ARIA-TRE
YELLOW BOOK

RESPIRATORY INNOVATION WALES

WIDI
Wales Institute of Digital Information

Aria-Tre Yellow Book

Respiratory Innovation Wales

WIDI Wales Institute of Digital Information

RESPIRATORY INNOVATION WALES

ISBN: 978-1-7394621-3-0

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

Clinical Archiving – A system that stores patient information for the purpose of delivering healthcare

Functional Product – Software or Hardware offered to the market as a commercial product

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 The Vision for the ARIA-TRE?

We envisage ARIA-TRE being a collaborative innovation ecosystem that creates a secure and safe environment to study the use of voice in chromic respiratory disease. The focus will be on respiratory diseases such as asthma and COPD, however the ‘building blocks’ of the ARIA-TRE can be applied to any chronic condition where relevant.

We define an innovation ecosystem is a set of interconnected elements that work together to facilitate and promote the development of innovative ideas and products. It includes people, technology, knowledge, resources, and processes that all interact and influence one another in order to create an environment that encourages innovative thinking and product development. The goal of an innovation ecosystem is to foster collaboration and creativity among its members and to create a culture of innovation and experimentation.

Any organisation that wants to work and use any element of the ARIA-TRE ecosystem must commit to the following:

- Work to improve the health and wellbeing outcomes of people with chronic conditions.
- Adhere to strict data protection and information governance standards
- Embrace the collaborative, multiple organisation approach so that UK leads on the development of voice analysis.

The ARIA-TRE ecosystem doesn't seek to be a single tool or process for the use of voice related analysis but will provide the environment for clinical, academic and industry colleagues (as appropriate) to explore.

ARIA-TRE is a research tool to support researchers and developers at the initial stages of exploring voice as an innovative tool. The system does not support clinical archiving or functional products. Software being developed within the system will not be classified as a medical device or subject to medical device regulations.

2.0 The Structure of ARIA-TRE Eco System

1.1 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.

1.2 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.

1.3 The role of the Blue Book

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

3.0 Scope and Objectives of ARIA-TRE

3.1 Why is Voice Analysis Important?

Many people in the UK live with a chronic condition or disease. A chronic condition can be anything that affects how their body works for over a year or requires ongoing medical care. These people can have symptoms that they experience every day, which can limit and impact their daily lives. Often these symptoms can change over time and can have no cure, but can be managed with the right treatment.

Unfortunately, many people with chronic conditions also experience loneliness and reduced interaction with others as it affects their normal routine. This often means that when disease symptoms get worse, it only becomes apparent at a late stage. Naturally, people can detect differences in conditions that have effects on the voice when speaking to each other. This could be repeating certain words, a style or

manner of speech or sounds made when speaking. However, it can be difficult to accurately measure these changes in voice and detection relies on having people to talk to and be listened to.

The British Lung Foundation estimates that Chronic Obstructive Pulmonary Disease (COPD)costs the NHS £1.9 billion annually, the majority of which relates to individual's symptoms getting worse, so that they require treatment and support from health and care services and may need to go into hospital. If individuals could become aware of their symptoms worsening early on each time, this could help to reduce the need for this and help them to manage their COPD themselves. This could mean they could be treated before they become very ill and this could be more cost-effective for the NHS as a whole.

Advances in Information Technology (IT) and Artificial Intelligence (AI) provide new potential to develop clinical information from this approach Unusual breathing and speech patterns when talking or exercising or when the patient is stressed can be identified particularly when they are getting worse. Chronic respiratory diseases often manifest with abnormal breathing and speech patterns that become more pronounced under periods of physiological stress, such as exercise or talking. With increasing severity these are more evident and subjectively recognisable but can be difficult to accurately quantify.

Similar work has been undertaken on human speech breathing patterns in diseases such as Parkinson's and associated muscle wasting conditions, and following stroke, but these focus on breathing patterns whilst speaking. Current advances in speech generation, speech analysis and transcription now allowing breath timing, breathing patterns, voice generation (ease and fluency of speech) and language production to be analysed at the same time in a clinically meaningful way.

These other areas of work have tended to use word structure and formation in chronic diseases such as Parkinson's or muscle wasting conditions. There has been little investigation in the use of a combined approach and data collection across speech, voice generation pattern and breathing dynamics.

With the rapid expansion of different approaches and the variety of elements involved, a key limitation has been the lack of a single process or system that can capture, store and analyse voice files particularly across multiple organisations.

3.2 What is the strategic context?

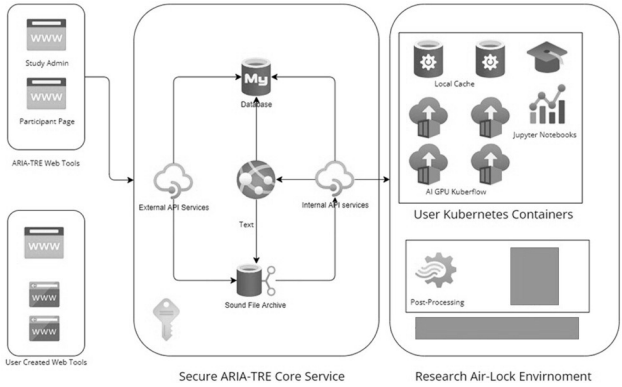
ARIA-TRE fits two of the key objectives in the UK Governments **Long Term NHS Plan**[1], to invest in supporting and treating lung conditions early to prevent 80,000 stays in hospital and to prevent illness and tackle health inequalities in the NHS. Health inequalities mean that some people access

and benefit from services less easily than others, for example because of where they live or for many other reasons, for example language difficulties or because they have a disability, and this can affect their long term health and well-being. This will meet many of the UK government's wider priorities to reduce the burden on the NHS by using an industrial focus to deliver these outcomes.

The **British Lung Foundation's Respiratory Health of the Nation**[2] project estimated that, in 2016, 1.2 million people in the UK—4.5% of all people aged over 40—were living with diagnosed COPD. Whilst the use of voice is a rapidly developing but new area of technological innovation, most of the research is in the early stages of development. However; given the number of people affected by COPD in the UK there is a clear clinical need to support and manage these conditions and other chronic repository diseases.

The **National AI Strategy**[3] being developed by **NHS England - Transformation Directorate** seeks to see AI development carried out in an ethical and efficient way. It highlights the need for a safe environment to support researchers through identifying questions for research through to undertaking and completing it. Through common data sets AI learning can be validated with efficiency and ethically whilst maintaining the confidence and trust of patients and the public generally.

3.4 What are the components of ARIA-TRE?



3.5 Why do we need ARIA –TRE?

ARIA-TRE is the first trusted research environment dedicated solely for voice files. It will be a repository (data bank) of population voice samples that commits to being a safe and secure research environment; founded on strong ethical principles, that will give patients and the public confidence in how their voice samples will be used.

Compliance with the General Data Protection Regulations (GDPR) and research ethics is essential for organisations looking to carry out research into patient bio-metric data. ARIA-TRE aims to provide a community of practice of researchers that can be assured that the ethical collection and storage of data is the foundation of the system, removing the

burden and barriers to individual biometric data research i.e. the automated process of recognising individuals by unique physical characteristics.

3.5.1 Safe & Secure

ARIA-TRE has been designed with the needs of researcher carefully balanced with the rights of patients and pubic to ensure that study participants can share their biometric data with confidence and trust. ARIA-TRE Is hosted in secure Microsoft Azure cloud services and is regularly audited for security compliance. All access to data is audited and has full traceability for the duration of the time it is held (data life cycle). Data entrusted into ARIA-TRE never leaves the system and all software development is carried out within a highly secure ‘Air Lock’ environment under the control of ARIA-TRE. No individual identifiable information is stored within ARIA-TRE and all data is encrypted when it is transferred to the system and whilst it is held.

Software development and AI learning is undertaken in a way that t only allows data to be shared within the ARIA-TRE system to minimise additional time and resource costs to researchers. ARIA-TRE is able to control all input and outputs from the system.

3.5.2 Principles of Trusted Research Environment (TRE)

ARIA-TRE has been built to comply with and exceed the principles of Health Data Research, HDRUK specifications.

'TREs provide researchers with a single location to access valuable datasets. Data and analytical tools are located in one place, similar to a secure reference library. Rather than extracts of individual level data being 'released', TREs provide access to a secure analytics environment ("safe setting") so researchers bring analysis (such as algorithms) to the data. Making data available through a TRE provides confidence to patients and the public that their personal health data is accessed securely, and their privacy is protected.'[4]

The Five Safes[5]

Safe data: data is treated to protect any confidentiality concerns.
Safe projects: research projects are approved by data owners for the public good.
Safe people: researchers are trained and authorised to use data safely.
Safe settings: a Secure Lab environment prevents unauthorised use.
Safe outputs: screened and approved outputs that are non-disclosive.

3.6 Ethical approach

3.6.1 Ethics guidelines for research

ARIA-TRE has adopted the principle and guidelines of the Organisation for Economic Cooperation and Development's (OECD) Principles and Guidelines for Access to Research Data from Public Funding [6] and Good Clinical Practice (GCP)[10]. Researchers using ARIA-TRE will be expected to have completed GCP training or equivalent prior to being accepted into the eco-system. The ARIA-TRE eco-system has been developed to offer co-production in research studies which means that all participants work together on the research on an equal basis. It offers best practice routes to supporting ethics approval processes within research institutes. ARIA-TRE will use the UKRIO(UKRI)[11] good practice checklist to support study applications.

3.6.2 Data protection

While research ethics are clearly defined, data protection is also a core issue.. GDPR gives individuals control over the way that data about them is collected and used, and it puts responsibility on those that collect and use data to ensure that it is properly protected. 'Data Protection by Design' is a fundamental principle of GDPR; but by addressing data protection issues at the design stages of the

Aria-TRE and each research project, both ethical considerations and legal requirements can be addressed. Having clear, agreed, well-defined and documented ‘purpose’ can provide transparency to both researchers and research participants enabling clear demonstration of consent both in terms of research ethics and data protection.

3.6.3 Data sharing practice

All new collaborators will sign ‘non-disclosure agreements’ to protect the contributions of all partners. Different types of collaborators are set out in section 6.0 below. How we approach data sharing is primarily focused on ensuring everything is open and clear to study participants regarding how their data is collected, stored and analysed. and ensuring they know which organisations are accessing the data. All research institutions will be publicly listed at all times to provide participants with a member status of all users of ARIA-TRE. There will be a publicly maintained table (matrix) of all studies and which research institutions and studies are accessing which data. At present we do not expect to share ARIA-TRE data with other organisations as it is unlikely that they would share the same ethical considerations as ARIA-TRE. We envisage data libraries will be made available within the ecosystem to support study building and this is detailed in a separate guide, the ‘Blue Book’ which contains more technical detail.

4.0 Data Protection By Design

4.1 General Data Protection Regulation (GDPR)

The GDPR's consists of a number of controls and measures to ensure that the privacy of individuals is respected. The regulations cover a number of areas, for example Data, Consent, Physical and Technical security controls, each of which must be addressed and documented for Aria to demonstrate compliance. For example, what personally identifiable information will be stored and what personal health information? What data is required to perform the research and what is needed to meet data protection requirements? What classification of data will be collected and processed? Voice data, as a biometric identifier comes under a GDPR 'special' category and it may give rise to higher ethical risk. The GDPR controls for the Aria TRE are listed in a separate guide, the 'Green Book'.

4.2 Value Sensitive Design

Value Sensitive Design recognises that while no technology is inherently good or bad, its use can bring harm as well as benefits. The approach requires initial discussion around stakeholder values; i.e., what is important to each of the individuals involved in the research. These values can be categorised loosely into ‘benefits’ and ‘harm’, and it is likely that there could be differences of opinion here where a perceived benefit to one stakeholder may seem potentially harmful to another. Despite the upfront costs involved in a stakeholder analysis (a process of basically finding out the views of key people involved), the benefits are many; the purpose of the system and data use can be better defined, expectations and concerns can be better managed, and trust can be enhanced.

Development of research studies within ARIA-TRE will focus on the researcher understanding the value of the data to the research participant and how using that data impacts them. Data collection consent must explicitly describe how data is going to used, why it is needed and if beneficially to other research how this data could be shared with other research studies or organisations. The design of a study will ensure that the participant has control over their data and suitably informed how they can remove their data or view how it's been used.

4.3 Responsibilities Matrix

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 The Role of Respiratory Innovation Wales

5.1 Data Controller

“data controller” means a person who (either alone or jointly or in common with other persons) determines the purposes for which and the manner in which any personal data are, or are to be processed. [7]

Respiratory Innovation Wales retains the data controller function of ARIA-TRE for the data stored within the central database, data “blob” storage or data lake and associated operational logs files generated by internal cloud services.

5.2 Data processor

“data processor”, in relation to personal data, means any person (other than an employee of the data controller) who processes the data on behalf of the data controller. [7]

Respiratory Innovation Wales, will act as a co-data processor, maintaining responsibility for study management data and delegating responsibility for processing of collected data to the individual research organisations. The Aria Eco System Steering committee will ensure that necessary governance arrangements are in place during application processes.

5.3 Processing

“processing”, in relation to information or data means obtaining, recording or holding the information or data or carrying out any operation or set of operations on the information or data, including— a) organisation, adaptation or alteration of the information or data, b) retrieval, consultation or use of the information or data, c) disclosure of the information or data by transmission, dissemination or otherwise making available, or Data controllers and data processors, d) alignment, combination, blocking, erasure or destruction of the information or data. [7]

Respiratory Innovation Wales will hold responsibility for the processing of internal operational data required to maintain a functional system with compliance to the GDPR. ARIA-TRE will not process person identifiable information within or externally linked to the participants of studies.

ARIA Eco-system Steering committee will ensure that governance is in place for all research organisations

that are processing data within secure containers and they maintain suitable governance arrangements validated during their application process.

Research institutions will hold responsibility of processing research participant information and maintaining a record of unique identifiers for their own studies. Respiratory Innovation Wales will accept processing requests from researchers based on a unique identifier subject to following processes outlined in the green book.

5.4 Future Sustainability of ARIA

RIW will manage the ongoing sustainable management of the ARIA-TRE ecosystem. Researchers will be required to fund their research within the ARIA-TRE system. The green book will provide details of rates of access and services provided.

6.0 Collaborative partners

6.1 Academic Partners

Academic partners will form part of the steering committee to assist in the continuing development of ARIA-TRE, maintaining the academic relevance of the research and supporting funding applications.

6.2 Industry Partners

Industry partners are important to the ARIA-TRE eco-system and provide opportunities for academic research to be scaled up to commercial products. Industry partners are key to continuing funding and the re-licencing of the ARIA-TRE technology within commercial products as an exemplar of best practice for voice data storage.

6.3 Technical Partners

Respiratory Innovation Wales will utilise technology partners to support the hosting of ARIA-TRE and ensure that the system remains current and protected from threats. Technical partners will assist in the development of DRPs and system upgrades as we evolve.

7.0 References

1. https://www.longtermplan.nhs.uk/
2. https://statistics.blf.org.uk/lung-disease-uk-big-picture
3. https://transform.england.nhs.uk/ai-lab/ai-lab-programmes/the-national-strategy-for-ai-in-health-and-social-care/
4. https://www.hdruk.ac.uk/news/new-principles-published-to-improve-public-confidence-in-access-and-use-of-data-for-health-research-through-trusted-research-environments/
5. https://htn.co.uk/2021/12/14/uk-health-data-alliance-publishes-principles-for-forming-trusted-research-environments/
6. https://www.oecd.org/sti/inno/38500813.pdf
7. https://ico.org.uk/media/for-organisations/documents/1546/data-controllers-and-data-processors-dp-guidance.pdf
8. https://www.snomed.org/snomed-ct/five-step-briefing

9. https://www.nhsbsa.nhs.uk/pharmacies-gp-practices-and-appliance-contractors/dictionary-medicines-and-devices-dmd
10. https://www.nihr.ac.uk/health-and-care-professionals/learning-and-support/good-clinical-practice.htm
11. https://www.ukri.org/councils/esrc/guidance-for-applicants/research-ethics-guidance/our-policy-and-guidelines-for-good-research-conduct/#:~:text=We%20have%20a%20duty%20to,maintain%20these%20standards%20is%20unacceptable.