



# 1 Patron Project & Program of Work Guidelines

1	Patron Project & Program of Work Guidelines	1
1.1	Purpose of this document	2
1.2	Principles underpinning the use of Patron data	2
1.2.1	Acquiring and Curating Patron Data	2
1.2.2	Practice Agreements for Use of the Data	2
1.2.3	Data Governance	3
1.2.4	Funding Support for Data Use	3
1.2.5	Patron Annual Review Process	4
1.3	Program of Work - Using Patron Data across Multiple Subprojects	5
1.3.1	Program of Work Structure	5
1.3.2	Program of Work Scope	6
1.3.3	Program of Work Subproject Feasibility Analysis	6
1.4	Applying to use Patron Data	7
1.4.1	Application for a Patron Project	7
1.4.2	Application for a Patron Program of Work	7
1.5	Data Export Requests	8
1.5.1	Exporting Research Results	8
1.5.2	Exporting Derived Datasets	8
2	Patron Research Dataset Terms of Use	10
2.1	Dataset Ownership	10
2.2	Approvals	10
2.3	Payment	10
2.4	Research Environments	10
2.5	Externally Hosted Datasets	10
2.6	Data Linkage	11
2.7	Program of Work	11

## 1.1 Purpose of this document

This document provides information for users of Patron data about:

- How Patron data is collected, managed and funded
- How to determine whether you require Patron data as a Project or as a Program of Work
- Applying for a Project or a Program of Work
- What data can be exported from the secure research environment
- Patron data terms of use

## 1.2 Principles underpinning the use of Patron data

Data for Decisions, incorporating the Patron Program of Research, is a University of Melbourne research initiative undertaken through the University's Department of General Practice and Primary Care. With the permission of collaborating general practices or other primary care providers, we collate data from de-identified medical records to support medical research to increase knowledge, quality improvement and improve healthcare practices. The data is stored in a University of Melbourne managed primary-care data repository called Patron (**P**rimary Care **A**udit, **T**eaching and **R**esearch **O**pen **N**etwork Program).

Researchers can apply to access subsets of data from the Patron data repository to answer research questions.

### 1.2.1 *Acquiring and Curating Patron Data*

The Patron repository is a curated dataset derived from Electronic Medical Record (EMR) data contributed by over 140 general practices in Australia. Once practices are onboarded, the Patron technical team liaise with these practices and ensure regular data extracts are undertaken following agreed guidelines. They manage the secure acquisition of the data and any changes required to de-identify the data being extracted before it is stored on a secure server. They ensure that the data is curated to the highest standards and then clean, code and phenotype it to prepare it for research. The data is curated within a cybersecurity-approved and penetration-tested secure research environment and is available to researchers via secure research environments for example the UoM Secure Research Environment (SRE) platform.

### 1.2.2 *Practice Agreements for Use of the Data*

Participating general practices agree to the use of EMR data from their practice for research. Depending on the wishes of a practice, this can be conducted in different ways:

- Investigator driven research funded by research grants
- Commercially funded research, where a research organisation undertakes the research independent of the funder. The funder in this case would not have access to the data and would only be provided with deidentified aggregated project outputs.

- Commercially funded research, where the funder is provided with access to the data for analysis. Usually this would be within a UoM SRE and data would not be provided externally to the funder.

The Patron technical team collaborates with approved researchers to ensure that projects and programs are restricted to data corresponding to the agreements held with participating practices. There is a careful review of each sub-project within an approved program of research to ensure on-going compliance with agreements and expectations.

The Patron data set includes some data fields that have been determined to be sensitive. These include geographical indicators and fields such as occupation, which, taken together, may increase the ability to re-identify individuals and determine their ethnicity. The Patron review process considers whether sensitive fields are appropriate for a project and may not approve their inclusion.

Research that focuses specifically on Aboriginal and Torres Strait Islander (First Nation) people or that differentiates such sub-populations may be permitted but requires special consideration within ethics and in the Patron review process and requires the involvement of First Nations people in the research design/leadership and/or community oversight.

### **1.2.3 Data Governance**

All Patron project and program applications are governed by an Independent Data Governance Committee (DGC), and the Patron team works with the DGC to ensure that appropriate oversight is applied, respecting the rights of consumers and stakeholders.

- The DGC requires data released for research to always be held within a Patron-approved Secure Research Environment (SRE). A suitable SRE will ensure researchers are unable to download any data from the secure research environment except through a formal and mutually agreed process. An SRE would typically prevent internet access and would implement multi-factor authentication for all user access.
- Data must stay in the institution and environment originally approved in the program application
- Linking Patron data with a second data set must be approved in the Patron application for a project or program, and no subsequent new linkage can be applied to a subproject. Any changes to data linkage would require an amendment or a new application and agreement.

### **1.2.4 Funding Support for Data Use**

Patron does not receive core funding support. As a result, all activities including keeping data from all participating GP surgeries current, data curation, data usage and cybersecurity must be supported by fees charged to users. Researchers seeking access to Patron data for approved purposes will be provided with a detailed quote, which will include some fixed

costs for the services required to provide the approved data to a Secure Research Environment (SRE) and may include fixed or annual fees for ongoing access to the data.

Researchers and research groups must engage with the Patron team prior to grant submission to ensure that all data access, extraction, governance, and delivery costs are appropriately scoped and fully costed within funding applications. The Patron team will work with researchers to define a fit-for-purpose service that aligns project requirements with available resources; however, Patron cannot absorb unfunded project costs or be held accountable for funding shortfalls.

Research funding levels vary considerably across projects. Where funding constraints exist, researchers may consult with the Patron team to re-scope or prioritise project components to remain within the available budget. If project funds are insufficient, elements of the proposed work may need to be reduced, deferred, or removed to ensure delivery remains financially viable.

Supporting student-led research is a core objective of the Patron initiative. However, student projects are not exempt from cost recovery requirements. Patron incurs fixed and ongoing costs associated with data governance, infrastructure, security, quality assurance, and staff expertise. As such, support for student projects is subject to available capacity, agreed pricing, and confirmed funding, and may be limited where resources are constrained.

### ***1.2.5 Patron Annual Review Process***

Patron implements an annual review process to ensure that all projects, programs, and subprojects are monitored, comply with ethics, have paid agreed-upon fees, and that any Patron-related publication has been reviewed by a Data for Decisions academic and/or has included them as an author or in acknowledgements.

All project and program ethics must include a statement ensuring that any changes to ethics are also submitted as an update to the corresponding Patron application. This includes any modifications resulting from changes to the research scope or personnel. Ethics must also state that publications will be sent to the Patron team for review and to ensure acknowledgement before publication.

All projects and programs have a single responsible individual who ensures compliance with national research requirements. This person must be the researcher accountable for the ethics of the project or program. This individual will be responsible for projects in accordance with ethics requirements and for all operational compliance matters relating to Patron on behalf of the contracting organisation.

### 1.3 Program of Work - Using Patron Data across Multiple Subprojects

Most Patron applications are for a single project where the project dataset provided is used for a specific one-off piece of research analysis.

However, there are also initiatives that request access to Patron data to answer multiple research questions that fall under an intended scope of work. In this case, it is preferred that an application be submitted as a Program of Work, with each subproject submitted as a separate application. This allows the Program to reuse a Patron dataset for multiple subprojects, provided they are within the agreed Program scope of work, and is a more cost-effective approach.

The Patron team attempt to implement efficient processes to manage subprojects. The Patron online application portal enables subprojects to be easily input under the overarching Program of Work application, avoiding duplication of data entry. All subprojects need to be approved by the DGC.

#### 1.3.1 Program of Work Structure

A Patron 'Program of Work' is defined as a request for Patron data that responds to an overall research aim/question, with the intention of submitting subsequent requests for the use of the same data set for related research questions or sub questions. A Program of Work should consider these core principles:

- Scope of the data set
- Scope of the research question/s
- Scope of the research team

Further to these principles, applications for a Program of Work also:

- Will be limited to 3 years, then will require renegotiation
- Will require consideration for funding. Patron fees will include a per-program and per-subproject cost. The funding mechanism must be agreed upon at the time of application
- Depending on the type of program, there may be a limit to the number of subprojects that can be negotiated at the time of application
- The Project Leader will be consistent across the Program of Work and each Patron data request within the program of work.
- Access to the whole program dataset should be limited to a small group of the program researchers. Generally, this should be limited to 3 researchers to ensure the Five Safes People are met and to ensure data minimisation principles apply in subprojects (Safe Data). It is envisaged this group would be responsible for Subproject Feasibility Analysis, which is discussed further below.

### **1.3.2 Program of Work Scope**

A Program of Work needs to be agreed upon with strict parameters preventing uses that may be beyond the initial intended scope.

**Parameter 1:** In a program, the users of the data need to be defined upfront and typically represent a particular collection of users working in a common area, such as cancer research.

**Parameter 2:** Any use of the data for groups or organisations outside the original approved Patron Program of Work scope CANNOT be undertaken without being documented and receiving approval through a new Patron application. Where the intent is to reuse a dataset a research group already has access to, the Patron team will need any agreements with external groups to be approved through the normal Patron application process. This will include a quote based on the Patron cost recovery policy to ensure the long-term viability of Patron.

### **1.3.3 Program of Work Subproject Feasibility Analysis**

Within a program of research, the research group may wish to explore the ability of the provided data to answer new research questions. Such a use is permitted subject to the following definitions and rules.

- Patron DEFINITION: A 'Feasibility Analysis' is where analysis is undertaken to determine whether the data held in a collection (i.e. Patron Program of Work dataset) is sufficient to answer a proposed research question. It must not ANSWER a research question.
- If a Feasibility Analysis results in any form of research publication, then it is RESEARCH and not a Feasibility Analysis.
- Feasibility Analysis need to be registered with the Patron Program of Work but does not need a sub-project application. It can be registered against the overarching Program application.
- A maximum of 3 individuals are allowed to determine research Feasibility Analysis using a Patron Program of Work dataset. This is because of the scale and sensitivity of the full dataset in a Program of Work, which in some cases is linked to other datasets can be extensive, and hence needs to be limited.
- Should the Program wish to use other individuals (for example students) to help with tasks like coding as part of a feasibility analysis, then this data should be partitioned and provided to the student away from the main data collection for example in a secured sub-folder.
- If the student expects a research publication from their work, then this is research and must be converted to be a subproject going through the usual subproject application process.

## 1.4 Applying to use Patron Data

### 1.4.1 Application for a Patron Project

The on-line Patron application process should be completed and the Patron team will work with you to review project feasibility and provide a cost estimate. If you think you may have a series of on-going research questions, then you may need to apply for a Patron Program of Work. Determining this early can save work later – the Patron team are here to help.

### 1.4.2 Application for a Patron Program of Work

The on-line Patron application process should be completed and the Patron team will work with you to review program feasibility and an appropriate management and cost model.

A Program of Work will be costed by 1) an ongoing subscription for continued use of the program dataset, including subproject feasibility analysis, data curation, and the preparation of subproject datacuts, and 2) a small cost per subproject dataset.

- A program can be funded to allow a research group access to Patron data (including data updates) for an agreed time interval (payment made in advance or as an annual fee). Some permitted subprojects may be agreed.
- A program can be funded to grant users within a research group access to Patron data, including data updates, with payments made on a per-subproject basis.
- Limitations on the scope of the delivered Patron dataset may be applied.
- A program can be designed for long-term implementation; however, the model and fees must be reviewed every three years.

Things to consider when assessing whether a Patron data request should be regarded as a Patron Program of Work.

- The Project Leader should identify within the initial application that the data request is submitted as a Patron Program of Work request, with related individual data requests intended to follow using the same dataset. Applicants must be aware that any requests for Patron data are assessed against the Five Safes<sup>1</sup>

The Five Safes are:

1. Safe Projects - Is the use of the data appropriate?
2. Safe People - Can the researchers be trusted to use it appropriately?
3. Safe Data - Is there a disclosure risk in the data itself?
4. Safe Settings - Does the access facility limit unauthorised use?
5. Safe Outputs - Are the statistical results non-disclosive?

A request for Patron data as part of a Patron Program of Work will be assessed against whether the use of the data is appropriate for the project, and consideration will be given to

---

<sup>1</sup> Ritchie, Felix. (2017). The 'Five Safes': a framework for planning, designing and evaluating data access solutions. 10.5281/zenodo.897821.

limiting the purpose for which the recipient can use the data. Ethics approvals for the Program of Work should include the application and approval of any subprojects by the Data Governance Committee.

- Any individual request for data within a program of work that requires additional data fields or additional years of data, beyond the original data request, will be considered as a significant change, requiring a new Patron data request (within the program).
- The project leader is required to complete a Patron annual progress report for each research project.
- A program of work should be considered by researchers so that it can be identified at the time of putting forward the Patron data request.
- Each subproject within the Program of Work must be covered by the Program of Work ethics and will require a subprogram application online. The Data Governance Committee will need to be notified of each subproject within the Program of Work and approval confirmed by the Patron Administrator. Subprojects may require individual ethics approval.
- Costs and funding mechanisms for Patron Programs of Work need to be negotiated in advance. Maintaining Patron is only funded through research projects that utilise it. Patron Programs of Work will usually be based on a subscription model (for example, an annual fee for the program) and an individual sub-project fee basis.
- Data must stay in the institution and environment originally approved in the program application
- For linked data, it must be approved for linkage in the approved Program of Work, and no subsequent new linkage can be applied to a subproject. This would require a new Program of Work application.

## **1.5 Data Export Requests**

### **1.5.1 Exporting Research Results**

A Patron project / program of work is likely to wish to export research results for the purposes of publication. Such data releases require review by a Patron team member to ensure results are aggregated and Patron individual patient-level data is not extracted. This is a routine process, and the Patron team are here to help review the data release.

Review of publications utilising Patron data also need Patron review in advance and appropriate acknowledgement of the contribution of Patron.

### **1.5.2 Exporting Derived Datasets**

Where planned-in-advance as part of a Patron project or program application, a request to extract a more comprehensive derived dataset for further use may be agreed.

*Definition: **Derived data** is new information created by processing and combining existing data sets to reveal new insights. It is not simply a summary but involves transformations, calculations, or aggregations on raw or other existing data.*

A derived dataset is your IP and can be exported in the same way as research results, however if ANY Patron individual level de-identified data is included, then strict rules apply:

### **Exporting Derived Datasets Containing Patron Data**

Two principles should be noted when considering generating a derived dataset containing any amount of individual patient level Patron data:

1. GP surgeries that have agreed to share their data for Patron research do so on the understanding that their patient data remains safe and under the custodianship and direction of the Patron Independent Data Governance Committee. This means derived datasets that also include de-identified individual-level Patron data must remain under Patron scrutiny and processes.
2. A derived dataset may contain a substantial fraction of Patron repository data. Patron functions through research projects paying for data access – we need to keep our network running. Creating a derived dataset is not a ticket to avoid paying for the on-going use of Patron on different projects or sub-projects.

The Patron team will work with you to determine the feasibility of allowing derived datasets containing Patron data to leave a Patron secure environment. It is likely that your on-going uses of data will remain indefinitely registered under a Patron program of work with associated review and annual review. A Patron representative will need to be involved in approving any further intended use of the derived dataset and a further Patron application or sub-project application may be required.

Further costs are likely to apply. The cost model should be discussed at an early stage. Costs may be to cover administrative overheads (application and Data Governance Committee sitting fee and costs) or may be more substantial if Patron data is playing a significant role in your on-going research

To assess the proportion of derived data versus Patron source data requires an understanding of the transformation and aggregation processes that have been applied to the source data. Some questions for consideration - is the derived data:

- highly dependent on the source Patron data
- directly linked to specific source Patron data records
- a direct transformation of individual Patron source data points (e.g., converting units, formatting strings)
- information that has been added enabling new analytical capabilities
- aggregated data

## 2 Patron Research Dataset Terms of Use

The Patron Program of Research hosts and manages a primary-care data repository from which researchers can request a dataset for a research Project or Program of Work. A dataset is created based on project/program custom requirements with specific inclusion and exclusion criteria that meet data minimisation principles (i.e. only data that is relevant to the research project/program is provided).

### 2.1 Dataset Ownership

- All ownership, governance and intellectual property rights in any Patron-provided data remains vested in Patron – a researcher or research group is a user - not an owner.

### 2.2 Approvals

- Researchers using data from Patron must gain approval from an NHMRC approved ethics committee, the independent Data Governance Committee, and commit to a legal agreement stipulating terms of data use.
- All Projects, Programs of Work and Program of Work Subprojects are subject to approval by both the independent Data Governance Committee (DGC) and the Patron Management Group Committee (PMGC).
- All Projects and Programs of Work must have a Data Risk Management Plan (DRMP) which has been approved by the Patron Data Steward.
- All Projects and Programs of Work are bound by the DGC approvals and the terms of their DRMP.

### 2.3 Payment

- Projects or Program of Work Subprojects must be applied for, approved and paid for before commencement.
- All payments are to be made in advance of any works commencing.

### 2.4 Research Environments

- Patron data is only approved for release to Secure Research Environments. University of Melbourne Secure Research Environments are the default location however external research groups may not have permission to use these environments.

### 2.5 Externally Hosted Datasets

- Projects requesting to host a Patron dataset externally will need to provide justification for consideration before being approved. There are risks to the Patron Program of Research associated with this scenario due to the loss of data control. These include data security, data breaches, unauthorised data access, data compliance issues with linked datasets, and undisclosed projects being undertaken.
- Externally hosted Patron datasets remain under the governance and management of the DGC and PMGC. The researcher(s) associated with the dataset are users of the data, NOT the owners.

- An externally hosted Patron dataset must be stored on secure servers in accordance with the DRMP that has been agreed.
- Additional Patron management is required to keep oversight of the usage of an externally hosted Patron dataset. This includes the Annual Patron Review and ongoing 6-monthly check-ins.
- A Patron dataset MAY NOT be moved to another storage entity without management and agreement from Patron.

## 2.6 Data Linkage

- Projects or Programs of Work that wish to link a Patron dataset to other datasets must have received all approvals (ethics, DGC and PMGC).
- Fields from linked datasets must not override any restrictions that DGC has mandated. For example, if the DGC has declined the inclusion of Indigenous status in the Patron dataset, then this field may not be included from another dataset.
- Program of Work Subprojects must have data minimisation principles applied across the linked dataset.

## 2.7 Program of Work

- A Program of Work is normally granted a subscription-based licence permitting continuous use of the dataset for activities including project feasibility analysis, data curation, and the preparation of Subproject datasets.
- Feasibility analysis can be undertaken to determine if a Subproject is viable, however a Subproject must be approved and paid for before a Subproject dataset is created and the Subproject commences.
- Feasibility analysis is to be limited to a small group of the program researchers that are provided with access to the whole program dataset. Generally, this should be limited to 3 researchers to ensure the Five Safes People are met.