ARIA-TRE
GREEN BOOK

RESPIRATORY INNOVATION WALES

WIDI
Wales Institute of Digital Information

Aria-Tre Green Book

Respiratory Innovation Wales
WIDI Wales Institute of Digital Information

RESPIRATORY INNOVATION WALES

ISBN: 978-1-7394621-2-3

Glossary

TRE - Trusted Research Environment
GDPR - General Data Protection Regulation
Data Controller – Entity responsible for the storage and maintenance of data
Data Processor – Entity responsible for processing of user data
Data Subject – Entity that provides their data to the Data Controller
COPD – Chronic Obstructive Pulmonary Disease
Voice Cassette - A set speech pattern that forms the voice data
API – Application Programming Interface, intermediate process between two systems
AI – Artificial Intelligence
Kubernetes – an open source container orchestration platform for containerised applications
Kubeflow – open-source machine learning platform
Docker Container – an open source software development platform

Azure – Microsoft's public cloud computing platform
MySQL – an open-source relational database management system (RDBMS)
Air lock - builds a centralised repository of all executable files handled in the system
WAV file – Electronic file format for sound recordings
OECD – Organisation for Economic Cooperation and Development
GCP – Good Clinical Practice

1.0 Introduction: The Structure of ARIA-TRE Eco System

1.1 The Role of the Green Book

The ARIA-TRE green book is to support applications and applicants join the ecosystem and start researching voice analysis using ARIA-TRE.

This book outlines the documentation that is required and the process of application to ARIA-TRE. Whether you are a first time researcher or an experienced researcher, compliance with the five safe principles of trusted research environments will enable you to write research proposals that ensure participant data is correctly managed and ethically used.

1.2 The role of the Yellow book

Is to set out the vision and rationale for the development of the ARIA-TRE and provides the context for the innovation ecosystem it creates. The

yellow book outlines the processes and documentation that is required for membership of ARIA-TRE and ensure the 5s principles are being followed.

1.3 The role of the Blue book

Is to set out the technical details of the ARIA-TRE design and how they align to the values of the stakeholders using the system.

1.4 The Steering Committee

The ARIA-TRE Steering Committee will provide the following functions.

1. Governance and Audit
2. Membership Applications
3. Ethical oversight
4. Removal of membership
5. Dispute resolution
6. System development roadmap

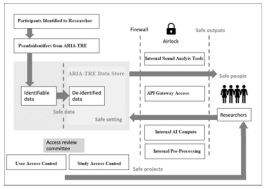
2.0 What ARIA-TRE can do to support you?

By following the processes within ARIA-TRE, your study will be constructed to ensure that the necessary data protection requirements are considered and put in place at the design stage of your study. ARIA-TRE can provide tools to aid your research and offers consistency between studies.

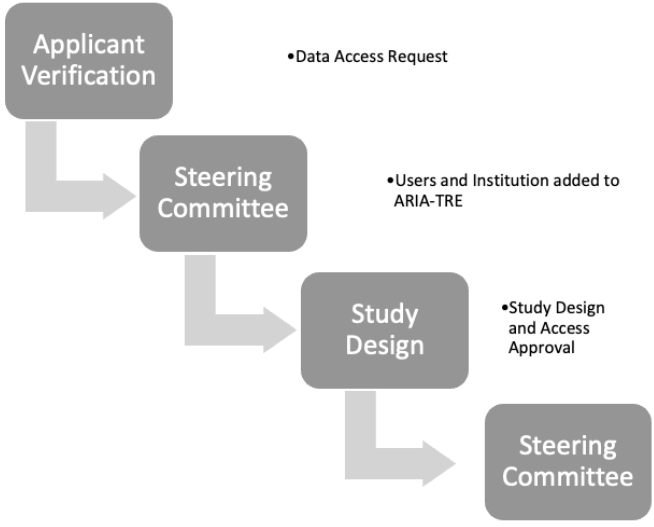
By applying the Five Safes, your study will be able to demonstrate:

- **Safe People** – Researchers must be trained and approved
- **Safe Projects** – Projects using our data are ethical and approved
- **Safe Settings** – Secure technology means data never leaves the safe ARIA-TRE
- **Safe Data** – Data used by researchers must be de-identified at source

- **Safe Outputs** – Outputs are checked so they cannot be used to identify an individual



3.0 Process Outline



4.0 Applicant application

The data access request must clearly describe all individuals that are requesting access to ARIA-TRE with their research justification outlined. Access will be granted at an institutional level only. The data access request requires a sponsor institution to be approved by the steering committee, the applicant must be a substantive employee of that institution. We cannot access applications for new institutions from honorary or visiting members of the institution.

The sponsor institution must declare that they understand their responsibilities as a data controller for their research studies and provide the name of the Data Protection Officer for their institution.

The sponsor institution must declare a senior researcher appropriately qualified to supervise research as the principal researcher representing their institutions interested within ARIA-TRE. The principle contact for the institution must be able to

act upon any concerns raised by the ARIA-TRE steering committee.

The sponsor institution must declare that all researchers have undertaken information governance training.

The sponsor institution must provide details of all collaborations, expressions of interest, declarations of funding or commercial interest, that may hold an interest or potential conflicts with access to ARIA-TRE.

Applications will be vetted prior to presentation to the ARIA-TRE steering committee, a schedule of committee meetings will be made available upon request.

4.1 ARIA-TRE Application review by steering committee.

The ARIA-TRE steering committee will review applications for compliance against the five safes principles of Trust Research Environments. The steering committee may decide that an application is approved with limitations to protect the security of ARIA-TRE. A provisional approval will require sponsor institution to provide an action plan to address issues highlighted by the steering committee within a specified time period. Membership of ARIA-TRE may be revoked upon failure to provide and action plan or failure to comply

with requirements specified by the steering committee.

The steering committee meets at set intervals per year, applications will only be accepted for review up to three weeks prior to the next steering committee meeting.

What is the purpose of your project?

Once your application to join the TRE has been approved, you are then able to submit a study design and data access request.

The *6th* principle of the ARIA-TRE is stakeholder engagement, based on the principles of Value Sensitive design. The ARIA-TRE platform development phase included a stakeholder analysis which clarified roles and responsibilities (see figure 3) and identified the range of values and expectations from the system.

While research ethics and data protection have slightly different focus, they share the requirement that the purpose for collecting and storing the data is clear to both researchers and participants, and this forms the basis of informed consent, which is a legal and ethical requirement. Undertaking a values-based stakeholder analysis can spark conversations around a common understanding of the purpose of the research project and how the data will be collected and used. This exercise can also be used to highlight

and resolve any tensions arising from differing perceptions of benefit / harm and to identify potential risks.

We will look for this approach in your application, from this we will be able to ensure that participants data shared with ARIA-TRE is acquired, stored, and processed, in an ethical manner. The data collected will retain auditable data of all uses and original consent information acquired. This transparency we define as a “Glass Door Vault” and ensures ethical use, privacy and security are the fundamental building blocks of ARIA-TRE and all studies that run within ARIA-TRE. Instructions for designing your data model for data collection or understanding data models stored within ARIA-TRE are defined within the Blue Book (Technical Specifications).

Stakeholder roles and responsibilities

Stakeholder	What is their Role?	What are their Responsibilities?	Governance & Audit
Participant	Sample providers	N/A	N/A
Researcher	Development and analysis	Data Controller Data Processor	HEI
RIW	System Host	Data Controller	Steering Committee
HEI	Ethics & Governance	Data Controller (Joint)	Steering Committee
Industry Partner(s)	Research Developer	Data Controller (Joint) Data Processor	Steering Committee
WIDI	Developer and architect	Software	Steering Committee
HDR	TRE Reference & Membership	TRE Specification	N/A

5.0 Study Design and Data Access Application

Completed study design applications must be submitted to the ARIA-TRE steering committee for approval. The committee will request a technical evaluation of the study if pre-processing and / or air lock applications are required.

The application process will review the follow items:

a. Consent collection
* Metadata template for consent
* Metadata example for consent

b. Participant Information Sheets
* In English Language
* In Country Native Language

c. Data Controller agreement

d. Study Metadata from Participant

* Metadata template for data collection
* Metadata example for data collection

e. Study Analysis Metadata
* Meta data template for analysis return
* Meta data example for analysis return

f. Study Pre-Processing Metadata
* Metadata template for pre-processing return
* Metadata example for pre-processing return

g. Study Pre-Processing Metadata
* Metadata template for pre-processing return
* Metadata example for pre-processing return

h. Study Cassette design

i. Outcomes export metadata template

j. Study Test Strategy

k. Participant well-being strategy

l. Safety Protocols

m. Pre-processing applications code review

n. Post-processing applications code review

o. AI model design

p. Participant data removal methodology

q. Trusted users accessing the study

r. Lay summary for using PseudoIdentifiers and ability to identify patients within their institution

s. Identification method for training and testing data with metadata template

The steering committee may decide that an application is approved with limitations to protect the security of ARIA-TRE. A provisional approval will require sponsor institution to provide an action plan to address issues highlighted by the steering committee within a specified time period. Study participation within ARIA-TRE may be revoked upon failure to provide and action plan or failure to comply with requirements specified by the steering committee.

6.0 Governance

Institutions are required to manage their study participation within their local rules and procedures. Joint data controller responsibility (data custodian) limits ARIA-TRE to data provided via metadata templates from qualifying studies. Voice data is protected in law as a special category data and ARIA-TRE and your institution acknowledge a lawful basis under Article 6 of the GDPR and a separate conditioning for processing under Article 9. Institutions are also required to meet additional conditions and safeguards by UK law, in schedule 1 of the DPA 2018.

The following consideration must be identified:

* Collection

    - Consent - Process belongs to the research project

- Designing the Voice Cassette – ARIA-TRE rules compliance
- Difference Tool

* Use and Disclosure

- How is the data going to be used
- Repurposing (consented branch not assumed)
- Disclosure who sees the data

* Retention

- Term of storage (given in the research protocol)
- Location of storage (ARIA-TRE Safe Location)
- Data Security
- Deletion methods ‘Right to Forget’
- Disaster Recovery

* Software Development

- Algorithm
- Sub-contracting
- Training
- Documentation
- Version control and deployment

* Study Management

- Risk review strategy
- Quality review process
- Complaints

ARIA-TRE does not store or accept Personable Identification Information about the study participant or information about the recruitment process used within the study. This information is the sole responsibility of the researcher's sponsoring institution and ARIA-TRE will not accept a data controller or processor role for this information. All transactional data between a research software platform and ARIA-TRE API gateways will be conducted via a pseudoidentifier provided either manually or electronically from ARIA-TRE when a participant is enrolled into ARIA-TRE as a new participant. ARIA-TRE allows the same participant who may have participated in multiple studies to exist within the TRE as a separate entity for each study.

The research institution is responsible for correlation between their research enrolment records and ARIA-TRE pseudoidentifiers used to identify participants within ARIA-TRE.

7.0 Special Category Data

The UK GDPR defines special category data as:

- personal data revealing racial or ethnic origin;
- personal data revealing political opinions;
- personal data revealing religious or philosophical beliefs;
- personal data revealing trade union membership;
- genetic data;
- biometric data (where used for identification purposes);
- data concerning health;
- data concerning a person's sex life; and
- data concerning a person's sexual orientation.

Special category data includes personal data **revealing or concerning** the above types of data. Therefore, if you have **inferred or guessed details** about someone which fall into one of the

above categories, this data may count as special category data. It depends on how certain that inference is, and whether you are deliberately drawing that inference.

Article 9 lists the conditions for processing special category data:

(a) Explicit consent
(b) Employment, social security and social protection (if authorised by law)
(c) Vital interests
(d) Not-for-profit bodies
(e) Made public by the data subject
(f) Legal claims or judicial acts
(g) Reasons of substantial public interest (with a basis in law)
(h) Health or social care (with a basis in law)
(i) Public health (with a basis in law)
(j) Archiving, research and statistics (with a basis in law)

If you are relying on conditions (b), (h), (i) or (j), you also need to meet the associated condition in UK law, set out in Part 1 of Schedule 1 of the DPA 2018.

If you are relying on the substantial public interest condition in Article 9(2)(g), you also need to meet one of 23 specific substantial public interest conditions set out in Part 2 of Schedule 1 of the DPA 2018. This is relevant for any studies that provide counselling during or after participation.

7.1 Article 22 EU GDPR

"Automated individual decision-making, including profiling"

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.
2. Paragraph 1 shall not apply if the decision:
3. is necessary for entering into, or performance of, a contract between the data subject and a data controller;
4. is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or
5. is based on the data subject's **explicit** consent.
6. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.

7. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

7.2 Checklist

The following checklist is required to be completed as part of the application process on behalf of the sponsor institution.

□ We have checked the processing of the special category data is necessary for the purpose we have identified and are satisfied there is no other reasonable and less intrusive way to achieve that purpose.

□ We have identified an Article 6 lawful basis for processing the special category data.

□ We have identified an appropriate Article 9 condition for processing the special category data.

□ Where required, we have also identified an appropriate DPA 2018 Schedule 1 condition.

□ We have documented which special categories of data we are processing.

□ Where required, we have an appropriate policy document in place.

□ We have considered whether we need to do a DPIA.

□ We include specific information about our processing of special category data in our privacy information for individuals.

□ If we use special category data for automated decision making (including profiling), we have checked we comply with Article 22.

□ We have considered whether the risks associated with our use of special category data affect our other obligations around data minimisation, security, and appointing Data Protection Officers (DPOs) and representatives.

8.0 Designing A Voice Cassette

The Voice Cassette (what you want people to say) is the choice of study. ARIA-TRE reserves the right to refuse voice cassettes that it considered reputationally damaging to ARIA-TRE or likely to bring research studies into dispute.

Examples of voice cassettes that may be refused are:

- Use of numbers
- Long passages of text i.e. Rainbow text
- Names
- Common password examples
- Offensive words
- Words that can't be translated into English
- Sounds or whistles

ARIA-TRE supports:

- Single voice cassette entry
- Multiple single voice cassette entry
- Timed repetitive voice cassette entry

9.0 Research Outcomes

Looking at how your work is shared supports the ARIA-TRE develop and increase the quality of research for all researchers.

ARIA-TRE support analysis and findings to be recorded against participants data. ARIA-TRE allows the sending of a metadata pre-processing and post-processing report back to the system as either private or public within ARIA-TRE.

Researchers utilising AI for model training are required to return to ARIA-TRE a unique identifier within their system that allows ARIA-TRE to record in the participant audit data that the participant data is existing within a trained AI model. Details of the data usage to be described in the metadata returned. Outputs from AI training analysis can be sent back to ARIA-TRE for individual participant data and study results for whole or partial training of studies.

Any modification of original metadata and voice cassette files is not permitted by studies. Participant ‘right to forget’ or consent revocation requests will be processed by ARIA-TRE steering committee on request.

All data removed from the secure application areas, must pass through our air lock security API gateway and must be pre-requested via your outcomes reporting metadata template.

ARIA-TRE is an air lock environment, any attempts to remove participant data in whole or part from the safe location storage will be considered a criminal activity and reported to the necessary bodies as such.