer and efficient product delivery as drivers.

To keep maximise value of our platforms, people and processes, while supporting our data processes (Figure 4), we should maximise reuse of our platform and storage components, introduce a central event streaming platform, create a central data catalogue and adopt a decoupled architectural style. This will reduce reliance on point-to-point integration which in turn will enable more organic and fluid systems development, enhancement, replacement or retirement. This is needed to support an agile way of work where product teams can work independently, but still share and re-use common capability and standards.

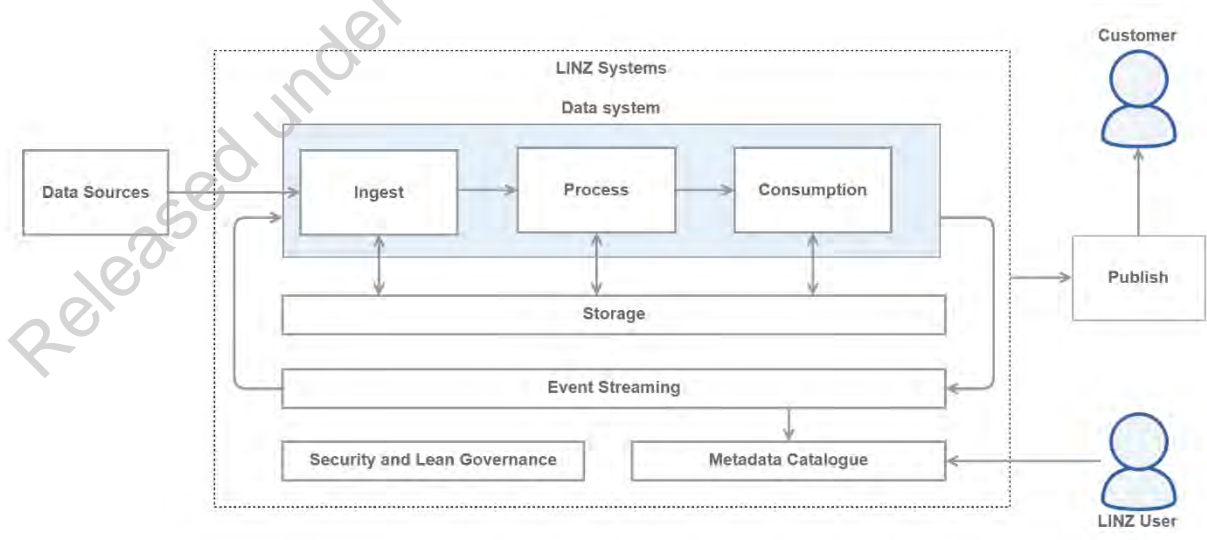


Figure 4 Data Processing Landscape

We need principles not to constrain us, but to guide our decision making:

- Decoupled event driven design over point-to-point integration
- Guided rails over governance
- No sharing of product databases, even during product replacement
- Principles over policy
- Reusing services over libraries

The message to get across and be clear about is : “We will fail to benefit from agile practices if our system designs are monolithic and interdependent”. We need to use cloud technology and best practices to the fullest extent. To do this and to get this right, we need to enable our teams with a clear understanding of their product purpose and boundaries, with cloud training and the proper tooling and platforms that support and encourage decoupled system design. Terms like microservices, learning, autonomy, events, serverless, DevOps, etc., should not be scary but be part of our daily vocabulary.

Exploring better ways to develop, deliver and operate should be encouraged and supported, with mistakes and failure normal and accepted as part of an evolving and innovating practice. We are not innovating if we are not experimenting. We are not experimenting if all attempts succeed. The correct principles and decoupled systems will remove significant risks from failure. We want to learn, progress, succeed and deliver in small, regular and incremental steps.

5.1 Cloud Enablement

EPIC

Embrace cloud platforms enabling operational excellence through new capabilities and self-service.

Our teams do not have the capability/tools to maximise the benefit of cloud technology. Many of our people are self-taught and don't understand how to draw value from instantly scalable and serverless systems. Some discussions and concerns around the cloud are hampered by limited understanding and assumptions that cloud technology is just another data centre. Systems are to some extent governed, designed and used in the same way as traditional systems. Cloud provides complete event-driven and serverless options, and powerful built-in tools and services, where infrastructure and platforms require no maintenance. Many best practices, guardrails, monitoring, reporting and management functions can be configured as part of the cloud environment and, for the most part, be transparent to data analysts and product teams. Pay-per-use systems can be operated at nothing more than the cost of the data storage if the system is not running. Cloud systems are usually designed in an event-driven fashion where services respond to events as needed and incur no or little cost while idle.

We are hamstrung by some datasets not being in the cloud. There is confusion around data sovereignty regulation. Assumptions are getting in the way of progress and creating hesitancy to commit to a cloud provider. LI has made a significant investment in AWS, and our products running on AWS are proving to be highly successful. We should further enable our teams with training and remove any question that systems need to be re-migrated to other cloud providers, i.e. Microsoft Azure. We also need an enablement

squad who can quickly guide and support product teams in setting up their environments consistently and for success.

We need better support for remote working. Cloud technology is remote by nature and needs to be supported as such. This further aligns with the way our partners and customers operate and use technology. Whether we should provide better support for the way we use the cloud is not up for consideration. Cloud is here to stay; the sooner we act to support it the better; the sooner we, our partners and our customers will benefit.

Ultimately, we want to view and use cloud technology as our data infrastructure as a service. In this state, we will have central access to data and data services. We will also be able to self-manage and provision any components we need to get our work done safely and efficiently.

5.2 Cloud Migration

EPIC

Accelerate and complete migration to AWS to enable centralised data access and reap the full benefit of cloud enablement.

New Zealand government [requires agencies to accelerate their adoption of public cloud services](#)² — in a balanced way — so they can drive digital transformation. This includes:

- enhancing customer experiences
- streamlining operations
- creating new delivery models.

The Cloud First policy requires government organisations to:

- adopt public cloud services in preference to traditional IT systems
- make adoption decisions on a case-by-case basis following a risk assessment
- only store data classified as RESTRICTED or below in a cloud service, whether it is hosted onshore or offshore.

LI data does not have jurisdictional and social licence concerns, unlike STEP. In our context, the balance is we don't have top-secret data but that our data is open or sensitive at most. Streamlining and improving our delivery model is top of the list.

We will never be able to gain full benefit from cloud technology if some of our systems still reside in local data centres with poor turnaround times on support. We have already moved a lot of our data and systems into AWS and must continue the effort for the remaining datasets. Commitment and accelerated migration of critical data systems to the cloud are needed to fully capitalise on the benefits of cloud technology. This migration will allow us to take full advantage of modern technology while removing the constraints associated with hosted infrastructure. Completing our migration to the cloud will have many benefits including but not limited to:

² See DIA <https://www.digital.govt.nz/standards-and-guidance/technology-and-architecture/cloud-services/>

- Facilitated access to bulk datasets
- Better alignment with customers who are also adopting cloud technology
- Reduced handling of customer requests
- Reduced costs
- Enabled automation
- Access to a myriad of hosted services with operational excellence already built-in
- Consolidation of some publication channels
- Self-service without long lead times for infrastructure and platform level changes
- Simplified integration and access to all systems
- Consistency of data storage/ interoperable access and use of data

We need our systems co-located to lower cost, enable faster delivery and quality, promote central data access and use, and maximise our talent. We accept that the hybrid cloud will be part of the LINZ future state, but that should be considered a last resort for core in-house systems. Therefore, it is recommended that we complete our shift to AWS to reap the full benefits of cloud adoption, our investment, intellectual capital, and skills. AWS is a leader in the global cloud market and has strong support in New Zealand. Furthermore, high costs and effort have already been committed to our AWS adoption and data migration. AWS is the only platform with in-house enterprise support, and the Survey and Title Enhancement Programme (STEP), with which we interact, has also invested heavily in AWS. We understand and use the technology with great success and have a solid internal AWS practice. This is where leadership needs to take a firm stance and not change cloud providers unless driven by customer or product requirements. It must be understood that every system migrated to a different platform than AWS (where most data is already stored and processed), will increase our dependency on external providers, lower our delivery rate and reduce central accessibility to data while increasing cost.

5.3 System Integration

EPIC

Decouple systems by introducing better integration technology, practices and patterns.

We have a complex network of point-to-point integrations and dependencies, either direct or indirect, across many systems. For example, some systems are dependent on the same underlying databases or database schemas. This environment makes it difficult for product teams to work independently and makes system replacement a challenge. By keeping systems decoupled and isolated, LINZ will ensure its agility by having systems that can be updated and modified separately.

To enable team and product autonomy and reap the associated benefits, we need to move away from point-to-point integration where publish and subscribe models are more suitable. This is currently challenged by the absence of a central event streaming platform. While cloud platforms provide many options for transferring messages between systems, a formally adopted central platform through which domain events pass will enable a single point where all systems can tap into. This will create a highway of information that will support our efforts for decoupled system development, standardisation, business intelligence, real-time metrics, data validation, data enrichment and much more. In addition, such a

platform will be well-served by a community of practice or product team to give direction and support teams as they adopt the technology.

Some aims for improved integration mechanisms are:

- Lowered risk of changes or failures in one system disrupting another
- Low dependency between systems, so change, agility and continuously evolving systems are easy and natural
- Standardised connections between datasets
- Improved and continued/continuous? sharing of data
- Visibility and control over information shared between applications and services
- Updates to one dataset can timely be reflected or responded to in another system

5.4 Automation

EPIC

Automate manual processes for improved efficiency, predictability, quality and supportability.

We have processes that manage data that vary from highly manual to partly automated. We should automate these processes to improve efficiency, consistency, recovery, agility, predictability, quality and supportability. Automation in line with agreed practises will also help transparency and understandability for other technical people, and aid when onboarding new staff.

This again highlights the importance of changing the way we work. We do not have the skills or capacity across LI to take full advantage of automation. This is especially true concerning a shift in technology and to the cloud. For example, historic directory monitoring scripts may now be replaced with S3 events. Correct use of cloud technology and understanding of event-driven design would change the solution completely.

5.5 Platform Services

EPIC

Provide centralised system services to enable teams with consistent, standardised and high-quality capability and data management workflows.

Platform services are enabling technology building blocks that are managed outside of the system that uses it. This includes cloud technology as the platform we build on and other software services like data warehousing and data lakes.

Each business unit has its systems and processes for everyday data management activities, which sometimes misses critical features. Providing centralised data infrastructure platform services provides teams with high-quality data management workflows without creating services and tools themselves.

Where these services are in-house (e.g. Geostore), they should be owned and operated by the respective product teams. Where these services are part of the cloud platform (e.g. AWS Relational Database as a Service) they should be supported by platform teams.

Using existing tools smartly is the highest and most efficient form of re-use. We should constantly look for opportunities to use services provided by the platform itself or extract common service capability from our products and establish standalone services. Doing so allows consumer services to benefit from improvements made to the platform service.

Platform Service – An enabling technology building block that is managed outside the system that uses it. This includes cloud technology as the platform we build on and other software services like data warehousing and data lakes.

5.6 Domain Modelling

EPIC

Document and confirm products and systems are split along domain boundaries, creating cohesive and loosely coupled solutions. This allows product teams to work independently.

We should be decomposing our IT systems into distributed services built around business domain capabilities and try to keep them in manageable sizes. As explanation of the problem, we may consider that Landonline manages both Survey and Title together in one system. Where this happens, we should break our systems up to allow more agile and rapid development. We are already on this path (i.e Addressing and Geodetic are out of Landonline now) but we need to sharpen our effort even more within LI. We should not have single databases that have multiple systems relying on them.

It will be valuable for us to capture and model the real-world entities we work with and their relationships. Domain modelling will present our systems as a set of domain objects that collaborate to fulfil our data management requirements. It will help solidify our domain terminology and provide a basis for discussing the events and data in our systems. This will give us a better understanding of our problem, solution and information spaces. Identifying domain entities and their relationships provides an effective basis for understanding.

Once we have created our domain models, this will help drive independent products and services and how to develop, deliver, and operate them. This, in turn, is foundational to enabling rapid and incremental development of products by product teams.

6 Improving Business Intelligence

THEME

Create a capability to support business intelligence in LINZ to drive better decision making.

To make good decisions, we need good supporting analytics. However, in many cases, such as customer usage metrics, this information is not available or hard to access and process in a timely fashion. In addition, we simply don't have a complete view of how our systems and data is used, which creates inefficiencies in the way we invest our time and money.

6.1 Business Analytics

EPIC

Create a capability to enable smart strategic investments, data acquisition choices and decision making.

Channels metrics are not centrally available or standardised. As a result, we do not understand what channels to grow, how to programme our work on an effective on-going basis, and what products to retire. Our disparate data delivery channels and lack of service analytics make it challenging to make data-informed decisions when we need.

We need to have key enabling data management components in place before establishing an efficient and complete business intelligence capability. This may in part be enabled by converging data channels, standardising data distribution metrics, central logging and a single data landing page. Of course, we will not solve the entire problem at once, but progress towards better metric storage will enable better analytics.

Cloud platforms, appropriate analytics tooling, and the adoption of an event streaming platform are vital enablers to collecting data across disparate systems and data sources, for both spatial and business systems, into a data lake or data warehouse.

Central access to system usage metrics will enable high-value business analytics opportunities to ensure we invest efficiently, spending less time on systems and data that have little value to our customers.

6.2 Data Warehousing

EPIC

Converge effort and resource on establishing a data warehousing capability in LINZ

Unlike data lakes, which mainly aim to store vast amounts of unstructured data, a data warehouse stores structured data, often containing large amounts of historical and summarised data from multiple sources used for analysis.

We have fragmented data warehousing and decision-making capability, which in part overlaps with services that Property Rights or the enterprise could provide. Therefore, there is an opportunity for creating a common capability in LINZ. We need to set clear goals and identify business value and datasets that would most benefit from representation in a data warehouse. This may include broader considerations such as the inclusion of survey and title data, property rights, and corporate data. A decision is needed on converging effort and resources around this in LINZ.

7 Next Steps

The strategic review captured strategic themes and epics on data management. However, some actions are required in a broader sense, such as improving the way we work. The first action is to create a single portfolio view starting with the themes and epics in this document, adding in the overall group value streams and epics. This will provide a consolidated portfolio that we can effectively size and prioritise during the planning phase. This will enable a balanced view for delivering the data management recommendations together with other business priorities.

7.1 Improving the Way of Work

LINZ has partially adopted some agile practices. However, our way of working is still an ineffective mixture of agile and legacy practices.

Our recommendation is that immediate attention and action are needed to change the way we work (see section 4). With Organising Ourselves Phase 2 already in full swing, we must use this opportunity to align the way we organise ourselves with agile enablement practices.

The following epics will provide immediate benefit to our structure, processes and capability to deliver effective data management.

Epic	Description
Organisational Structure	Align organisational structure with agile roles and responsibilities. Lay the foundation to prioritise with purpose, communicate better, enable product teams, focus on value and deliver efficiently.
Community of Practice	Establish, shape and drive our communities of practice to become tools for collaboration, standardisation and knowledge sharing.
People & Capability	Create a shared common capability around platforms, people and squads to support product teams.

7.2 Improving Data Management

A key component of Fixing the Foundations is to improve our data management practice. The **Improving Data Management** theme (section 5), along with **Improving the Way We Work** will set LI down the path to success. The table below lists recommendations for the first epics to consider.

Epic	Description
Data Governance	Create a framework to assess our maturity and discover gaps in data management.
Master Data Management	Improve master data management to ensure control, uniformity, accuracy, stewardship, consistency and accountability of LINZ's master data assets.
Metadata Management	Implement modern metadata standards and user interfaces for these standards into a data catalogue to manage all metadata centrally.

8 Appendix – Backlog

* For prioritisation as part of a broader portfolio view

Strategic Theme	Creating Customer Value	Improve customer experience. Respond to the changing needs of customers. Prioritise on customer value
Epics	Customer Insights	Collect metrics, improve understanding of customers and the services they use and incorporate the understanding into our product decisions.
	Data Publishing	Improve data delivery and discovery to align with changing customer needs and strategic programs of work.
Strategic Theme	Improving the Way We Work	Aligning to agile management, data and development practices to respond quickly to change, focussing on continually delivering value to our customers, making our work open and transparent and building a learning culture.
Epics	Organisational Structure	Align organisational structure and ownership with agile roles and responsibilities. Lay the foundation to prioritise with purpose, communicate better, enable product teams, focus on value and deliver efficiently.
	Community of Practice	Establish, shape and drive our communities of practice to raise our level of maturity by becoming tools for collaboration, standardisation and knowledge sharing.
	People & Capability	Create a shared common capability around platforms, people and communities of practice to support product teams. Upskill our existing capability in product teams to work with our modern platforms.
Strategic Theme	Improving Data Management	Improving how we manage our data
Epics	Data Governance	Create a framework to assess our maturity and discover gaps in data management to support data owners with consistent data management principles and practices.
	Master Data Management	Improve master data management. Ensure access, standardisation, completeness, accuracy and ownership of LINZ's master data.
	Metadata Management	Implement modern metadata standards. Create a data catalogue and manage all metadata centrally.
	Common Data Management Tools	Establish common tooling for data management to reduce the complexity of supporting multiple technology sets and support convergence in data management processes and standards.

	Data Supply	Extend our data catalogue to allow suppliers to upload data and capture metadata to reduce delays and double handling of data by LINZ.
	Data Quality	Implement and automate data quality management. Review and establish quality best practices and data quality metrics.
Strategic Theme	Designing Systems for Agility	Evolve architecture, design thinking, integration and platform capability to enable and align with autonomous teams and agile practices.
Epics	Cloud Enablement	Embrace cloud platforms enabling operational excellence through new capabilities and self-service.
	Cloud Migration	Accelerate and complete migration to AWS to enable centralised data access and reap the full benefit of cloud enablement.
	System Integration	Decouple systems by introducing better integration technology, practices and patterns.
	Automation	Automate manual processes for improved efficiency, predictability, quality and supportability.
	Platform Services	Provide centralised system services to enable teams with consistent, standardised and high-quality capability and data management workflows.
	Domain Modelling	Document and confirm products and systems are split along domain boundaries, creating cohesive and loosely coupled solutions. This allows product teams to work independently.
Strategic Theme	Improving Business Intelligence	Create a capability to support business intelligence in LINZ to drive better decision making
Epics	Business Analytics	Create a capability to enable smart strategic investments, data acquisition choices and decision making.
	Data Warehousing	Converge effort and resource on establishing a data warehousing capability in LINZ

Released under the Official Information Act 1982

9 Appendix – Data Management Principles

LINZ leverages guiding data management principles to influence data objectives and essential capabilities. These guiding principles are foundational to all data efforts.

The sections below provide guidance in meeting key principles. They generally relate to key themes from dataset review recommendations and are designed to be measurable. Meeting principles through these objectives will often involve a broader LINZ effort and should not be seen as a series of tasks that individual data managers need to tackle in isolation.

9.1 Background

This appendix sets objectives that would help LINZ to meet its [Data Management Principles](#) provided by the Geospatial Data Management Review (GDMR), with a particular focus on addressing the Related [Key Themes](#) that emerged from the recommendations.

The Requirements for Measures are provided to consider what progress LINZ needs to measure for each objective. The identified priority is for a LINZ data catalogue that allows data to be collated, discovered and monitored more easily. This data catalogue should be presented in a way that makes reporting, resourcing, decision-making and prioritisation easy, fast and efficient.

In some cases, objectives do not relate to Key Themes. This may be because LINZ is already performing well against these objectives or because this aspect of data management was not considered in the GDMR. These objectives are still included here as recognised as being common and important objectives out of similar pieces of work that we should be aligned to. References to? are linked throughout.

9.2 Data is a Strategic Asset

Guided by our whakataukī and purpose, [LINZ Kaupapa](#) is to deliver high value geographic and property information, provide a world-class property system, make the best use of Crown Estate and manage overseas investments to serve New Zealand's changing needs. Data is at the centre of these outcomes and LINZ is committed to providing products, services, tools and systems that deliver fit-for-purpose datasets to support decision-making and meet customer needs.

LINZ will know it is meeting the “Data is a Strategic Asset” principle when:

- **Objective 1: LINZ understands the value of its datasets by using analytics and customer insights.**
 - Related Key Themes
 - There is no portfolio view of the datasets, and no way to assess how they contribute to the overall goals or value propositions of LI.
 - Requirements for Measures
 - LINZ has a portfolio view of its datasets.
 - LINZ monitors customer usage (from channel analytics), queries, survey results and use cases over time.

- **Objective 2: LINZ has a portfolio view of its datasets, and prioritises effort based on how they contribute to LINZ's desired outcomes.**
 - Related Key Themes
 - Most teams are being left to prioritise work themselves. This leads to silos, inconsistent prioritisation and inconsistent allocation of resources.
 - Most datasets do not have consistent governance for data management systems or data management teams.
 - Requirements for Measures
 - LINZ has a method for dataset maintenance and dataset improvement prioritisation based on demonstrated value and contributions to desired outcomes.
 - LINZ monitors resourcing against prioritised datasets.

- **Objective 3: LINZ datasets have a customer-centric product backlog.**
 - Requirements for Measures
 - Every LINZ dataset has a location in JIRA where a product backlog is maintained.
 - Epics, stories and tasks are actively tagged / curated / triaged / refined before being processed.
 - LINZ monitors activity in JIRA and ensures that work being completed is also being logged and tracked.

9.3 Data Stewardship

To ensure data reaches its full potential, LINZ is defining data and information roles and responsibilities and embedding these in operations. LINZ will assign domain stewards, data managers, and a set of functional data maintainers to achieve accountability throughout the entire data lifecycle. Domain stewards lead the case for investment in data and establish policies governing data co-design, access, use, protection, quality, dissemination and depreciation. Data managers are responsible for promoting the value of data and enforcing policies, ensuring data is fit-for-purpose, and functional data maintainers implement the policies and manage day-to-day quality. More details on these roles and responsibilities can be found in the [LINZ Data and Information Management Policy](#).

LINZ will know it is meeting the " Data Stewardship" principle when:

- **Objective 1: LINZ datasets have a data management plan.**
 - Related Key Themes
 - Most datasets do not have a fully formed [data management plan](#). Introducing a practise of creating and maintaining data management plans would help data managers consider what and how the data will be collected, stored, licensed, protected, what standards to use and when ethical issues are addressed.
 - Requirements for Measures
 - LINZ has a data management plan template.
 - LINZ monitors how many datasets have an actively maintained, peer reviewed data management plan.

- **Objective 2: LINZ datasets have identifiable and accountable Domain Stewards, Data Managers and Data Maintainers embedded in operations.**
 - Related Key Themes

- Most roles and responsibilities are only informally established and are not embedded in operations. This increases the likelihood of unmanaged risks (e.g. data loss), sub-optimal business continuity, compromised security or ineffective data discovery.
 - Requirements for Measures
 - Completed data management plans include formally identifying the Domain Stewards, Data Managers and Data Maintainers.
 - Who is accountable for these roles for a specific dataset can be easily identified by anyone at LINZ.
 - LINZ monitors how many datasets have identified Domain Stewards, Data Managers and Data Maintainers.
- **Objective 3: LINZ datasets have a disaster recovery plan, a business continuity plan (BCP) and a risk register.**
 - Related Key Themes
 - Most datasets have a disaster recovery plan, however, it is generally untested, or the responsibility is assumed by and/or put onto a third party (e.g. Datacom). An untested disaster recovery plan is unlikely to work as intended on the first attempt, introducing a high degree of uncertainty into our capability to recover from data loss.
 - Few datasets have no disaster recovery plan. A disaster recovery plan helps to reduce the potential for data loss and quickly restore services when an incident occurs.
 - Most datasets do not have a BCP. If the LINZ office/network becomes unavailable for a large time, these datasets will be unable to be modified.
 - Most datasets do not have a risk register. Maintaining a risk register helps to identify, assess and mitigate any potential risks, reducing the likelihood or severity of any incidents that impact the dataset.
 - Requirements for Measures
 - LINZ has guidance, templates or standardised methods for creating disaster recovery plans, business continuity plans and risk registers.
 - Completed data management plans include links to a disaster recovery plan, a BCP and a risk register.
 - LINZ monitors how many datasets have regularly reviewed and tested disaster recovery plans, BCP and risk registers.
- **Objective 4: LINZ implements master data management, ensuring consistency and accountability over shared data assets.**
 - Related Key Themes
 - Check how often master data management recommendations were surfaced - add a key theme if we missed it earlier.
 - Requirements for Measures
 - LINZ has a portfolio view of its datasets.
 - LINZ monitors occurrences of duplication of effort, particularly with data collection and ingestion.

9.4 Data Ethics

Data ecosystems should be designed and function in ways that consider the rights, views and decisions of all relevant stakeholders, strengthen Māori-Crown relations and enable all New Zealanders to derive clear and collective benefits from the data. Analytical processes are tools to support decision-making and should never entirely replace human oversight.

LINZ will know it is meeting the “Data Ethics” principle when:

- **Objective 1: LINZ considers the privacy and ethical implications of any data collected about people.**
 - Related Key Themes
 - Refer to Privacy Commissioner / Stats New Zealand principles for safe and effective use of data and analytics.
 - Requirements for Measures
 - LINZ has a Privacy Officer, whose role is...
 - LINZ monitors how it handles privacy information against the [Privacy Maturity Assessment Framework](#).
 - LINZ monitors datasets with ethical considerations in alignment with the [Principles for Safe and Effective Use of Data](#) and the [Algorithm Charter](#).
- **Objective 2: LINZ consults with stakeholders and Māori as partners to ensure data is highly trusted, inclusive and appropriately protected.**
 - Related Key Themes
 - Refer to Privacy Commissioner / Statistics New Zealand principles for safe and effective use of data and analytics and the Principles of Māori Data Sovereignty.
 - Requirements for Measures
 - LINZ monitors external stakeholder engagement across its datasets.
- **Objective 3: LINZ datasets deliver clear public benefits for all New Zealanders.**
 - Related Key Themes
 - Refer to Privacy Commissioner / Stats New Zealand principles for safe and effective use of data and analytics.
 - Requirements for Measures
 - LINZ data management plans include a clearly defined purpose and benefit analysis.
 - LINZ monitors how many datasets have documented benefits.

9.5 Data Collection

Data is collected or generated for specified public policy, operational business, or legislative purposes, with an aim to continuously add value to best inform the customer. Consequently, LINZ must enable electronic collection of data at the point of creation and ensure traceability of that data at all times. The moment data is created, it should be tagged, stored, and catalogued. When the data is combined or integrated, the resulting product must also be immediately collected, tagged, curated, and appropriately secured. To expedite these processes and to minimise the risk of human error, these steps should be automated to the maximum extent possible.

LINZ will know it is meeting the “Data Collection” principle when:

- **Objective 1: LINZ utilises common platforms and services to create, store, retrieve, share, utilise, and manage data.**
 - Related Key Themes
 - Most datasets have manual ingestion processes. This can take a lot longer (depending on the size of the dataset) and introduces risks of data loss.
 - Some datasets have manual publishing processes with repetitive tasks that could be automated.
 - Most datasets rely on ad-hoc scripts that have been created in an unstructured fashion. This makes it difficult for developers or data analysts to easily understand and get up to speed on processes in other areas.
 - Requirements for Measures
 - LINZ creates common platforms and services that datasets should use for key common processes.
 - LINZ monitors how many datasets use these common platforms and services.
- **Objective 2: LINZ utilises common open metadata standards that allow data to be joined and integrated.**
 - Related Key Themes
 - Most datasets store and publish metadata in an outdated format (e.g. ISO 19115). This standard is no longer current and is not being updated to respond to changes in customer needs.
 - Requirements for Measures
 - LINZ identifies common metadata standards that should be adopted.
 - LINZ monitors how many datasets use these common metadata standards.
- **Objective 3: LINZ data has access, lineage, and data quality metadata bound throughout its lifecycle.**
 - Related Key Themes
 - Most datasets do not store lineage or processing metadata. This makes it difficult or impossible to track where data came from, what software was used to process it and what processes or modifications were applied.
 - Requirements for Measures
 - LINZ specifies a standard way to record and store access, lineage and data quality metadata.
 - LINZ monitors how many datasets have compliant metadata.

9.6 Data Access and Availability

LINZ works with users to deliver the information they need, enable data to be connected, make data easier to find and use and to provide nationally significant geospatial data, information, products and services. This is enabled by successful implementation of enterprise capabilities, such as an enterprise cloud, identity, credentials, and access management and associated data-sharing tools. Data and information held by Government should be open for public access unless grounds for refusal or limitations exist under the Official Information Act or other government policy. For instance, Personal, confidential and classified information are protected.

LINZ will know it is meeting the “Data Access and Availability” principle when:

- **Objective 1: LINZ datasets are catalogued in a central repository including location and access methods for shared data.**
 - Related Key Themes
 - Some datasets do not have metadata accessible outside of their team. This makes data discovery impossible by others.
 - Some datasets do not store their metadata in a reusable and open format. This means that the metadata is not currently consistently stored, to enable tools and APIs to be built that can traverse our various datasets.
 - Requirements for Measures
 - LINZ has a central metadata catalogue.
 - LINZ monitors how many datasets are catalogued and whether they include access metadata.
- **Objective 2: LINZ datasets are marketed to our customers based on a customer segmentation model.**
 - Related Key Themes
 - Refer to actionable recommendations in the [Channels Review Final Report](#).
 - Requirements for Measures
 - LINZ has a customer segmentation model.
 - LINZ monitors how it is marketing datasets for each customer segment.
- **Objective 3: Data access and sharing is controlled through standardised reusable open APIs.**
 - Related Key Themes
 - Refer to [New Zealand Government Data Toolkit Principles](#).
 - Requirements for Measures
 - LINZ has common data sharing platforms that provide data using standardised reusable APIs.
 - LINZ monitors the APIs that are used to share datasets and ensures that they are standardised and reusable.
- **Objective 4: Data and information held in proprietary formats are also released in open, non-proprietary formats.**
 - Related Key Themes
 - Refer to [New Zealand Government Data Toolkit Principles](#).
 - Requirements for Measures

- LINZ has common data publishing platforms that provide data in open, non-proprietary formats (in addition to common proprietary formats where necessary due to customer needs).
 - LINZ monitors the formats that datasets are available in and ensures all datasets available in open, non-proprietary formats.
- **Objective 5: Personal, confidential and classified data and information are protected.**
 - Related Key Themes
 - Refer to [New Zealand Government Data Toolkit Principles](#).
 - Requirements for Measures
 - LINZ identifies which datasets contain personal, confidential and classified information.
 - LINZ datasets that contain personal, confidential and classified information have management plans for how that data is handled.
 - LINZ monitors how this information is protected.

9.7 Data is Fit for Purpose

Fit-for-purpose data is quality data that is readily discoverable, easily interpreted and meets the needs of the customer. Data is curated and published by considering the breadth of customer needs, in order to generate maximum value. Ensuring that data is fit-for-purpose should include careful consideration of any ethical concerns in data collection, sharing, use, representation of the information intended, rapid data integration, and minimisation of any sources of unintended bias.

LINZ will know it is meeting the “Data is Fit for Purpose” principle when:

- **Objective 1: LINZ executes automated data quality management techniques to assess and enhance data quality.**
 - Related Key Themes
 - Most datasets do not track the [6 standard data quality metrics](#). Tracking these metrics would instil confidence that data is fit-for-purpose and enable reporting required for good prioritisation at a senior level.
 - Most datasets have some data validation as part of their publishing, but this is generally not monitored. Monitoring these would provide customers with more confidence in our data quality.
 - Requirements for Measures
 - LINZ has a standardised way of storing data quality metadata.
 - LINZ monitors how many datasets have data quality metadata and the data quality itself.
- **Objective 2: Data is presented in a way that preserves semantic meaning and is expressed in a standardised manner throughout LINZ.**
 - Related Key Themes
 - Refer to [Department of Defense Data Strategy](#).
 - Requirements for Measures
 - LINZ operates using a shared vocabulary.

- **Objective 3: All LINZ data and systems issues are tracked in JIRA and monitored to determine opportunities for improvement.**
 - Requirements for Measures
 - Every LINZ dataset has a location in JIRA where issues are logged.
 - Issues are actively tagged / curated / triaged / refined before being processed.
 - LINZ monitors activity in JIRA and ensures that work being completed is also being logged and tracked.

9.8 Design for Data Compliance

LINZ must implement system solutions that automate the data and information management lifecycle, properly secure data, and maintain end-to-end records management. Compliance with data management policies is a critical success factor.

LINZ will know it is meeting the “Design for Data Compliance” principle when:

- **Objective 1: Only authorised users can access, edit and share data.**
 - Related Key Themes
 - Refer to [New Zealand Government Data Toolkit Principles](#).
 - Requirements for Measures
 - LINZ has a centralised logging solution.
 - An audit log is persisted using the centralised logging solution.
 - LINZ monitors access, editing and dissemination of data and unauthorised actions are alerted.
- **Objective 2: LINZ data systems and data maintenance processes are well documented.**
 - Related Key Themes
 - Some datasets have data maintenance processes that are not well documented and rely on single staff. This creates a single point of failure that may jeopardise our ability to deliver the datasets.
 - Requirements for Measures
 - LINZ has templates/examples of what good data documentation looks like.
 - LINZ monitors how many datasets have process documentation.
 - LINZ monitors how many datasets have only one Data Maintainer.

10 Appendix – Review Team

Name	Position	Role
Jeremy Palmer	Director Data	Product owner
Richard Fone	Senior Project Manager	Project Manager and Scrum Master
Adam Greenland	National Hydrographer	Business Owner
Anna Meissner	Senior Marine Geospatial Data Specialist	SME - Hydrography
Jonathan Ball	Manager Open Data and Reuse	SME - Publishing
Daniel Silk	Technical Leader	SME - Topography
Jack Reyneke	Solution Architect	Architect
Blayne Chard	Solution Architect	Architect
Bill Nelson	Manager Data Infrastructure	SME - Data
Trent Gulliver	Manager Addressing	SME - Addressing
Nic Donnelly	Manager Positioning	SME - Positioning

Released under the Official Information Act 1982