NEW ZEALAND
RESEARCH
INFORMATION
SYSTEM

# NZRIS AND DATA MANAGEMENT

### Get in touch with the NZRIS team

If you have any questions at any time, email us at nzris@mbie.govt.nz – we're here to help.

### For more information

See our website at www.mbie.govt.nz/nzris

MINISTRY OF BUSINESS,
INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI

New Zealand Government

# OVERVIEW

## WHY GOOD DATA MANAGEMENT IS IMPORTANT

The way data is managed in the New Zealand Research Information System (NZRIS) is important as we are required by law to protect the information we collect. However, we also need to ensure that organisations providing data have trust and confidence that their data is being handled appropriately, and that people accessing information through NZRIS know it is consistent and accurate.

NZRIS will hold a diverse range of information from New Zealand's research, science and innovation organisations about research funded by the New Zealand Government and potentially privately-funded research in future. In the initial stages, much of the data held in NZRIS will already be accessible to the public. Currently this information is held in a variety of places and formats through public websites, or it is the type of information that can be requested from the source organisation. NZRIS will make this information more readily accessible to a wider audience as part of an open data approach to research information.

While we support this approach, we also recognise that there will be cases where submitted data is confidential and does need to be protected. An example of this includes information about research applications where the decision on the application has not yet been made.

NZRIS incorporates a number of systems and processes that are designed to achieve a good balance between openness and transparency of data, the protection and security of data that is commercially sensitive or raises privacy issues, and ensuring data is as accurate as possible.

## APPROACH TO DATA MANAGEMENT

The NZRIS data management approach aligns with the Government's approach to data and information management, and similar approaches used by highly trusted organisations such as Stats NZ.

### Data principles

Underpinning NZRIS are seven principles that have guided the development of NZRIS. These principles are to:

1. provide a system-wide view of research, science and innovation information
2. ensure open data which is easily accessible and widely used
3. protect personal and commercially sensitive data
4. enable the re-use of data
5. reduce collection and reporting burdens
6. ensure data is trusted, authoritative and well-managed
7. enable easy and automatic movement of data between systems.

### Two key approaches

NZRIS uses two key approaches to ensure data is managed effectively in the system:

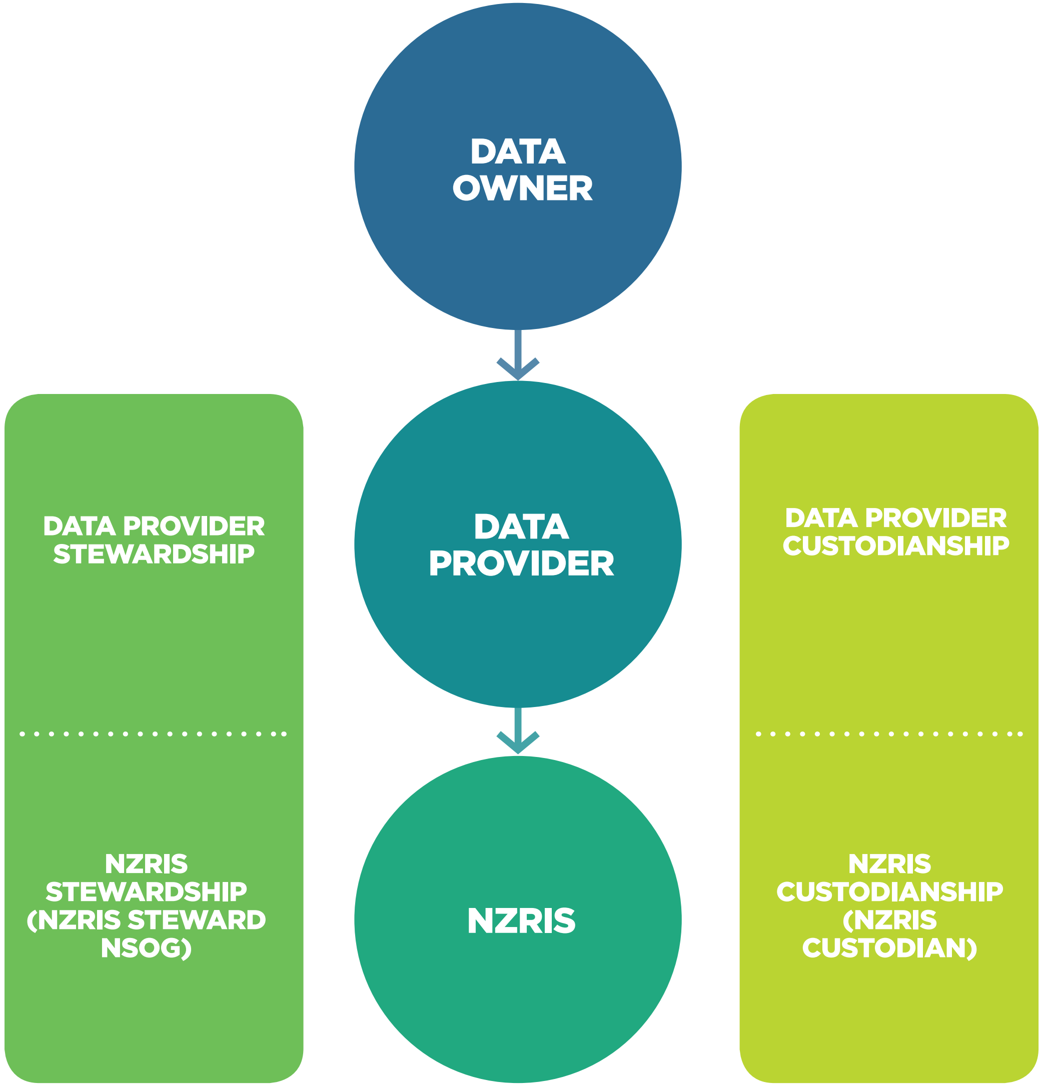
1. the adoption of well-defined and clearly agreed **roles and responsibilities**
2. the implementation of robust **systems and processes** for handling data within NZRIS.

# ROLES AND RESPONSIBILITIES

It is important that any organisation involved in submitting or managing 'NZRIS data' has a clear understanding of the different roles and responsibilities in relation to data. We have defined four key data management roles – data provider, data owner, NZRIS custodian and NZRIS steward.

| ROLE | RESPONSIBILITY |
|---|---|
| Data Provider | A data provider supplies data to NZRIS. A data provider could be: an Asset Pool Manager (for example, a research funder such as MBIE or the Health Research Council); or a Research Science and Innovation Manager (for example, a Crown Research Institute). Some organisations are both. A data provider may also be the data owner, but not always (see "data owner" below). Where the provider is submitting data they don't own, they may need to check with the data owner to ensure the correct data protection has been applied.<br>A data provider is responsible for:<br>• submitting data to NZRIS<br>• working with NZRIS team on data quality issues<br>• requesting suppression of previously approved data<br>• approval of data progressing to the public environment<br>• responding to data concerns and feedback raised by public users of NZRIS<br>• working with data owners to resolve data quality issues, when feedback is received.<br>The data provider organisation also has responsibility for custodianship and stewardship of their own data, data management and processes whilst the data is within their organisation. This means they are also responsible for the administrative care of data, and ensuring it remains secure, accessible and well-managed. |
| Data Owner | A data owner is responsible for the production, maintenance and quality of their data. A data owner may also be a data provider.<br>Specific responsibilities of a data owner include:<br>• informing the data provider of data that should be kept protected, and how<br>• working with data providers to resolve quality issues, when feedback is received.<br>Note that research organisations or individuals that are only data owners generally do not interact with NZRIS directly, but rather through a data provider. |
| NZRIS Custodian | The NZRIS custodian has administrative care of data, and ensures it remains secure, accessible and well-managed. In the case of NZRIS the data custodian role sits with the NZRIS team.<br>Specifically, the NZRIS custodian is responsible for:<br>• working with data providers to resolve data quality issues<br>• approving data before it is published.<br>As NZRIS is not a master data source but instead includes exact copies of data submitted by data providers, the NZRIS custodian is not responsible for the accuracy, integrity, and timeliness of the submitted data, and is unable to make any changes to the submitted data. This responsibility sits with the data owner and data provider. |
| NZRIS Steward | The NZRIS steward is responsible for data rules and agreeing quality parameters. NZRIS stewardship will ultimately be the responsibility of the cross-sector NZRIS Stewardship and Oversight Group (NSOG).<br>Key responsibilities of the NZRIS steward include:<br>• setting data policies and rules and ensuring the NZRIS custodian follows these<br>• monitoring of good practices to manage the data and information across its life-cycle. |

# ROLES AND RESPONSIBILITIES


DATA OWNER
DATA PROVIDER STEWARDSHIP
DATA PROVIDER
DATA PROVIDER CUSTODIANSHIP
NZRIS STEWARDSHIP (NZRIS STEWARD NSOG)
NZRIS
NZRIS CUSTODIANSHIP (NZRIS CUSTODIAN)

This diagram builds on the information in the table on the previous page. It illustrates how the various roles relate to each other. You can see that a data provider still needs to have custodianship and stewardship roles over its own data, and that NZRIS has custodianship and stewardship roles over the data it holds.

# SYSTEMS AND PROCESSES FOR MANAGING DATA

NZRIS has two key ways of managing data effectively through the systems and processes. These are:

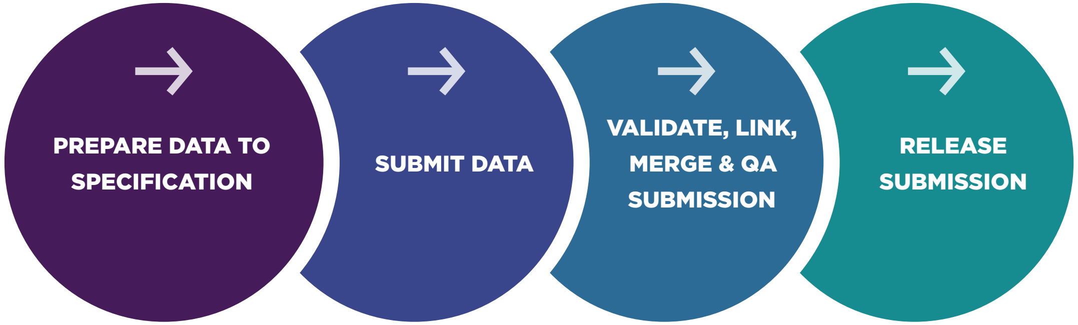
1. the data submission and publication process
2. the data protection principles and patterns.

## DATA SUBMISSION AND PUBLICATION PROCESS

All data in NZRIS is linked to the submission it came from, for approval, update and feedback purposes. The key features that underpin the data submission process and support the protection, security and accuracy of data submitted to NZRIS are set out below.

- NZRIS will only accept data that aligns to the NZRIS data specifications. At the point of submission records will be validated against this specification.
- An automated process will notify the NZRIS support team and data provider if a record has been rejected because it does not meet the specification, or if there are data quality warnings.
- When validation checks identify any records that don't conform to the required specification, the data provider must submit updated data before it can be approved for release.
- The data provider is responsible for the quality of the data; however the NZRIS custodian and system tools will assist with this.
- The NZRIS custodian and authorised data providers are jointly responsible for authorising the public release of data.

The diagram below sets out the key stages of the data submission process.


PREPARE DATA TO SPECIFICATION
SUBMIT DATA
VALIDATE, LINK, MERGE & QA SUBMISSION
RELEASE SUBMISSION

These stages and the defined steps in the submission process help to ensure that the responsibilities of the different roles are clear, and support data accuracy.

## DATA PROTECTION PRINCIPLES AND PATTERNS

In order for NZRIS to be successful, in both the short and long term, it's essential to have the trust and confidence of data provider and owners. This means ensuring the privacy, confidentiality and security of NZRIS data.

In practical terms, this has meant developing a principle-based approach to data protection and developing protection patterns that apply to the data that is submitted into NZRIS.

### Protection principles

The NZRIS data protection principles set out the framework for how data that needs to be protected is managed. The details of these five principles are on page 6 of this document, but at a summary level these principles are:

1. NZRIS will not cater for or accept any data that is classified above Sensitive (based on the Guidelines for Protection of Official Information).
   *This means NZRIS will only accept data classified as Public Information, In-Confidence, Restricted and Sensitive.*

2. The data owner is responsible for establishing the level of protection required for the data.
   *This means that while the data provider needs to apply the appropriate level of protection to submission data, if they are not also the data owner they may need to consult with the owner to ensure the correct level of protection is applied. They may collect this information from the data owner early in the funding process (for example, as part of the application process).*

3. The NZRIS user access profile defines the data that can be seen by the user.
   *This means that access to the data in NZRIS is specific to a role. For example, within the NZRIS custodianship application, data providers will only be able to see data that they have submitted (both protected and public), while people using the public interface will only be able to see public data.*

4. The protection patterns will control data and aggregate data availability.
   *This means that the patterns applied to the data will control which data is allowed to progress to the public interface.*

5. Protection will be categorised into three distinct types of implementation.
   *This means that some data types will have default or mandatory protection applied; while other types will have optional protection applied that the data provider may be able to change.*

### Protection patterns

The NZRIS data protection patterns are the specific technical protections that are applied to NZRIS data. These can be applied to data by the data provider or data owner. The details of the patterns and how to use them are set out in the NZRIS Data Specification document; however examples of these include (amongst others) protection for:

- application review results
- commercially sensitive financial information
- personal demographic information.

These patterns are applied during submission and ensure that data that needs to remain confidential is suppressed rather than published. All records submitted to NZRIS must have a protection pattern applied.

# DATA PROTECTION PRINCIPLES

Details of the five data protection principles are set out below.

## 1

### NZRIS will not cater for or accept any data that is classified above Sensitive (based on the Guidelines for Protection of Official Information).

- NZRIS will only accept data that is classified as Public Information, In-Confidence, Restricted and Sensitive.

NZRIS will not accept data that is classified as Confidential, Secret or Top Secret.

## 2

### The data owner is responsible for establishing the level of protection required for the data.

- The data provider is accountable for providing the protection level of the data when they submit it to NZRIS. If they are not the data owner, they will need to collect this information from the data owner (for example, as part of the application process).
- The data provider will identify the protection patterns (set out in the NZRIS data specification document V2.0) that are to be are applied at entity/record level.
- If no pattern is identified, the data will not be accepted into NZRIS.

If none of the standard specific patterns are appropriate, but the data is still considered protected, either the Unknown Protection or the Work in Progress Protection pattern should be applied.

## 3

### The NZRIS user access profile defines the data that can be seen by the user.

- NZRIS Custodians and NZRIS system administrators will be able to access all NZRIS data.
- The data provider's authorised users will be able to see all data provided by their organisation.
- NZRIS may provide approved researchers with access to sensitive data, at some point in the future.
- All other access will be controlled by patterns that restrict access.

Accessible data will reflect all the protection patterns applied to the underlying data, and the access profile of the specific user.

## 4

### The protection patterns will control data and aggregate data availability.

- The protection will be enforced at entity level and controlled at pattern level.
- Protection pattern enforcement is the sum total of, and most restrictive of, all applied patterns from all requesting parties.
- Where data has been provided more than once with conflicting protection patterns, all impacted records will be protected and treated as sensitive data while the conflict is resolved.

## 5

### Protection will be categorised into three distinct types of implementation.

#### Default

- Standard for certain types of data
- May be overturned if data provider identifies another pattern on submission.

#### Optional

- Identified by data provider on submission
- Options for data providers to choose:
  - No expiry date: data will be released to the public
  - Specific date expiry: data provider enters a specific date when supplying data to NZRIS, then data will be released on that date.

#### Mandatory

- Standard for certain types of data
- Cannot be overturned by request from data provider