

National Transfusion Dataset

ARDC AUSTRALIAN DATA PARTNERSHIPS PROGRAM PROJECT FINAL REPORT

Prepared by: Prof Erica Wood, Dr Karina Brady, Dr Fiona Chen, Prof Zoe McQuilten, Ms Tina van Tonder,
Mr Neil Waters, Dr Cameron Wellard

30/11/2023

CONTENTS

| | | |
|-----|---|----|
| 1 | PROJECT INFORMATION | 2 |
| 1.1 | Background | 2 |
| 1.2 | Achievement of project aims | 3 |
| 2 | DESCRIPTION OF PROJECT OUTPUTS | 4 |
| 2.1 | Achievements against project work packages: | 4 |
| 2.2 | Project outputs | 8 |
| 2.3 | Outreach and training activities | 8 |
| 3 | SUSTAINABILITY PLAN | 9 |
| 4 | FAIR | 10 |
| 4.1 | Implementation of FAIR Data Guidelines | 10 |
| 5 | PROJECT IMPACT | 13 |
| 5.1 | Communications and Engagement | 13 |
| 5.2 | Research Outcomes Planning | 16 |
| 6 | LESSONS LEARNED | 17 |
| 6.1 | What Went Well? | 17 |
| 6.2 | What Could be Improved? | 17 |

1 PROJECT INFORMATION

| | |
|------------------------------------|---|
| INVESTMENT ID | DP708 |
| PROJECT START AND END DATES | Jan 2021 - Dec 2023 |
| LEAD ORGANISATION | Monash University |
| PARTNER ORGANISATIONS | Ambulance Victoria (AV) Blood Synergy Myeloma and Related Diseases Registry (MRDR) Lymphoma and Related Diseases Registry (LaRDR) Australian and New Zealand Intensive Care Society (ANZICS) Pre-hospital Emergency Care Australia and New Zealand (PECANZ) Aplastic Anaemia and Other Bone Marrow Failure Syndromes Registry (AAR) |
| PROJECT CONTACT PERSON | Prof Erica Wood |

1.1 Background

Blood transfusions, used wisely, save lives. Blood is designated an essential medicine by the World Health Organization and is a precious national resource fundamental to modern medical systems. It is used in a wide range of clinical settings from critical care and trauma to obstetrics, surgery, and cancer care. Transfusions must be used judiciously: blood is donated by volunteers, and transfusions carry risks to patients and great cost to the community – ***Australia spends >\$1.3 billion every year on blood products. Despite the essential role of blood transfusion in health care, Australia lacks the infrastructure to provide fundamental data on:***

- How blood is used (who needs it, when, where, how much, and why);
- Outcomes for transfused patients, including adverse events (“haemovigilance” reporting); and
- Costs to the Australian health system.

Governments have identified these data gaps as national research priorities. To address these important unmet needs, the National Transfusion Dataset (NTD) is bringing together information on blood product

use from prehospital and hospital datasets and linking it with patient outcomes collected by clinical registries to establish new national transfusion research infrastructure.

1.2 Achievement of project aims

The overarching objectives of National Transfusion Dataset (NTD) are to:

1. expand transfusion data coverage to include prehospital transfusion and create the first comprehensive national dataset of transfusion practice linked with transfusion laboratory data and clinical outcomes
2. strengthen data quality
3. streamline workflows for data collection to ensure an ongoing up-to-date resource
4. improve access to transfusion data for stakeholders and researchers
5. improve Australian transfusion research capacity and efficiency
6. create new research opportunities (e.g. data for health economics analyses) to inform national transfusion policy and practice, improve blood utilisation and patient management and outcomes

The NTD aims, ultimately, to form a vein-to-vein picture of blood use in Australia, linking donor, prehospital and hospital data to determine when, where and why blood is used and what the outcomes are for patients. By bringing these datasets together a national dataset of transfusion practice can be formed. With the funding received from the ARDC, the NTD was established and the process of obtaining governance and collating the initial, pilot datasets commenced. The NTD is an ongoing project and the objectives are therefore also ongoing.

Data are collected from many different sources with no common data format. The NTD strengthens data quality by harmonising these datasets into one format in order to link information. During the data harmonisation process, the Data Team work closely with the data custodians to understand the data and ensure accurate interpretations are made, where necessary.

The collection of large data sets directly from health services is complex due to the sensitive nature of the data and the variety of health databases where data are housed. Data extraction and harmonisation workflows have been established for the prehospital and hospital sites that are currently providing data to the NTD under the ARDC funding. Streamlining of the data pipeline is an ongoing process and we are working to automate data harmonisation.

In addition to providing a data resource for research, the team produces hospital reports which are shared with participating hospitals every 12 months to provide a summary of transfusion practice at the institution. As the NTD grows, these reports will provide information on how the hospital

compares to all the hospitals in the dataset, providing an opportunity for benchmarking. In addition, data access requests can be submitted through the NTD website (<https://www.transfusiondataset.com/>).

The NTD is governed by a Data Partnership Group with experience covering many different fields including donor, prehospital and hospital services, as well as clinical registries and health economics. The NTD creates opportunities for researchers to collaborate and for young researchers from non-transfusion backgrounds to learn about and participate in transfusion research.

In 2022 the NTD was awarded a MRFF Research Data Infrastructure Grant to expand the number of participating prehospital and hospital services, undertake further data linkages, such as with donor data, add unstructured data from medical records and perform health economic analyses to better understand the cost of blood products and blood product administration. The groundwork laid in the ARDC-funded project formed the foundation for the MRFF-funded expansion which will greatly increase the capacity of the NTD and create new opportunities for transfusion research.

2 DESCRIPTION OF PROJECT OUTPUTS

2.1 Achievements against project work packages:

| WORK PACKAGE | DETAILS INCLUDING EXPLANATION OF ANY VARIATION | COMPLETION DATE |
|--|--|-----------------|
| 1.1 Project planning | The Project Plan was submitted to the ARDC and approved with minor updates to Section 8. | 31/03/2021 |
| 1.2 Appoint project manager | Dr Kim Huynh was appointed as Project Manager for the NTD. | 29/03/2021 |
| 1.3 Secure ethics approval for data linkages | For the DP708 to build on the Australian and New Zealand Massive Transfusion Registry (ANZ-MTR) it was incorporated with the Transfusion Database (TD) and renamed the National Transfusion Dataset (NTD). Following the advice of the Alfred Ethics & Research Governance Office, ethics approval for the | 16/11/2021 |

| | | |
|--|---|------------|
| | <p>incorporation was targeted first with ethics approval for the data linkages subsequently secured (approvals attached). Approval for the linkages of immunoglobulin data from selected clinical registries were also included in the subsequent submission. Although the two-step ethics approval led to a delay with completing work package 1.3 it allowed for an earlier completion of work package 2.4 and commencement of work packages 2.5 and 3.2.</p> | |
| 1.4 Governance approval at pilot hospital sites | <p>We obtained governance approval at Alfred Health (AH) and Flinders Medical Centre (FMC). There was a minor delay with the governance submission at FMC due to the transitioning of a new Principal Investigator (PI), with the previous PI retiring.</p> | 31/01/2022 |
| 1.5 Assessment & selection of data harmonisation tools, processes & plan | <p>Assessment of data harmonisation tools complete. Stata was selected to clean and organise the data to ensure format consistency between all sites and subsequent integration.</p> | 30/05/2021 |
| 1.6 Data harmonisation tool build and testing | <p>Codes for data harmonisation have been developed and tested on ANZ-MTR data. Harmonisation has been piloted on AH and FMC data, and tool refinement is ongoing.</p> | 31/05/2023 |
| 1.7 Data entry from pilot hospital sites | <p>Data extraction from AH and FMC and integration into the NTD is complete.</p> | 31/05/2023 |
| 1.8 Data entry from Ambulance Victoria | <p>There was a delay with data extraction at Ambulance Victoria (AV) due to the flow-on effect of work package 1.3. AV data was successfully extracted and received in March 2022.</p> | 31/03/2022 |
| 2.1 Inclusion of data from | <p>South Australia Ambulance Service (SAAS) MedSTAR was selected as an additional prehospital service for</p> | 30/11/2022 |

| | | |
|---|--|------------|
| additional prehospital services | inclusion. Prehospital data was extracted and added to the NTD in November 2022 after governance approval was confirmed. | |
| 2.2 Governance for additional hospital sites | Governance approval for the NTD has been submitted and received at 19 additional sites located in Tasmania, Queensland, South Australia and Victoria. | 31/10/2023 |
| 2.3 Data entry from expanded hospital sites | Data extracts have been requested from The Royal Melbourne Hospital, Launceston General Hospital, all NTD hospitals located in Queensland and all NTD hospitals and health services located in South Australia. The availability of data extracts for integration into the dataset is dependent on hospital resourcing and, as such, it can take time for data extract requests to be actioned. Once extracts are received, they will be harmonised and integrated into the NTD. | Ongoing |
| 2.4 Ethics for linkage of immunoglobulin data from selected clinical registries | Ethics approval secured with work package 1.3. | 27/10/2021 |
| 2.5 Pilot integration of clinical registry immunoglobulin data (MRDR, LaRDR) | Immunoglobulin linkage has been piloted with the MRDR and LaRDR. The linkage pool was very small as the linkage only included the first two hospital sites. | 31/05/2023 |
| 3.1 Pilot linkage with ANZICS APD data | Data from the ANZICS-Adult Patient Database (APD) were shared and initial linkages undertaken. Along with Dr Andrew Flint we are looking at blood product use in intensive care and patient characteristics. Dr Flint is a specialist trainee in intensive care medicine and PhD student with | 11/05/2023 |

| | | |
|---|--|------------|
| | Monash University. | |
| 3.2 Pilot integration of AAR data | Data linkage piloted and successfully demonstrated the usefulness of linking NTD data with the AAR. | 30/04/2022 |
| 3.3 Continued expansion of hospital sites | Per 2.2, the NTD has expanded the number of hospital sites with governance approval beyond the two initial pilot sites, and data have been requested from all approved sites. The NTD will continue to expand into the coming years, aiming for 100% coverage. Governance applications are in progress in the Northern Territory and additional governance applications will be submitted in Western Australia and selected sites within Victoria in 2024. | 30/11/2023 |
| Reconfiguring the NTD to allow tracking of dataset downloads by users | Access to the closed dataset is tracked through the NTD data access approvals and by monitoring of the Secure eResearch Platform. | 31/12/2023 |
| Integrating automated responses to provide timely and informative usage summaries to the user | Dataset summaries will be provided to users via analytics as described below. | 31/12/2023 |
| Implementing live analytics for users to view interactive outputs and generate summary data reports | Interactive outputs and summary data report options using either RShiny or PowerBI for live analytics are currently being explored. This will be implemented as more data are added into the NTD. | 31/12/2023 |

| | | |
|---|---|------------|
| Upgrading the NTD Data Access Request process to an electronic submission procedure | An online data access application submission procedure was developed using REDCap. It is available via the Resources page of the NTD website. | 31/12/2023 |
|---|---|------------|

2.2 Project outputs

The infrastructure built by the NTD project and the resulting dataset is the major output of this project. This dataset combines transfusion data from hospitals and health services and links with ambulance and retrieval services to incorporate prehospital use of blood products. The dataset has been linked with clinical registries including those in ICU, multiple myeloma, lymphoma and aplastic anaemia. These linkages have demonstrated the possibility of developing operational and research projects and reports to further investigate how, when and why blood products are used in these settings and what the outcomes of blood product use are for these patients.

From this pilot work, the NTD was awarded an MRFF grant to further expand the number of hospital sites across Australia, and to incorporate donor services as well as additional ambulance and retrieval services. Further linkages, health economic analyses and a clinical trial will be undertaken as part of this MRFF grant.

Using the current hospital and health service data available, the TRU data team have developed a hospital report to be shared with participating hospital sites. This report provides an overview of the blood product use at sites along with patient characteristics and outcomes. The report is still a work in progress and will be further reviewed by the hospital transfusion staff before publication. As more data are incorporated into the registry, the reports will be used for benchmarking, comparing the individual site against all contributing sites and tracking changes in practice, including uptake of new products.

2.3 Outreach and training activities

| ACTIVITY | ACTIVITY DESCRIPTION | NO. OF PARTICIPANTS | DATE OF ACTIVITY |
|----------|----------------------|---------------------|------------------|
| NTD open | A meeting with NTD | 25 | 7 Nov 2023 |

| | | | |
|--|---|------------------------------|----------------|
| <p>meeting and discussion, Blood 2023</p> | <p>contributors, users and stakeholders (including the National Blood Authority, Lifeblood and clinicians) providing an update on the NTD, followed by a discussion on what the parties would like to see done with the NTD and how the data extraction and access procedures could be further improved.</p> | | |
| <p>Meetings with hospital information systems (HIS) and laboratory staff</p> | <p>The NTD Team have been meeting with HIS and laboratory staff, working closely with them on data extractions for the NTD. These meetings help to understand the nuances and differences between hospitals and ensure that they are involved in development of data extracts as well as helping to resolve queries and issues as they arise.</p> | <p>Approx. 5 per meeting</p> | <p>Various</p> |

3 SUSTAINABILITY PLAN

Additional funding for a further five years of activity was obtained through a MRFF Research Data infrastructure grant in 2022.

4 FAIR

4.1 Implementation of FAIR Data Guidelines

FAIR Data Actions

| # | FAIR ACTION | DETAILS OF IMPLEMENTATION |
|---|---|---|
| 1 | All data outputs are assigned appropriate PIDs, preferably a DOI | All project activities have been uploaded into the Monash Bridges project repository where PIDs (DOIs) are generated. Landing page: https://doi.org/10.26180/c.6001192.v4 |
| 2 | All data outputs have metadata to enable discovery | The dataset is on Monash Bridges and metadata are generated via the project repository to enable discovery (https://bridges.monash.edu/articles/dataset/National_Transfusion_Dataset_NTD_/22151987). Metadata can be exported as DataCite or Dublin Core. Metadata on project activity (i.e. publications, presentations) can be downloaded via the Monash Bridges collection, accessible from the landing page. |
| 3 | All data outputs have a record in Research Data Australia | The Monash Bridges repository is harvested to <u>Research Data Australia</u> , so any record created in Bridges is discoverable via Research Data Australia. |
| 4 | All data outputs are registered with relevant discipline-specific discovery aggregators (Recommended , if they exist) | Transfusion discovery aggregators were explored, none exist. |
| 5 | The persistent identifier for the data output being described is included in the metadata | The PIDs (DOIs) for the project activity are included in the metadata via the Bridges repository. The PID (DOI) for the dataset is available from the dataset entry in the repository, per line 2. |

| | | |
|----|---|---|
| 6 | All data outputs are made as openly available as possible; they are only closed where necessary | Due to the identifiable and sensitive nature of the data, the NTD Dataset is closed and accessible via an application process detailed and made available on the dataset Monash Bridges repository page and the NTD website. Project activity with outputs (i.e. publications) that are not sensitive are openly available. |
| 7 | All data outputs are made available through a repository | Project activity with outputs that are not of a sensitive nature are published on the NTD website and in the Monash Bridges repository and are available via the Monash Bridges landing page. |
| 8 | All data outputs are available as a download and/or accessible through an open, documented API (where data is not closed) | The Dataset is closed but available through an access request. |
| 9 | If the data outputs are not openly available, there is a clear description on the landing page on how to request access to the data outputs and the conditions that need to be met | Data access (via application and Steering Committee approval) details are available on the project repository landing page. The link to request data access is also available on the NTD website. |
| 10 | The persistent identifier for the data output points to a landing page about the data output, even if the data output is not public (open). | The PIDs (DOIs) for the Dataset and related project activity (i.e. publications) point to the project repository landing page or the specific entry in the Monash Bridges repository. |
| 11 | If the data output is not openly available there is an authorisation and authentication procedure to provide access to the data | The NTD Data Access and Publication Policy and related data access request application is available on the NTD website (https://www.transfusiondataset.com/projects/data-access) |
| 12 | The persistent identifier for the data output continues to point to a landing page, even if the data output is no longer available, and there is a policy to maintain these landing pages | The PIDs (DOIs) for the Dataset and NTD project activity will point to the NTD Monash Bridges project repository landing page indefinitely. The NTD Data Management Plan details the NTD policy regarding the |

| | | |
|----|--|---|
| | | maintenance of the project repository landing page. |
| 13 | Data outputs use community-agreed standard data formats (where such agreed formats exist) | Data from the NTD are available in the following community-agreed standard data formats: .csv. |
| 14 | Metadata for the data output uses community-agreed standards (where such agreed standards exist) | Transfusion schemas were explored, there are no existing standard schemas. |
| 15 | Data and metadata use community-agreed vocabularies, data models and ontologies (Recommended , preferably internationally agreed ones where they exist) | Transfusion vocabulary standards were explored, there are no existing vocabularies. However, standard transfusion product and other terms are used, where these exist. |
| 16 | Metadata contains persistent identifiers for research objects and entities (people, organisations) linked to the data outputs (including ORCIDs, grantIDs, RAIDs, DOIs, IGSNs) | All metadata will contain PIDs (DOIs) generated via the Monash Bridges project repository. Where appropriate, ORCIDs and grant IDs are also included. |
| 17 | All data outputs are assigned a machine readable licence (preferably CC-BY 4.0) | Licensing options have been explored, and an Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) will be assigned to the Dataset and all related outputs using the data. |
| 18 | The licence information is available in a machine readable form on the landing page that the persistent identifier (for the data output) refers to | The licence information is detailed in the Dataset project repository on Monash Bridges. |
| 19 | There is a citation statement for the data output on the landing page that the persistent identifier refers to | The citation statement is part of the NTD Data Access and Publication Policy and is available on the landing page. |
| 20 | Provenance information on the data output is attached alongside the data (Recommended) | Provenance information (in the form of a short statement outlining where data was obtained from) will be attached to the data. |

| | | |
|----|---|--|
| 21 | Relevant discipline-specific metadata to enable reuse is captured and presented alongside the data output following research community best practice (Recommended) | There is no existing discipline-specific metadata. |
|----|---|--|

5 PROJECT IMPACT

The ARDC and the Government wish to demonstrate the impact on researchers, industry and the general public of the NCRIS investment.

5.1 Communications and Engagement

| ACTIVITY | DETAILS | LINK TO MATERIALS |
|-----------------------|---|---|
| Presentations/Posters | Poster: Huynh et al. (2022, September 11-14) Building a National Transfusion Dataset (NTD) for Australia. BLOOD 2022 Meeting, Sydney, Australia | Abstract available via: https://onlinelibrary.wiley.com/doi/epdf/10.1111/tme.12933 |
| | Invited presentation: Wood E (2022, September 13: Transfusion clinical research in Australia and New Zealand: the 'big picture'. BLOOD 2022 conference, Sydney, Australia | No link available. National Transfusion Dataset included as example of critical infrastructure required and being developed for transfusion research in Australia |
| | Invited presentation: Wood E: Registries for clinical transfusion research. Japanese Society for Transfusion Medicine | http://www.jstmct.or.jp/annua170/ (Website in Japanese) |

| | | |
|--|---|--|
| | and Cell Therapy, 2022 May 27-29 | International conference presentation |
| | <p>Presentation:</p> <p>Busch, Michael: Transfusion medicine research - big studies, big data, big impact. National Institutes of Health (USA) 2022 State of the Science in Transfusion Medicine Symposium, 2022 August 29-30</p> | <p>https://www.nhlbi.nih.gov/events/2022/2022-state-science-transfusion-medicine-symposium</p> <p>Australia's National Transfusion Dataset included in presentation of international efforts in 'big data' for transfusion practice and research</p> |
| | <p>Invited presentation: Wood, Erica: Massive transfusion practice and outcomes in trauma: Update from the Australian and NZ Massive Transfusion Registry. Trauma 2022, Australasian Trauma Society, 2022 September 1-4</p> | <p>https://www.traumaconference.com.au</p> <p>Invited presentation at Aust/NZ trauma society annual congress, including overview of reconfiguration of Massive Transfusion Registry into new National Transfusion Dataset</p> |
| | <p>Presentation:</p> <p>Huynh (2022, May 3) Blood Synergy Open Meeting, Melbourne, Australia</p> | <p>https://bloodsynergy.org/outcomes/meetings</p> <p>or</p> <p>https://www.youtube.com/watch?v=g2xLZRTfisk&t=6s</p> |
| | <p>Invited presentation on the National Transfusion Dataset to the National Blood Authority (Blood and Data and Patient Blood Management teams), 2023, January 17</p> | <p>No link available. Invited (closed) meeting with National Blood Authority on datasets to understand national transfusion practice and outcomes</p> |

| | | |
|----------------------------|--|---|
| | <p>Presentation: Van Tonder (2023, May 2) Blood Synergy Open Meeting, Melbourne, Australia</p> | <p>https://bridges.monash.edu/articles/presentation/Blood_Synergy_Open_Meeting_2023/22735163</p> |
| | <p>Presentation and discussion: The NTD was presented by Erica Wood and discussed as part of the 'Big Data Workshop' at International Society of Blood Transfusion (ISBT) Conference in Gothenburg (2023, June 18)</p> | <p>Closed to ISBT members. https://www.isbtweb.org/resource/big-data-workshop.html</p> |
| <p>MRFF Grant Received</p> | | <p>https://bloodsynergy.org/outcomes/news https://twitter.com/erica_m_wood/status/1572558394892091394 https://ardc.edu.au/article/national-transfusion-dataset-to-expand-with-2-9-million-mrff-grant/ (tweet: https://twitter.com/ARDC_AU/status/1581812282358497281) https://www.monash.edu/news/articles/mrff-backs-monash-health-and-medical-research-with-\$28m-in-funding (tweets: https://twitter.com/erica_m_wood/status/1576884145787047936 https://twitter.com/ProfSueElliott/status/1576799836833140736) https://www.monash.edu/medicine/news/latest/2021-</p> |

| | | |
|---|--|--|
| | | articles/new-transfusion-dataset-a-first-step-in-nationally-streamlined-blood-use (tweet: https://twitter.com/Monash_SPHPM/status/1574609069549096961) |
| Outline of NTD project in the Blood Synergy Research Report | 2020-2021 Published May 2022 2022-2023 Published Nov 2023 | https://bloodsynergy.org/assets/Synergy-Research-Report.pdf https://bloodsynergy.org/assets/Synergy-Research-Report-2023.pdf |

5.2 Research Outcomes Planning

| ACTIVITY | DETAILS |
|--|---|
| Establishment of a monitoring framework (for project outcomes and impacts) | <p>Annual monitoring of the five indicators detailed in the DP708 Project Plan with impact categories: Scientific, Technology and Social & Societal commenced this year.</p> <p>A data management and monitoring policy with details regarding the monitoring framework has been developed.</p> |
| Inputs from research users in design of the infrastructure | <p>A Steering Committee (SC) has been established with members representing prehospital services, hospital, blood disease registries and research collaboratives to oversee the project outcomes and impacts. The SC members last met in May 2023 and were presented with the project updates and progress and asked for input.</p> |
| Partnerships with research translation specialists | <p>Close links and discussions are ongoing with clinicians through the participating registries which will ensure rapid translation of data into practice.</p> |
| Establishment of policies, systems and workflows to track uptake (of project outcomes) | <p>Policies, systems and workflows to track uptake of project outcomes are being furthered as data integration has started and will be updated in parallel with the data analysis and reporting.</p> |

6 LESSONS LEARNED

6.1 What Went Well?

The NTD leverages existing infrastructure in the form of the Australian and New Zealand Massive Transfusion Registry (ANZ-MTR) and the Transfusion Dataset (TD). The ANZ-MTR was an investment in research infrastructure established for over 10 years with approvals in place at more than 50 sites. The existing ethics and governance were amended to transition existing sites to the NTD. As the ethics was already under the National Mutual Acceptance (NMA) scheme, we were able to obtain ethics approval for the NTD relatively quickly. Getting governance amendments through at each existing ANZ-MTR site can be a slightly longer process but, since the start of funding, we have gained governance approvals from 21 hospitals and health services.

We were able to build on our relationships with national clinical and research networks to establish the NTD Data Partnership Group (DPG). This group includes members from Ambulance Victoria, Lifeblood and the Australian and New Zealand Intensive Care Society (ANZICS). The DPG has offered essential support and guidance in the governance of the NTD and has championed the establishment of the Dataset and the initial data linkages.

The Transfusion Research Unit, who manage the NTD, runs a variety of clinical registries and other research projects, working with large amounts of data on a daily basis. As such we had ready access to staff who are very experienced in research design and conduct, as well as data management.

6.2 What Could be Improved?

There is a high degree of variability in content and format between and across existing hospital and pathology datasets, which means that data harmonisation is labour-intensive and time-consuming. Whilst the challenges of data harmonisation were anticipated given our experience with the ANZ-MTR, the expanded dataset collected by the NTD also increased volume and complexity of the challenge. We're constantly working with the sites to review the data dictionaries, documents and processes to guide sites on how best to provide the data, with minimal additional work on their side.

Gaining access to hospital and health service data and databases is a lengthy process. Data are often housed on different platforms even within a single health service. A variety of stakeholders within the hospital and health service must be approached and engaged to perform the work of extracting the data. Additionally, hospital and lab staff, including pathology IT staff, are already extremely busy with their 'day jobs' so these projects are seen as an extra task or burden, and even where individual

scientific or IT staff are keen to be involved, managers are often wary of the additional workload. The NTD does offer payment for staff time at sites to perform the data extractions, and our data team are available to assist with linking siloed datasets.

Once the data from the hospitals and health services are received, a data harmonisation processes is undertaken prior to linking with prehospital and registry datasets. In this process, data are cleaned and allocated to the NTD fields. No hospital dataset is the same and, as such, this task has a highly manual component making it very time-consuming for our data team. In 2023 we were able to employ a Data Analyst who has worked on automating some of this harmonisation process to improve its efficiency.