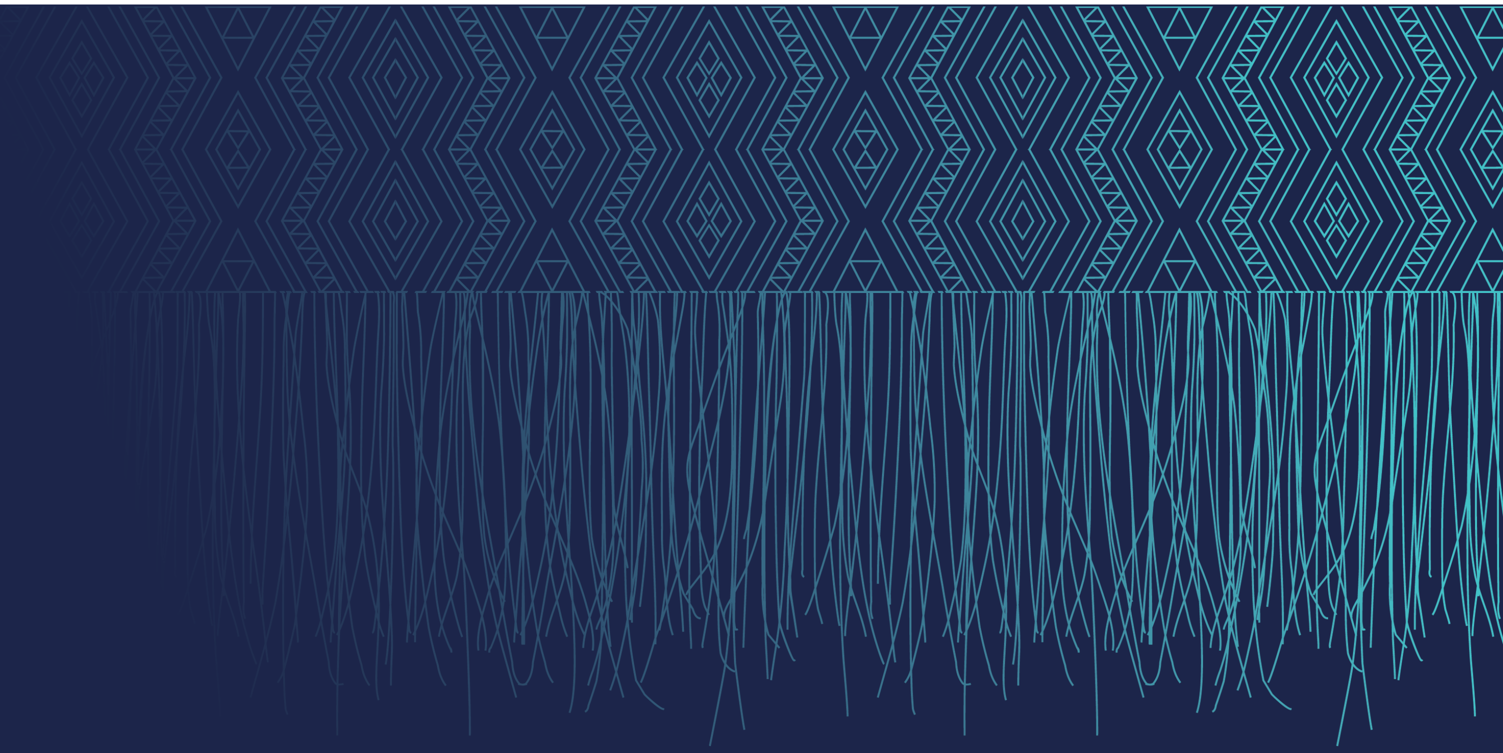
Te Whatu Ora
Health New Zealand

Primary Health Data Set Programme (PHDP)

Phase 2 Summary – Establishment of a Foundational Primary Care Data Set for New Zealand



Te Kāwanatanga o Aotearoa
New Zealand Government

Citation: Te Whatu Ora – Health New Zealand. 2023. *Primary Health Data Set Programme (PHDP) Phase 2 Summary – Establishment of a Foundational Primary Care Data Set for New Zealand*. Wellington: Te Whatu Ora – Health New Zealand.

Published in May 2023 by Te Whatu Ora – Health New Zealand
PO Box 793, Wellington 6140, New Zealand

ISBN 978-1-99-106721-0 (online)

This document is available at **tewhatuora.govt.nz**

Foreword

The strategic direction for the establishment of the Primary Health Data Set Programme (PHDP) was agreed and the PHDP programme was initiated in February 2022. PHDP Phase 1 - Discovery concluded in August 2022 and then design work was progressed across five domains:

- Data Governance and Māori Sovereignty.
- The data classes and data elements to populate a nationally consistent primary care data set.
- The requirements of a technical platform to support the acquisition, storage, processing and usage of the data set.
- A costing model to inform investment in PHDP Implementation, as well as informing any procurement options.
- The needs of the analytics community.

A series of co-design working group meetings were held from October 2022 through to March 2023 - PHDP Phase 2 - Design. This has led to a comprehensive high-level approach for the establishment of nationally consistent foundational primary care data set for New Zealand.

The document informs the implementation of a nationally consistent foundational primary care data set for New Zealand, essential not only for the health system to plan, commission and evaluate health services effectively, but most importantly, to support the health system reform's focus on equity, partnerships, sustainability, person-and whānau-centred care and excellence.

PHDP is complementary to the following initiatives: commissioning priorities of Te Whatu Ora 2022, Te Pae Tata Interim New Zealand Health Plan 2022, Data & Information Strategy for Health & Disability (DISH), Data and Digital Strategy and 2 Year Roadmap 2022, Data Tripartite Agreement between Te Whatu Ora – Health New Zealand, Te Aka Whai Ora – Māori Health Authority and Manatū Hauora – Manatū Hauora (MoH) 2022, the Rapid National Data Automation Project (RNDA) and National Data Platform1 (NDP).

The Te Whatu Ora Commissioning refers to the need to address funding models:

"There is much work ahead over the coming two years to address historic funding imbalances, improve our efforts to better integrate services across primary and community care, and update funding models in an effort to support improved access and equity of outcomes for key population groups."

This message is reinforced in the recent Sapere review:

“A future capitation funding approach, 2021: - We recommend an approach to funding general practice services based upon age, sex, ethnicity, deprivation, and ***morbidity****. This appears to give results that are adequately correlated to health need.”*

PHDP aligns to:

- Pae Ora; healthy futures for all New Zealanders. The reforms have an absolute focus on equity, partnerships, sustainability and the development of person-and whānau-centred care.
- Health Services and Outcomes Kaupapa Inquiry (WAI-2575). The findings are compelling and clear – consolidation of policy, strategy, and operational direction for the system pointing to improved outcomes for Māori with data being a key enabler.

Governance of this data will be crucial, chiefly towards improving patient outcomes through ensuring that patient data is accurate, secure, and accessible to those who need it. Essential to this will be the ability to respect and maximise the strategic value and cultural significance of data and information across the health sector, including supporting how Māori can assert tino rangatiratanga and mana motuhake as the Te Tiriti o Waitangi is honoured and upheld.

There is huge value and significant support for the development of a primary care data set. This report clarifies what data should be included, how it will be governed, the technology requirements and the likely cost.

The Programme Team would like to thank the many sector leaders and their teams for working with them on this programme. It is a complex piece of work that has significant potential to transform planning and investment in services to improve the health outcomes of New Zealanders.

Contents

List of Figures

List of Tables

Acknowledgements

Author(s):

Stephen Lavery, Programme Co-Lead, Primary Health Data Set Programme, *Programme Manager Primary Health Care System Improvement, Innovation and Commissioning, Te Whatu Ora*

Tony Cooke, Programme Co-Lead, Primary Health Data Set Programme, *Primary Health Care System Improvement, Innovation and Commissioning, Te Whatu Ora*

Contributor(s):

Gayatri Sevilla Fernando, Senior Business Analyst, *Primary Health Care System Improvement, Innovation and Commissioning, Te Whatu Ora*

Andrew Weaver, Senior Project Manager, *Primary Health Care System Improvement, Innovation and Commissioning, Te Whatu Ora*

Roshini Subramaniam, Project Coordinator, *Primary Health Care System Improvement, Innovation and Commissioning, Te Whatu Ora*

Simon Royal, Māori Sovereignty Consultant, *Simon Royal Consulting*

Fiona Wakefield, Barrister and Solicitor, *Fiona Wakefield Barrister & Solicitor*

Simon Ross, Manager, *Data Governance, Data & Digital, Te Whatu Ora*

Robyn Kamira, Chief Advisor*, Data & Digital, Te Aka Whai Ora*

Neil Annenberg, Programme Consultant, *Data & Information Strategy, Health & Disability, Te Whatu Ora*

PHDP Governance Group members

Refer to Appendix G.1

PHDP Working Group members

Refer to Appendix F.1

Other acknowledgements

Peter Dolan – Group Manager, *Evidence, Research and Analytics*, *Ministry of Health*

Namoe Tuipulotu – Interim National Pacific Intelligence Lead, *Te Whatu Ora*

Laura Cleary – Manager, *National Collections Reporting, Data & Digital, Te Whatu Ora*

Kendra Telfer – Principal Analyst, *Data Governance, Data & Digital, Te Whatu Ora*

Kevin Maloney – Principal Architect, *Data Governance, Data & Digital, Te Whatu Ora*

Disclaimer

The insights in this document are based on the engagements held from October through to April 2023 with PHDP stakeholders and findings from Phase 1 Current State Analysis.

Executive summary

The Primary Care Data Set Programme (PHDP) is a partnership between Te Whatu Ora, Te Aka Whai Ora, general practice and the wider primary care sector to establish a foundational primary care data set.

New Zealand does not currently have a nationally consistent foundational primary care data set and as a result is unable to produce a nationally consistent and timely view of the health status of New Zealanders.

The collaborative approach that underpins PHDP to date can lead to the establishment of a primary care data set that is co-designed with the primary care sector. It is clear that the data can be sourced from national health data sets, general practice Patient Management Systems (PMS) and Primary Health Organisations (PHOs).

There is huge value and significant support for the development of a foundational primary care data set. It is accepted that the combination of the skills within the primary care sector and the skills and resourcing that sits within Te Whatu Ora, Te Aka Whai Ora and Manatū Hauora will allow for rapid progress.

Governance of this data will be crucial and has to ensure that people's data is sourced respectfully and within the law. Essential to this will be the ability to respect the cultural significance of data and information across the health sector, including supporting how Māori can assert tino rangatiratanga and mana motuhake as the Te Tiriti o Waitangi is honoured and upheld.

The PHDP design phase has clarified what data is important, how it will be governed in principle, the technology requirements and the modelling of potential cost of implementation, specifically:

- PHDP, its co-design approach and its objectives are supported by the primary care, Te Whatu Ora, Te Aka Whai Ora and Manatū Hauora representatives on the working groups.
- Data governance approaches and the purpose statement of the data set have been agreed.
- A nationally consistent primary care data set has been agreed encompassing 14 data classes and a core set of data elements which need to be sourced from the sector.
- The requirements for a technical platform on which the data set can be supported have been agreed.

- Development of a dynamic costing model estimates costs associated with the acquisition, processing, usage, disposal and other costs associated with the implementation of PHDP.

Through completing the PHDP design phase, the following has been identified and outlines the strategic direction for the PHDP implementation phase:

- Continued partnership between Te Whatu Ora, Te Aka Whai Ora, Manatū Hauora and the primary care sector to harness the skills that exist across the sector through PHDP implementation essential.
- Establishment of shared governance model with general practice, Te Whatu Ora, Te Aka Whai Ora and Manatū Hauora for the usage of primary care data including strengthening of data governance frameworks to support the trusted relationship between people and their general practice team.
- Investment to rapidly establish consistent procedures for access, curation and transformation of primary care data from agreed sources.
- Establishment of analytical approaches applicable to primary care data alongside other data sets to support Pae Ora, as well as supporting key stakeholders (general practice, Iwi-Māori Partnership Boards (IMPBs), localities, Te Whatu Ora, Te Aka Whai Ora, Manatū Hauora and PHO operational teams).
- Leveraging the developments from DISH, Data and Digital Strategy, Data Tripartite Agreement, National Data Platform (NDP) and Hira to PHDP and vice versa.
- Investment in PHDP Implementation.

PHDP is a long-term programme and will require steady investment over years. There are a number of foundational standards that need to be addressed over time to improve data standards and data governance.

1 PHDP scope

This document summarises the findings from the Phase 2 - High-Level Design for the establishment of a nationally consistent foundational primary care data set for New Zealand.

'Primary care data' is defined as any data collected in or needed for use in a primary and community care setting. This includes data held within primary and community care Practice Management Systems (PMS), Primary Health Organisations (PHOs) and data held by Te Whatu Ora – Health New Zealand across its National Health Data Sets.

This broad definition also includes patient data which exists in other primary and community-based systems, such as community pharmacy, maternity, Well Child Tamariki Ora, aged care, ambulance and other provider networks. However, the scope of the PHDP programme for the time being is focused on general practice activity which holds the longitudinal clinical "system of record" for a person. Such data may already be held nationally in a national collection or locally within a general practice or primary care organisation setting.

Phase 2 - High Level Design focuses on how to inform high-level scope, options and considerations to progress an implementation phase of the programme. Key problem statements requiring a design solution are:

- how to bring together a nationally consistent foundational primary care data set, which comprises:
    - the specification and sourcing of data elements which are not held nationally
    - the specification and availability of data elements that are held nationally and
    - how to select a platform which can join up disparate sources and types of data and make it available for use by the many stakeholders that are informed and supported under the Pae Ora[1](Healthy Futures) Act 2022 strategies.

In order to inform these high-level design objectives, five working groups (refer to Appendix F.1 Working group membership register for the list of members) were established in the period October 2022 through to April 2023. These groups informed and co-designed five key streams of work:

- The data governance rules, standards and support structures required for the nationally consistent foundational primary care data set;
- The list of the components of the data set itself and the potential sources of data;

---

[1] https://www.legislation.govt.nz/act/public/2022/0030/latest/LMS575405.html

- The purpose statement supporting the transparent governance and use of the data set, including an approach which meets the obligations under Te Tiriti o Waitangi;
- The requirements of a technical platform to support ingestion, curation and usage of the data set; and
- A costing model to support implementation options.

This report synthesises the findings from these groups, including potential sources of data, the subsequent data management and transformation that is required, as well as considering detailed implementation options and the potential costs associated with implementation.

2 Introduction

On 1 July 2022, the commissioning and delivery functions of the health system were organised into two national entities, Te Whatu Ora – Health New Zealand (HNZ) and Te Aka Whai Ora – Māori Health Authority (MHA) with Manatū Hauora – the Ministry of Health (the Ministry) retaining responsibility for policy and its role to monitor and report on performance of the health system.

To support the ability to plan, commission and evaluate health services effectively, access to nationally consolidated, population level primary health data is critical.

The creation of a nationally consistent foundational primary care data set has been discussed widely since the Primary Health Care Strategy was published in 2001[2]. However, robust primary health data to support decision-making and population health planning at national, regional, and local levels has remained fragmented and incomplete since then.

The challenges in creating a nationally consistent foundational primary care data set have persisted due to the innate complexity of data sources, the transformation required as well as the structure and nature of the organisational and contractual relationships that underpin the primary care sector. This includes the autonomy of and variation in systems between general practitioners (GPs); the business and commissioning arrangements between funders, Primary Health Organisations (PHOs) and GPs; and the disparity in data and digital maturity across the locations, districts and regions.

2.1 Summary of recent developments

In 2019 a proposal for a National Primary Data Service (NPCDS) was developed by the Ministry in order to create an extract of primary health data[3]. However, the proposal did not make progress due to competing demands, including the COVID-19 response, and resources were directed elsewhere. In order to support population health and service planning through the COVID-19 response, manual and imperfect processes were adopted as a proxy of national primary health data. This approach was not systematic or enduring.

In early 2021, a public health information strategy was commissioned by the then-Deputy Director-General Public Health and Primary Care Transformation as part of work to support the health reforms and the establishment of the Public Health Agency. The objective of the strategy was to identify the key data initiatives required to support a robust

---

[2] https://www.health.govt.nz/publication/primary-health-care-strategy

[3] Stephen Lavery, 2019, National Primary Care Data Service, Andrew Inder, Gabrielle Roberts & Darren Douglass

public health function. This identified the gap in the provision of national primary health data for public health purposes.

Later in 2021, a combined effort between the Public Health and Primary Care Transformation team and Primary Health System Improvement and Innovation (PHSII) team saw the preparation of an options paper, Primary and Community Care Data for Population Health[4] outlining the possibilities for development of a national view of primary and community health data. This paper was supported by relevant members of the Ministry's executive leadership team and was followed by a request for funding in February 2022.

In February 2022, following on from the development of the options paper, a memo outlining a recommended way forward was signed off by then-Deputy Director General, Data & Digital, the Ministry and the then-Director, interim Te Whatu Ora. This approved interim development of a national foundational primary care data set in support of the health system reforms and the establishment of Te Whatu Ora and Te Aka Whai Ora[5]. The Primary Health Data Set Programme (PHDP) was established, and programme resources recruited.

In May 2022, a briefing was presented to the interim-Boards of Te Whatu Ora and Te Aka Whai Ora by the then-Director, interim Te Whatu Ora, stating that "[PHDP] seeks to bridge the data gap by working in partnership with the primary health sector for the provision of the 'missing data' and then combining it to create population health analysis at national, regional, and locality levels".

The memo summarised that the resulting data analysis will be used for:

- Population health planning, policy development and implementation across the health system.
- Delivery of nationally and locally relevant public health knowledge and surveillance.
- Planning and development of localities.
- Insights and monitoring of progress for Te Whatu Ora, Te Aka Whai Ora and Manatū Hauora[6].

---

[4] Cooke, Lavery, Horsfall. 2021. Primary and Community Care Data for Population Health. In Confidence Briefing: Access to Aggregated, Population Level Primary Health Data. Health New Zealand. Wellington. (Page 5)

[5] https://www.health.govt.nz/new-zealand-health-system/overview-health-system/health-and-disability-system-reforms

[6] Hefford, Hunter. 2021. Primary Health Data Set Programme Brief. Health New Zealand. Wellington.

In September 2022, after completing the Phase One current state analysis of data capability within the primary sector, seven recommendations were adopted by the PHDP Governance Group[7]. These recommendations formed the basis for the establishment of five sector working groups and planned programme activity through to June 2023. The seven recommendations across five workstreams are described in Table 1 - PHDP Governance Group recommendations.

Table 1 - PHDP Governance Group recommendations

Data Governance
1. To establish a data governance authority which will be governed in partnership with the sector, and which will work with the Data Tripartite Agreement, Māori data framework and primary health data governance provisions to build consistent and transparent data governance processes across the sector. 2. To co-design the data governance provisions for primary health data with expertise from primary care clinicians, general practitioners, PHO data managers, analysts, National Collections staff, Māori data sovereignty designers, and privacy specialists. Noting that access to a consumer reference group is recommended – potentially the same one used by HQSC. 3. To develop a best practice privacy and consenting framework through a review of existing privacy statements and data sharing agreements. Noting that a Privacy Impact Assessment on the agreed approach to data collection and governance of the primary health data set will need to be undertaken.
Data Standards and Requirements
4. To implement agreed current data standards (based on approved HISO standards) for the curation of primary care data captured locally and held nationally, to ensure that the work of general practice and PHOs is recognised and aligned nationally. Noting that work on developing a new standard for the transfer of primary care data is in progress and has been circulated for consultation (NZ International Patient Summary) and that the Hira programme preference is for this data standard to support its development of real time interfaces.
Analytics Framework
5. To establish a whole of sector analytics framework to create a community of practice that combines the strengths of all parties to develop and apply consistent standards of analytic outputs and methodologies on a standardised set of data that includes primary and community data for use by health sector participants. Noting that analytics to improve equity of health outcomes for priority populations and to meet obligations under the Te Tiriti o Waitangi is critical.

---

[7] https://www.tewhatuora.govt.nz/publications/primary-health-dataset-programme-current-state-analysis-report/

Technology Platform
6. To set up the technology platform to support the primary health data sets allowing it to be part of a wider data ecosystem which will be available for authorised analytical users from primary care, Manatū Hauora, Te Whatu Ora, and Te Aka Whai Ora. Noting that migration of National Collections to the Snowflake platform is already a Data & Digital committed programme of work.
Commercial Model
7. To develop a costing model for the provision of data management services to support the primary data set for (a) Te Whatu Ora platform and support costs and (b) primary health data providers costs to supply the data. Following that to select, procure and contract data providers appropriately.

2.2 Programme vision

The vision of the health system reform is to build a new system that achieves Pae Ora (Healthy Futures) Act 2022 strategies; for all New Zealanders. To achieve this the health system will need to focus on equity, partnerships, sustainability, person-and whānau-centred care and excellence. A critical focus is to strengthen the primary and community healthcare sector to improve outcomes.

The report on stage one of the Health Services and Outcomes Kaupapa Inquiry (WAI-2575)[8] further consolidates policy, strategy, and operational direction for the system in terms of improved outcomes for Māori. Comprehensive population data and knowledge relating to Māori is fundamental to support this.

PHDP will support this vision by allowing a more comprehensive set of health data to be available across the system for:

- Enabling the Pae Ora (Healthy Futures) Act 2022 strategies legislation.
- Population health planning, policy development and implementation.
- Public health knowledge and surveillance.
- Planning and development of localities.
- Insights and monitoring of progress for Te Whatu Ora (Health New Zealand), Te Aka Whai Ora (Māori Health Authority), and Manatū Hauora (Ministry of Health).

The PHDP approach also recognises the unique position that general practice has in the New Zealand health system and general practices rights and obligations related to the

---

[8] https:/www.health.govt.nz/our-work/populations/maori-health/wai-2575-health-services-and-outcomes-kaupapa-inquiry

stewardship of the 4,871,757 enrolled (April 2023) people’s health data that not only supports clinical care but is fundamental to the success of the National Primary Care Data Set being established in the first place.

2.3 Governance, co-design, decision-making and advisory arrangements

Shortly after the programme was initiated, the PHDP Governance Group was established. The purpose of the governance group is to provide oversight and guidance on the work this programme is delivering and to:

- share knowledge and expertise regarding the state of primary health sector information
- review and/or endorse final reports and recommendations
- facilitate linkages with related projects and stakeholders from the wider sector
- keep other colleagues and leaders informed
- support engagement and communication activities, and
- advise on and assist with the management of any associated risks and constraints.

The PHDP Governance Group members can be found here: **PHDP Governance Group**

Throughout 'Phase One Current State Analysis' there was agreement by Te Whatu Ora, Te Aka Whai Ora and Manatū Hauora to work in partnership with the sector to co-design the next phase of PHDP. Five working groups (Data Governance Requirements, Primary Care Data, Analytics Leadership Group, Technical Platform Requirements Group and Commercial Requirements Group) were set up to support the co-design. These groups included representatives from central agencies, PHOs, MSOs, Digital Health Association (DHA), GenPro, RNZCGP and PMS providers.

Table 2 - PHDP working groups and deliverables describes the deliverables across each of the PHDP working groups as documented in each of their Terms of Reference documentation (refer to Appendix A.2 Terms of Reference).

Table 2 – PHDP working groups and deliverables

Data Governance Requirements
1. To provide standardised guidelines and procedures to ensure the safe collection, storage, and use of primary care data nationally.

2. To provide standardised guidelines and procedures for the collection, storage and use of data relating to the principles and obligations of Te Tiriti o Waitangi in line with the emergent Māori Sovereignty framework that is being developed concurrently.
3. To develop access protocols for end users of data (elemental data or aggregated) in line with the legal and contractual obligations that exist in relation to the purpose and permitted use of primary healthcare data
4. To determine the different levels of de-identification needed for different purposes/end-users.
5. To prepare a code of conduct for end-users who are exposed to identifiable data in the trusted analytical environment.
6. To describe on-going operational requirement for the management of roles, responsibilities, and permissions.

Analytics Leadership

1. To identify end user groups who may initially be interested in this type of data.
2. To ensure that the deliverables of this programme will meet the needs of end-users across a broad range of analytics perspectives.
3. To provide recommendations and ideas to help facilitate the most cost-effective, easy-to-use, and comprehensive solutions for use by end-users, for example, repeatable reports or ad hoc analytical approaches.
4. To recommend requirements for analytical services and outputs for the first tranche of analysis from a ‘joined up’ data set.
5. To advise on data standards and data quality.
6. To provide access to training and proven methodologies.
7. To build a resource library of repeatable analytical processes which have been peer reviewed and accepted.

Technical Platform Requirements

1. To advise on the capability required of a technical platform to support reporting, analysis and queries for data of interest to primary care.
2. To assess any approach to platforms that exist or are being developed across the PCC (Primary and Community Care) sector and within Te Whatu Ora.
3. To suggest options for how such a platform might be realised.

Commercial Requirements

1. A working commercial approach based on best practice and learning from previous models.
2. To advise on the commercial or agency arrangements that may emerge in the implementation of the platform and the data that will sit across it.
3. To scope and scale both initial development and ongoing costs of delivery.
4. To advise on process steps including procurement and contracting for services if these are applicable.

Primary Care Data
There are two components or outputs required 1. A primary care data schema and definitions for each of the data elements and data classes that need to be sources from general practice and PHOs 2. A prioritised list of data classes held by Te Whatu Ora in the national health data sets that are of interest to the primary and community sector.

2.3.1 Approach

Core design principles were outlined in the original approval memo Primary Health Data Set Proposal[9] dated 9 February 2022 as below.

Table 3 – PHDP core design principles

PHDP core design principles
1. Data governance to define and agree how the Primary Care Data Set will be used and maintained;
2. Ongoing engagement with primary and community care providers;
3. Improved standardisation and quality of data at the point of creation;
4. Straightforward tools and processes to store and use the Primary Care Data Set in a repeatable manner;
5. Technology to support virtual data warehousing so that data is not unnecessarily moved around or transformed;
6. Improved efficiency and ease of access for legitimate end-users; and
7. Ability to “add value” to the [nationally consistent foundational primary care data set] by linking and combining it with other data.

One of the key challenges for the programme has been balancing the needs of multiple stakeholder groups who seek access to sensitive health information. It is important to ensure that the overarching principles that underpin the privacy and confidentiality of personal health information, held by general practice on behalf of their patients, is respected and enhanced in any high-level design and subsequent implementation.

Te Whatu Ora, as well as PHOs are already the long-term custodians of some sensitive patient information held in national and local data sets. The programme approach has

---

[9] Cooke, Lavery, 2022 Memo – Primary Health Data Set Proposal, prepared for Shayne Hunter & Martin Hefford

been to build on existing trusted relationships and reputations to co-design a nationally consistent foundational primary care data set.

The programme is adopting the Health Information Standards Organisation (HISO) standards for the format of the data.

2.3.2 Stakeholder engagement

In the past, PHOs and other primary health representative groups have expressed their concerns about how Te Whatu Ora will use primary health data once they receive it and want strong safeguards around its use – as there is for PHOs currently. There is an inherent distrust of government health agencies from past experience and past impositions of requirements without consultation or recognition of the associated costs of compliance. Similar concerns have been raised by Māori stakeholders.

To progress PHDP, it was essential that a co-design with expertise from primary care clinicians, PHO data managers, analysts, National Collections staff underpinned Phase 2.

To date the programme has attempted to role model an inclusive approach to engagement and design. The partnership between Te Whatu Ora, Te Aka Whai Ora, Manatū Hauora and the primary care sector is essential for any implementation, particularly in terms of leveraging governance principles, existing consolidated data sources, the skills, and the knowledge and methods that the primary sector have built up over many years.

The primary care sector agrees with this aim and the objectives of PHDP.

Approximately 62 organisations (for both phases) have been consulted in the development of the programme to date, these organisations are listed here (refer to Appendix F.6 PHDP workshop engagement register and Appendix H.1 PHDP engagement register).

2.4 Population health management pain points

At a high level, the identified problem is that currently there is not a nationally consistent and complete collection of primary care data for New Zealand[10]. This means the New Zealand health system often does not know about:

- **Health conditions**: Primary care systems provide a longitudinal repository for diagnosis, health condition status and treatment for most people in New Zealand. This localised data supports clinical point of care. Localised data does not support a

---

[10] Pigden A et al. 2022. Developing a national primary care research network: a qualitative study of stakeholder views. Journal of Primary Health Care. doi:10.1071/HC22081

nationally consistent view of burden of disease in New Zealand, and how many people have which health conditions. This means that there is an inability to shift investment to the right place at the right time to improve health outcomes for people who need it the most.

- **Health interactions from all primary care providers:** e.g. knowing when and how often people visit different practitioners, laboratories or other health providers.
- **Risk factors and behaviours**: e.g. diet and exercise, smoking, etc. of its population.

Te Whatu Ora spends nearly $8 billion annually on primary and community care services provided by community-based organisations and yet there is limited information to manage investment or support outcomes for that spend.

Note that the New Zealand Health Survey numbers are dependent on the coverage of the data when the population sample is small. Some IMPB, locality or PHO populations are likely to be too small to obtain statistically reliable data from the survey. There are numerous examples of how a nationally consistent foundational primary care data set could be used:

Health interactions: e.g. people visiting their general practitioner, a diabetes clinic, or a nurse in a GP practice.

- Measure equity of access: This would require understanding which people access which services, and how frequently, by demographics. For example, the Pacific women receiving diabetes clinic support, compared with women of other ethnicities, and with men of different ethnicities. This could help to understand gaps, and inequities by identifying groups of people of high need who lack access or don’t connect with services. This information could be used to prioritise resources and programmes to address unmet need, and support equity of access.
- Measure use of primary care services: This information would identify how frequently people in New Zealand access primary care services, by gender, age, ethnic group(s), location, health status, etc. This will help to understand where services are being provided, and to whom. It could assist to identify if there are gaps in location, age groups or ethnicity in terms of where the services are available, or if there are people who aren’t receiving services, who need services. This could assist to identify the best targeted investment required to support primary and community care.
- Measure specific conditions by access to services: This would enable an understanding of unmet need for particular conditions, or which conditions are generally well managed at a population basis. An example could be to identify the number of people who have Hepatitis C (from lab tests data). This could be compared to other information to see if they have received appropriate treatment (from pharmaceutical dispensing), and if they have visited their primary care provider.

There are also a number of publicly available reports (refer to Appendix B.5 Existing reports – public interaction with the health system) on people's journey view in terms of interaction with health system framed in a way to highlight the potential for service improvement.

Health conditions:

- Measure burden of disease in New Zealand: This would involve identification and understanding of which are the most common conditions in New Zealand (and track this over time). This information could be linked to health outcomes, e.g. mortality, morbidity and hospitalisations to see which conditions are causing the highest health burden. This would be analysed by demographics: which ethnic groups, ages, genders, and people living in which regions have a higher burden of specific diseases, and which groups have higher mortality and morbidity from these conditions. This information could be used to guide policies and programmes to address equity. For example, those of Māori ethnicity are more likely to be diagnosed with diabetes than non-Māori, and at younger ages. They are more likely to be hospitalised from this condition, and to die. Programmes are needed to ensure better access to primary care services for Māori with diabetes, to reduce the impact of this condition. This information could help identify those high-risk areas and ensure appropriate resources and training developments occur to best meet those needs on a national basis.
- Understand the conditions which are increasing or decreasing over time. Interventions, programmes, and policies can be planned to address these changes.

Risk factors and behaviours:

- Understand the drivers of health outcomes for New Zealanders. Factors known to impact on health outcomes could be collected and analysed to identify the health impacts and measures that may reduce those risks. Analysis of known risk factors such as: who is smoking or who is a healthy weight, will be helpful to determine the impact on the New Zealand population disease burden. Differences by age, gender, ethnic group(s), and region could also be analysed.

3 PHDP data sources

In the PHDP context, the proposed data sources are either the originating source system for the data, e.g. General Practitioners capturing patient data and saving them into a Practice Management System (PMS), or from a secondary data source e.g. from National Collections where the original data has been collected from either hospitals, the PMS, from Primary Health Organisations (PHOs), or a single source from a nationally operated system such as the NHI.

This section summarises the sources of data that could form a nationally consistent foundational primary care data set, based on the recommended data classes and elements by the PHDP Primary Care Data working group.

Domains of data

The definition for ‘Primary care data’ the programme is using is ‘*any data collected in or needed for use in a primary and community care setting*’. This means a nationally consistent foundational primary care data set will include data held by general practice PMSs, PHOs and data held by Te Whatu Ora in its 37 National Health Data Sets (within the boundaries of the current governance of each of these national data sets).

The Primary Care Data Working Group focused mostly on the key elements of a foundational primary care data set which are not held nationally, in other words the primary source being general practice PMSs. The group also explored data held within the National Health data sets.

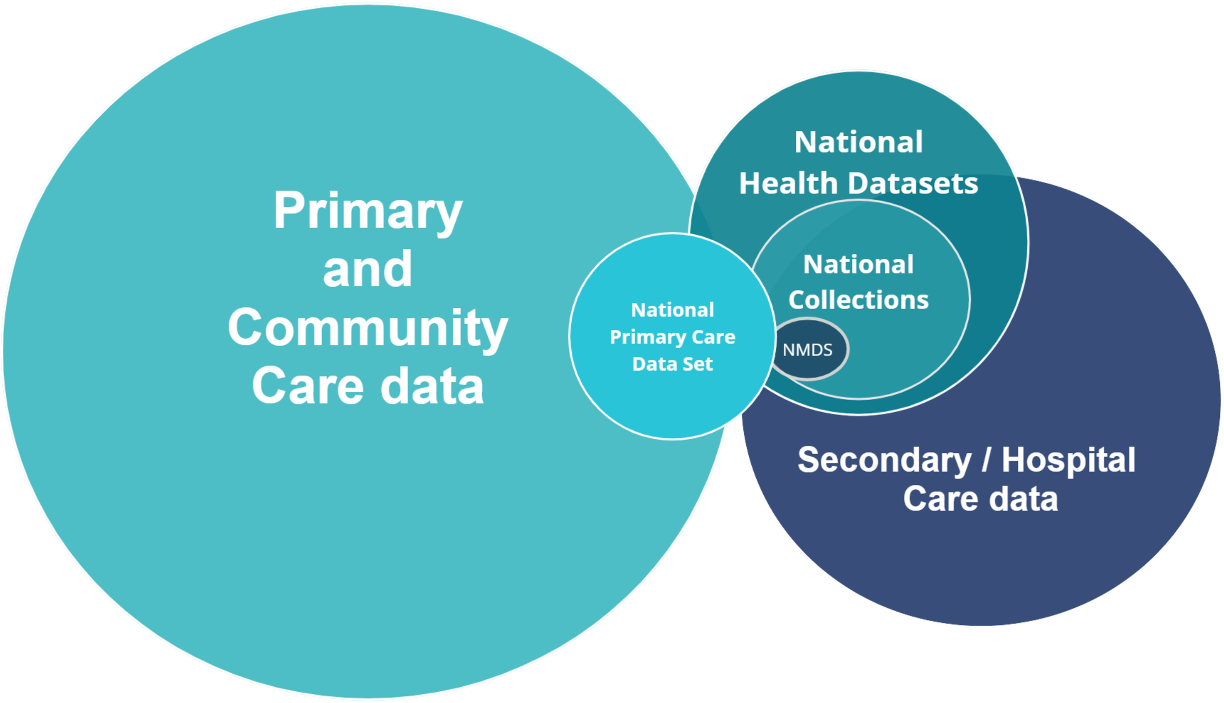
Data required will consist of unit record level transactional data associated with the patient identifier National Health Index (NHI). The resulting core set of primary care data proposed is called the National Primary Care Data Set. Note, previously this had been referred to as the PMDS (Primary Minimum Data Set) but this term has been replaced with National Primary Care Data Set.

Data held in the Te Whatu Ora 37 National Health Data Sets will complement data that is currently not held nationally. For example, the National Minimum Data Set (NMDS) is comprised of data sourced from hospital level care and this combined with National Primary Care Data Set data could be available (with appropriate governance) to all key stakeholders in the health sector. This would include providers internal and external to Te Whatu Ora, in accordance with those governance approvals.

Nationally consistent and joined-up data will specifically support the establishment of Localities and Iwi Māori Partnership Boards (IMPBs) to support whanau and community wellbeing, academic research and innovation as well as for public health surveillance purposes.

Over time the National Primary Care Data Set will also be able to promote new treatments, improve transparency in medication uptake, etc., and link supporting direct patient care, clinical audit and clinical risk management (within appropriate data governance provisions).

Figure 1 – Data domains



The Primary Care Data Working Group, with feedback from all other working groups co-designed the National Primary Care Data Set. There are multiple sources of data for some classes of the National Primary Care Data Set. Further analysis of sources for some data classes will need to be explored in terms of accuracy, consistency, privacy and security in the implementation phase.

The National Primary Care Data Set was developed with a view of the population health approach(es) described in section 2.4 Population health management pain points. The development also addresses **Pae Ora (Healthy Futures) Act 2022 strategies** , a **population health approach in primary health care** in line with Health Quality and Services Commission (HQSC) triple aim for quality and safety[11], as well as policy question

---

[11] https://www.hqsc.govt.nz/our-work/leadership-and-capability/about-us/#:~:text=The%20Triple%20Aim,system%20with%20a%20common%20purpose.

recommendations from Amy Downs[12] , Primary Care Research Network[13], and from Gentil, et al[14].

3.1 Data sources for the National Primary Care Data Set

The recent Current State Analysis[15] proposed the data classes and data elements making up the National Primary Care Data Set (reference Table 4 - National Primary Care Data Set potential data sources, or Appendix B.1 National Primary Care Data Set workbook for a detailed overview).

Working group discussions through the PHDP Phase 2 - High Level Design confirms the majority of data classes originate from general practice PMSs[16]. Some data could come from data sources already held by Te Whatu Ora, while some could come from community laboratory systems, also noting the plans to develop a national laboratory systems repository within the Hira programme.

The Primary Care Data Working Group recommended that the data can be sourced weekly or preferably daily in most cases.

Table 4 - below - summarises the most likely data classes and sources.

Table 4 – National Primary Care Data Set potential data sources

Source	Data class(es)
Practice Management Systems and PHO Data Sets	• Conditions • Immunisations (primary care) • Medications (prescribed and some dispensed) • Smoking • Measurements and vital signs • Diagnostic results (numeric, coded)

---

[12] Downs, Amy. 2017. Ian Axford (New Zealand) Fellowships in Public Policy, Fulbright New Zealand. Wellington, New Zealand. https://www.fulbright.org.nz/wp-content/uploads/2017/09/DOWNS-From-Theory-to-Practice-The-Promise-of-Primary-Care-in-New-Zealand-.pdf (accessed December 2022)

[13] Pigden A et al. Developing a national primary care research network: a qualitative study of stakeholder views. Journal of Primary Health Care 2022 doi:10.1071/HC22081

[14] Gentil, M et al. 2017.Factors influencing the development of primary care data collection projects from electronic health records: a systematic review of the literature. BMC Medical Informatics and Decision Making. 17:139 DOI 10.1186/s12911-017-0538-x

[15] Fernando, Gayatri. 2022. PHDP Current State Analysis. Wellington, Health New Zealand. (p.14) https://www.tewhatuora.govt.nz/publications/primary-health-dataset-programme-current-state-analysis-report/

[16] Ibid.

	• Encounters • Referrals (incl. radiology orders) • Procedures • Pregnancy • Vaping • Care Plan • Functional status
Medicines Data Repository	• Medications (dispensed and prescribed)
National Immunisation Register	• Immunisations (public schedule, not adult flu)
Laboratory Systems	• Diagnostic results (numeric, coded)[17]
National Enrolment Service (NES)	• Qualifying Encounters (GPQED) (under PHOSA rules)
NHI	• Patient demographics
National Minimum Data Set	• Conditions (hospital discharge diagnosis only)
e-Referrals Management System	• Referrals (incl. radiology orders – coded)

All data sources have been found to have inconsistencies, which will require a curation process to resolve. The level of transformation is variable based on each element. Current State Analysis was undertaken with PHOs in Phase 1 of the PHDP programme. This work by the group, identified at a high level that consistency is an issue in 100% of data classes, followed by data completeness issues at 89% of data classes. Uniqueness, timeliness and validity issues, on the other hand, are found in 22% of the data classes[18] respectively (refer to Table 5).

---

[17] The Data Set group recommended going directly to labs, however that is a whole project in itself which will involve changes to lab systems and notification of results. In the interim, it was agreed to source results from PMS/PHO systems.

[18] Excluding the following data classes not introduced into PHDP's Current State Analysis survey: Referrals, Procedures, Pregnancy, Functional Status, Patient Demographics.

Table 5 – National Primary Care Data Set quality issue ratings

Data quality issue	Definition	Data classes affected (out of 9*)
Completeness	Refers to whether all required data is present	**8** (89%)
Uniqueness	The degree to which there are no redundant occurrences of the data	**2** (22%)
Consistency	The extent to which data values are consistently represented	**9** (100%)
Timeliness	Measure of time between when data is expected versus made available	**2** (22%)
Validity	Refers to whether data values are consistent with a defined domain of values.	**2** (22%)

*Based on results of analysis and engagement with PHOs via the Current State Analysis. This excludes 5 data classes added during the course of the Primary Care Data Working Group meetings.

A more detailed summary of the data quality issues across data classes is elaborated in Appendix B.4 National Primary Care Data Set – data quality issues (detailed view).

3.2 “New” data sources

It is clear that the data classes and elements proposed for the National Primary Care Data Set exist and can be sourced from single or multiple sources. From the PHDP perspective, “new” data sources described in this document are those that have not been used as a data source by any National Collection, specifically for the data elements that are collected to develop the National Primary Care Data Set.

The transformation required for some data elements may be significant but in many cases, this is currently being done for existing consolidated collections. From the current state analysis of PHO data management approaches, this transformation capability already exists within the sector to varying degrees, based on the scale and scope of PHOs or their Management Service Organisations (MSOs).

Through the course of the Primary Care Data Working Group discussions, new data sources have been identified and prioritised including addressing critical gaps vs. quality of data, ease of sharing in the short term (for the two years to 2025) and for up to five years

(2028). This has been done in line with the Data Governance Requirements Working Group proposed purpose-use matrix (refer to Table 6) as well as the Analytics Leadership Group's use case recommendations (refer to Appendix E.1 Population health approach to understanding primary care: analytical questions).

3.2.1 Short term (within first two years) – general practice activity data

It is critical that access to and consistency in defining national health needs and conditions data are available to begin to answer and support the bedding in of the new commissioning framework and various processes to related to the priorities of Pae Ora (Healthy Futures) Act 2022 strategies.

The focus for the first two years is to understand primary care through general practice activity. This involves analysing community health needs and conditions across 10 of National Primary Care Data Set's 14 data classes: Problems / Conditions / Diagnosis, Medications, Immunisations, Smoking, Measurements and Vital Signs, Diagnostic results – Numeric, Coded, Encounters, Vaping, Functional Status and Patient Demographics.

Table 4 identifies data class elements and potential sources. For the first two years, only the PMS sourced data can be considered 'new' relative to the proposed National Primary Care Data Set. All other sources align to existing available national sources such as National Health Data Sets or other national data e.g. the National Enrolment Service (NES).

For some data elements or classes, significant data quality improvements will be required over time as evidenced in Table 5 .

3.2.2 Concurrent data source development

Data sources for some classes or elements will require more specific investigation or progress over the implementation phase:

- Functional Status – Functional Status, will be a field derived from conditions coding to be supplied in the National Primary Care Data Set. Over time this could be replaced by a more detailed and reliable Disabled Status data element. This would be a self-identified and moderated 'disabled status' initiated as part of the Patient Profile (PP)/National Health Index (NHI) Disability Data programme currently underway. When implemented, the Disabled Status field will be captured on the Health Identity Platform[19].

---

[19] To be delivered as part of Horizon One - July 2022 – 2024, Identity and Access Management workstream, Te Whatu Ora Digital Strategy

- Measurements and vital signs – More measurements (e.g. waist circumference) can begin to be added. The Cardiovascular Disease Risk Assessment (CVDRA) measures could be made accessible with appropriate data governance provisions for wider analytical purposes.
- Diagnostic results – Efforts to standardise these across New Zealand are in progress as well as for these to be made available through one national laboratory results repository[20]. It is also recommended for this repository to include hospital laboratory tests as well as community laboratory tests.
- Encounters – It is recommended that encounter type and visit mode are standardised within this timeframe and implemented via the National Enrolment Service (NES).
- Vaping Status and Pregnancy status – Currently, this data is not consistently captured or made visible. It is recommended, however, that both fields are included in the National Primary Care Data Set within the first two years.
- Immunisation – Public schedule immunisation data will be available nationally through the implementation of the Aotearoa Immunisation Register (AIR) system.

3.2.3 Medium term (up to five years)

The Primary Care Data Working Group have identified data sources that are deemed important for understanding primary care in the medium term (up to five years). These are data classes such as Procedures, Referrals, Diagnostic results from a National Laboratory Repository[21] and Care Planning that currently is treated differently within primary care systems.

Recommended data capture and quality progress areas during this five-year timeframe include:

- Enabling Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) coding of diagnosis and clinical findings to be the dominant coding system for General Practice. This will enable a finer level of detail and relationships between clinical findings and disease to be more fully realised.
- Encounter diagnosis codes coming through as part of NES.
- Minor procedures done in a general practice setting captured and included in the National Primary Care Data Set.
- Referral types, ‘services’, ‘reasons’ reference set (in SNOMED CT) developed and incorporated into PMS software in a clinical workflow friendly manner so that this data can then be extracted and used for analysis as part of the National Primary Care Data Set.

---

[20] https://www.tewhatuora.govt.nz/our-health-system/digital-health/hira-connecting-health-information/

[21] *Ibid*.

- Care plan summary coded data.
- Development of microservices[22] in relation to Read to SNOMED CT translation and vice versa.
- Encounter data from community-based systems e.g. pharmacy services (non-prescribing services), physiotherapy services, counselling services coming through to be included into the Encounter type data element.

During this time, some population health management questions with no previous consistent national understanding (see section 2.4 Population health management pain points) can start to be addressed. Some key themes of descriptive analytical needs identified by the Analytics Leadership Group can also be progressed (refer to Appendix E.1 Population health approach to understanding primary care: analytical questions).

---

[22] An architectural pattern that arranges an application as a collection of loosely coupled, fine-grained services, communicating through lightweight protocols

4 National Primary Care Data Set data governance

The health data ecosystem is complex (see Appendix A.5 Primary Care Data Set draft data ecosystem diagram). Governance across and within this system is equally complex, and there remains variation of approach within and external to Te Whatu Ora.

The national health data sets in operation prior to the establishment of Te Whatu Ora, Te Aka Whai Ora and Manatū Hauora are in the process of having their data governance frameworks updated. Similarly, there is a natural variation of data governance within the primary and community care provider community.

The PHDP Data Governance Requirements Working Group (including Māori Sovereignty) recognised that PHDP can act as an enabler to establish parameters for information management improvement amongst Te Whatu Ora, Te Aka Whai Ora and the primary and community care sector. It is acknowledged that transparent and easily understood data governance is critical to PHDP's success. It plays a crucial role in building trust and improving patient outcomes through ensuring that patient data is used transparently, is accurate, secure, and accessible to those who need it. This extends from planning and resource management to direct provision of healthcare to people.

Similarly, adherence to applicable laws and regulations, administrative drivers to ensure the efficient use of finite health resources, and collaboration between healthcare providers are key drivers that underpin the data governance approach.

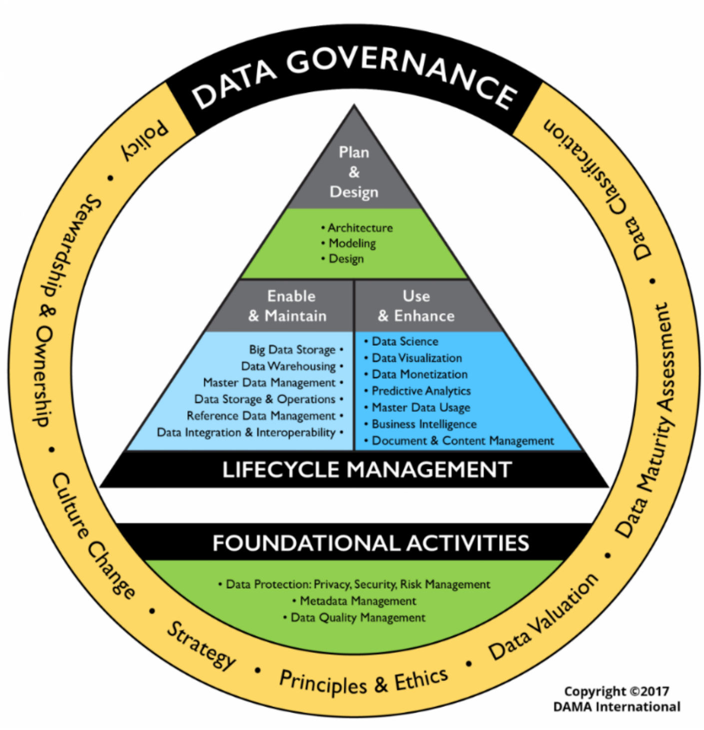
Global industry bodies such as the Data Management Association (DAMA) International[23] also recognises data governance as a critical function from which all other data management knowledge areas are supported (refer to Figure 2 - Data Management Book of Knowledge (DMBOK) Wheel) [24].

---

[23] A vendor-independent, global association of technical and business professionals dedicated to advancing the concepts and practices of information and data management

[24] DAMA. Earley, S., & Henderson, D., Sebastian-Coleman, L (Eds.). 2017. The DAMA Guide to the Data Management Body of Knowledge (DAMA-DM BOK). Bradley Beach, NJ: Technics Publications, LLC.

Figure 2 – Data Management Book of Knowledge (DMBOK) Wheel



4.1 Data governance drivers to underpin National Primary Care Data Set

Governance of a National Primary Care Data Set will need to recognise respective rights, obligations and decision structures across and between organisations. More broadly, it will need to respect the strategic value and cultural significance of data and information across the health sector (government and non-government). This will be about what data gets collected, how it is used, who gets to use it, who makes those decisions, and how Māori can assert tino rangatiratanga[25] and mana motuhake[26] as the Te Tiriti o Waitangi is honoured and upheld.

When data is not governed properly, not only does it risk a breakdown of trust between the patient and their provider it also risks regulatory non-compliance. The current state demonstrates inconsistency in data integrity, i.e. the methods and standards for how data is collected, curated, stored, and analysed as reported in PHDP's Current State Analysis[27]. If effective governance remains unaddressed it can lead to discrepancies in reporting and decision-making based on incorrect or incomplete data, impacting patient care and

---

[25] https://www.health.govt.nz/system/files/documents/pages/whakamaua-tiriti-o-waitangi-framework-a3-aug20.pdf

[26] Ibid.

[27] https://www.tewhatuora.govt.nz/publications/primary-health-dataset-programme-current-state-analysis-report/

satisfaction. Poor data governance can also lead to data breaches and unauthorised access, which can be even more costly and damaging towards maintaining or gaining the public's trust at this time of unprecedented transition within Te Whatu Ora, Te Aka Whai Ora and Manatū Hauora.

Finally, data governance will be critical towards enabling the use of data to develop strategic insight that supports commissioning and planning, policy development, service development, management and improvement towards achieving Pae Ora (Healthy Futures) Act 2022 strategies.

4.2 Factors impacting primary care data Governance

The health reforms strongly signal the targeted and equitable delivery of health services. Data informs and enables priority outcomes of access and equity, requiring a more cohesive, consistent, and transparent approach to data governance across the health system. This means:

- increased connectedness of data, information and insights
- improved data quality, accessibility and flow to support integrated care
- greater participation, transparency & trust for consumers and data users.

A partnership approach to primary health care data governance is critical in supporting Te Whatu Ora data strategies. There is a strong vision for the mutual value that can be created by connecting detailed, patient-level data from primary care with data from other parts of the system – secondary care, tertiary care, population health and determinants of health from sources outside of the health sector.

There are some constraints on data uses based on the purpose of collection – there are legal or privacy obligations only to use data for the purpose(s) that is collected, or with patient authorisation, or where a relevant exception applies (e.g. research and statistical purposes). The programme will retain an awareness that the health consumer remains at the centre of considerations when using this sensitive health information about them. Over time the PHDP approach allows for expansion the ways data can be used only if privacy statements are updated to provide increased clarity to consumers about the ways Te Whatu Ora or the sector plan to use their data to benefit them directly and people like them.

Different stakeholders would like to use those connected data sets for different purposes. In some cases, there is a desire to use connected national data about individuals for indirect patient services, however there are many other secondary uses for the data – for policy, planning, service improvement, quality improvement, whole-of-system performance monitoring, research and evaluation, for which PHDP sourced and connected data sets could provide system and individual benefits. Governance provisions for each use case would need to be explored.

There are also some options about what data sources are used and what processes can be developed to build trust relating to the use of linked data. There is a role for the PHDP programme, the Hira programme, the National Data Platform (NDP) programme, Data & Information Strategy for Health & Disability (DISH) implementation, existing primary care summary records, existing shared care records here as well as linkage involving national data sets.

4.3 Data & Information Strategy for Health & Disability (DISH)

PHDP is proposing the establishment of the National Primary Care Data Set. Running in parallel is the establishment is Te Aka Whai Ora’s Māori Sovereignty Framework[28], Te Whatu Ora’s Data & Information Strategy for Health & Disability (DISH)[29], and the health agencies’ Data Tripartite Agreement which proposes tiered models of data governance and oversight.

The tiers span executive leadership of strategic data governance across the sector, through to operational groups driving a consistent and cohesive approach across the motu. The Data Iwi Leaders Group/Stats NZ Māori Data Governance project also informs an all of government approach to data governance.

Figure 3 - Building blocks of data governance for Data & Information Strategy for Health & Disability (DISH) below illustrates the in-scope components of data governance that have been identified as the most important to support the implementation of DISH and supported by the Data Governance Requirements Working Group to address specific gaps in the data and information ecosystem and support a range of strategic outcomes.

Desired improvements resulting from DISH include matters such as freeing-up the flow and use of data and information through the health system and increasing data maturity and capability. The data governance improvements envisaged by DISH are primarily concerned with principles such as rangatiratanga, data sovereignty, rights, roles and responsibilities, equity, accessibility, transparency, participation, and consistency. This has strong alignment to PHDP and the establishment of the National Primary Care Data Set.

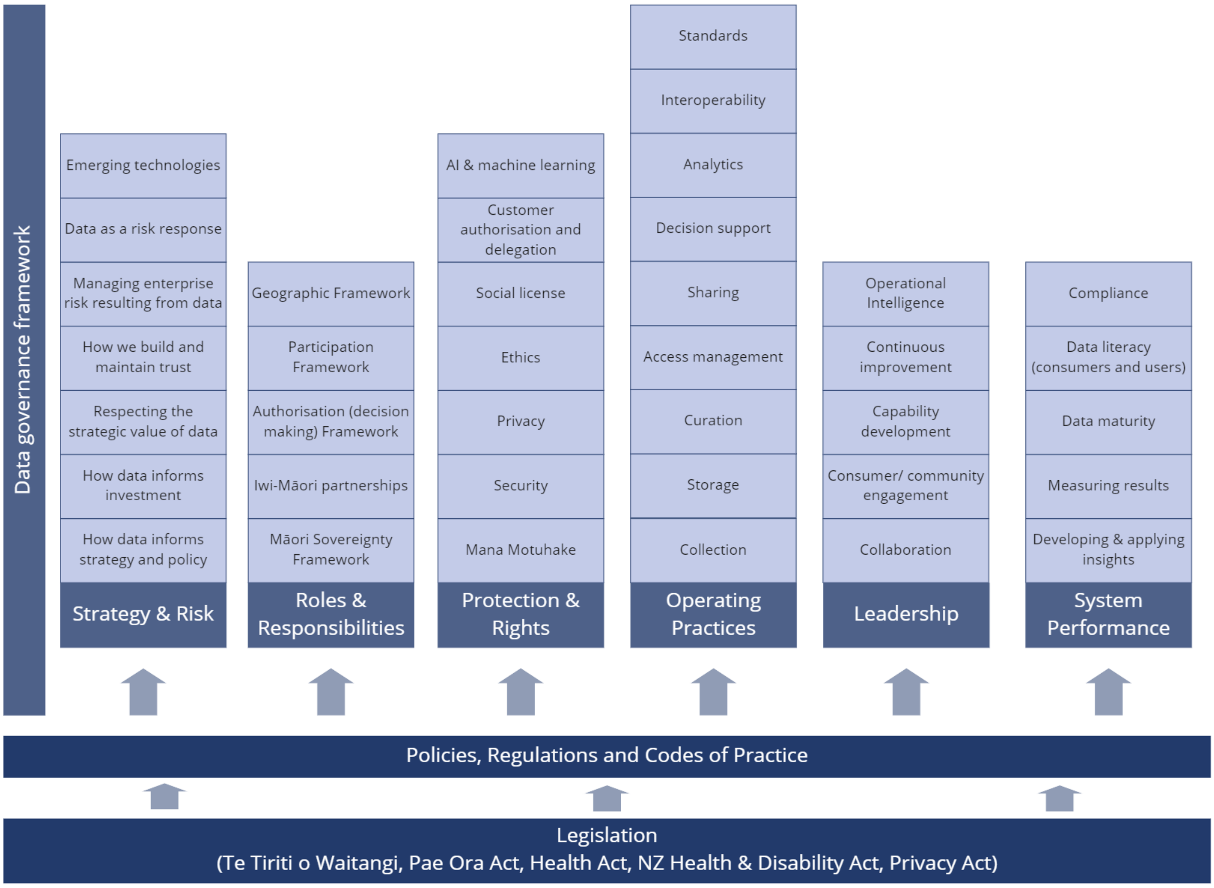
---

[28] Te Pou Tarawaho te Rangatiratanga Maori; Raraunga-oranga, mohiohio, noho matatapu me te whakahaumaru – A Maori Sovereignty Framework: Health Data, information, privacy and security (pre-draft), 2022, Robyn Kamira, Paua Interface Ltd, for Joint Boards.

[29] https://www.health.govt.nz/system/files/documents/publications/13308_data_strategy_document_final_9_dec.pdf

Overall, DISH is concerned with building on the commonly applied semantic and technical elements of data governance, by creating greater cohesion and connectivity in how we treat data as taonga, and enhancing the more ‘human’ aspects that respect decision rights (e.g. mana motuhake), increase participation, ensure transparency and build trust.

Figure 3 – Building blocks of data governance for Data & Information Strategy for Health & Disability (DISH)



The proposed ‘building blocks’ structure places the planned in-scope data governance components for National Primary Care Data Set within six categories[30] as follows.

- **Strategy & Risk** – how data, information, and insights contribute to the sectors strategic outcomes and risks, protect the integrity of the health system (government and non-government), builds trust between the primary care sector and government agencies, and informs our strategic investments in the health system. It includes the contribution made by data, information and insights to mitigating risks to health outcomes, our desire and ability to make data accessible across the sector, the use (or

---

[30] Inland Revenue’s Data & Information Policy identifies similar themes, which have been amended to reflect the health context. Correspondence via Neil Annenberg, former Inland Revenue Employee, March 2023, via email.

absence) of data and information to inform strategic and policy decisions, and its quality (timeliness, accuracy, completeness, relevance).

- **Roles & Responsibilities** – who and how we mutually govern data, information, and knowledge. This includes decision-making, expert and advisory bodies with appropriate representation for the nature of decisions and subject matters being considered. It defines the roles, accountabilities, and responsibilities for data, information, and knowledge.
- **Protection & Rights** – standards that regulate our stewardship and use of data, information, and knowledge. These standards include references to and alignment with national and sector specific legislation, policies, and published standards.
- **Operating Practices** – how the day-to-day management and use of data, information, and knowledge for clinical and non-clinical purposes is applied, by consumers and data users. This includes how data and information is collected, used, accessed, shared and analysed within and beyond the health sector, for a wide range of direct (e.g. health care), indirect and secondary purposes. It includes the protection and quality assurance of data, information and insights. For example, the development of data standards and resources to help address issues of bias in the collection and use of health data.
- **Leadership** – how leaders collaborate across the health system, develop workforce capabilities, collaborate between sectors, partner with health care providers (government and non-government), facilitate ongoing improvements to the use of data and identify and apply intelligence and insights.
- **System Performance** – measurement and management of the contribution data, information, and knowledge make to achieving shared objectives. It includes how data literacy and system capabilities are ensured to be in place across the sector to collect, use, analyse and share data, information and insights. It also includes how well the system is performing around shared objectives.

4.4 National Primary Care Data Set governance oversight

A documented, robust, useful and transparent data governance and management approach across all aspects of the data is required for PHDP. DISH will inform this overarching governance including a series of formal (recognised and visible) interrelated groups, standards, guidelines, rules and tools.

A joint data governance approach between health agencies and primary sector providers respects the independent organisational rights, needs and accountabilities, and provides a framework for each parties' common interests in and ambitions for data and information to be fulfilled. This will likely include a combination of:

- a health sector national data governance body that establishes common principles, expectations and positions on the collection, use, analysis and sharing of data and information;
- organisational and/or geographic decision making and/or advisory bodies, that apply data governance in the context of their specific circumstances;
- networks of data consumers and practitioners on specific subject matters in clinical and operational domains, e.g. data analytics, digital innovation, service design;
- documented frameworks for broad application (mandates, structures, expectations, criteria, protocols) and that require specific and formal 'rules and tools' to manage participation, optimise value, promote compliance and consistency;
- published artefacts that can be applied in a number of circumstances to aid participation, capability development, and monitoring, e.g. guidelines, standards, templates, measures.

4.5 A National Data Platform and its relationship to the National Primary Care Data Set

Te Whatu Ora is progressing the establishment of a National Data Platform to enable the technical mechanism for data and information to be shared, stored, accessed, and curated for multiples agencies and organisations across the health sector. For the primary and community care sector and the health agencies, there is huge opportunity for the health agencies to leverage mutual value for the benefit of health consumers and the improved effectiveness of connected services. Information would be more accessible, more secure, and more transparent. Sharing data and insights would reduce duplication and accelerate innovation.

It is recognised that the primary care sector, largely through PHOs and MSOs has developed platforms locally and, in some cases, regionally. The establishment of a National Primary Care Data Set will require joining of resources to establish a shared platform that can meet day-to-day demands of the sector.

This requires an agreed set of rules and tools about how, what and where data is stored, shared, and used, for what purpose, and under what conditions and controls. As a minimum, it is anticipated this means:

- A common set of core ‘ground rules’ and principles for any consumer whose data is stored on or accessed via the data platform.
- Partitioned zones to enable access by organisations, respecting rights, needs and accountabilities of each entity.
- Agreements and/or protocols for how, when and what data is stored, shared, with whom, and how value is derived.
- Standards, rules, and tools to ensure data conformity, integrity, quality, and assurance, so that it can be relied upon, is secure, and useful.

4.6 Maturity and evolution of data governance capability

The state of maturity of data governance processes varies widely internationally depending on factors such as regulatory environment. In general, developed countries and industries such as finance and healthcare tend to have more mature data governance processes[31].

However, even within developed countries and industries, there can be significant variation in maturity levels. Some organisations have robust data governance frameworks in place that are integrated into their overall business strategy, while others may have ad-hoc or reactive approaches to managing data. Additionally, the increasing importance of data sovereignty, social licence, data privacy and security regulations in recent times have led to a greater focus on data governance processes, with many organisations working to improve their maturity levels in this area.

Within the New Zealand health system, there is acknowledgement of the need to improve the state of data governance across healthcare organisations. PHDP’s Current State Analysis report has already previously revealed data governance as a capability with the least confidence rating from across the PHOs (at 76%, compared to an average 83% self-rating) relative to other data capabilities[32].

Within the context of the PHDP working groups it was acknowledged by all stakeholders that all are living in ‘glass houses’ to some extent and there is work to do, both within and external to Te Whatu Ora, to evolve data governance approaches for the benefit of the

---

[31] CMMI Institute, (2019), “Data Maturity Model at a glance”, accessed 15th of August 2022, [*https://stage.cmmiinstitute.com/getattachment/cb35800b-720f-4afe-93bf-86ccefb1fb17/attachment.aspx*] Elias Baltassis, Anne-Douce Coulin, Antoine Gourévitch, Yassine Khendek, and Lucas Quarta, (2019), *“A Rough Road to Data Maturity”, accessed 15th of August 2022, [https://www.bcg.com/publications/2019/rough-road-to-data-maturity]*

[32] Fernando, Gayatri. 2022. PHDP Current State Analysis. Wellington, Health New Zealand. https://www.tewhatuora.govt.nz/publications/primary-health-dataset-programme-current-state-analysis-report/

people. There are pockets of expertise but also “deserts” where there is low maturity or attention in terms of data governance. This will take time and all stakeholders acknowledge that all organisations will have to work in partnership on improvement of data governance.

Consistent with this, the programme team developed and presented Appendix A.2 Draft early thinking around the secondary use of primary health data. This was first shown to the Data Governance Requirements Working Group on 2 February 2023, and then subsequently shared with all other working groups. It documents an inventory of prompts and placeholders to elicit data governance policy and process, as well as strategic positions and operational requirements across categories to use and gain access to the National Primary Care Data Set.

This was based on a review and analysis of various sources. These sources included a primary care data development project from the Australian Institute of Health and Welfare[33], global standards such as International Standards Organisation (ISO) / International Electrotechnical Commission (IEC) 38505-1 Application of ISO/IEC 38500 to the governance of data[34], Control Objectives for Information and Related Technology (COBIT) governance standards[35], lessons from the programme teams’ discussion with the Metro Auckland Data Sharing group and a view of their regional data governance processes (refer to Appendix A.6 Metro Auckland Data Sharing Group conversation notes, as well as a review of some international precedents (refer to Appendix A.4 PHDP references on international precedents).

4.6.1 National Primary Care Data Set data governance constructs

In context of working across health agencies and the primary and community care sector, a common data governance framework for National Primary Care Data Set needs to operate at two distinct but interdependent constructs that are equally important:

- Thinking of and treating data as taonga – deals with decision rights, values, expectations and principles over the appropriate safe, transparent and respectful collection, sharing, curation and use of data. Good data governance should respect organisational sovereignty, mandates, and accountabilities. It should build trust and care for the enduring value of data and information for the overall benefit of the sector; it is very much about stewardship of data and information.

---

[33] Commonwealth of Australia as represented by The Australian Institute of Health and Welfare. 2019. National Primary Health Care Data Asset Data Development Plan. Canberra. The Australian Institute of Health and Welfare.

[34] https://www.iso.org/obp/ui/%23iso:std:iso-iec:38505:-1:en

[35] https://www.isaca.org/resources/cobit

- Working with data – involves design and application of the rules, tools, methods, structures, standards and processes that enable data and information to be collected, stored, maintained, shared and transformed. This also includes how people, processes and systems interact with operational services and is very much about custodianship of data and information.

Data governance models also typically refer to governors and participants as stewards (kaitiaki) and/or custodians of data and information, sometimes using the terms interchangeably. A draft working definition of the difference between stewardship and custodianship proposed by the Te Whatu Ora data governance team in August 2022 is as follows:

- **Stewardship** (kaitiakitanga): acknowledging that people are the true kaitiaki of their information, health agencies are stewards of government and health sector-held data and information. Agencies must provide and require good practices which manage the data and information over their lifecycle, including catering for technological obsolescence and long-term preservation and access. Good practices include collaborating with other agencies, providers, iwi-Māori, communities and the public, facilitating access, strengthening awareness, building trust, and supporting cooperation within and beyond the health sector.
- **Custodianship**: custodians implement stewardship practices on a day-to-day basis. Stewards may delegate specific responsibilities to custodians. For example, by providing access to national health data sets, Te Whatu Ora may set expectations and obligations on a data consumer to manage and use information within agreed standards.

4.6.2 National Primary Care Data Set and data "ownership"

It is common for people to wonder who owns their health information. However, ownership is not necessarily the best way to think about health information. It is more accurate to say that "People have rights over health information about themselves". Rule 6 of the Health Information Privacy Code (HIPC)[36] gives individuals the right to access information about themselves and Rule 7 (HIPC) gives them the right to seek correction of that information if they think it is inaccurate or misleading.

---

[36] https://www.privacy.org.nz/privacy-act-2020/codes-of-practice/hipc2020/

Health agencies have obligations over the health information they hold. These obligations are set out in the 13 rules of the Health Information Privacy Code.

4.6.3 National Primary Care Data Set purpose statement

The PHDP Data Governance Working Group agreed to the following purpose statement (refer to Appendix A.1 Purpose statement documentation for more detail) to underpin data governance arrangements and ancillary processes including privacy provisions:

"*The National Primary Care Data Set approach will gather, analyse, and when appropriate, share information about the health conditions and encounters experienced by all New Zealanders.* In Aotearoa New Zealand, people have differences in health that are not only avoidable but unfair and unjust. Equity recognises different people with different levels of advantage require different approaches and resources to get equitable health outcomes.

This new collection is designed to provide more complete national information to improve national service delivery and share experiences of success. Early identification of areas of risk to our population, and comprehensive management of issues of concern to us all will lead to improved health outcomes.

This information will be held securely, and not used in an identifiable format unless specifically agreed by the data governance group.

This information will:

- *enable understanding of the population groups and regions most at risk of not receiving the health care they need to meet equity needs*
- *help develop national, regional and local policy that will best meet the needs of all New Zealanders*
- *enable targeted support with the right resources at a population level*
- *check how effective health intervention strategies have been*
- *enable early identification and management of infectious diseases and health conditions*
- *be shared with health system funders, Iwi Māori Partnership Boards, healthcare providers and their support networks as anonymised data sets and analytical results".*

4.7 Implementation considerations

As part of the implementation phase there will be the development of Privacy Information Statements to support people and clinicians when discussing the use of data associated with National Primary Care Data Set.

Some input from the PHDP Data Governance Working Group identified the following matters could be considered in further implementation planning:

- An operational framework be developed that preserves the privacy and autonomy of personal information and is consistent with the principles and obligations of Te Tiriti o Waitangi, Māori data sovereignty, privacy legislation, security standards, and all ethical and best practice guidelines.
- Pae Ora (Healthy Futures) Act 2022 health sector principles. The importance of, and the need to address issues of equity and Tiriti obligations must be addressed, including mauri ora (healthy individuals), whānau ora (healthy families), and wai ora (healthy environments).
- Decisions related to consenting frameworks, including opt out or opt in to preserve consumer choice in relation to identifiable data (on the assumption that use of de-identified data is acceptable in terms of social licence – but the disclosure of identifiable information from the collection source requires an opt out option).
- Access to information to allow health providers in our communities to work in partnership with Te Whatu Ora, Te Aka Whai Ora and Manatū Hauora to improve the delivery of targeted healthcare to the New Zealand community within the boundaries of the purpose of National Primary Care Data Set.
- Protocols to support national data sets to be more readily accessible to Māori providers, and Iwi Māori Partnership Boards to ensure the health system is able to respond to Māori sovereignty in the ownership, management and access to data.
- The analytical anonymised data sets will be shared with national, regional and local health service providers to enable data comparison and help to improve service delivery in their own regions or to share their experiences of success with other providers to improve national service delivery.
- Use of identifiable data sets to be subject to prior review and authorisation by a National Primary Care Data Set data governance group with expertise from both health agencies and the primary and community care sector. Use cases that might be considered include:
    - Possible introduction of a pathway for consideration of individual research projects - subject to appropriate consideration of Māori sovereignty, preserving privacy of information, and requirements to subsequently obtain research ethical approvals

etc., via established Health and Disability Ethics Committees (HDECs) channels etc.).

  - Other appropriate uses identified that are fully reviewed from a privacy perspective via the National Primary Care Data Set data governance group.
- Addition of section about what will not be done with the information.

Additional considerations will include a comprehensive communication strategy and will include collaboration with Health Quality and Safety Commission (HQSC) voice of the consumer representative, and overall Te Whatu Ora strategy.

Similarly, implementation will require a review and alignment of documentation at original source points – e.g. general practice and PHO enrolment or laboratory tests etc will need to be reviewed and potentially updated to ensure transparency for health consumers and compliance with Rule 3 of the HIPC.

4.8 Overarching direction

The case for use of National Primary Care Data Set must be compelling for the custodians of the information (currently the healthcare providers) for them to be comfortable sharing (disclosing) the most sensitive of information about the health consumer.

Similarly, it needs to be compelling to the person themselves that their information is secure and will be used within the context of the purpose of National Primary Care Data Set. The trust of the health consumers must be retained – or there is a risk of negative health outcomes (distress or unwillingness to share information with a health professional for fear it may be repeated in a way the person does not understand or agree with), and diminishing or lost trust in the digital initiatives to support health care.

4.9 National Primary Care Data Set assurance: protection of the privacy of individuals and providers

The Data Governance Requirements Working Group was able to tease into, validate and explore the associated processes and high-level requirements that managing a data set like the National Primary Care Data Set might need that includes the processes and controls mentioned below (see also Appendix **Error! Reference source not found.**. REF _Ref134005542 \h **Error! Reference source not found.**. for more details).

The establishment of a National Primary Care Data Set will require best practice and enduring privacy protection mechanisms. These already exist to protect the privacy of consumers and healthcare providers. Some of these critical processes and controls include:

- **The Privacy Act 2020**[37]. This Act provides the legal and policy foundation for the rules and regulations that govern the collection, use, and disclosure of personal information by government agencies and private sector organisations.
- **Health Information Privacy Code (HIPC)**[38]. Provides a set of guidelines developed by the Office of the Privacy Commissioner to protect the privacy of personal health information. This HIP code sets specific rules for agencies in the health sector. It covers health information collected, used, held and disclosed by health agencies and takes the place of the information privacy principles for the health sector.
- **Informed collection**. In line with both the Privacy Act and the HIP Code, healthcare providers must ensure the health consumer is appropriately informed before collecting or using their personal health information (typically via the Primary Health Organisation’s enrolment form in the primary care setting). Patients must be fully informed about the purpose and scope of the information that is being collected. With this new National Primary Care Data Set all initial source collection points will need to be reviewed, and potentially updated.
- **Information security**. Healthcare providers are required to implement appropriate security measures based on the HISO 10029:2022 Health Information Security Framework (HISF)[39], in line with the Protective Security Requirements[40] to protect the privacy and confidentiality of patient information. This includes the use of technologies such as encryption, firewalls, and access controls to prevent unauthorized access to health information.
- **Privacy Impact Assessments**[41]. Healthcare providers are also expected to ensure privacy impact assessments have been conducted to help those agencies identify the potential risks arising from their collection, use or handling of personal information, and to confirm that they are meeting their legal obligations.
- **Data breach reporting**. Healthcare providers are required to report any data breaches that have caused or are likely to cause serious harm to the New Zealand Privacy Commissioner, and affected individuals, as soon as possible. This ensures that patients are informed of any potential privacy breaches and can take steps to protect their personal information.
- **Ethics and code of conduct**. All public servants, regardless of their department or agency, must act with a spirit of service to the community and meet the standards set by the Public Service Commission (previously called the State Services Commission)

---

[37] https://www.legislation.govt.nz/act/public/2020/0031/latest/LMS23223.html
[38] https://www.privacy.org.nz/privacy-act-2020/codes-of-practice/hipc2020/
[39] https://www.tewhatuora.govt.nz/publications/health-information-security-framework/
[40] https://protectivesecurity.govt.nz/
[41] https://www.privacy.org.nz/responsibilities/privacy-impact-assessments/

which issued “The Standards of Integrity and Conduct” (the Code) under section 57, of the State Sector Act 1988. This continues to have effect under the Public Service Act 2020 and applies to all employees of the organisations defined in section 17(2) of the Act. Breaches of the Code are managed through employment arrangements at the agency level, in the first instance. By maintaining the expectations set out in the Code, public servants enable the Public Service to maintain the trust and confidence of people in New Zealand.

- **Professional Obligations**. Health practitioners working in health agencies have ethical and professional obligations and are subject to responsible authorities under the Health Practitioners Competence Assurance Act 2003.

4.10 Accountability: controlled management and use of data

Management policies and controls that currently exist or those to be developed are expected to be in line with the national health data sets data governance framework as it is developed. Areas to be established or developed further include:

- **Data governance**. A data governance framework to oversee the management and use of National Primary Care Data Set data. This framework includes policies and procedures for National Primary Care Data Set data collection, storage, access, use, and sharing. It will also specify the roles and responsibilities of individuals and organisations involved in the handling of National Primary Care Data Set data, clearly delineating management versus governance processes and the roles and responsibilities of National Primary Care Data Set.
- Direction and establishment of **procedures for obtaining consent and ensuring National Primary Care Data Set data privacy and security**.
- Alignment with a **risk management framework**[42] to identify and manage potential risks associated with the collection, use, and sharing of health data. This framework includes procedures for risk assessment, risk mitigation, and risk monitoring.
- **National Primary Care Data Set data security**. Appropriate technical and organisational measures such as Health Information Standards Organisation security framework[43] that ensures the security of health data. This may include access controls, encryption, and data backup and recovery procedures.

---

[42] https://mohgovtnz.sharepoint.com/sites/moh-ecm-PubDocs/SitePages/CDocViewer.aspx?CDocID=77
[43] https://www.tewhatuora.govt.nz/publications/health-information-security-framework/

- **National Primary Care Data Set audit and monitoring**. Regular audits and monitoring conducted to ensure compliance with legal and regulatory requirements, data governance policies and procedures, and data security measures.
- **Transparency and Accountability**. Demonstrating transparency in the handling of health data, and how individuals are informed about how their data is collected, used, and shared. Accountability is also ensured through clear lines of responsibility and consequences for breaches of data privacy and security that will be referenced in the upcoming data governance framework. Various artefacts include:
- A National Primary Care Data Set data access framework[44] describing the principles, rules and processes for who/how can access national health data (typically data sets) so that what's available is known, accessible and useful to those with a legitimate purpose and the appropriate mechanisms to be good custodians.
- A National Primary Care Data Set data sharing framework that captures the expectations for how, by whom and with whom data is shared across the sector so that the flow and application of data is transparent, legal, ethical, secure, appropriate, trustworthy, and adds value to sector outcomes. Pending design, this may be combined with the data access framework, and a National Primary Care Data Set consumer authorisation and delegation framework speaking to the mechanisms, tools and options that should be available to health care consumers. Includes expectations of how consumers will be informed and empowered to understand why and how their information is used, to permit the collection and use of their information according to their rights and preferences, and how access rights can be delegated to other people.
- National Primary Care Data Set training and education. Those involved in the handling of National Primary Care Data Set data receive appropriate training and education on legal and regulatory requirements, data governance policies and procedures, and data security measures to ensure that health data is handled in a responsible and accountable manner[45].

The National Primary Care Data Set, among other national data sets, will be governed by a group of experts drawn from the sector that includes data, consumer and clinical expertise and operates as primary care data specialists in line with the data governance framework and under the direction of the proposed data governance structures.

While a data governance structure has yet to be finalised, examples of data governance models such as federated data governance approaches[46], or Robert Seiner's non-invasive data governance model[47] suggest distinctive structures and responsibilities typically at

---

[44] https://www.health.govt.nz/publication/current-data-access-policy currently being updated as at March 2023

[45] https://www.health.govt.nz/nz-health-statistics/access-and-use

[46] Ladley, John, Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program 2nd edition, 2019, Academic Press

[47] used as a reference by MBIE, as per Gayatri Fernando, a former MBIE employee until June 2022.

three levels – a tactical / custodian, domain / stewardship and executive / strategic level to effectively manage and govern data.

4.11 Harnessing the power of the National Primary Care Data Set

Key to harnessing the power of the National Primary Care Data Set involves clear data governance, supported by a purpose-use framework. Integrated data sources by way of clear analytical approaches would allow deeper understanding of health needs, outcomes and trends, the use of advanced analytics and data visualisation, greater collaboration across the system and beyond it. This would be underpinned with strong privacy and security of the data, and robust stakeholder engagement to understand the needs of the consumers and data users.

While there may be data available for regional or national population analysis for PHOs[48], many of the existing data sets are not necessarily consistent, or standardised at a national level. The PHDP intends to help address this, for example, as the Primary Care Data Working Group agreed to recommend adoption of the Health Information Standards Organisation (HISO) New Zealand International Patient Summary (NZIPS) standards[49].

NZIPS is largely based on globally recognised Systematized Nomenclature of Pathology Clinical Terms (SNOMED CT)[50] standards, and the use of its data class and element definitions will introduce national consistency across newly linked data that was previously unavailable or unlinked (e.g. National Primary Care Data Set and Medicines Data Repository). Feedback had also been sought and collected from the Analytics Leadership Working Group surrounding the National Primary Care Data Set itself, as well as future data quality and standards efforts.

The recommended high level purpose and use framework is illustrated in Table 6 (refer to Appendix A.9 Purpose-use framework documentation for a detailed view). The framework presents the various stakeholders expected to use and gain insight and value from the National Primary Care Data Set and for what purpose.

This framework has been developed based on the Health Safe[51] and Metro Auckland Data Sharing Group framework, with indications to highlight what is new or modified from a National Primary Care Data Set context.

---

[48] Correspondence with Wing Cheuk Chan, 14 April 2023, via email.

[49] https://consult.health.govt.nz/hiso/hiso-10099-2022-nzips/

[50] https://www.health.govt.nz/nz-health-statistics/classification-and-terminology/new-zealand-snomed-ct-national-release-centre/snomed-ct

[51] Based on the analysis of Health Safe framework documentation provided to the PHDP Programme team

It is important to note that while PHDP's scope has been to prioritise analytical, de-identified and anonymous data mainly, it is crucial to state that identifiable data can also be shared as long as, where appropriate, this request has been approved by national data governance structures that are established to support National Primary Care Data Set and the required privacy requirements have been met. "Opt out" provisions for this use of data will be explored.

Table 6 – PHDP stakeholder recommended purpose-use framework

Purpose category	**Active patient care**	**Indirect patient services**			**Health service support**					**Research and Innovation**	
Use of Health Information - Consumer view	(a) To make decisions about your health care now				(e) To make decisions about improving the health service						
	(b) To make decisions about your health care in the future				(f) To investigate how to better understand the population and their needs by combining information on the whole population.						
	(c) To share with other health professionals involved in your care within the same organisation				(g) To continue to help others even once you have died or have moved out of New Zealand where your information continues to be useful and contributes to the full picture for points (e) and (f).						
	(d) To share with other health professionals involved in your care in other organisations										
Stakeholder Group	**Direct Patient Care**	**Clinical Audit**	**Clinical Risk Management**	**Public Health Surveillance***	**Service Management and Development**	**Monitoring & Accountability**	**Commissioning and Planning**	**Improve (Community) Well-being***	**Public Health Intelligence**	**Research**	**Research (ethics approved)**
Health Care Providers (incl PHOs and PHUs)	**Y**	**Y**	**Y***	**Y***	**Y**	**Y**	**Y**	**Y***	**N/A**	**Y**	**Y**
Funders/ Central Agencies	**N/A**	**N/A**	**Y***	**Y***	**Y**	**Y**	**Y**	**Y***	**Y**	**Y**	**Y**
Pae Ora - IMPBs, Localities*	**N/A**	**N/A**	**Y***	**Y***	**Y**	**Y**	**Y**	**Y***	**Y**	**Y**	**Y**
Individual, Whanau, Communities*	**Y**	**N/A**	**Y***	**Y***	**Y**	**Y**	**Y**	**Y***	**N/A**	**N/A**	**Y**
Privacy Requirement	Direct or indirect patient consent (Health Information Privacy Code Rule 3) or as required by law				Consumer is informed about the purpose of use (Health Information Privacy Code Rule 3 and data may be used in a non-identifiable way Rule 10(e)						Individual consent and Ethics approval
Level of identifiability	All identifiable (NHI level) data*				Unit level de-identified data used for analysis Outputs and reports are anonymised (not able to re-identify individual from the outputs)*						Identifiable if direct patient consent given

* indicates modifications or additions to the HealthSafe or MetroAuckland Data Sharing Group Purpose-use matrix

5 Māori Sovereignty Framework: Health data, information, privacy and security

5.1 Māori Sovereignty Framework (MSF) and Primary Health

Background and Context

Te Aka Whai Ora, Te Whatu Ora and Manatū Hauora have agreed to the establishment of a singular approach to addressing Māori rights and interests in the management of health data and information held across the various repositories in the health sector. It is agreed by the Boards and the Executive Leadership Team (ELT) that Te Aka Whai Ora will lead the development and operationalising of the Māori Sovereignty Framework (MSF) for health data and information on behalf of the three health agencies alongside the Data Tripartite Agreement.

The MSF builds upon the already significant work undertaken by Māori thought-leadership in the crafting of the United Nations Declaration on the Rights of Indigenous Peoples and sets in place two critical concepts that form the basis of the MSF. The first is that indigenous peoples are the exclusive owners of their cultural and intellectual property, and the second is that the first beneficiaries of indigenous knowledge (cultural and intellectual property rights - traditional, evolving and contemporary) must be the direct indigenous descendants of such knowledge[52].

The PHDP programme has acknowledged Māori rights in data as expressed in the MSF and seeks to address Māori interests by making national collections of primary care data more readily accessible to Māori stakeholders – Iwi, Hapū, whānau, Māori health providers, Iwi-Māori Partnership Boards (IMPB's), etc. Furthermore, the PHDP programme in affirming the principles of the Tiriti of Waitangi, recognises the application of these rights

---

[52] Te Pou Tarawaho te Rangatiratanga Maori; Raraunga-oranga, mohiohio, noho matatapu me te whakahaumaru – A Maori Sovereignty Framework: Health Data, information, privacy and security (pre-draft) 2022 Robyn Kamira, Paua Interface Ltd, for Joint Boards.

in the implementation of PHDP will significantly assist the vision set out in the Pae Ora legislation for Māori health equity.

Tiriti of Waitangi based data governance capability and structures

The MSF points strongly to the adoption of a Te Tiriti partnership model of data governance supporting good decision-making and strengthening Māori stakeholder confidence. Key to this is the adoption of rangatiratanga and mana motuhake in the context of health data and information associated with Māori people.

The articles of the Tiriti when applied to the health data, information, privacy and security provisions of the MSF will enable any PHDP data governance body to:

- Protect Māori intellectual and cultural property;
- Collect and make accessible data and information that is relevant to Māori values and models of healthcare;
- Act in the roles of kaitiakitanga, custodianship, stewardship, guardianship in relation to data and how they interact; and
- Set criteria for collection, analysis, reporting and publication of information about Māori.

The creation of an interim establishment data governance group will enable some of the other critical interdependencies to gain some maturity prior to finalising what a Te Tiriti responsive data governance arrangement might look like. Some of these interdependencies include development and acceptance of data sharing agreements with providers that provide for Māori data and information sovereignty, the evolution of the roles of IMPBs and the emerging commissioning models, the maturing of Iwi-based data capability and capacity, the resourcing by government of that capability, and the maturing of internal government agency capability to support data analytics for Māori kaupapa.

Data governance principles and supporting enablers

The key principles that apply to Māori data and information stored or moving through the health system and collections, and where applicable data stored externally by other parties with whom health agencies contract with or have sharing agreements with, may be summarised as:

- No harm to individuals and collectives.
- Ensure first beneficiaries of data and information sourced from, or about Māori, are Māori.
- Give effect to Te Tiriti of Waitangi.
- Embed Tikanga and Matauranga Māori.
- Fulfil kaitiakitanga responsibilities.

- Dismantle siloes.
- Recognise and respect collectives.
- Improve Māori health outcomes.
- Be compliant and consistent[53].

Other enablers and supporting documentation include:

- He Korowai Oranga: The Māori Health Strategy is the overarching framework that guides government and the health and disability sector to achieve the best health outcomes for Māori with the ultimate aim of "Pae Ora" – healthy futures for Māori.
- Whakamaua: Māori Health Action Plan 2020 -2025 guides the whole health and disability system detailing how the government intends to give effect to He Korowai Oranga. Whakamaua is underpinned by the Ministry's new Te Tiriti o Waitangi Framework which provides a tool for the health and disability system to fulfil its stewardship obligations and the special relationship between the Crown and Māori.

Assurance: Protection of the privacy of individuals and providers

The MSF provides a specific focus on the Privacy Act 2020 and the Health Information Privacy Code 2020 (HIPC) which sets out the controls for the management of data and information about individuals. Importantly, privacy policy standards and compliance to the HIPC and legislation amongst provider groups must be a criteria for acceptance of sharing of data. Where no specific statutory provision applies the MSF helps determine the decisions for data and information access and release.

---

[53] Ibid Page 20 :for further definition of principles see Māori Sovereignty Framework

6 Data management

Each organisation in the health system will benefit from access to more complete information about what they and other organisations do, and this likely requires access to data not available to them now. The Pae Ora (Healthy Futures) Act 2022 strategies legislation sets down the way things will work in the health system, the health system priorities, and the partnership expectations and outcomes that are expected to be delivered.

A vital component of progressing these aims is a comprehensive understanding of the populations served by the health sector. The health 'system' needs to be seen as one multi-faceted consumer focussed service. To achieve this the consumer's health interactions with the system and where those interactions take place need to be known. It is important to understand the patient journey and support continuity of care to help manage demand. It is also important to understand risk factors and behaviours of consumers to help target preventive measures. Much of the data required to meet these needs sits within a primary care context.

The Primary Care Data Working Group worked on defining a National Primary Care Data Set to support these needs. It looked at two main sources, firstly, from general practice and/or PHO data sets, and secondly, from national health data sets which contain a mixture of data sourced from primary and from secondary care settings.

6.1 Data user(s)

Users and their use of the National Primary Care Data Set can be categorised into four major user types. These user types are Health Care Providers; Funders/ Central Agencies; Iwi Māori Partnership Boards (IMPBs) and Localities; Individual, Whānau, or community stakeholder groups. Data users and definitions use by the programme for this document are as follows:

- **Health care providers** or their support organisations (PHOs for example). These are the health care professionals and/or organisations that deliver health services.
- **Funders / Central Agency** – Government organisations which fund health services, principally Manatū Hauora, Te Whatu Ora, Te Aka Whai Ora.
- **Iwi-Māori partnership boards (IMPBs), Localities** – Iwi-Māori Partnership Boards and Localities as defined in the Pae Ora (Healthy Futures) Act 2022.
- **Individual, whānau, and communities** – Individual health consumers, their family group(s) and the communities they identify with.

6.2 Application for the use of the National Primary Care Data Set

Currently identifiable data within the National Health Data Sets is carefully managed within overall governance procedures established within Te Whatu Ora.

There are two key constructs that were explored within the working groups. The tension between the use of identifiable data for analytical purposes and its provision for clinical care (as well as standard section 22F Health Act provisions). The purpose statement for the National Primary Care Data Set centres on anonymised outputs but also allows for the possibility of identifiable data being used, subject to full governance processes to enable this prior to the release of identifiable data.

The National Primary Care Data Set will require strengthening of the provisions to protect people's information and also to assure source providers of this information (the person's chosen clinician) that their rights and obligations to their "patients" and their practice are meeting their legal and social licence obligations.

A proposed high-level process diagram for a data application and governance process for de-identified or anonymous data can be found in Appendix A.10 Draft high-level application process for the use of the National Primary Care Data . High-level analysis[54] of this process provides some future implementation considerations in line with the early thinking around the use of primary health data (refer to Appendix A.2 Draft early thinking around the secondary use of primary health data). Table 7 outlines national primary care data governance and data management considerations for the future.

Table 7 – National Primary Care Data Set data governance and data management future considerations

Prerequisites	Future Considerations
1.0 Governance model for secondary use of primary health data	• Health system data governance and structures in partnership with the sector. • Data sharing business rules are documented, and a decision-making model exists - documented as part of the data governance framework.

---

[54] based on the Australian Institute of Health and Welfare model https://mohgovtnz.sharepoint.com/:b:/r/sites/PrimaryHealthDatasetProgramme-INTTeam/Shared%20Documents/General/International%20Examples/Australia/framework-to-guide-the-secondary-use-of-my-health-record-system-data.pdf?csf=1&web=1&e=8W1Qwg , however customised for the New Zealand context with the expectation of being adjusted and mapped in further detail during implementation phase as applicable e.g. to incorporate the role of IMPBs.

	• Nationally consistent privacy and consenting framework aligned with the Privacy Act and Pae Ora (Healthy Futures) Act 2022 strategies legislation. • Exploration of data governance accreditation.
2.0 Consumer control of primary health data (incl. patient preferences) in the appointed data platform	• Transparent and more prescriptive guidelines to be defined by the data governance framework, including any opt out or special treatment of sensitive data, data with confidentiality flags as required.
3.0 Applying to access primary health data for secondary use	• Updated application forms and a data application process in line with the data governance framework guidelines and processes, and with a Pae Ora (Healthy Futures) Act 2022. • Any exceptions to data access for secondary use are clear and documented.
4.0 Access to, or release of, primary health data for secondary use	• Shared within a nationally recognised platform, except under national data governance structures agreed conditions, where outputs can be shared in pre-determined file format(s) – ensuring security and privacy settings are consistently applied. • Increased national health data sets available, some of which can be joined with the National Primary Care Data Set as determined by the national data governance structures. • Data sharing will be aligned with Māori Sovereignty or other requirements relative to priority population requirements, as determined by the national data governance group. • National health data sets consideration or development of guidelines for its use for commercial purposes.
5.0 Process for requesting and accessing primary health data	• Clear condition(s) for when ethics approval is required in line with appropriate National ethics advisory structures and data governance framework guidelines. • Alignment to purpose and use of PHDP data. • Enrolment, data sharing and a 'conditions of use agreement' or similar is required for data outputs shared that includes encrypted identifiable data.

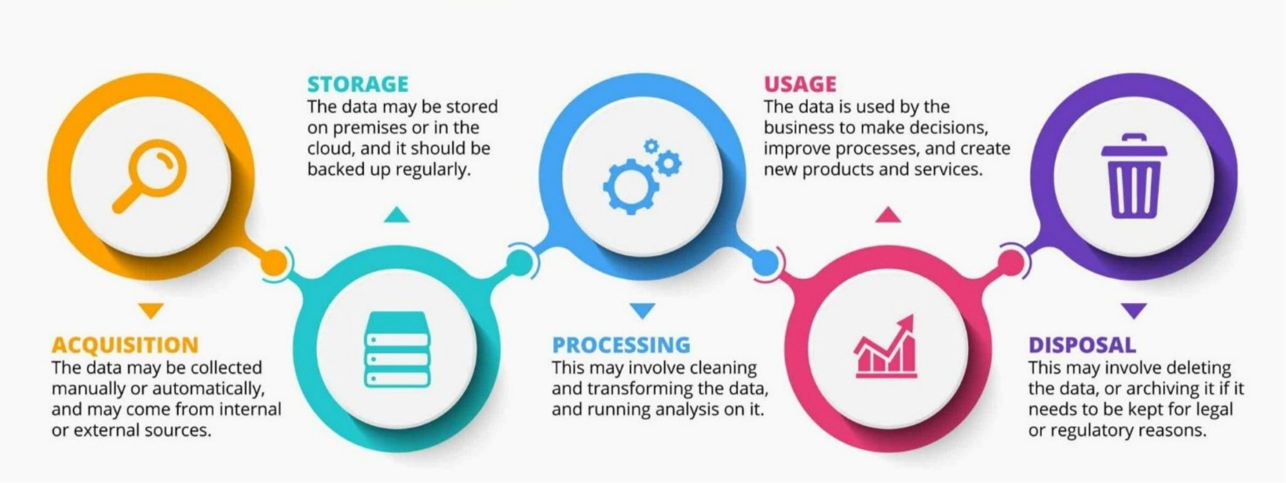
	• Data requestor held to account on a code of conduct and on 'conditions of use agreement' or similar. • Establishment or development of a data quality improvement program or plan in line with agreed data governance strategy and delivered across data governance structures.
6.0 Data linkage	• Data that can be joined up for National Primary Care Data Set, and otherwise are clear and documented (and specific about the limited circumstances in which identifiable data is to be made available). • Data linkage permissions aligned with national data governance group defined purpose-use matrix and national health data sets that will be available. Exceptions assessed, documented and re-applied as applicable. • Data linkage rules will be aligned with Māori Sovereignty or other requirements apply relative to priority population requirements, as determined by the national data governance body.
7.0 Processes to ensure protection of the privacy of individuals	• Proven methods that reduce the risk of privacy breaches include built-in capabilities within the data platform e.g. ciphered data, row, column and cell level access controls, k-anonymity, oversight of a Privacy Officer, and data masking. • Establishment of a specific privacy officer for data set to ensure any potential breaches or concerns are resolved. • Monitoring and evaluation processes are conducted regularly to ensure privacy protection processes around secondary use are fit-for-purpose as well as in the review of the purpose-use matrix. • Data governance structures with sector involvement stays abreast with changes in technology and data science; anticipating new privacy threats to inform assessment of applications for the use of data. • Data management standards and processes in line with HISO Information Governance Guidelines to preserve patients' privacy. Analytical pipelines disclose process, code and technical methods being used.
8.0 Preparing and making data	• Updated data quality framework in line with the data governance framework. Data quality assurance enforced

available, and data quality	in line with data governance framework across data governance structures. • A data platform that ensures privacy is protected by design. Data outputs stay within the trusted environment and are otherwise bound by a ‘conditions of use agreement’ or similar. • Data pipelines within a national data platform expected to exist for reuse. • Various privacy enhancing mechanism options exist for use within the national data platform.
9.0 Monitoring and assurance processes	• Reporting and communication as well as review and monitoring of data governance and data management outputs or processes to be conducted regularly, in line with the data governance framework. • The platform that contains feature(s) that support assurance and monitoring of data pipelines and data.
10.0 Risk mitigation strategies and imposed penalties	• Risk management monitoring and guidance with thresholds defined and agreed by the national data governance group. Key risks and objectives are known and a risk appetite guidance for projects and data and information exists.

6.3 Functional technical platform requirements

A data lifecycle approach (refer to Figure 4) has been used to help define the boundaries and technical requirements to support the collection and use of the National Primary Care Data Set.

Figure 4 – Data lifecycle diagram



Across several meetings, the Technical Platform Requirements Working Group worked on defining the features and capabilities needed to support primary care data from its acquisition, storage, processing and usage to its disposal.

Technical platform capability that has been identified to be required is that of a data platform to support online analytical processing (OLAP). This typically involves having a centrally controlled repository for storing and managing large amounts of data from various sources for analysis and reporting.

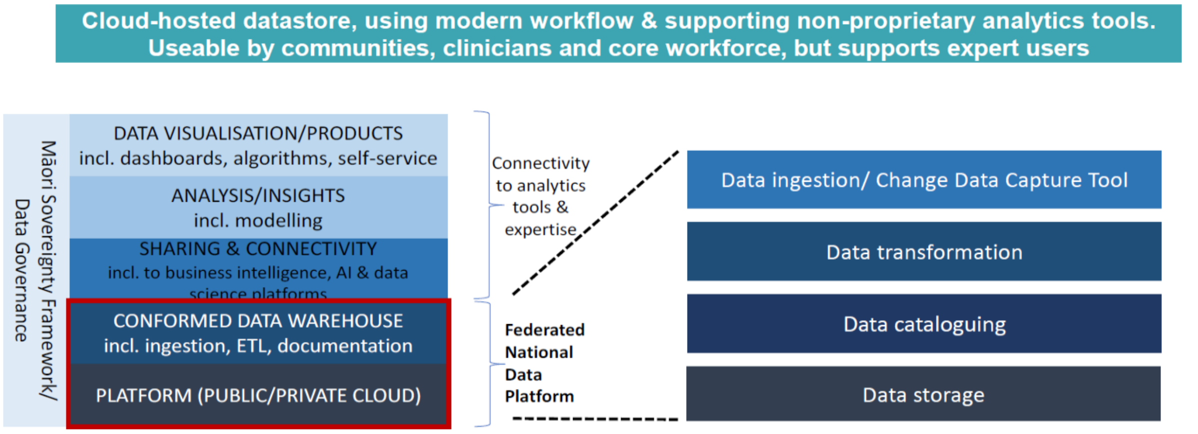
It is optimised for fast querying and analysis, enabling organisations to make informed decisions by providing a single source of truth for data. Data warehousing typically involves transforming and integrating data from multiple sources into a unified, organised, and consistent format[55].

It is useful to refer to a functional model showing how each component can work together - the focus of the Technical Platform Working Group was on the data warehouse and storage platform component as illustrated below. The National Data Platform (NDP) programme is establishing a nationally consistent approach to the many data platforms that already exist in Te Whatu Ora and the core functions offered in the NPD programme has relevance to the PHDP technical platform requirements. Figure 5 - National Data Platform - functional model.

---

[55] GeeksforGeeks.com. Characteristics and Functions of a Data warehouse. https://www.tutorialspoint.com/characteristics-and-functions-of-data-warehouse accessed Feb 2023

Figure 5 – National Data Platform – functional model



*Supplementary slides include a draft list of specific functions & capabilities within each function

Informed by the NPD process, the requirements for a proposed platform to support a National Primary Care Data Set have been established based on the following assumptions:

1. That data will be ingested daily from around 50 different sources.
2. That the data from each source will be provided within standardised schema in terms of format and content.
3. That the data from these sources will create a National Primary Care Data Set which includes relevant data held within national health data sets.
4. That usage will be restricted to authorised data analysts (up to 200 users, 30-50 users concurrently) and that permissions are built into the system.
5. That the disposal of data is governed by the General Disposal Authority for health, that is, data is retained for 10 years minimum.
6. That export of data out of the platform will be limited for specific purposes, for example, matching with external locally held data sources.

6.3.1 Data acquisition and consolidation

The first part of the data lifecycle is to acquire data from the external sources, cleanse it and then repackage it to make it analytics ready. Recommended requirements are as follows:

1. Data from multiple sources must be ingested on to the platform.
2. Initial checks must be done to identify data integrity issues (such as duplicate records), weight issues according to their severity and reject batches if they will compromise the overall integrity of the warehouse.

3. Any data transformations/translations/mappings must be done to convert data into a single unified and consistent data set with the correct indexes and key fields to enable efficient links to other data sets on the same platform.
4. These processes need to be repeatable, and errors need to be identified and followed up with such processing automated as much as possible.
5. Incorporate translation tables where required.
6. Only update the changes (the delta) to the data in the data warehouse from each data load.
7. Optimise design to support fast querying and analysis.
8. Provide user friendly interface to easily manage and configure data acquisition, consolidation and reporting.

6.3.2 Data storage, disposal and analytics processing

The second, third and fifth parts of the data lifecycle are to store and retain data in a time series and non-volatile way to ensure results of analysis are repeatable and consistent over time. Recommended requirements are:

1. Data must not be altered, updated or deleted when new data is inserted (non-volatile) so that the history can be maintained (a version history of 60-90 days is desirable).
2. Allow permanent snapshots of data to be retained for longer periods, for example, monthly snapshots for up to 3 years.
3. Allow data to be tagged and processed to generate derived data.
4. Allow use of Python, R, Structured Query Language (SQL) and other programming languages in common use.
5. Allow integration with third party data providers (incoming and outgoing).
6. Allow the storage of the results of data analysis, that is, tables or graphs to allow for re-use/recall by the creator of the analysis.
7. Allow data to be summarised and aggregated for faster querying and analysis.
8. Provide built-in tools to help manage bias and ensure compatibility or inclusion of privacy enhancing technologies such as k-anonymity (validates outputs as not being able to be re-identified).
9. Capture, maintain and display data lineage including any transformation or changes to each data element over time.
10. Store and maintain library of repeatable analytics procedures/pipelines for re-use by other analysts.
11. Allow for independent Artificial Intelligence (AI) plug-ins in the future e.g. to support natural language processing of free text fields.

12. Automatic deletion of audit logs, loads, or duplicated data after a set amount of time.
13. Supports workflow/job management and allows one procedure to trigger another.
14. Allow use of third-party visualisation/analytic/reporting tools to data on the platform.

The disposal of health information from the platform will be governed by the General Disposal Authority for Health under the Public Records Act 2005. Minimum disposal is typically set out to be 10 years from the date of ingestion.

6.3.3 Data access and usage

The fourth part of the data lifecycle is to provide appropriate access controls and usage rights to authorised users. Recommended requirements for this part of the lifecycle are:

1. Data governance rules should be enforced at the technical level where possible (including data sovereignty rules as they emerge from further consultation with Māori stakeholders).
2. Allow encryption of identifier fields e.g. NHI, Health Provider Index (HPI) Facility Code, and masking of other direct or indirect identifiers.
3. Provide group access rights based on users being assigned to defined roles.
4. Provide cell, object, row, column and table level security/masking.
5. Provide affordable and flexible user licensing/subscription charges for common usages.
6. Ability to segment access rights based on enrolled population or other forms of population health data sharing agreements.
7. Allow export of analytical results or raw data selection (with appropriate permissions and limitations) in industry standard formats.

6.4 Non-functional requirements

Other non-functional requirements which apply across all areas of functionality are:

1. Dynamic scalability as load comes on or off the platform to ensure user response times are reasonable.
2. Platform must be highly reliable for analytics (12 hours per day) and outages limited to weekly/monthly outages at pre-arranged times.
3. Meets optimal security standards for maintaining in-confidence and sensitive level data.
4. All access and activity is audited and logs are made available for review and analysis.

5. The platform must be cost-effective for a range of classes of end-user from a light touch to dedicated end-users.
6. User friendly interface to manage the on-going configuration and administration of the platform.
7. Provision of versioning and testing environments to support fixes, changes and release cycles.
8. Data security and privacy enhancing features must be built into the native capability of the solution.
9. Preference for federated data warehouse configuration which is in the Cloud and operates independently of the supporting database management system.
10. Data definitions are understandable, aligned with current standards and are applied consistently throughout the platform.

6.5 National Primary Care Data Set data element selection and methodology

In April/May 2022, discussions were held with senior data analysts in National Collections and Reporting[56] and in Health and Disability Intelligence to help identify the most important primary and community care data to complement data held at a national level. From these discussions, and with input from PHOs, a selection of 9 data classes were included in the Current State Analysis - Data Availability Survey[57].

This was reduced to 7 data classes after the survey found that two data classes (Vaping and Care Plans) were embryonic in their collection. This revised initial scope ensured that the workload for the working group was manageable and focused.

The final set was agreed at the PHDP Governance Group meeting in September 2022. The 7 data classes, containing a total of 30 discrete data elements, are (refer to Appendix B.1 National Primary Care Data Set workbook for a detailed view):

1. Problems/Conditions
2. Medications
3. Immunisations
4. Smoking status

---

[56] The department previously within the Ministry of Health but now within Te Whatu Ora that maintains and manages data and access to 37 national health data sets.

[57] Data Availability Survey https://www.tewhatuora.govt.nz/publications/primary-health-dataset-programme-current-state-analysis-report/

5. Measurements and vital signs
6. Diagnostic results – laboratory
7. Encounters.

The initial data definitions and framework was taken from the draft NZ International Patient Summary (NZIPS) standard prepared by HISO in April 2022. The purpose of this standard is to enable interoperability and transfer of data between systems at a summary level and this therefore aligns well with the National Primary Care Data Set. The underlying coding standard is SNOMED CT which was adopted by the NZ health sector in 2010. The use of this framework was strongly affirmed by the Primary Care Data Working Group during the co-design process, even as SNOMED CT still requires further adoption in the health system.

Results from co-design with the Primary Care Data Working Group

Based on the Primary Care Data Working Group terms of reference (refer to Appendix A.2 Terms of Reference), the criteria for selection of a data item included:

- Alignment to the Pae Ora (Healthy Futures) Act 2022 strategies legislative intentions, its importance for analytics use, and its ease of availability via PHO data sets.
- Additional criteria were added by the group as part of the co-design process:
    - scope and completeness of the data,
    - degree of transformation required, and
    - conformance to accepted data standards.

During the co-design process, the metadata elements for each data class and data element were discussed and modelled. After an initial comprehensive list of metadata items (refer to Appendix B.3 National Primary Care Data Set data dictionary (full template)), the metadata was reduced to a manageable quantum in the time allowed. The following table is the revised metadata schema with minimum attributes which was agreed (refer to Appendix B.2 National Primary Care Data Set data dictionary (minimum attributes) for more details).

Table 8 – National Primary Care Data Set proposed data dictionary (minimum attributes)

Attribute	Description
Data Class	
Name	Name of data class or domain of primary care data of interest.

Description	Description of the data class, including * references to related data and documents, * table attribute(s).
Frequency	Desirable frequency of update from the curated data source
Depth of history	Desired period (in years) of event data.
Data source(s)	Possible data sources. e.g. dispensing data has several data sources.
Agreed data source(s)	Agreed data source(s) noting future preferred options.
Data quality caveats	Known data quality limitations or issues from each source. i.e. Completeness, Uniqueness, Timeliness, Validity, Accuracy, Consistency.
Data sharing limitations on availability and use rating	From a scale of 1 to 5.
Data Element	
Mandatory to National Primary Care Data Set	Specifies whether the element is an agreed minimum requirement.
Data domain	Acceptable values and ranges that are acceptable for the data element. Each coded data element has a specified code set.
Data quality caveats	Known data quality limitations or issues from each source i.e. Completeness, Uniqueness, Timeliness, Validity, Accuracy, Consistency, as well as key dates to consider when treating or interpreting the data, and any known data lag(s) where applicable.

Data Set Schema Analysis

The data classes and elements selected for the National Primary Care Data Set have a primary use by General Practice to deliver patient care and are used in real-time. The focus of the Primary Care Data Working Group was on the secondary use of data which does not require real-time access. The secondary uses of data were split into two categories (1) analytical uses of data where data can be held and processed in a de-identified way and no individual’s data is disclosed, and (2) operational uses of data (these must be covered within the purpose statement and supported by a governance agreement and privacy frameworks).

Each of the initial 7 data classes and 30 data elements were reviewed by the Primary Care Data Working Group. In each case, the following aspects were discussed (refer to Appendix B.2 National Primary Care Data Set data dictionary (minimum attributes):

- The level of transformation required[58] (from local source to proposed nationally consistent format).
- The data quality assumptions and caveats – completeness, uniqueness, consistency, timeliness and validity.
- The relative value of each data class and data element in relation to its analytical purpose and its operational purpose.
- The sizing and effort required to source the data.

The Primary Care Data Working Group agreement was that, for secondary use, daily update and access is optimal – so the data could be fixed at a point-in-time each 24-hour period at midnight. This frequency allows data to be used for operational and surveillance purposes. From an analytical usage perspective, however, The Primary Care Data Working Group deemed weekly or monthly updates as acceptable.

The table below is a summary of the proposed set of National Primary Care Data Set classes and data source findings. The detailed table with the full list of data elements for each data class and additional commentary can be found in Appendix B.1 National Primary Care Data Set workbook.

Table 9 – National Primary Care Data Set data source findings

Data Class	History required	Data source(s)
1. Problems/Conditions	All Active	Consolidated sources available
2. Medications	Current Medications	National source available
3. Immunisations	All Immunisations	Mixed – consolidated and national sources available
4. Smoking status	Last recorded status	Consolidated sources available
5. Measurements and vital signs	Last recorded measurements	Consolidated sources available

---

[58] Informed by the Current State Analysis

6. Diagnostic results – laboratory	3-year history	Consolidated sources available
7. Encounters	3-year history	National source available for some
8. Vaping	These data classes (10-14) were not considered in more detail when the Current State Analysis survey was completed. Data dictionaries for these have yet to be completed.	Data classes 8-12 are only partially collected, and data elements remain unspecified.
9. Care Plan		
10. Referrals		
11. Procedures		
12. Pregnancy		
13. Functional status		
14. Patient Demographics		

See also Appendix B.1 for the detailed overview of the National Primary Care Data Set.

6.5.1 Problems/conditions

The problem code is a SNOMED CT code representing the problem or condition – medically speaking this is equivalent to a diagnosis. This is one of the foundational data elements collected in a primary care setting as it strongly indicates the health status or morbidity of the individual. This is closely associated with clinical findings which represent the signs or symptoms experienced with a given problem. Also useful is the severity of the problem and whether it is likely to be a long-term problem.

This data class is recognised by the Primary Care Data Working Group as one with the highest priority for analytical purposes and is currently available through consolidated and curated sources but is not available nationally.

The smoking status, vaping status and functional status (a person's functional status if affected by a disability) has been agreed by the Primary Care Data Working Group to also be recorded here[59].

6.5.2 Medications

This data class contains the active list of prescriptions and medications dispensed to the patient in a community setting. The Primary Care Data Working Group sees this as a high

---

[59] Although under FHIR conventions it is considered to be an "observation" and therefore may be better placed in the Measurements data class

priority for analytical purposes as it describes the therapy or treatment a patient is currently receiving – and therefore it is also indicative of health status.

It is also a priority for operational purposes as PHO services would like to know patients who have not picked up their prescriptions within accepted timeframes – as this is indicative of lack of treatment risk. Any such subject use cases would be progressed through privacy consideration process / PIA etc before being finalised, particularly if they involve identifiable information.

When the programme first started in February 2022, the medication prescriptions and dispensed data were not available nationally, but since November 2022, there has been a daily update of prescribed and dispensed medications from the NZ electronic Prescribing Service (NZePS) to the Medications Data Repository (MDR) making this data available for analytical purposes. 94% of general practice prescriptions are done through NZePS and 100% of prescriptions and dispensing are recorded by community pharmacies at the time of dispensing[60].

6.5.3 Immunisations

This data class aims to contain the active list of vaccinations given to the patient. This has been deemed by the Primary Care Data Working Group as a medium-high priority from an analytical perspective but is especially useful for identifying coverage gaps in populations for public health surveillance. This is a form of preventative care and is important for equity considerations as well.

There is currently no single (digital) source of truth for a patient's vaccinations – the vaccination record is spread across (and duplicated within) both the National Immunisation Register (NIR) and the PMS vaccination record. It is recognised that the historical position enabled people to opt out of their information being shared with the NIR. This would require PIA process to look at each of these issues and worked through how to mitigate the risks. PHO services have reported the need to combine data from both these sources to achieve close to a single source of truth for their enrolled population and so a regular feed from the NIR is a high priority from an operational perspective.

In the medium to long term (2023 – 2024) the replacement for NIR, the Aotearoa Immunisation Register (AIR), aims to resolve many of the current issues the primary sector faces around the data quality of the immunisation record. Also, as patients expect their immunisation record to be available on-line through their health portal, the accuracy of the record will improve.

---

[60] https://www.tewhatuora.govt.nz/our-health-system/digital-health/emedicines-and-the-new-zealand-e-prescription-service/eprescriptions/nzeps-uptake/

6.5.4 Smoking status

This is one of the system-level measures and PHO performance indicators. Smoking is considered one of the top 5 modifiable public health risk factors. This is high priority and has been included in the Problem/Condition list.

6.5.5 Measurements and vital signs

High BMI and high blood pressure are also considered in the top modifiable public health risk factors. There are many measures in this category but only a few which are regularly recorded in a primary care setting.

The measures which are taken and are of value as identified by the Primary Care Data Working Group are -- height and weight (combined to create body mass index (BMI)), and blood pressure. Waist circumference has also been raised as an important health risk indicator but is not always measured. These selected measures have been identified as medium-high priority for analytical purposes and are available through consolidated sources.

6.5.6 Diagnostic Results

Diagnostic test results (primarily from laboratories) are objective indicators of a person's health condition and often link to the drug therapy being provided. This data class has been identified as a high priority for analytical purposes and for assessing clinical risk and follow-up by the Primary Care Data Working Group.

The NZ standard for diagnostic test results is New Zealand Pathology Observation Code Set (NZPOCS) which uses Logical Observation Identifiers Names and Codes (LOINC) that are currently in common use. The Primary Care Data Working Group has reported PMS systems storing lab results in dissimilar ways and that significant curation is required when consolidating the data. The working group's preference was to source lab results directly from the laboratories themselves and have wider adoption of the NZPOCS/SNOMED CT standard for reporting results. Note that access to laboratory data directly, would require full analysis of the privacy impacts.

6.5.7 Encounters

The Encounter data class has been deemed a high analytical priority by the Primary Care Data Working Group. This class includes a record of each interaction with the patient in a primary care setting – this may be an in-person visit, a telehealth session, a phone call or an email interaction. The type of encounter may be a doctor's consult, a nursing intervention, a vaccination, counselling or physiotherapist appointment, or any other type of clinical or community health service. These activities are important in assessing the overall workload in primary care and therefore relate strongly to capacity and workforce planning in a whole of sector context.

The National Enrolment Service provides a proxy for encounters currently, as each “qualifying encounter date” (QED), which includes in-person visits, is recorded in this national system in near real time.

However, there is no currently agreed national standard for encounter type and mode of encounter, and it is not clear whether PMSs will handle such a standard through configuration or whether a new workflow involving changes to the user interface will be required. The collection of diagnosis at the time of the encounter will also require significant analysis, user engagement and change.

Encounters are also important to obtain visibility on activity within community-based practice systems where patients interact, and services are delivered. In the future, encounters and encounter types recorded in non-General Practice settings will be important for capacity planning and funding models.

6.5.8 Other data classes

Other data classes considered for inclusion in the National Primary Care Data Set that have been added during the course of the Primary Care Data Working Group meetings are referrals, procedures, pregnancy and care plans. These data classes have been identified as more difficult to collect partly because of lack of standards and in the case of care plans, they have not been reported to be widely used or available through PMSs. Note that vaping and functional status are covered off in the Problem/Conditions list.

The programme recognises Referrals (including those to radiology services) as the most important priority from an analytical perspective after Problems/Conditions because they enable more complete tracking of the patient journey through the health system, whereas procedures and pregnancy data will also enable a view of the additional work going on in primary care.

6.6 Other National Health Data Sets

The National Health Data Sets consists of a set of 37 data domains with data collected from both primary and secondary care sources. The programme’s co-design approach with the sector enabled findings and PHO data preferences from PHDP’s Current State Analysis report[61], discussions from working group workshops and the current usage by PHOs to be brought together. Results of this demonstrated a good degree of consistency, allowing clarity in priorities to have emerged. It is noted that data related to cervical

---

[61] https://www.tewhatuora.govt.nz/publications/primary-health-dataset-programme-current-state-analysis-report/

screening is subject to specific legislation and any sharing of data will need to meet restrictions, such as compliance with one of the categories of s112J of the Health Act.

Table 10 – Important National Health Data Sets

National Health Data Set	Current State Survey preference	Currently provided to PHOs*	Workshop preference
National Immunisation Register (NIR)	Y	Y	Y
Cervical Screening Programme	Y		Y
National Minimum Data Set (NMDS)	Y	Y	Y
Breast Screening	Y		Y
PHO Enrolment	Y	Y	Y
National Non-Admitted Patient Collection (NNPAC)	Y	Y	
Laboratory Claims	Y	Y	
NZ Cancer Registry	Y		
General Medical Subsidy Collection	Y		
New-born Hearing Screening	Y		Y
Mortality	Y	Y	Y
Pharmacy Claims		Y	Y
Well Child/B4SC			Y
Maternity			Y

*Selected PHOs

The key issues PHDP need to address for implementation of access to relevant National Health Data Sets are:

- To establish common rights of identifiable access for preferred National Health Data Sets applicable to all PHOs.

- To establish which preferred National Health Data Sets PHOs do not have identifiable access to and why. Define what processes need to be followed to gain access in the future.
- To establish which National Health Data Sets should be included in the National Primary Care Data Set.

6.7 Data management summary

Table 9 - National Primary Care Data Set data source findings in section 6.5 National Primary Care Data Set data element selection and methodology summarises the data source findings, however, details surrounding the relative priorities of each data class and the relative effort required to source from data not held nationally based on the Primary Care Data Working Group discussions can be found in Appendix B.1 National Primary Care Data Set workbook.

7 PHDP costing model

The PHDP Commercial Requirements Working Group was formed to develop guidance on the scale and scope of investment required to partner with the sector on the development of a National Primary Care Data Set and estimating the Te Whatu Ora and primary and community care sector impacts in terms of both the initial set-up and the on-going costs required to meet the objectives of PHDP.

The PHDP Commercial Requirements Working Group has a responsibility to:

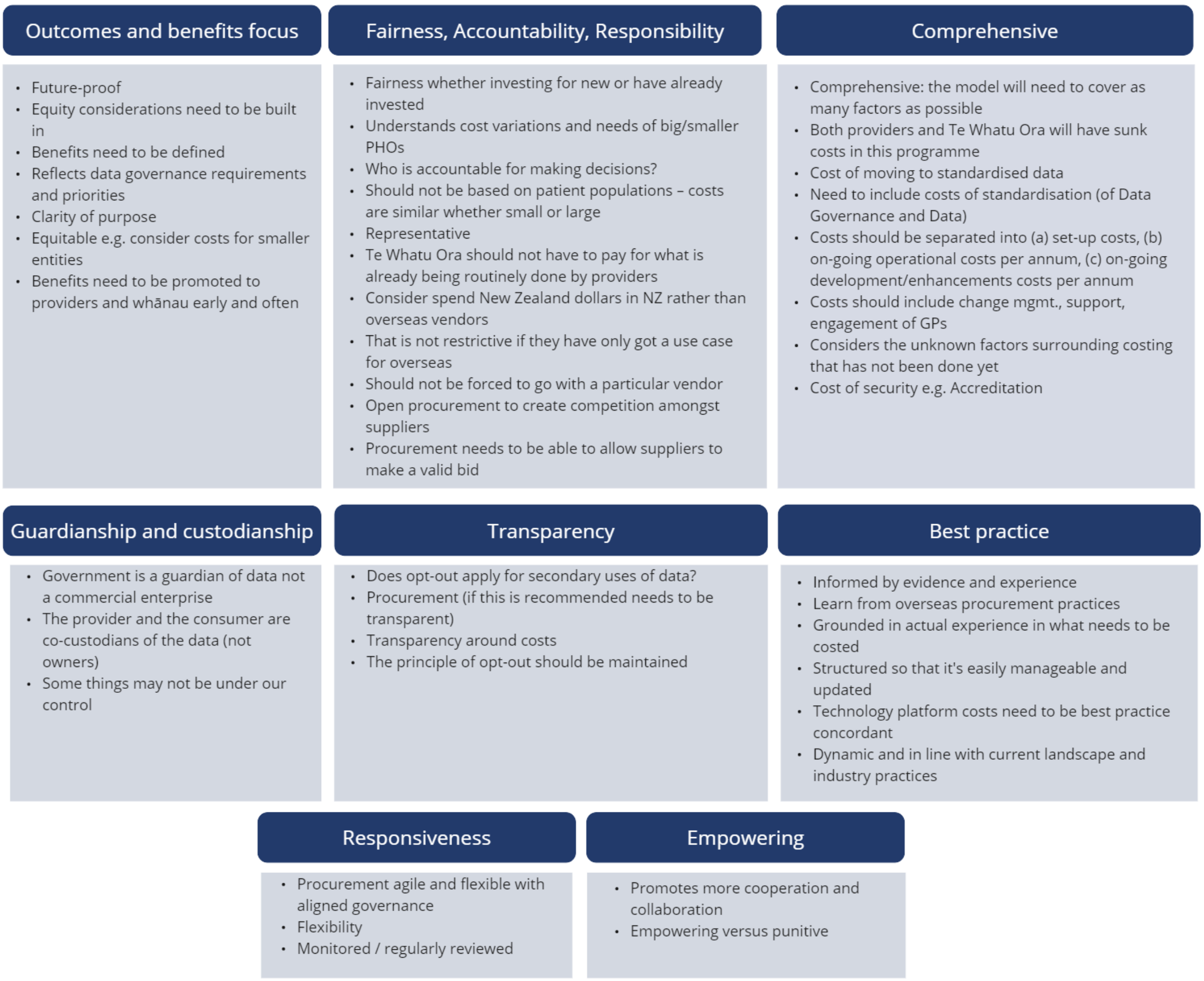
- give effect to the Te Tiriti articles and principles
- work to our fullest extent to protect and achieve equitable health outcomes for Māori, as pro-equity champions
- ensure whānau remain at the centre of health planning, policy and delivery
- honour the unique and emerging relationship between Te Whatu Ora and Te Aka Whai Ora and will operate in accordance with their shared values and principles.

The approach was developed transparently and objectively in partnership with the sector using a dynamic modelling and options process, in effect a broad costing model. Potential costs include those applicable to the general practice and PHO sector, Te Whatu Ora and Te Aka Whai Ora for the necessary investment in data governance, data management, curation, technical platform, and outputs in line with the objectives of the PHDP. They are dependent on the technical platform(s) deployed, the number of data elements and size as defined in the primary care data schema, the number of national health data sets used and other factors as they emerge.

7.1 Principles

The following costing model design principles were identified:

Table 11 – PHDP costing model design principles



Outcomes and benefits focus

- Future-proof
- Equity considerations need to be built in
- Benefits need to be defined
- Reflects data governance requirements and priorities
- Clarity of purpose
- Equitable e.g. consider costs for smaller entities
- Benefits need to be promoted to providers and whānau early and often

Fairness, Accountability, Responsibility

- Fairness whether investing for new or have already invested
- Understands cost variations and needs of big/smaller PHOs
- Who is accountable for making decisions?
- Should not be based on patient populations – costs are similar whether small or large
- Representative
- Te Whatu Ora should not have to pay for what is already being routinely done by providers
- Consider spend New Zealand dollars in NZ rather than overseas vendors
- That is not restrictive if they have only got a use case for overseas
- Should not be forced to go with a particular vendor
- Open procurement to create competition amongst suppliers
- Procurement needs to be able to allow suppliers to make a valid bid

Comprehensive

- Comprehensive: the model will need to cover as many factors as possible
- Both providers and Te Whatu Ora will have sunk costs in this programme
- Cost of moving to standardised data
- Need to include costs of standardisation (of Data Governance and Data)
- Costs should be separated into (a) set-up costs, (b) on-going operational costs per annum, (c) on-going development/enhancements costs per annum
- Costs should include change mgmt., support, engagement of GPs
- Considers the unknown factors surrounding costing that has not been done yet
- Cost of security e.g. Accreditation

Guardianship and custodianship

- Government is a guardian of data not a commercial enterprise
- The provider and the consumer are co-custodians of the data (not owners)
- Some things may not be under our control

Transparency

- Does opt-out apply for secondary uses of data?
- Procurement (if this is recommended needs to be transparent)
- Transparency around costs
- The principle of opt-out should be maintained

Best practice

- Informed by evidence and experience
- Learn from overseas procurement practices
- Grounded in actual experience in what needs to be costed
- Structured so that it's easily manageable and updated
- Technology platform costs need to be best practice concordant
- Dynamic and in line with current landscape and industry practices

Responsiveness

- Procurement agile and flexible with aligned governance
- Flexibility
- Monitored / regularly reviewed

Empowering

- Promotes more cooperation and collaboration
- Empowering versus punitive

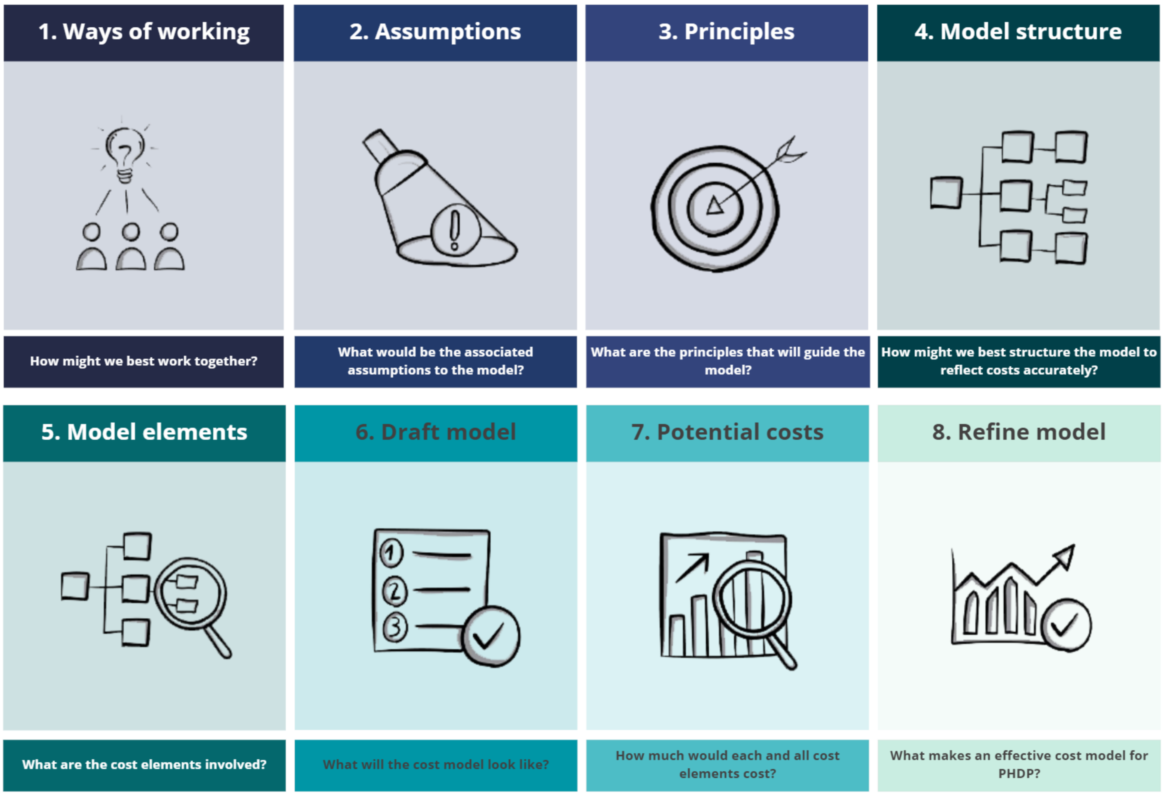
7.2 Methodology

Previous attempts to work with the sector in 2016 and 2019 on the establishment of the National Primary Care Data Set did not scale or scope the funding required to achieve this aim. It is not enough for all parties to agree with the direction of PHDP - there are obvious benefits to all parties, and it was critical to estimate the material impact of implementation for all parties.

The intention of this work is to provide broad guidance to Te Whatu Ora and the sector on costs of establishment of a National Primary Care Data Set, and to establish boundaries to any implementation. It provides a “reality check” in the funding required to do so.

Procurement processes were also canvased to define minimum standards of procurement in line with government procurement requirements. The data lifecycle (Figure 4) also underpinned the structure of the costing model as did the findings from the Current State Analysis of PHOs access, curation and transformation of data that in many cases is already occurring.

Figure 6 – PHDP costing methodology overview



7.3 Factors explored

The following was agreed on how a costing model would be developed.

1. A dynamic costing model would be produced that can accommodate differing options.
2. A technical platform would be costed – including access to national data sets (recognising that the NDP is progressing a similar programme of work).
3. Consideration of indirect costs for general practice – data governance improvements would be estimated.
4. Exploration of the concepts of Intellectual Property (IP) rights and other commercial costs as articulated by the sector.
5. Consideration of data curation costs estimated from the current state analysis of PHOs as well as GPs, other providers.

6. Coding standards implementation would be explored.
7. “Sunk” costs based on the work undertaken by the sector over many years accommodated.
8. Analytical costs.
9. Direct costs for PHOs.
10. Impact for PMS vendors.
11. Data Governance requirements and provisions.
12. Te Whatu Ora costs.

7.4 Agreed elements

Through the working group process, the following costing elements were agreed (Table 12).

Table 12 – PHDP costing model elements

Costing elements
Acquisition costing elements
A❶ Initial data extraction costs
A❷ Ongoing data extraction costs
A❸ Considerations of economy of scales for smaller entities
A❹ Data manipulation and transformation costs
A❺ Charges and software licence costs of extracting data out of PMS/PHO/MSO systems
A❻ Cost of developing standards for data not adhered to SNOMED / FHIR / IPS standards
A❼ Costs of aligning to data standards e.g. HISO and compliance costs
A❽ Intellectual property rights of PMS vendors/PHOs/MSOs
A❾ System change and test management costs of obtaining data
A❿ Changes to data consent process and enrolment forms
A⓫ Data support costs for providers (GP/PMS vendors/PHOs/MSOs)

A⓬ Meeting data governance / data sovereignty (cloud data outside NZ and Māori data sovereignty) requirements
A⓭ Manage commercial risks and liabilities
A⓮ Sector training costs relating to data extraction
Storage costing elements
SC❶ Technical platform and data warehouse design costs
SC❷ Data storage and data warehouse costs
SC❸ Cost of security and accreditation including regular testing
SC❹ Data backup costs
SC❺ Costs of technology upgrades
SC❻ Costs of privacy enhancing technology feature requirements
Processing costing elements
PC❶ ETL and data transformation costs within the technical platform
PC❷ Cost of de-identifying data
PC❸ RAM and processing capability for analytical use
Usage costing elements
UC❶ Cost of software licences of the technical platform
UC❷ Reporting functionalities
UC❸ Any reporting enhancement - current state vs desired future state
UC❹ Data visualisation component / tool
UC❺ Analytical functionalities
UC❻ Technical support for users
UC❼ Technical platform user training
UC❽ Data consumptions via API
UC❾ Costs of access and segmentation controls and guideline development
UC❿ Costs of user administration
UC⓫ Cost offset for consuming data from the new platform

UC⓬ Cost offset of providing useful reports to the sector
Disposal costing elements
DC❶ Data archive and deletion costs
DC❷ Data removal considerations (due to people changing privacy consent etc)
Other costing elements
OC❶ Cost offset of providing national collections data back to sector
OC❷ Cost model vs cost-benefit model considerations
OC❸ Cost of supporting comms, timing and sector engagement of PHDP
OC❹ Cost of project initiation, resourcing, set-up, consultation, project management, administration, change management
OC❺ Governance costs
OC❻ Contingency cost as % overall cost
OC❼ Enforce penalty?
OC❽ Audit and legal costs
OC❾ Publicity campaign (sign up to new enrolment form etc)
OC❿ Integration of National Collections data with PHDP data

7.5 Dynamic modelling

The model can approximate (within known limits) level of investment to establish nationally consistent foundational primary care data set for New Zealand and more importantly inform Te Whatu Ora, Te Aka Whai Ora and Manatū Hauora decision makers on the reality of investment required and for how long.

The model makes assumptions based on known development within Te Whatu Ora, likely costs to the sector but remains agnostic on who will provide the services for establishment. The model stands alone and estimates minimum investment in a foundational primary care data set.

The group traversed proprietary interests, intellectual property, investment horizons, likely investor, timing, data access, curation and transformation costs, change management, costs to the sector and costs to Te Whatu Ora.

The group agreed that a data lifecycle stage model would underpin an approach.

The cost of transformation can be a key driver of cost per data class or element, based on the variability of data quality that exists across the sector. The PHOs quantified these realities in PHDP - Phase 1.

Fixed, Variable Costs and Transformation

The costing model has a simple fixed and variable cost approach for the majority of the costing elements being considered. Feedback from the working group suggested that funding to acquire data should be calculated based on number of PMS vendors and general practice clinics each organisation has as these two factors significantly affect the efforts required to set up and provide required data from general practice sources.

The balance between fully costed and marginal costing based on ‘sunk’ costs within Te Whatu Ora and the sector can be explored in the more detailed peer review of the model.

The model is built to be dynamic and adjustable based on core assumptions and can be changed based on new factors or assumptions that may be defined.

Investment across years

The model estimates overall cost per year across a five-year horizon.

8 PHDP Analytics and Reporting

Of critical importance to the objectives of PHDP is to provide analytical outputs to many stakeholders; not just data, but knowledge that can inform and enable improved health outcomes for people.

The establishment of a nationally consistent foundational primary care data which is the overarching objective of PHDP requires advanced analytical approaches to inform:

- Enabling the Pae Ora (Healthy Futures) Act 2022 legislation.
- Population health planning, policy development and implementation.
- Public health knowledge and surveillance.
- Planning and development of localities.
- Insights and monitoring of progress for Te Whatu Ora (Health New Zealand), Te Aka Whai Ora (Māori Health Authority), and Manatū Hauora (Ministry of Health).

Figure 7 - Population health approach to understanding primary care: descriptive analytical questions - key themes - illustrates the key themes of analytical needs identified by the Analytics Leadership Working Group anchoring on analytical questions, indicators or reports across health needs, accessibility, and other areas (see Appendix E.1 Population health approach to understanding primary care: analytical questions for the detailed view of themes and entries across diagnostic and predictive analytical themes as well).

While Te Pae Tata[62], Whakamaua[63] Ola Manuia[64] and a number of top-level health strategies (interim, or otherwise) exist to support the health and disability system reforms[65], further development and alignment of reporting requirements at capability or functional levels are still underway.

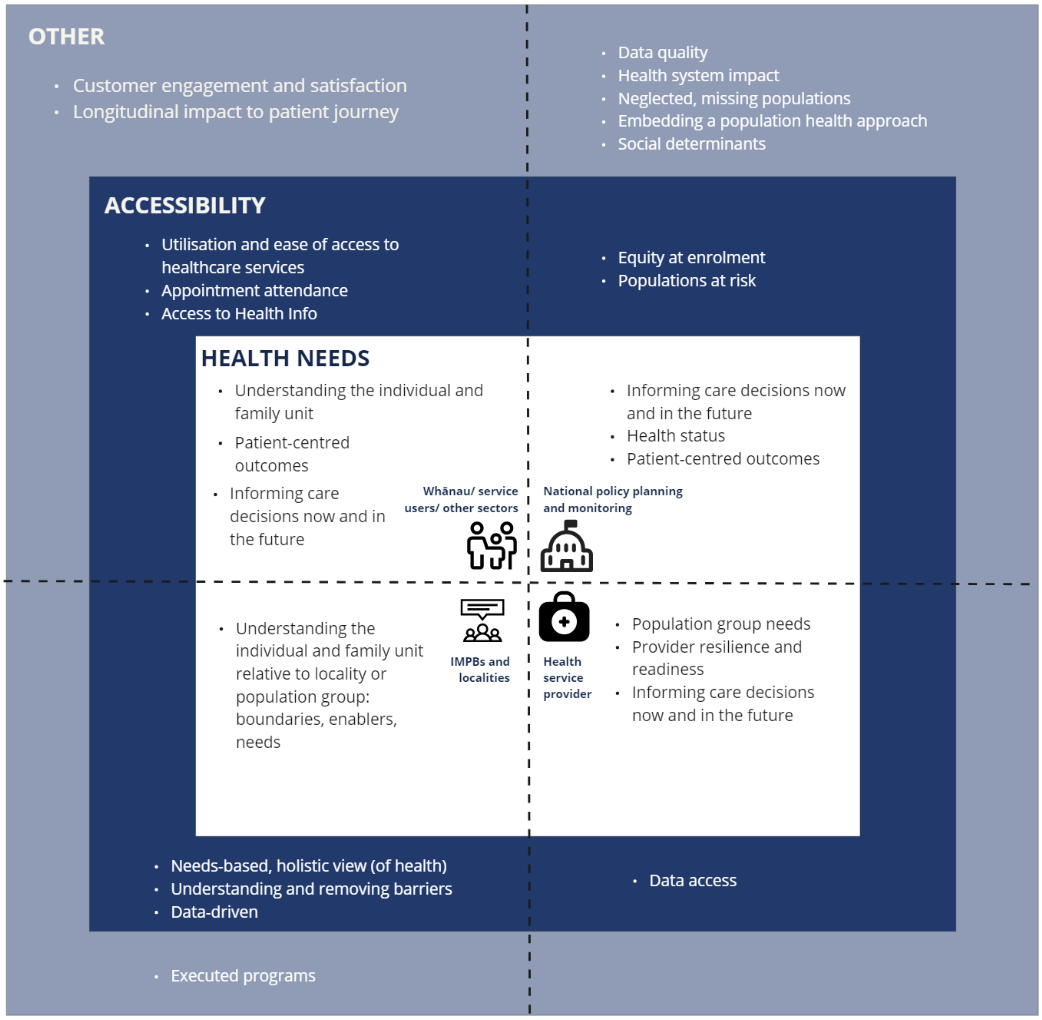
The programme team in collaboration with the Analytics Leadership Working group chairs recognise this and continue to coordinate with various health agency representatives as well as other stakeholders to ensure that business strategies are directly supported and aligned with the outputs of the PHDP program. Reporting requirements are also expected to arise from the data governance framework when it is developed, among which would

---

[62] https://www.tewhatuora.govt.nz/whats-happening/what-to-expect/nz-health-plan/
[63] https://www.health.govt.nz/publication/whakamaua-māori-health-action-plan-2020-2025
[64] https://www.health.govt.nz/publication/ola-manuia-pacific-health-and-wellbeing-action-plan-2020-2025
[65] https://www.health.govt.nz/new-zealand-health-system/overview-health-system/health-and-disability-system-reforms

typically require, among others, some data quality, data usage, compliance, performance and various trend analysis reports as well.

Figure 7 – Population health approach to understanding primary care: descriptive analytical questions – key themes



9 Appendices

A. Data Governance Requirements Working Group outputs

1. PHDP Purpose statement documentation

2. Draft early thinking around the secondary use of primary health data

3. PHDP references on international precedents

4. Primary Care Data Set draft data ecosystem diagram

5. Metro Auckland Data Sharing Group conversation notes

6. PHOs enrolment forms analysis

7. PHOs Data Sharing Memorandum of Understanding (MoU) analysis

8. Purpose-use framework documentation

9. Draft high-level application process for the use of the National Primary Care Data Set

B. Primary Care Data Set Working Group outputs

1. National Primary Care Data Set workbook

2. National Primary Care Data Set data dictionary (minimum attributes)

3. National Primary Care Data Set data dictionary (full template)

4. National Primary Care Data Set – data quality issues (detailed view)

National Primary Care Data Set Data class	Completeness	Uniqueness	Consistency	Timeliness	Validity
Conditions	Y	Y	Y	Y	
Medications	Y		Y		Y
Immunisations	Y	Y	Y	Y	
Smoking	Y		Y		
Measurements and vital signs	Y		Y		
Diagnostic results (numeric, coded)			Y		
Encounters	Y		Y		Y
Vaping	Y		Y		

Care Plan	Y		Y		
Referrals	These data classes were introduced after the Current State Analysis survey was completed. Unlike all other National Primary Care Data Set data classes initially proposed, data dictionaries for these have yet to be completed. See also Appendix B. 2 National Primary Care Data Set data dictionary (minimum attributes or Appendix B.1 National Primary Care Data Set workbook for the detailed overview of the National Primary Care Data Set.				
Procedures					
Pregnancy					
Functional status					
Patient Demographics					

5. Existing reports – public interaction with the health system

C. Technical Platform Requirements Working Group Outputs

1. Technical platform requirements workbook

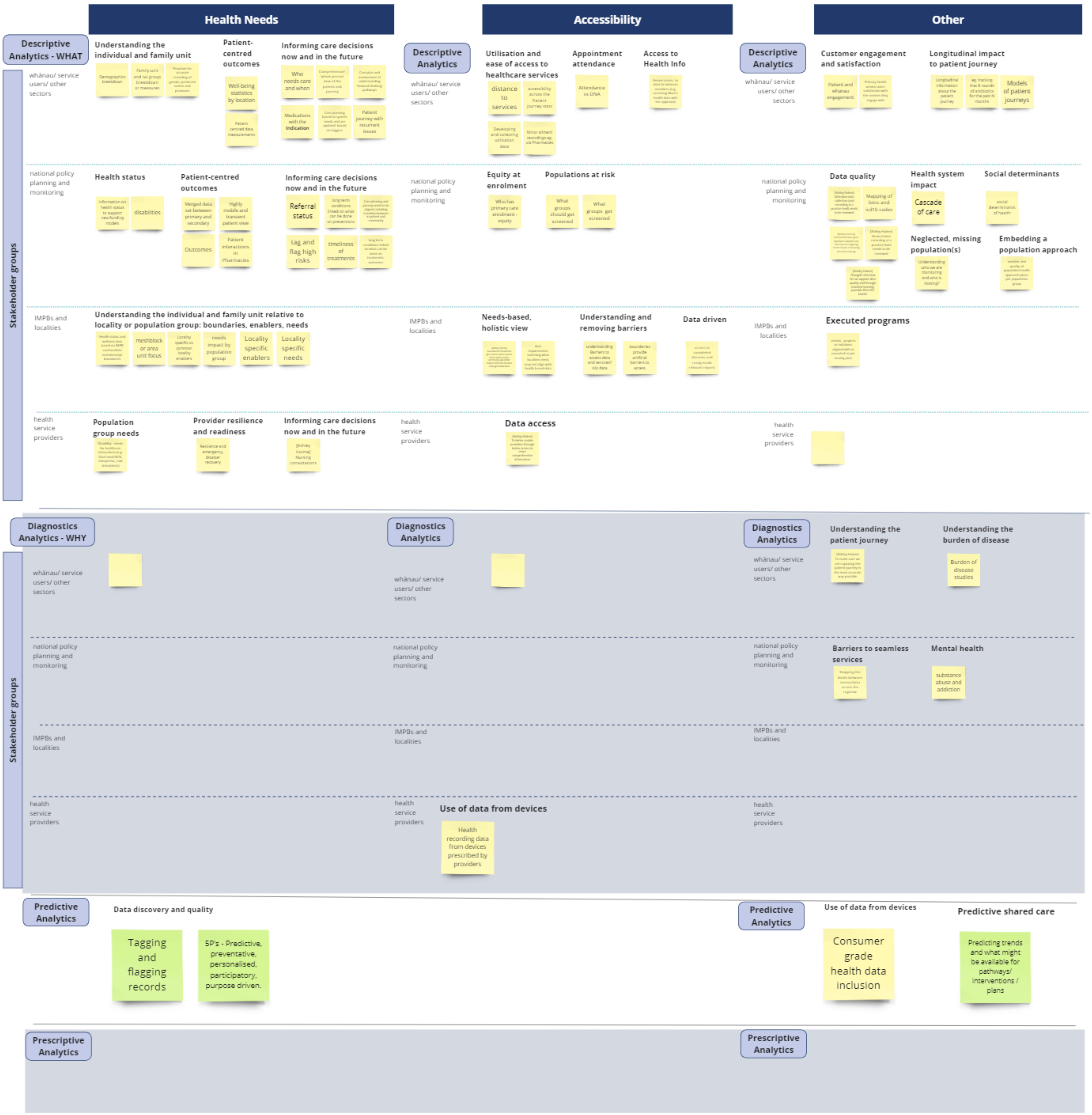
D. Commercial Working Group Outputs

1. Costing model elements

E. Analytics Leadership Group Outputs

1. Population health approach to understanding primary care: analytical questions

Population health approach to understanding primary care: descriptive analytical questions



F. PHDP Working Groups – general documentation

1. PHDP working groups membership register

2. Terms of Reference

3. PHDP working groups – ways of working

4. Collaborative online whiteboard (Miro) link

G. PHDP Governance Group – general documentation

1. Governance group membership register

H. PHDP engagement

1. PHDP engagement register