Waipapa Taumata Rau Research Data[1] Management Capability Maturity Model

	Level 1 INITIAL *Process disorganised & ad hoc*	Level 2 DEVELOPMENT *Process is under development*	Level 3 DEFINED *Process is standardised and communicated*	Level 4 MANAGED *Process is managed and measured*	Level 5 OPTIMISED *Focus is on continuous improvement*
Strategy, Policies, & Governance	- Strategy, policies and procedures may be underdeveloped, not up to date, and/or inconsistent - Institutional awareness is low - Institutional governance of research data are ad hoc or absent	- Strategy, policies and procedures, governance mechanisms are developing and harmonised for specific tasks - Institutional awareness of roles and responsibilities is increasing	- Strategy, policies and procedures, and governance are defined and operationalised into research practice - Widespread institutional awareness of roles and responsibilities	- Strategy, policies and procedure, and governance are integrated and accepted as part of culture and subject to audit	- Institution is a recognised leader following international best practice in capability maturity roadmapping - Strategy, policies and procedures, and governance are accepted as BAU, audited and regularly reviewed to align with current best practice
Māori Data Sovereignty	- Little or no recognition of Māori data - Māori data stored without consideration of sovereignty issues	- Recognition of Māori interests in data - Māori participation in a data access committee(s) - Some integration of Māori principles (e.g. kaitiakitanga)	- Māori governance of data and protocols for data access - Māori data definition is consistent - Māori can access raw data about their collective on request - Full range of storage options, including NZ based, as required	- Māori ownership of data - Māori data access committee - Māori have full access to data about their collective -Widespread use of Māori principles and DM protocols	- Māori Data sovereignty principles are intrinsic to RDM at an institutional level - RDM is aligned with the principles of rangatiratanga, whakapapa, whanaungatanga, kotahitanga, manaakitanga and kaitiakitanga
Data Management	- Metadata management is chaotic and understood by only a few - Data quality measures are ad hoc or absent	- Responsibilities are defined and skills are developed - Simple DMPs available - Data collection guidelines in development	- Processes are standardised and integrated - All data are assigned an appropriate globally unique persistent identifier (e.g.DOI)	- All datasets described and metadata shared - DMPs used for provisioning and active management of data	- Continuous improvement applied to processes and capabilities - Integrated research ecosystem (e.g. leveraging multiple persistent identifiers to enable seamless RDM)
Data Operations	- Simple data sharing can be a challenge - Curation and preservation services absent or disorganised	- Project-based data sharing services become available - Curation practices are developing and awareness of the value of curation is increasing - Value and practice of preservation is not recognised	- Project-based data management practices are developed and documented - Data sharing is straightforward - Curation practices are standardised and widely understood - Preservation practices for selected data	- Data are FAIR: sharing becomes commonplace and embedded in practice; curation understood as critical; and, data are being preserved in alignment with policy	- In addition to data being FAIR, data management operations (sharing, curation and preservation) are fully automated and machine accessible throughout the organisation
Platforms, Infrastructure, & Architecture	- IT infrastructure is acquired, deployed, and managed inconsistently, not easily discoverable, supportable, nor documented as services	- Responsibilities for provisioning, maintaining, and lifecycle management are defined - Beginning to integrate instruments,, storage and transfer services, and research compute	- Widespread availability of data platforms and tools, including analysis, visualisation - Facilities are well-defined, standardised, and integrated - Automated provisioning of project infrastructure in development	- Funding adapts to platform needs - Platforms are well-managed within a defined research-delivery architecture - Auditing of platforms and architecture in place	- IT infrastructure management optimises the IT infrastructure evolved in previous levels through continuous focus on management and improvement of data assets
Skills & Support	- Data management planning is unsupported - Training is ad hoc or missing - QA is ad-hoc or absent - Staff are unable to locate required support or documents	- Investment in skills and processes - Data management planning is used on projects, documentation and training are developed	- Widespread availability and uptake of training and skills development in data management - QA becomes feasible on training and support services	- QA is routinely applied to processes, results feed into future planning	- Support services and training processes are optimised and periodically refined

---

[1] *Data* is used as an umbrella term for digital, print and physical datasets, artefacts, cultural taonga, research evidence and also covers the digital representation of a physical item used in research, that may need to be managed and shared.

Strategy, Policies, & Governance

1. Institutional RD strategy, Including Māori Data Sovereignty
2. Policies (institutional and small research unit, direct roles and responsibilities e.g. Code of Conduct, data classification, social licence)
3. Governance, including Māori Data Sovereignty, long term stewardship
4. Institutional awareness and culture

Level 1 INITIAL	Level 2 DEVELOPMENT	Level 3 DEFINED	Level 4 MANAGED	Level 5 OPTIMISED
Process disorganised & ad hoc	*Process is under development*	*Process is standardised and communicated*	*Process is managed and measured*	*Focus is on continuous improvement*
1.1 Strategy does not include or inform management of research data	1.1 Strategy is developing and beginning to direct institutional activity	1.1 Strategy is defined and operationalised	1.1 Strategy is accepted as part of culture	1.1 Strategy is regularly reviewed, improved and aligned with current best practice
1.2 Strategy does not recognise Māori interests in data	1.2 Strategy recognise Māori interests in data	1.2 Some use of Māori Data Sovereignty and CARE principles (guidelines) in strategy	1.2 RDM roadmap is informed by the institution's strategies and its researchers' priorities	1.2 Māori Data sovereignty principles are intrinsic to RDM at an institutional level, as evidenced in strategy, policies and governance
2.1 RDM policies and procedures are underdeveloped and/or out-of-date and/or inconsistently applied and do not include compliance measures	2.1 RDM policies and procedures are developed and some activities are coordinated or aligned	1.3 RDM roadmap in place - compliance-focussed and defined by funder requirements	2.1 Following RDM policies and procedures is normal behaviour across the University and compliance is checked	1.3 Strategy (roadmap) seeks to derive competitive advantage from RDM support. It aims to be sector-leading and innovative
2.2 RDM policies and procedures do not recognise Māori interests in data	2.2 Recognition of Māori interests in data within policy and procedures	2.1 RDM policies and procedures are widely followed	2.2 Institutional policies with a bearing on RDM (e.g. OIA, ethics, privacy, IP, research conduct, etc.) are joined up and complementary	2.1 RDM policies and procedures direct standard behaviours across the University and are reviewed regularly to align with good practice
3. RDM governance is absent or emerging in a handful of projects/programmes	2.3 Policies articulate roles and responsibilities for researchers, other staff and students to comply with legal and regulatory obligations and key external funders' RDM policy expectations	2.2 RDM provisions are integral to the Research Code of Conduct	2.3 Institutional strategy and policy is operationalised with policies/procedures within smaller business units	2.2. Faculty/LRSI/School specific roles deliver accountability of RDM framework
4. Institutional awareness of FAIR, CARE, Māori Data Sovereignty, and Privacy data principles is mostly absent	2.4 Research data policies are promoted to all relevant staff, students and researchers	2.3 Clearly defined operational practices and standards in place for PIs, including stewardship and Māori Data Sovereignty	2.4 Faculty/LRSI/School specific roles with responsibility to implement RDM framework	2.3 Feedback from partners and key stakeholder communities is actively pursued to improve policies
	2.5 Developing University research data classification, including sensitive data	2.4 Policies publicly articulate institution's social licence	2.5 Policies are externally shared and promoted, aiming to push the sector forward	3.1 Governance Framework is in place that supports all data interests with regular review and proactive adaptation in line with international best practice
	3.1 Data governance is developing - exists within specific projects, programmes, services	2.5 Institutional strategy and policy is operationalised within some policies/procedures within smaller business units	3.1 Governance Framework is in place	4. Strategy, policies and procedures are embedded
	3.2 Māori governance over Māori data is developing within specific projects and programmes	2.6 Guidance on how to apply all relevant policies to the institutional context is provided and promoted to all relevant staff, students and researchers	3.2 Māori governance over Māori data is established in strategy and policies and checked	
	4. Awareness of RDM policies, procedures, roles and responsibilities is increasing	2.7 Research data classification is aligned with administrative data classification	4.1 Policies are promoted by the institution through channels designed to engage with staff, student and researcher groups' specific interests	
		2.8 Clearly defined RD operational practices and standards in place for PIs		
		3.1 Scalable data governance structures and processes in place for research data		
		3.2 Māori governance over Māori data is defined in strategy and policies		
		4. Awareness of RDM policies, procedures, roles and responsibilities is the norm		

Māori Data Sovereignty

Māori data refers to digital or digitisable information or knowledge that is about Māori people, their language, culture, resources or environments
Māori Data Sovereignty refers to the inherent rights and interests that Māori have in relation to the collection, ownership, and application of Māori Data

Māori Data Sovereignty Principles Te Mana Raraunga | Māori Data Sovereignty Network https://temanararaunga.maori.nz
1. Rangatiratanga | Authority: 1.1 Control 1.2 Jurisdiction 1.3 Self-determination
2. Whakapapa | Relationships: 2.1 Context 2.2 Data disaggregation - collection and coding 2.3 Future use
3. Whanaungatanga | Obligations: 3.1 Balancing rights - individuals' rights and collective 3.2 Accountabilities - individuals and organisations
4. Kotahitanga | Collective benefit - 4.1 Benefit - individual and collective 4.2 Build capacity - workforce development 4.3 Connect
5. Manaakitanga | Reciprocity 5.1 Respect - dignity of Māori communities, groups and individuals 5.2 Consent
6. Kaitiakitanga | Guardianship: 6.1 Guardianship 6.2 Ethics - tikanga, kawa (protocols) and mātauranga (knowledge) 6.3 Restrictions

Level 1 **INITIAL**	Level 2 **DEVELOPMENT**	Level 3 **DEFINED**	Level 4 **MANAGED**	Level 5 **OPTIMISED**
Process disorganised & ad hoc	*Process is under development*	*Process is standardised and communicated*	*Process is managed and measured*	*Focus is on continuous improvement*
1.1 Little or no recognition of Māori data	1.1a Recognition of Māori interests in data	1.1a Māori data definition is consistent with, for example, the Te Mana Raraunga charter	1.1a Māori ownership of data is recognised	1.1a Māori Data sovereignty and CARE principles are intrinsic to RDM at an institutional level
1.2 Māori data is stored overseas and there is a poor regulatory environment	1.1b Awareness of CARE principles for indigeneous data governance, and relationship to Māori data sovereignty	1.1b Some use of Māori and CARE data principles (guidelines) in institutional strategy	1.1b Māori governance over Maori data is defined in strategy and policies	1.1b Māori rights to data and interests in data are recognised - exclusive rights /shared rights /shared interests
1.3 Some consideration of risks for Māori but no specific access arrangements	1.2 Storage overseas – strong regulatory environment	1.2 Storage primarily in Aotearoa NZ, or as advised though Māori governance	1.3 Māori have full access to data about their collective (copies	2.1 Māori metadata (including origin, consent and availability) is machine-readable and available for federation with other registries
2.1 No data on *whakapapa* - ethnicity data and iwi affiliation	1.3 Māori can access summary data about their collective	1.3 Māori can access raw data about their collective on request	2.2 Māori data is coded using standardised metadata categories that prioritise Māori needs and aspirations (eg. TKL)	3.2a Cultural licence: Māori/iwi have confidence in the ability of the organisation to use and share data in a legitimate and culturally acceptable way
2.3 No consideration of future use	3.2 Institutional-wide governance body, appropriately resourced, in-development	2.1a All data has *whakapapa* - ethnicity data and iwi affiliation collected	2.3 Policy and guidelines in place for secondary use of Māori data	3.2b Operational practices, standards and exemplars are available to all, to meet Māori data sovereignty and CARE obligations
3. Little understanding of sensitivities in use of Māori data or accountability to iwi and individuals	4.2 Institutional workforce capability and/or capacity developing, some reliance on external expertise	2.1b Metadata is used to indicate Māori data origin (provenance, purpose and context of collection), collective consent and data availability e.g. Traditional Knowledge Labels (TKL))	3.2 Systems and processes are in place to ensure accountability to the communities, groups and individuals from whom Māori data derive	4.1 Māori data aligns with FAIR principles, enabling collective benefit.
4.1 No consideration given to Māori in the design and function of data ecosystems	5.2 Individual consent – for general use of Maori data	3.1 Sensitivities in the use of data are identified including privacy issues for individuals and communities	4.1 Data ecosystems are designed and function in ways that enable Māori to derive individual and collective benefit	4.2 Māori and other indigenous peoples have the capacity to fully manage data to meet aspirations and enable common goals (CARE)
4.2 No Māori research workforce capacity building to enable the creation, collection, management, security, governance and application of data	6.2 Some integration of Māori principles (e.g. kaitiakitanga)	3.2 Clearly defined operational practices and standards in place for PIs	4.2 Building both organisational and community capacity	5.1 Collection, use and interpretation of data uphold the dignity of Māori communities, groups and individuals
5.1 Little awareness or understanding of the potential harm and loss of dignity to Māori through some use and interpretation of Māori data	6.3 Māori participation in a data access committee	4.1 A process in place to ensure Māori data is used with a development focus rather than a deprivation focus	5.1 Systems and processes in place to identify potential harm to Māori through data collection and analysis	6.1 Māori data are protected against future harm
6.1 Māori data stored without consideration of the implications of location/security		4.2 Building organisational capacity to enable the creation, collection, management, security, governance and application of Māori data	5.2 Individual and collective consent processes in place for access to Māori data	6.2 RDM is aligned with the principles of whanaungatanga, rangatiratanga, kotahitanga, whakapapa, manaakitanga and kaitiakitanga
6.2 Ethics: some consideration of risks for Māori but no consent processes for Māori data		5.1 An awareness of the risk of collective or individual harm to Maori from data analysis that stigmatises or blames	6.1 Māori data may be stored in a Māori repository	6.3 Māori/iwi involved in decision-making for issues that have been escalated
		5.2 Processes in place for individual consent – for specific/defined use of Māori data	6.2 Widespread use of Māori principles and data management protocols	
			6.3 Māori data access committee is in place	

		6.1 Quality, security, privacy and confidentiality mechanisms are in place to protect Māori data		
		6.2a Māori protocols in place for data access		
		6.2b Some use of Māori principles (guidelines) in policies and strategies		

Data Management

1. DM processes, including README/active data metadata to enable FAIR - interoperable and reusable (whereas FAIR focused metadata, repository record/schema is Data Operations)
2. Roles and responsibilities
3. Data Management Planning
4. Quality Assurance
5. Institutional knowledge of data

Level 1 **INITIAL**	Level 2 **DEVELOPMENT**	Level 3 **DEFINED**	Level 4 **MANAGED**	Level 5 **OPTIMISED**
Process disorganised & ad hoc	*Process is under development*	*Process is standardised and communicated*	*Process is managed and measured*	*Focus is on continuous improvement*
1. DM is chaotic and understood by only a few (absence of metadata, data duplication, versioning issues)	1.1 Data management processes are created and documented	1.1 DM processes are standardised and integrated	1. Adherence to DM processes are reported	1.1 Continuous improvement applied to processes and capabilities
2. No defined responsibilities in relation to DM	1.2 Data is regarded as useful within the project and meets funder/legal requirements (documented eg. use of disciplinary metadata schema)	1.2 All new data are described using standard metadata schemas and vocabularies	2. Roles and responsibilities have additional focus on measuring adherence and QA	1.2 All datasets richly described and metadata shared to fully enable FAIR data
3. Limited awareness or use of Data Management Plans (DMPs) or activity of DMPing	2. Responsibilities are defined and skills are developing	1.3. All new data are assigned a unique persistent identifier, with standard data licences	3.1 DMP templates are shared	*2.* Roles and responsibilities have additional focus on continuous improvement alongside evolving requirements
4. Data quality measures are ad hoc or absent	3. Simple DMP template and guidance available	1.4 Selected existing high value data are identified/described with/assigned:	3.2 DMPs are shared and used as dynamic and machine readable tools to support research data lifecycle, incl. provisioning of storage and active management of data	3. Machine-actionable DMPs are dynamic and auditable throughout the research data lifecycle, linking to output and research impact
5. Little awareness, support or understanding of data within the institution	4.1 Data are rarely cleaned and include inconsistencies	- unique identifiers - standard metadata - standard licences	4.1 Data quality metrics are refined to be fit- for-purpose	4. Quality metrics are refined periodically, data collection guidelines are in continuous use-
	4.2 Some data quality metrics are used for specific tasks or projects	2. Roles and responsibilities are defined, standardised, and widely understood	4.2 Metadata (e.g. README enabling FAIR) is audited to assess quality and availability	5. Comprehensive data registry, supported and linked to outputs and impact used for institutional and national reporting
	5.1 Starting to know where data are but most data is still managed locally, uncatalogued, and with practices that do not ensure enduring availability or description of the data	3. Widespread awareness and use of DMPs tool(s) and template(s) enabling provisioning and active management of data	5.1 Information about what data exists and where it is located is available at an institutional level, for internal admin processes, such as research reporting or researcher profiles	
	5.2 Key data or collections are described (metadata enabling FAIR)	4. Data quality strategy is developed: quality collection guidelines are consistently applied; common data quality coding scheme is being used as part of the domain-specific data generation	5.2 Metadata is offered in such a way that it can be harvested and indexed, including externally (FAIR)	
		5.1 Systems in which metadata can be entered and/or gathered from are connected, so information is only collected once, in a formal structure that can be crosswalked to other schema		
		5.2 Knowledge of data held within the institution e.g. data registry		

Data Operations

1. Sharing (collaborating, transfer, data sharing governance/agreements, metadata-only and data publishing, back-up)
2. Tools
3. Curation (human and machine readable metadata to enable **FA**IR - findable and accessible
4. Preservation
5. Roles and responsibilities, including stewardship/ guardianship - Rangatiratanga and Kaitiakitanga

Level 1 **INITIAL**	Level 2 **DEVELOPMENT**	Level 3 **DEFINED**	Level 4 **MANAGED**	Level 5 **OPTIMISED**
Process disorganised & ad hoc	*Process is under development*	*Process is standardised and communicated*	*Process is managed and measured*	*Focus is on continuous improvement*
1.1 Lack of awareness of data storage options to meet different needs	1.2 Some project-based data sharing services available but only data custodians know where the data is stored and data accessible only to the researcher or small group of researchers	1.2a Data sharing becomes more straightforward with the development of processes and guidelines consistently followed.	1.2 Data sharing becomes commonplace and embedded in practice throughout the organisation	1.1 Data management operations are fully automated and machine accessible throughout the organisation and provide the basis for rapid discovery of data in all disciplines, both within the institution and beyond
1.2 Inability to share data between systems or sharing is cumbersome, time- consuming and requires specialist skills	1.3 Processes and environment for sensitive data collection, storage and sharing are developing or present in patches	1.2b Development of support and analysis tools to release the benefits of reusing and sharing the data is undertaken	1.3 Processes and environment for sensitive data collection, storage, sharing and breach reporting are present, supported, communicated and audited against security standards	1.2 Continuous improvement is applied to data sharing processes and capabilities
1.3 Sensitive data collection, storage and sharing is not secure	1.4 Rudimentary data transfer tools available.	1.3 Processes and environment for sensitive data collection, storage and sharing are present, supported and communicated	1.4 Secure data transfer services are available (supporting transfer of sensitive data)	1.3 Processes and environment for managing sensitive data are established with ongoing review of technology, auditability, and cybersecurity defense mechanisms
1.4 Data is transferred manually or using standard http	1.5 Data is published via publisher sites and disciplinary repositories	1.4 Efficient data transfer services available	1.5 Processes for packaging data for publication to increase Interoperability and Reuse (e.g. containerisation)	2. Processes for lifecycle management of tools and support to analyse, collaborate and/or communicate data/artifacts are regularly reviewed
1.5 Data publishing, incl. metadata-only records, is ad hoc	1.6 Automated backup of data in place	1.5 Institutional data publishing services are available, including DOI reservation	2.1 Tools have been developed to analyse and link data automatically and quickly between differing systems	3. Curation services are integral to research data management, metadata schema are understood and implemented across the organisation and training/awareness services in place.
1.6 Backup of data is ad hoc or disorganised	1.7 Data sharing agreements are ad hoc, commonly between individuals	1.6 Data backup and archiving services and policies defined	2.2 Mature instrument data workflow capability, including transfer compute and storage integration, and support	4. Data is being preserved as standard practice with preservation planning and regularly reviewed against policy
1.7 No data sharing agreements captured	2.1 Tools are available to meet researcher needs for analysis, visualisation and collaboration around active data, even if not yet been widely publicised, rudimentary and with limited ability to interoperate other tools and systems	2.1 Processes for acquiring or accessing tools and support to analyse, collaborate and/or communicate data/artifacts are defined and communicated	3.1 The data has associated metadata including provenance information automatically collected and is stored in formats and structures that enable retrieval and sharing across domains	5. Roles and responsibilities have additional focus on continuous improvement
2.1 Tools to analyse the data are often proprietary, limited access, the community is not aware of them, or they lack support	2.2 Tools are developed to support the automatic storage of data with the metadata from various systems and instruments	2.2 Instrument data is regularly captured and stored	3.2 Data conforms to FAIR data principles	
3.1 Curation services absent or disorganised	3.1 Appropriate open data standards, formats, protocols, ontologies and vocabularies are identified and begin to be applied to some data	3.1 Data begins to be managed using standard open formats, using machine-readable vocabularies and ontologies enabling links and sharing with other systems	3.3 Curation understood as critical to data management	
3.2 Standard formats are not applied and the potential for reusability is limited	4. Preservation services absent or ad hoc	3.2 Selected legacy high value data are identified and curated to enable FAIR	4. Preservation practices in place for most data, with preservation plans including file migration or normalisation.	
4. Value and practice of preservation is not recognised	5. Roles and responsibilities for stewardship, curation and preservation are defined, but not well resourced	3.3 Provenance information and standards begin to be routinely collected but may not be standardised		
5. Roles and responsibilities are unclear. Some groups or individuals may utilise good practices but are unsupported by institution-wide services		3.4 Data services of the individual domains within the institution are still being fully developed		
		4. Some initial data preservation processes around shared projects		
		5. Roles and responsibilities for are resourced		

Platform, Infrastructure, & Architecture

Encompassing data services and platforms, infrastructure, architecture, and the resources that enable them
1. Data platforms (a set of services, tools and technologies that enable a community e.g. clinical research to meet their specific needs) and services (what researchers consume, the 'front-door' to infrastructure)
2. Infrastructure, including storage, research compute - hardware and software
3. Architecture (the plan, criteria and connectors by which the research delivery ecosystem is developed and integrated)
4. Roles and responsibilities

Level 1 **INITIAL**	Level 2 **DEVELOPMENT**	Level 3 **DEFINED**	Level 4 **MANAGED**	Level 5 **OPTIMISED**
Process disorganised & ad hoc	*Process is under development*	*Process is standardised and communicated*	*Process is managed and measured*	*Focus is on continuous improvement*
1.1 Provision of platforms and services for data capture, organisation, analysis and visualisation is ad hoc	1.1 Some provision of standard platforms and services for data capture, organisation, analysis and visualisation but no defined process to access/acquire non-standard offerings	1.1 Widespread availability of data platforms and services to meet different research community needs	1.1 Data management platforms and services are fit for purpose, managed and auditable	1.1 Data platforms are continuously developed with user community and support international best practices
1.2 Data sharing and transfer processes are ad hoc, including email, portable drives, personal cloud-based storage and institutional offerings	1.2 Some provision of standard data sharing and transfer platform, services and tools, including internal and external data sharing processes, to meet most needs	1.2 Awareness and access to a range of data sharing and transfer platforms and services, including sharing clinical data with other organisation	1.2 Data sharing and transfer platforms and services, including sharing clinical data with other organisation, are managed	1.2 Data services/platforms automatically capture metrics and audit findings to inform service quality and drive change in rapidly changing technology support areas
2.1 IT infrastructure for RDM is acquired, deployed, and managed inconsistently, not discoverable, not supportable, and not documented as services	1.3 Developing integration of research compute, storage and transfer platforms and services	1.3 Automated provisioning of project infrastructure in development	1.3 Automated provisioning of project infrastructure becomes standard practice	1.3 Use of data platforms, services and tools is driven by measures of effectiveness
2.2 IT infrastructure is inadequate to support targeted research data needs	1.4 Institutional Data Management Planning tool available to some	1.4 Machine actionable DMP tool integrated with systems and researcher workflow	1.4 DMP adherence is measured	2.1 IT infrastructure management optimises the IT infrastructure evolved in previous Levels through continuous focus on management and improvement of data assets
2.3 Research data storage facilities may be rudimentary such as shared drives or free web services	1.5 Provision of an data publishing service (institutional data repository/repositories) with limited integration	1.5 Provision of an institutional data repository integrated with systems and researcher workflow	1.5 Institutional and national platforms and services fully integrated and available to all researchers	2.2 Concerted efforts to optimise platforms and architecture to fit emerging research needs
3.1 Lack of consideration of IT research delivery architecture	1.6 Services to support institutional, and community specific data registries are accessible to very few	1.6 Services to support institutional, and community specific data registries are accessible are widely accessible	1.6 Automated testing and auditing of platforms and services to support institutional and community specific data needs registries are accessible are widely accessible	2.3 All aspects of data IT infrastructure facilities and services are monitored through highly visible automated testing and feedback systems
3.2. Capability and processes to support minting of persistent identifiers are non-existent or limited	2.1 Institutional needs for storage of research data are recognised and procedures in place to ensure persistent storage for data, whether created in-house or imported	2.1 IT infrastructure needs for RDM are coordinated with capital expenditure planning and with technology transition planning	2.1 Scalable computational infrastructure available to support a range of targeted research compute services, including HPC, AI, machine learning, etc.	2.4 Automated provisioning of integrated research compute service with managed upgrades and migration
4. IT/research data platforms, services, and,infrastructure expertise is held by only a few individuals who may be unable to cope with the demand	2.2 Data storage may be in-house, shared with another institution or outsourced.	2.2 Some integration with national infrastructure services (e.g. HPC, storage, repositories and transfer)	2.2 Data treated as key organisational asset and supported by targeted IT infrastructure	3. Ecosystem approach to architecture supported by use of persistent identifiers and systems integration for seamless experience,fit-for-purpose, user centric experience that is continuously adapting
	3.1a IT infrastructure provisioning and IT equipment are recognised in research delivery architecture	2.3 Well-defined internal and external data storage facilities are available and in the process of being integrated with other data management services	2.3 Well-defined matrix of internal and external data storage facilities options to suit different needs and data classifications	4. Individual IT support roles enact well-publicised technology transition and process improvement plans
	3.1b Research compute service available but processes not integrated within research delivery architecture	2.4 Research compute service fully integrated within institution architecture	3.1 RDM platforms and services well managed within a defined architecture that incorporates University and external ecosystems	
	3.2 Increasing institutional ability to mint persistent identifiers across the research lifecycle	3.1 Facilities are well defined, standardised and integrated as part of architecture for research delivery	3.2 Managed consumption, linking, validation and curation of persistent identifiers (including for organisations, people, projects, instruments, research outputs) that are minted internally and externally	
	3.3 Institutional identity, authentication and access is defined	3.2 Ability to mint range of persistent identifiers across the research lifecycle	3.3 Identities, authentication and access management are embedded into standard RDM practice	
	4. Responsibilities for research data infrastructure, provisioning, support of	3.3 Identity, authentication and access management are standardised		
		3.4 Architecture supports agreed data classification		

enterprise hardware and software, and IT architecture are developing	4. Responsibilities for IT infrastructure, provisioning, support of enterprise hardware, software, and architecture are defined and clearly communicated	4. Roles and responsibilities in support of RDM are documented and training provided to enable workforce capability and continuity	

Support & Skills

1. Training
2. Skills (DMPing, organising, describing, accessing and utilising storage and compute, sharing data. Awareness of//practices to achieve FAIR, Māori Data Sovereignty, and CARE data principles)
3. Support services and resources
4. Quality Assurance (QA) of training, skills and support services
5. Roles and responsibilities

Level 1 **INITIAL**	Level 2 **DEVELOPMENT**	Level 3 **DEFINED**	Level 4 **MANAGED**	Level 5 **OPTIMISED**
Process disorganised & ad hoc	*Process is under development*	*Process is standardised and communicated*	*Process is managed and measured*	*Focus is on continuous improvement*
1.1 RDM training is absent or ad hoc	1.1 Investment in RDM skills and processes	1.1 Widespread availability and uptake of training and skills development in data management, including FAIR and Māori Data Sovereignty	1.1 Researchers, relevant staff and students are trained in skills required across the research data lifecycle, including data management planning, ethics/legal/privacy, collection, organisation, storage, documentation, sharing, archiving, stewardship, enabling FAIR, CARE and Māori Data Sovereignty data principles - training outcomes are recorded and offerings regularly reviewed	1.RDM training and educational resources are regular adapted to maximise learning outcomes
1.2 Limited workforce capability or capacity development	1.2 Training is available across institution but not offered across institution as part of induction or project onboarding processes	1.2a Resources to support defined operational practices and standards across different groups		2. RDM skills are regularly assessed aspect of the student, professional staff and academic staff according to role
2.1 Data management (DMPing, organising, storing, sharing) skills are mostly absent	1.3 Some alignment of training/ educational resources with data management within ethics processes and managing sensitive data, incl privacy	1.2b RDM training is provided as part of research project onboarding processes		3. Support services are proactive in their scope and design to meet evolving RDM best practice
2.2 Knowledge, understanding and practices to support	2.1 Increasing awareness about and development of RDM skills	1.3 Alignment of researcher ethics and integrity, Māori Data Sovereignty, and data management educational offerings across institution	1.2 RDM training is integral to researcher, relevant staff and postgraduate induction/ provisional year and project onboarding processes	4.1 QA is routinely applied to training and support processes, results feed into future planning
- FAIR data	2.2 Data management planning is used on projects, resources and training is developed but not widely known	2.1 Researchers, relevant staff, and postgraduate students routinely exhibit RDM skills across the research data lifecycle	1.3 Training is available in a variety of delivery modes, levels and to suit different needs (disciplinary, data type, research method, data classification)	4.2 Updates to recommended practices, support services, and training materials are made synchronously with changes to policy, other external requirements, and evolution of best practice
- Māori Data Sovereignty	2.3 Increasing awareness of Māori Data Sovereignty data principles with some resources/support available to operationalise	2.2.Research postgraduates are trained in data management planning	2.1 Recognition and reward (e.g. ADPR) for RDM leadership, exemplary practice	5. Training and support roles and responsibilities have additional focus on measuring adherence, QA and improvements to meet evolving landscape
- CARE data	3.1 Good practices are identified and some effort to share and scale across institution-wide services	2.3 Clearly defined RD operational practices and standards in place for PIs	2.2. Research postgraduates are trained in data management planning	
- privacy obligations, and	3.2 DMPing support/resources are available to some researchers and postgraduate research students	3.1 Support services are defined and communicated eg. ethics processes support and training are aligned with data management recommended practices	2.3 PI compliance with standard RD practices is measured	
- protective security	3.3 Increasing awareness and accessibility of DM support and resources	3.2 DMPing support and resources are available to some researchers and postgraduate research students	3. Support services provide resources for managed approach to RDM that can be reported	
are mostly absent	4. QA of some support and training	3.3 Support services are aligned to data classification	4. QA is routinely applied to support processes, results feed into future planning	
3.1 Some groups or individuals may utilise good practices but are unsupported by institution-wide services	5. Training and support roles and responsibilities across the institution are defined	4. QA becomes feasible on training, and support processes to share and curate data	5. Training and support roles and responsibilities have additional focus on measuring adherence and QA	
3.2 Support for Data Management Planning is absent or disorganised		5. Training and support roles and responsibilities are aligned, standardised and communicated across the institution, incl. enabling the institutional to operate as a Māori Data Sovereignty organisation		
3.3 Staff are unable to locate required support or resources				
4. QA of support and training is ad-hoc or absent				
5. Training and support roles and responsibilities are not defined				

This initial RDM CMM (v1.2) is a companion to the RDM Stocktake Report. It is expected to evolve particularly in regard to **Māori Data Sovereignty** under the leadership of a Māori-led governance structure for Māori data, as recommended by the report.

The model may be cited as follows:
RDM Project Team. (2021) Waipapa Taumata Rau Research Data Management Capability Maturity Model v1.2. Research Data Management Project, University of Auckland.

The companion report may be cited as follows:
RDM Project Team. (2021) Waipapa Taumata Rau Research Data Management Project Stocktake Report. Research Data Management Project, University of Auckland.

Glossary

ADPR means the Academic Development and Performance Review, University of Auckland, process.

Active data is data that is in frequent use, being added to and processed/manipulated.

CARE - Collective benefit, Authority to control, Responsibility and Ethics. Find out more here: https://www.gida-global.org/care

Containerisation is the process of packaging together data, code, software, etc. to enable interoperability and reuse e.g. re-running of models, simulations, exploration of data across relational tables, etc.

Curation is an active and ongoing process of organising and integrating data that may include annotation, publication and presentation of the *data* such that the value of the *data* is maintained over time.

Data Maturity Model (DMM) is a framework of data, not specifically research data, management practices in six key categories that helps organizations benchmark their capabilities, identify strengths and gaps, and leverage their data assets to improve business performance. Areas: Data Strategy; Data Governance; Data Quality; Data Operations; Platform & Architecture; and Supporting Processes. Find out more here: https://cmmiinstitute.com/dmm

Data publishing is the act of releasing research data in published form for access by others. Research data can be published as supplemental material to an article, thesis, etc., downloadable files from a public website or a data repository, or can be published as the primary research output in, for example, a data journal. All forms of data can be considered for publication - raw, processed (e.g. de-identified, translated), sample, and synthetic data. When the research data itself can not be made public, a descriptive or metadata-only record can be published to enable others to know that the data exists and how to access the data through a mediated process. Research funders and publishers have different levels of requirements and practices for publishing research data that vary by discipline.

DOIs - Digital Object Identifiers - for research data are persistent identifiers using DataCite's DOI registration service and tools. The University of Auckland is a founding member of this National Library of New Zealand managed group. Find out more here: https://natlib.govt.nz/collections/digital-preservation/new-zealand-digital-object-identifiers

FAIR - Findable, Accessible, Interoperable and Reusable. Find out more here: https://www.go-fair.org/fair-principles/

Machine-actionable, sometimes referred to as machine-readable, Data Management Plans (DMPs) are structured so they can be read and actioned by humans and computers or programmes to enable automation and integration. Find out more here: https://doi.org/10.1371/journal.pcbi.1006750

Māori data refers to digital or digitisable information or knowledge that is about or from Māori people, our language, culture, resources or environments. Find out more here: https://www.temanararaunga.maori.nz/nga-rauemi

Māori Data Sovereignty refers to the inherent rights and interests that Māori have in relation to the collection, ownership, and application of Māori data. Find out more here: https://www.temanararaunga.maori.nz/nga-rauemi

Persistent identifiers (PiDs): An identifier is any label used to name something uniquely (whether online or offline). URLs are an example of an identifier. So are serial numbers, and personal names. A persistent identifier is guaranteed to be managed and kept up to date over a defined time period. ORCID is an example of a persistent identifier for a person, a DOI is a persistent identifier for a research output. Find out more here: https://www.ands.org.au/guides/persistent-identifiers-awareness

Registry may refer to a set of metadata records of the institution's data assets from one or more sources, for example an institutional data repository. This type of institutional registry may form part of a national or international, generic or disciplinarily specific system of federated registries. Another type of registry common to Universities are those collated for healthcare patients that share a disease or set of feature of interest.For example, https://www.monash.edu/medicine/sphpm/registries

Repository is a central location where digital data and digital representations or records are stored and managed. Data repositories may be institutional, generic and disciplinary, and may be provided by funders, publishers or institutions.

Research Data is currently defined by the Centre for eResearch as anything that is created, collected or obtained in the course of research that underpins a research output. **Data** is used with this report and CMM as an umbrella term for digital, print and physical datasets, artefacts, cultural taonga, research evidence and also covers the digital representation of a physical item used in research, that may need to be managed and shared. The institutional definition will be developed as part of the RDM Policy (refer improvement opportunity 1.B).

Research Data Management (RDM) is the process of planning and undertaking the collection, organisation, management, storage, backup, preservation and sharing of data during and after the project is complete.

Researcher refers to all University of Auckland staff and postgraduate students undertaking research.

Static data (opposite of active data) is no longer changing and may be associated with a research activity that has ended or where the results are not being manipulated or changing.

Traditional Knowledge Labels (TKL) are an initiative for Indigenous communities and local organizations. Developed through sustained partnership and testing within Indigenous communities across multiple countries, the Labels allow communities to express local and specific conditions for sharing and engaging in future research and relationships in ways that are consistent with already existing community rules, governance and protocols for using, sharing and circulating knowledge and data. Find out more here: https://localcontexts.org/labels/traditional-knowledge-labels/

References

CAUDIT. (2020) The Higher Education Business Reference Model, v.2.5.0. Council of Australasian Director of Information Technology. Accessible here: https://www.caudit.edu.au/EA-Framework

FAIR Data Maturity Model Working Group (2020) FAIR Data Maturity Model. Specification and Guidelines. Accessible here: https://doi.org/10.15497/rda00050

Hudson, M. (2019) Māori Data Sovereignty: Implications for democracy and social justice. Ethics in Research: Governance, Democracy & Social Justice, Wellington School of Medicine. Presentation. Accessible here: https://www.otago.ac.nz/wellington/otago706724.pdf

Medyckyj-Scott, D., Stock, K., Gibb, R. et al. (2016). Our Land and Water National Science Challenge - A Data Ecosystem for Land and Water Data to Achieve the Challenge Mission.

Rans, J and Whyte, A. (2017). Using RISE, the Research Infrastructure Self-Evaluation Framework v.1.1 Edinburgh: Digital Curation Centre. Accessible here: www.dcc.ac.uk/guidance/how-guides/RISE

RDA International Indigenous Data Sovereignty Interest Group. (2019) CARE Principles for Indigenous Data Governance. The Global Indigenous Data Alliance. Accessible here: https://www.gida-global.org/care

Te Mana Raraunga. (2018) Principles of Māori Data Sovereignty. Brief, 1. Accessible here: https://static1.squarespace.com/static/58e9b10f9de4bb8d1fb5ebbc/t/5bda208b4ae237cd89ee16e9/1541021836126/TMR+Ma%CC%84ori+Data+Sovereignty+Principles+Oct+2018.pdf