Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Data Protection Impact Assessment Template

Research

Template Version Control

Reference	Date	Author	Comments
1.0	Oct 18	Jennifer Ryan	

DPIA Circulation

Name	Date	Reviewed/Consulted	Comments
Evelyn Fox DDPO	28/01/2020	Reviewed	*Reviewed and Advised - Actions included at end of DPIA*
	Click or tap to enter a date.		Some of the database (use by Firalis) is subject to a HRCDC application. As such amnesty is in place until

			declaration is made. No further action at this time.
Amendment made 10. 06. 2020			1. Ethics amendment given for HELICAL 2. Ethics amendment given for FAIRVASC
Amendment made 16. 06. 2020			1. Ethics amendment given for Decompress

Objective

The purpose of a data protection impact assessment (DPIA) is to assess and demonstrate compliance with data protection legislation. The DPIA also provides evidence that the risks to individuals have been considered and sufficient measures have been taken to protect those individuals. The DPIA should assess the activity to be carried out against all the principles of data protection and determine whether the processing of personal data is both necessary and proportionate or whether changes to the process or additional controls are required

Data Protection Checklist for Health Research

If your project relates to health research, then you must comply with the requirements of the Health Research Regulations 2018. You must carry out the following:

1. Obtain ethical approval for the health research by a research ethics committee. ☒
2. Identify and document the data controller, joint-controllers and data processors. ☒
3. Ensure relevant contractual arrangements are in place. ☐
4. Identify and document funding bodies. ☒
5. Identify third parties with whom data will be shared even if pseudonymised. ☒
6. Ensure all members of the research team have completed data protection training. ☒
7. Carry out a risk assessment of the data protection implications of the research. ☒
8. Carry out a DPIA if the research is a high risk to individuals or involves the use genetic data, monitoring of behaviors, large scale processing of sensitive data, use of the data for new purposes or the linking of datasets. ☒
9. Ensure you only use the minimum data necessary to carry out the research. ☒
10. Implement controls to ensure the integrity and accuracy of data and determine when the data has been altered, disclosed or erased and by whom. ☒
11. Implement security measures to protect the personal data e.g. device encryption. ☒
12. Ensure the data is archived, anonymized or destroyed when the research is completed. ☒
13. Ensure that participants are provided with sufficient information about the use of their personal data via participant information leaflets or the project website. ☒

14. Obtain explicit consent for the processing of personal data for the health research Including the screening of individuals for research purposes. ☒

Data Protection Impact Assessment Instructions

You should complete all the questions on the DPIA and forward it to the Data Protection Officer who will provide feedback on any risks identified and recommendations on the actions or controls needed to address those risks.
It is the responsibility of the project supervisor to ensure the required controls are put in place and to sign off on any risks arising from the processing.

The DPIA should be updated to reflect any material changes to the processing as the project or activity progresses.

DPIA Name: RKD (Rare Kidney Disease) Registry and Bioresource
Date: 10/06/2019
Owner: Professor Mark Little

1. Describe in detail the process, system or project to be assessed.

Name of Project: Rare Kidney Disease (RKD) Registry and Bioresource

The purpose of this study is to set up a central repository that can support future research aimed at improving the prevention, diagnosis and treatment of vasculitis.

There are two components of the project:
1) A registry containing patient healthcare data and demographics. This registry will be used to study outcomes for systemic vasculitis and associated kidney diseases.
2) A biobank/ bioresource containing stored biological samples (serum, urine, plasma and DNA samples) which is used alongside the registry data to study the disease in-depth and identify potential biomarkers for the disease by carrying out ethically approved research studies.

Primary Systemic small vessel Vasculitis (PSV) refers to a group of rare autoimmune multi-system disorders with an annual incidence of 15/million or 60-75 cases per year in Ireland and occurs primarily in adults with a median age of onset 62. This condition is accompanied by a wide range of medical complications which frequently presents a diagnostic dilemma as many clinicians will be unfamiliar with the condition, which delays diagnosis by 6-12 months. Consequently, organ destruction progresses such that 30% end up with end-stage kidney disease and intensive care unit level care is often required due to multi-organ dysfunction. Although, effective treatments are available, Systemic Vasculitis remains a largely fatal disease accompanied by a chronic relapsing condition with no cure that requires coordinated long-term multi-disciplinary input.

RKD is an initiative aimed at addressing this unmet need and comprises dedicated multi-disciplinary centres across the country that provide a coordinated care path from diagnosis to relapse and on to long term remission, and provide access to clinical trials for patients with PSV. To support high quality translational research this network has established the RKD Registry and Bioresource which has recruited 1900 cases, including 668 vasculitis patients, 315 age-matched healthy controls and 698 disease controls, with the aim to enrol all 900 vasculitis patients in Ireland, providing a rich longitudinal clinical database-which is linked to an extensive vasculitis biobank.

Eligible Vasculitis patients are identified by the lead study clinician at local hospital sites and recruited into the study by the local research team after obtaining informed consent. Every recruit is assigned a study number which is recorded on the patient medical record and consent form. Patient identifiable information is stored in local hospitals.

Biological samples (serum, proteomic grade plasma, urine, leucocytes, RNA and DNA) are processed and stored centrally at the TCD Biobank with archiving managed using industry standard Freezerworks software. All aliquots are bar coded and mapped to the freezer location. Every aliquot has a clearly documented provenance. Stored samples can be requested for research purposes by an academic or industry (commercial) lab by entering into a material transfer agreement (via tech transfer office) with TCD.

Linked coded clinical data are recorded (using the study number in place of name) by the research team in the RKD registry database, which is hosted in TCD (by IT Services) and secure access controlled.

The database system used for this study has been created using REDCap (previously 'Distiller' platform was used and managed by DCCR/CRDI - this has subsequently moved to REDCap, hosted by TCD IT Services). It has a web interface for building and managing online surveys and databases. The security and infrastructure of this database is managed by the IT Services department at TCD. Host server, main database server and the connection between the two is securely protected behind host and institutional firewall.

This project is organised by Professor Mark Little of Trinity College Dublin, in collaboration with Nephrologists and other specialists with an interest in Vasculitis from hospitals across Ireland.

Data controllers involved in the project are identified as follows:

Data controllers:

1) Trinity College Dublin (for RKD Biobank and Registry - coded samples and data)
Study Head: Prof. Mark Little
Database management and storage: TCD High Performance Computing
Lead Research Nurse: Jane Richardson (contracted via St James's Hospital/TCD clinical research facility; the assigned Lead research nurse may change throughout the course of the study)

Controllers of patient data - who share CODED data with RKD, are listed below.
(Note: in relation to the Checklist on Page 2 - item 3 - contractual arrangements -this is currently in progress. Data sharing agreements were not formally established at the start of this study as the relevant ethics committee and hospital approval was deemed adequate. The protocol and information leaflet is now undergoing substantial revision. Once the protocol amendment is finalised and approved by ethics, formal data sharing agreements will be established with each site - this document is in draft and pending protocol revision).

1) Tallaght University Hospital (for medical records of recruited participants)
Lead clinician: Prof. Mark Little
+ research nurse support (which may include TCD research nurse from the CRF at St James's Hospital)
See attached re ethics received for
a. HELICAL
b. FAIRVASC
c. DECOMPRESS

2) St James's Hospital (For medical records of recruited participants)
Lead clinician: Dr. Niall Conlon
+ research nurse support (which may include TCD research nurse from the CRF at St James's Hospital)

3) St Vincent's University Hospital, Department of Rheumatology (For medical records of recruited participants)
Lead clinician: Dr. Eamonn Molloy
+ research nurse support

4) Cork University Hospital, Department of Nephrology (For medical records of recruited participants)
Lead clinician: Dr. Michael Clarkson
+ research nurse support
Contract negotiation underway

5) University Hospital Galway, Department of Nephrology (For medical records of recruited participants)
Lead clinician: Prof Matthew Griffin
+ research nurse support

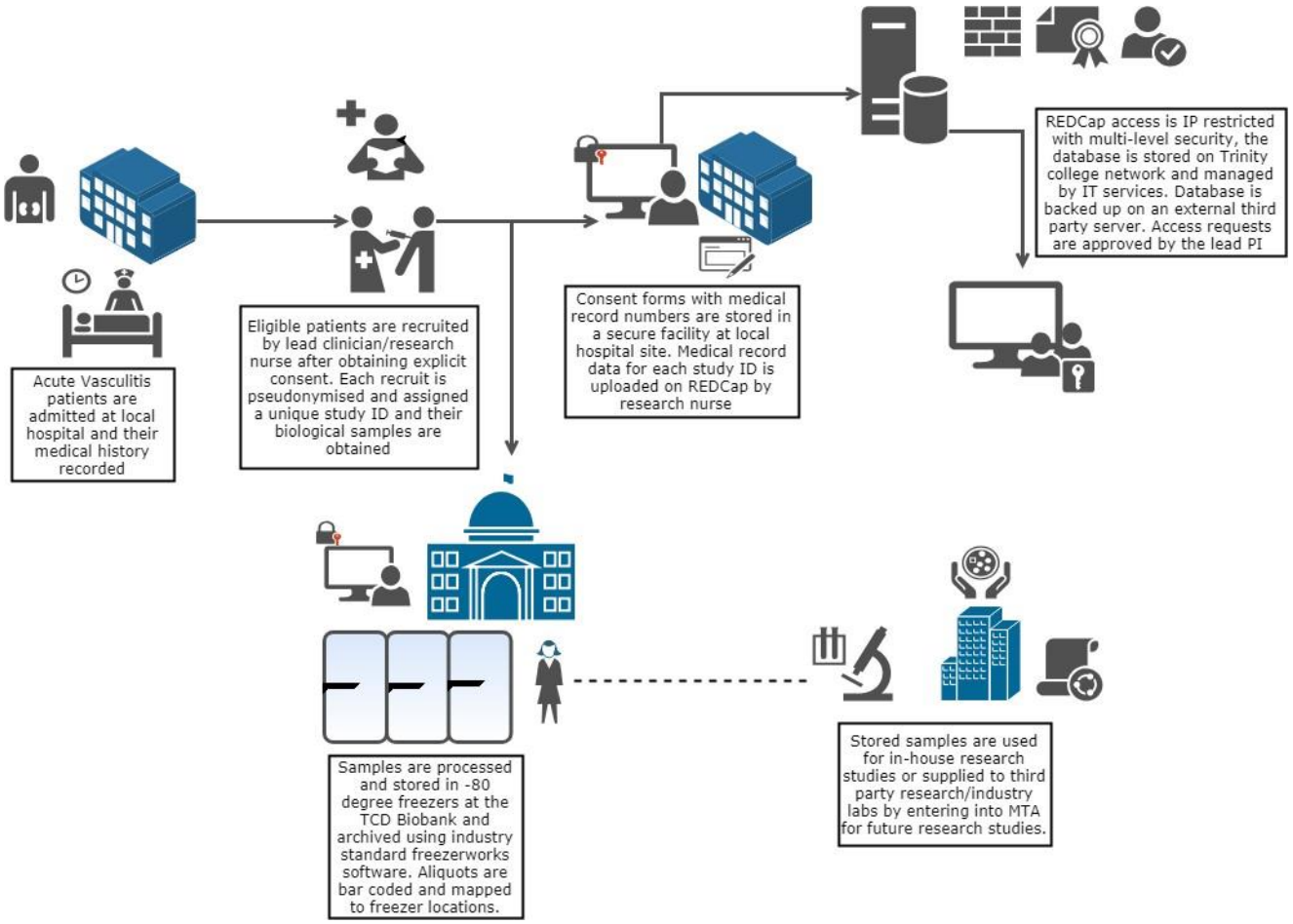
6) University Hospital Limerick, Department of Nephrology (For medical records of recruited participants)
Lead clinician: Dr Liam Casserly

7) Mater University Hospital, Department of Nephrology (For medical records of recruited participants)
Lead clinician: Prof Yvonne O'Meara

8) Beaumont Hospital, Department of Nephrology (For medical records of recruited participants)
Lead clinician: Prof Mark Little
+ research nurse support

Data processors:
TCD IT Service may subcontract a cloud storage vendor for data storage/ backup. If an international data centre is used, non EEA data processing may occur. TCD will only use services which are compliant with the European Data Protection Regulation and who satisfy the conditions for processing personal data outside the EEA. Appropriate data processing agreements will be used for third party vendors.

Data Flow:



2. List the types of personal data that will be collected, used, accessed or shared for the purpose of this activity?

This study involves collecting and processing sensitive personal data pertaining to study recruit demographics which includes gender, date of birth, ethnicity and employment status, and health data which includes detailed medical record describing disease clinical characteristics, biopsy characteristics, induction and maintenance treatment, details about co-morbidities, diagnosis and continuing treatment. Additionally, detailed disease assessment at subsequent encounters and, if applicable, transplant details including donor type and immunosuppressive medication. Only coded data (labelled with study id) is entered into the registry.

The study also involves collecting biological and genetic data which includes blood, urine, RNA and DNA samples which are centrally processed and stored in the TCD biobank until further investigations. Biopsy samples are stored at local hospital sites.

3. Have you ensured that you will only collect the minimum data that you need or that is necessary for the activity? Provide details

The case report form/data base fields are fixed in the RedCap database. These have been designed to collect the minimum data that is deemed to provide a complete picture of disease history to facilitate research of high impact. The study organiser has reviewed and tested this extensively throughout the duration of the project and is confident that only minimum required data is being collected. As this is a disease registry, the information on medical history and disease progression needs to be sufficiently extensive to support anticipated research at this time and in the future.

4. Health Research Only - Describe how you will you ensure that explicit consent is obtained for processing personal data for health research? Attach supporting documents.

If you intend to seek a public interest waiver from the Health Research Consent Declaration Committee please contact the Data Protection Officer at dataprotection @tcd.ie.

The attached participant information/consent form has been used for consenting participants to the research study. This document forms part of a wider consenting process, which includes a discussion of the study with a research team member, opportunity to ask questions, time to consider participation and explanation of the voluntary nature of the study and opportunity to withdraw at any stage. Explicit/ documented consent is obtained for all participants (signed and dated consent form - held at each respective site). This information leaflet and consent form has undergone a substantial revision to ensure full compliance with the requirements of GDPR and the Health Research Regulations.

5. Health Research – Describe how you will you ensure that the processing of the personal data is lawful?

Include whether you are relying on consent, statutory requirement, performance of a contract, public interest, vital interest or legitimate interest as a mechanism for processing. If using consent, then describe the consent process and attach supporting documents.

The lawful basis for processing data is:

Article 6(1)(e) - Public Interest and Article 9(2)(j) Scientific Research. Explicit consent is sought as an appropriate safeguard to rights of the data subject as mandated by the Health Research Regulations.

6. Does the research involve any of the following:

- **evaluating or predicting outcomes in individuals.**
- **decision making by automated means e.g. using algorithms.**
- **monitoring the behaviors of individuals.**
- **the surveillance of individuals, use of location or the use of biometric technology such as facial recognition?**

If so provide details and if so describe the impact to the individuals?

The RKD registry currently holds a large database of healthcare and medical information about 750 vasculitis patients, 315 age-matched healthy controls and 698 disease controls, linked to a Vasculitis biobank. The aim of this study is to provide a readily accessible resource of vasculitis patient samples to facilitate small or large scale research projects to identify novel disease biomarkers, develop better diagnostic tools and treatments for vasculitis patients. Immediate effects of this study cannot be addressed at the moment as it is a longitudinal study and the state of research and technological advancement will dictate the large scale impacts of this study.

112 of these recruits were included in the Firalis study, which is the subject of a HRCDC application.

However, as the data in this study is coded and only intended for research purposes, outcomes will not be used for monitoring individuals or making automated decisions that will affect individuals - this is not considered to constitute automated decision making or profiling.

7. Will the activity involve the use of special categories of data or sensitive personal data, if so provide details?

Sensitive personal data includes:
-Health & Genetic
-Racial & Ethnic
-Sexual

-Religious and Philosophical
-Political
-Biometric
-Trade Union

Special Categories of data include:
-Children's data
-Data relating to criminal convictions

Yes - Health, Genetic data and ethnic data - as part of the registry we will be collecting demographic data, date of birth, gender and detailed medical history.

As this is a biobank - samples can be used for a wide range of research studies. The research will look at the effects of family history, genetics and environmental impact on disease progression/ triggers etc. This will require genetic data. Pseudonymised samples and data (including genetic data) will be made available to other researchers (including with non-EEA research institutions) for further research.

1. The Firalis sub-study does not include genetic or ethnic data
2. HELICAL does not include genetic or ethnic data;
3. FAIRVASC does not include genetic data; it does include ethnic data:
4. DECOMPRESS does not include genetic data; it does include ethnic data

8. If using sensitive personal data which of the conditions for processing listed below will be used, provide details or state if unsure?

-Explicit consent
-Employment law
-Social protection
-protect the vital interests of an individual
-defence of legal claims
-medical diagnosis
-preventative or occupational medicine
-provision health or social care
-public health
- scientific or historical research
-archiving in the public interest
-legitimate activities of a foundation or not for profit with a philosophical, trade union on religious aim
-the data has previously been made public
-substantial public interest

Scientific research, Explicit consent

9. Will the data be shared with third parties including IT service providers, cloud services, sub-contractors etc.?

Provide details including information on the contractual arrangements in place and confirm what due diligence has been carried out.

The registry database (coded data only) will be shared with TCD IT Services Department as its security

and access permissions are managed by them. This database will be periodically backed up on an external third party server located in Dublin under management of IT Services Department. TCD IT Service may subcontract cloud storage vendor for data storage/ backup. If an international data centre is used, non-EEA data processing may occur. TCD will only use services which are compliant with the European Data Protection Regulation and who satisfy the conditions for processing personal data outside the EEA. Appropriate data processing agreements will be used for third party vendors.

Biobank samples will be shared with third party research/industry labs, such as Firalis, for research purposes with appropriate data and material sharing agreements in place (agreements are managed via the OCPKE office, TCD).

See Table at the end of this document, which is a living document and will be updated each time a MTA or MDTA is executed.

10. Will the data be transferred or stored outside the EEA at any point or placed with cloud providers that store data outside the EEA? Provide details.

Coded biobank samples and data may be transferred to research/industry labs outside the EEA for research purposes with standard material/ data transfer agreements in place.

11. If you are transferring personal data outside the EEA have you ensured that suitable conditions for transferring the data are in place? Provide details or state if unsure

These include :
-Adequate jurisdiction
-US Privacy Shield
-Standard Contract Clauses
-Binding Corporate Rules
-Authorisation from the Data Protection Commissioner

This will be in accordance with TCD policies on material and data sharing agreements. It is anticipated that standard contract clauses or US Privacy Shield will be used where there is no other inter-institutional agreement in place. For transfer to any non-EEA or commercial entity, legal support from TCD for the data sharing agreement will be sought.

12. Describe in detail the technical and organisational security measures which will be taken to protect personal data including but not limited to access controls, encryption, pseudonymisation etc.

Identifiable patient/control data is pseudonymised after recruitment by assigning a study ID; consent forms and recruitment log (which includes identifiable data) will be stored in a secure facility at the local hospital site under control of the site PI (depending on local institutional policies and approvals). The recruitment log will be stored in spreadsheet or database format on a password protected drive.

Pseudonymised data will be entered manually into the REDCap database, which will be mapped to a dedicated password protected computer using IP address. The database will be protected behind host and institutional firewall with access to dedicated personnel only. The database operates on multiple-access levels. Only the lead study PI and research nurse will have complete access to the database and permission

to modify, while other approved members will have limited access rights according to site and requirements; new access requests will be approved by the lead study PI and a log of approved users, with access privileges, will be maintained. New users are required to agree to a code of conduct (see below). A procedure is in place to remove users that have left the study.

Coded biological samples are processed and stored centrally at the TCD biobank and archived by the biobank technician using industry standard Freezerworks software (https://freezerworks.com/) to catalogue samples; only the biobank technician and Lead study PI have access to the software. This is hosted on a dedicated, password protected computer stored in a secure location with backup to a hard drive kept in a locked cabinet.

Code of conduct for new REDCap database users:
"I understand that as a XXXXXXXX at XXXXXX Hospital, you need read and write access to the RKD database on REDCap. This database gives you access to real clinical data about Irish vasculitis patients. Before we give you access to this data, you need to read the following and then confirm to me that you understand the issues and agree to the safeguards:

- ***Although the data under consideration is de-identified, due to its nature, in practice the data cannot be assumed to be anonymised.***
- ***Because of the rareness of the disease, factors such as clinic visit dates, unusual symptoms, etc., may allow specific patients' identities to be revealed if such data were known to third parties.***
- ***As such, further measures need to be undertaken to ensure in so far as possible that the possibility of the patients' privacy being compromised is minimised. We therefore require that:***
    - ***RKD data is saved, if this is necessary, on a password encrypted device.***
    - ***RKD data is not emailed to yourself or anyone, or stored on cloud services without being encrypted***
    - ***RKD data is not shared with anyone else, or discussed with anyone else.***
    - ***Demonstrations, reports and publications about the project will not display actual individual level patient data.***"

13. How long will the data be retained for and why? Provide details

We would like to consent for indefinite retention of data. The value of the samples in the biobank to research is implicitly linked to the data retained on the registry. In order to provide assurance that samples were collected with explicit consent, we will need to retain the consent forms - this will provide a link between the sample and the participant identity; hence data will be deemed 'pseudonymised' for the lifetime of the sample. As this study includes longitudinal follow-up of participants, it will be necessary to retain a link between the study code and individual at site level to allow for follow-up of disease progression. As this is a biobank, it is unknown at this stage what specific research projects will be performed and when - and it is currently unknown how long the biobank will remain active (this will depend partially on continued funding for the resource). Given the long-term value of the samples and associated data, we feel that potentially indefinite retention of data is justified.
We will be seeking approval for indefinite retention with each participating site ethics committee and DPO as part of the amendment.

1. HELICAL
   The samples and data shared with Firalis, KTH, ISGlobal, Medical University of Vienna and University of Glasgow will be retained by Firalis until the end of the HELICAL programme (31/12/2022), after which unused samples will be returned or destroyed and data will be irrevocably anonymised. This will allow completion of the EU-funded project.
2. FAIRVASC
   The data shared with accredited users of the FAIRVASC interface will be aggregated so that no individual level data will be shared. These reports will be held as long as required by the user.
3. DECOMPRESS
   DECOMPRESS data and samples will be managed in accordance with standard registry and biobank processes.

14. How will you notify participants about the data processing that will be carried out using their personal data? Provide details and attach a copy of the participant information leaflet if available.

This will be outlined in the information leaflet and consent form. It will not be possible to inform participants of processing that is carried out on individual results (this is also addressed in the consent form) as it is not known when such processing will occur and researchers will only receive coded or completed de-identified samples and data only and may analyse pooled data. Information on publications using the Registry and Biobank will be disseminated to participants via Tara open access publications repository http://www.tara.tcd.ie/.

15. What plans are in place for responding to a request from an individual in relation to their data protection rights?

These include:
-right of access;
-rectification;
-erasure;
-right to object to processing based on legitimate or public interest;
-right to data portability;
-right to object to profiling or making decisions about individuals by automated means.

Rights in relation to medical records will be the responsibility of each site and in accordance with hospital policy.

Rights in relation to personal data in the RKD Registry will be directed by the relevant study site PI (using study ID number) to the PI. Requests will be reviewed to establish whether or not they are reasonable and impact on the integrity of the study. Where possible, they will be honoured. A copy of data recorded on the registry for a study ID can be provided to the site PI in a portable format. Data can be erased and samples destroyed at the level of the Biobank and Registry only. Where samples have been shared with other researchers the PI can request that any unused samples are destroyed - if this is included in the MTA. Results that have been analysed cannot be deleted - this will be explained clearly in the study consent at the next revision.

16. What guidance and training will be provided to individuals involved in this project or activity to enable them to understand their data protection responsibilities?

All staff working in Trinity (study data controller) undergo mandatory GDPR training (online module). The study organiser (Mark Little) will ensure that members of the RKD research team have undergone any mandatory GDPR or health research data protection training that TCD provides.

Responsibility for training staff at each participating site remains with each data controller.

Feedback from Data Protection Officer

Summary of the key risks

Risk	Risk Rating	DPO comments and recommendations
Transparency	Low	*Consent and information leaflets (v.5) sufficiently explicit except for use of biopsy for future research, which has been advised to PI as of email on 14.01.20.*
Governance	Low	(Agreements need a bit of work - current template provided is sufficient for record keeping but no GDPR specific clauses re responsibilities of individual controller etc. Note that point 15 covers off process so this needs to be put into an Agreement. Processor agreement with Redcap also requires review by DPO's office (Rose Gaynor)
Combined Datasets	Medium	Data sets being combined across multisite in Ireland and EEA. Note that for Ireland, database is pseudonymised, and that for EEA – intention is that data will be anonymous to recipient
Click or tap here to enter text.		Click or tap here to enter text.
Click or tap here to enter text.		Click or tap here to enter text.
Click or tap here to enter text.		Click or tap here to enter text.

Changes or actions required based on DPO feedback

This section should be completed by the data owner based on the feedback from the DPO and the actions should be tracked to completion

Action	Owner	Comments
Consent will be amended in V. 6	PI	
Controller/Processor Agreement under Review	DPO's office	
FAIRVASC agreement under review	PI	
Click or tap here to enter text.		Click or tap here to enter text.

Table of Data or Material Transfers from RKD Registry

Date of Agreement	Purpose of sharing	Name of Institution or industry	Recipient location (Country)	List all samples/ variable of data shared	Data classification (Coded or anonymous to recipient)	Type of Agreement in place	Duration

Appendix 1

i~HD Data Sharing Asset Register

This template has been developed to capture and share information about data assets that are available for shared use within a collaborative initiative or as part of a federated network of data sources, for research. It focuses primarily on the governance aspects of data sharing.

The term data sharing is used here to refer to the access provided to external researchers and research organisations by the custodian of a data asset. This access might be granted through the direct provision of a relevant data set extract, by enabling on-site querying of the data asset, access through an intermediary safe haven repository or as part of a federated network that supports distributed querying.

This asset register aligns with the FAIR principles, which promotes that data assets should be:

Findable: easily discovered, normally through an online catalogue or internet searching
Accessible: there is relevant metadata to support understanding of the data asset
Interoperable: adopting standards for the data and for the descriptive metadata
Reusable: the terms of data reuse are transparently provided

However, this template elaborates on certain aspects that are needed to give information and assurance to potential research users about the legal basis (including GDPR compliance) and permissions (including ethical approvals) that are in place to permit data sharing. It does not go into detail on the data or metadata describing the suitability of the data for different kinds of research, since this is the role of other data cataloguing templates. It also does not seek to capture evidence of the GDPR compliance of the data asset itself, but only in relation to the reuse of the asset for research by external organisations.

Question	Asset response
Asset descriptors	
What is the name of the asset?	Rare Kidney Disease registry and biobank
What are the overall asset objectives (the purposes for holding the data)?	Research into pathogenesis, diagnosis and treatment of vasculitis
What type of cohort or population asset is it? *Options: Research study, Cohort, National registry, Healthcare provider registry, EHR extract*	Cohort
How long will the asset continue to be available for external research use?	No end date set
Organisational contacts	
What is the name of the organisation with legal responsibility for holding and maintaining the asset?	Trinity College Dublin
What kind of organisation is the asset owner? *Options: Academic, Governmental, Charity, Other not for profit, Industry, other entity*	Academic
How is the asset primarily funded?	Competitive grant funding
Administrative contact name	Dr Alan Kennedy
Administrative contact address	Trinity Health Kidney Centre, Lab/Office 1.06, Trinity Translational Medicine Institute, Trinity Centre for Health Sciences, St James' Hospital, Dublin 8, D08 W9RT, Ireland.
Administrative contact email	rkdbiobank@tcd.ie
Administrative contact phone	+353-1-896 2105
Scientific contact name	Prof Mark Little
Scientific contact address	Trinity Health Kidney Centre Room 1.07 Trinity Translational Medicine Institute Trinity College Dublin, the University of Dublin St James's Hospital campus Dublin D08 W9RT Ireland
Scientific contact email	mlittle@tcd.ie
Scientific contact phone	+353-1-896 2145
Technical contact / data manager contact name	As per admin contact
Technical contact / data manager contact address	As per admin contact
Technical contact / data manager contact email	As per admin contact
Technical contact / data manager contact phone	As per admin contact
URL of the asset website or Unit/Centre Web site	https://www.tcd.ie/medicine/thkc/research/rare.php
Data subject population	
What are the main inclusion characteristics of the asset population?	Diagnosis of **primary systemic vasculitis** **Healthy control** **Disease control** (mixture of chronic kidney disease, other autoimmune diseases)
What is the approximate number of subjects in the repository?	2000
What is the approximate number of subjects for whom data collection is ongoing	1000
What is the geographical area where are the data subjects located?	Republic of Ireland
What was the age range of your data subjects at the time of recruitment?	16-95
What are the overall time periods of the data held?	2012 to present
Are you recruiting new participants? *Options: Recruiting new participants, Continuing to collect data on existing participants, No active data collection, but continuing analysis/research, Frozen and held for data sharing*	Recruiting new participants
Which care setting data sources are incorporated? *Options: Birth and child health services, Primary care, Outpatient care, Inpatient care, Pharmacy, Management reporting, Billing, Disease registry, Health event registry, Procedure registry, Bio-Bank, Other*	Outpatient care, Inpatient care, Bio-Bank, Disease registry
Do you hold data about family or other people related to the subjects?	Yes (limited)
Data set and metadata	
What main categories of health data are held?	Diagnosis, treatment, disease activity, outcomes (death, kidney failure)
What kinds of genetic information are held?	None
What kinds of bio-samples are held?	Urine, serum, plasma, DNA, PBMC
Does the asset have a formally documented data dictionary?	Yes
In what data or database format are the data primarily held?	RedCap (cohort data) / Freezerworks (biobank data)

Question	Asset response
What data standards are adopted for organising, storing, managing or protecting the asset data sets?	In general, the asset adheres to principles of Good Pharmacoepidemiological Practices. However, due to resource scarcity, quality assurance of the entire dataset (particularly those sites outside Dublin) is patchy
Does the asset have a standard format or computing language in which data extraction queries are constructed on the data set?	Extractions can be conducted directly into R, CSV, XML and SPSS formats.
Does the asset have a process for documenting and saving locally-run queries for audit purposes and for potential future re-use?	Yes, incorporated into the RedCap structure
GDPR compliance	
Data Controller organisation details	Trinity College Dublin
Data processor organisation details (if applicable)	
Main purpose of processing	Research into pathogenesis, diagnosis and treatment of vasculitis
Legal basis for processing under GDPR Article 6	Article 6(1)(e) Processing is necessary for the performance of a task carried out in the public interest.
Legal basis for processing under GDPR Article 9 (if applicable)	Article 9(2)(j). Processing is necessary for scientific research purposes.
Geographic location of personal data (including pseudonym keys, if applicable)	Locked database on a restricted access server behind TCD firewall
Data Protection Impact Assessment (DPIA) status, with regard to data sharing	Reviewed and signed off by DPO
Are there any stipulations or obligations on data sharing arising from the DPIA?	No
Informed consent (if applicable)	
Has informed consent been given by the data subjects?	Yes
Is participant consent required for new research uses (e.g. new research topics)?	Not if covered by existing consent form (v4);
Is participant consent required for external organisations to reuse the asset data?	No
How does the asset custodian handle withdrawal requests by data subjects?	SOP for removing data from the database and returning to the individual (if desired). PIL includes a statement that data already used in research cannot be removed.
Approvals	
Which regulatory bodies have approved the collection and use of the data set?	Institutional ethics committees, TCD DPO
What additional approvals are required for external organisations to reuse the data?	Material and data sharing agreement
What research purposes have been approved for the asset?	Research into the pathogenesis, diagnosis and treatment of vasculitis
Is ethics board approval required for every additional study, only for new research areas, or is there broad approval for many kinds of research?	Chairman approval is required for each study. This usually has a 2-week turnaround.
Is ethics board approval required for new studies that only use an anonymous extract of the data?	Chairman approval is required for each study. This usually has a 2-week turnaround.
Are there any areas of research that are explicitly excluded?	For profit collaborations with industry require re-consent with V5 consent form.
Are there specific limitations on the time during which research can be done?	No
Is there a requirement for significant findings to be notified to data-subjects, and by whom?	Participants have the option of receiving information on significant findings. These are managed through a clinical encounter in vasculitis clinic or, if required, a clinical genetics clinic
How frequently and in what form must studies from external organisations report to the asset's internal board, ethics committee or other bodies?	No explicit requirement
What requirements are there for archiving or destruction of shared datasets?	No explicit requirement
Linkage and de-identification	
Are any standard demographic identifiers retained e.g. a national health number?	Only in recruitment log
Does the asset hold names, addresses and contact information of the data subjects?	Only in recruitment log
Does the asset hold the details of external data providers e.g. GP details?	Only in recruitment log
To which external data sources does this repository have linkage?	None
Are consistent pseudonymous IDs issued to permit longitudinal linkage of data sets?	Yes
What methods are applied to generate an anonymous data set for external users?	Manual removal of identifiers
Which variables are normally aggregated before sharing (e.g. by age-	None

Question	Asset response
banding)?	
Which variables are normally blurred before sharing?	Date of birth
Is the data filtered in any other way (e.g. removal of rare diseases or conditions)?	No
Are combinations of variables checked in case they may act as a pseudo-identifier?	No
Are small-number restrictions applied to query results?	No
Is there a tracking of serial queries to detect 'triangulation' attempts?	No
Data sharing processes	
Is there a published procedure for submitting a data sharing request?	Yes
Are any data access fees normally required from a data sharing party?	Cost recovery is usually requested
Are external requesters permitted to come to the asset site and view the data before finalising their data set request, or to perform queries themselves?	Yes
Are external parties permitted to access the data remotely, provided this is secure?	No
Does the asset have documented procedures for the methods by which data set extracts may be sent to data sharing third parties, including required safeguards?	Yes
Are there particular measures necessary in order to share genetic data?	Yes
Are there particular measures necessary in order to share biological samples?	Yes
Is a code of practice specified that parties reusing the data must adhere to?	Yes
Are there specified measures or standards for data protection and information security that the research user must adopt?	No
Does the asset keep a formal record of all disclosures made, including a reproducible definition of the data set that was disclosed?	Yes
Does the asset require data sharing collaborators to provide back any newly derived or quality-improved variables?	No
Is there a policy on how long after disclosure a collaborator may access or retain keep a data set?	No
Is there a publication policy or guidance indicating the kinds of authorship or acknowledgement that should be included in publications derived from the asset?	No

Entry completed by: Name Mark Little Role Principal Investigator Date 16/10/19