
Ulster
University

Ulster University

Prof Michaela Black

RAISE: Responsible AI iS an Enabler

Professor of Artificial Intelligence

School of Computing, Engineering and Intelligent Systems

Intelligent Systems Research Centre (ISRC)

@mmblack https://linktr.ee/sceis

Artificial Intelligence (AI)

Responsible AI Institute (RAII) use OECD definition:

"an AI System as a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments."

OECD: Organisation for Economic Co-operation and Development

What is Responsible AI?

So what is the difference between *Responsible, Ethical, and Trustworthy AI*?

Values-driven responsible actions taken to mitigate harm to people and the planet

The term "*Trustworthy AI*" is most often used to reference the technical implementation of AI, mainly on ensuring fairness through:

- the avoidance, detection and mitigation of bias and
- ensuring AI models are transparent and explainable

Essential part of Responsible AI: organizations define their own *AI ethics principles* and make these transparent to their employees and customers

Responsible AI operates under a framework of responsibility, trustworthiness, and ethics

RAII: *Responsible AI, Ethical AI, or Trustworthy AI all relate to the framework and principles behind the design, development, and implementation of AI systems in a manner that benefits individuals, society, and businesses while reinforcing human centricity and societal value*

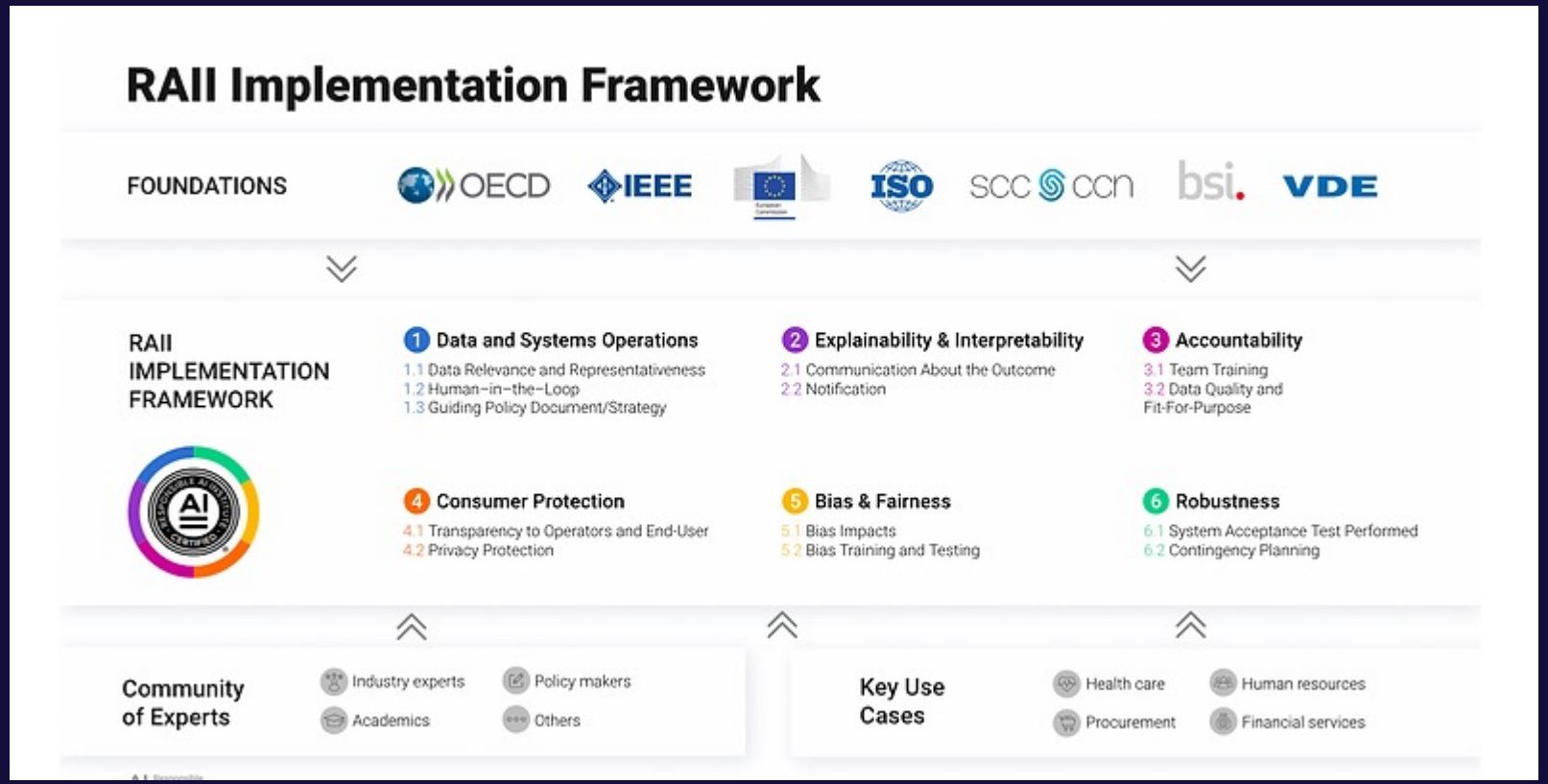
Artificial General Intelligence (AGI) Brain-inspired AI Neuroscience

Artificial General Intelligence (AGI) has been a long-standing goal of humanity, with the aim of creating machines capable of performing any intellectual task that humans can do

Brain-inspired AI is a field that has emerged from this endeavour, combining insights from neuroscience, psychology, and computer science to develop more efficient and powerful AI systems

Neuroscience is the study of the nervous system from structure to function, development to degeneration, in health and in disease

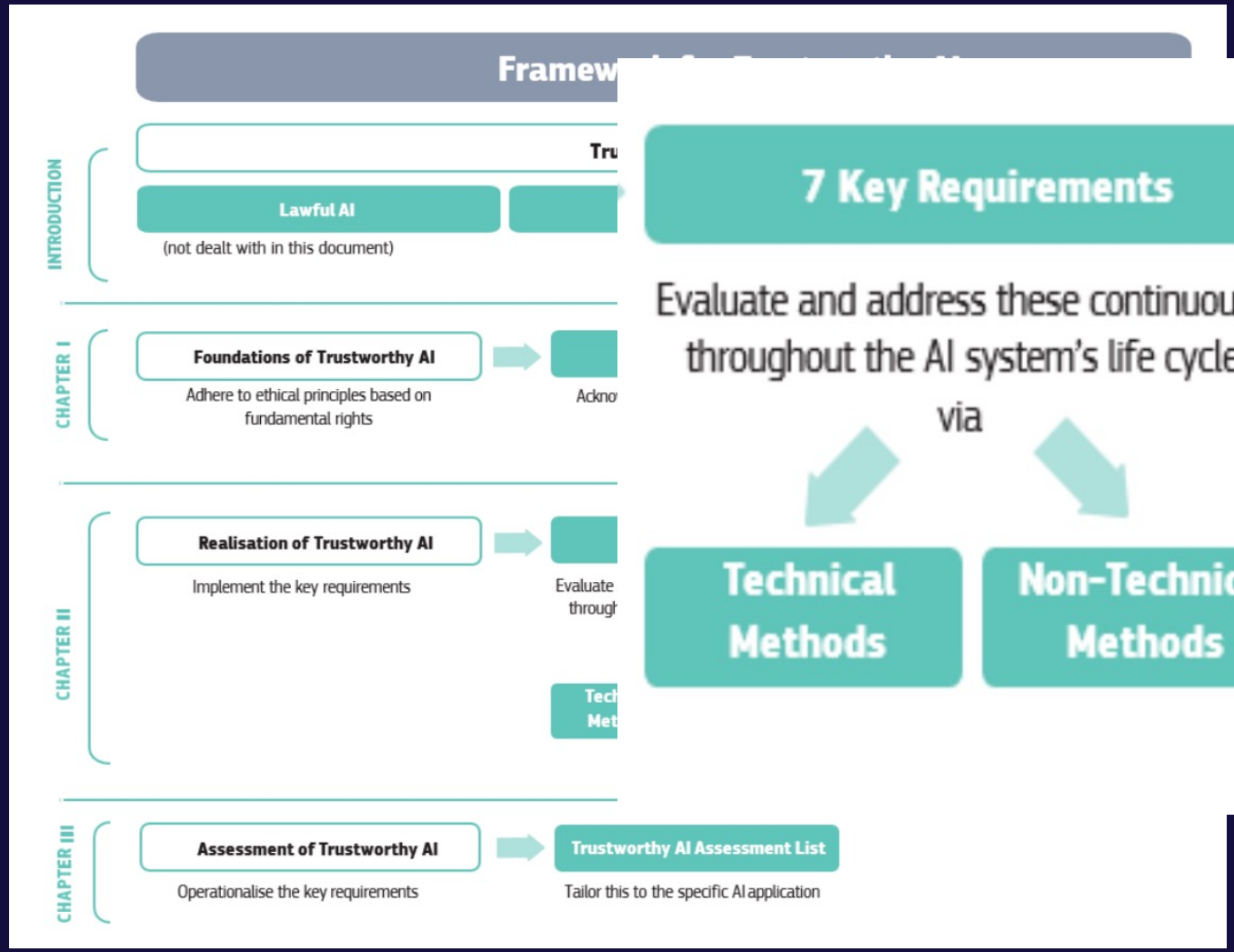
Responsible AI
Example Framework - RAII Implementation

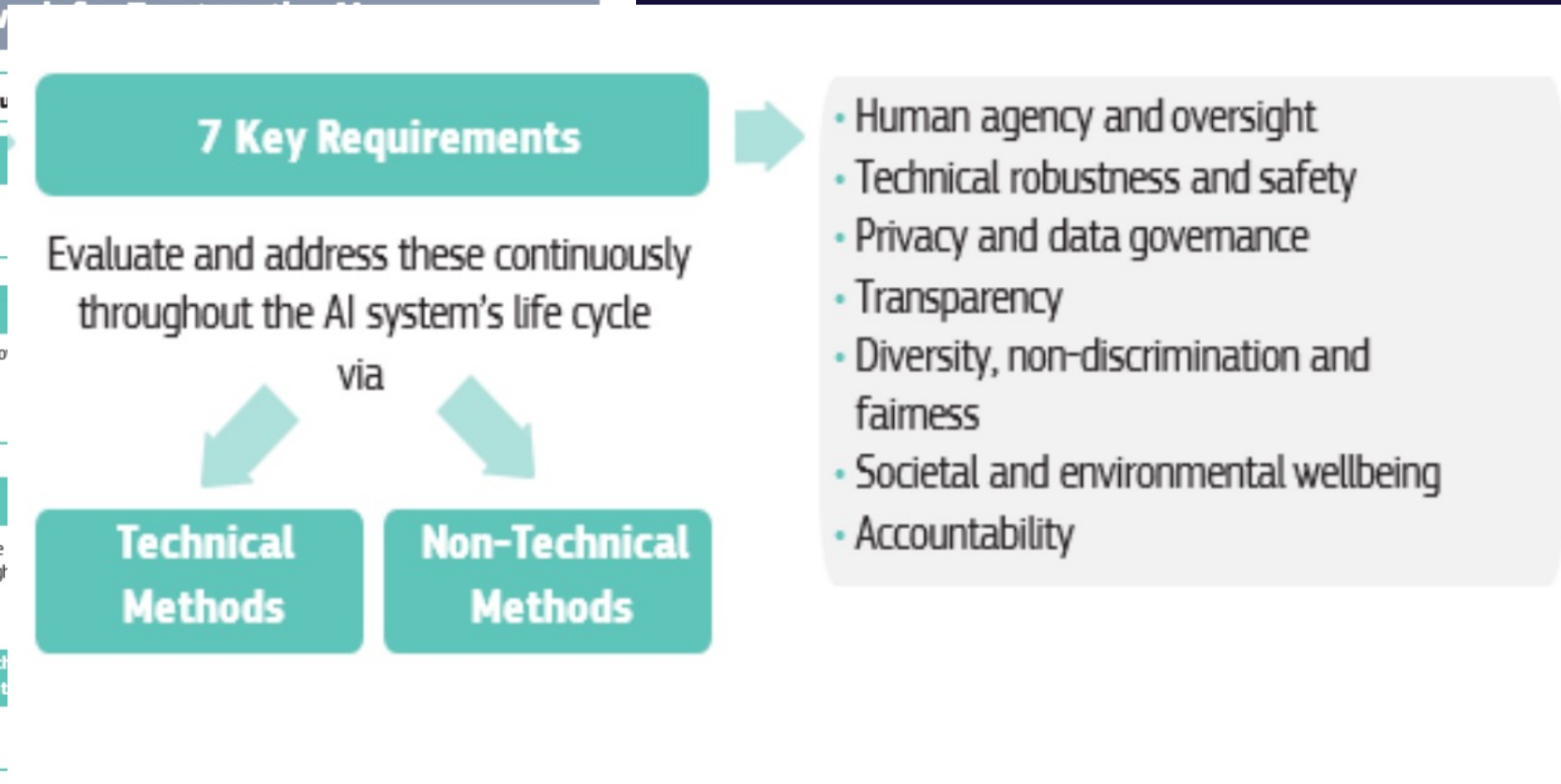


- An AI system might be designed to *process consumer data* to personalize ads, while the *responsible* design of that same AI system would have safeguards in place
- Such safeguards might provide: a notification to end-users about how and when their data is used; providing users opportunities to understand why the AI system is used; and providing avenues for redress in cases where the system goes wrong

https://20965052.fs1.hubspotusercontent-na1.net/hubfs/20965052/RAII%20Certification%20White%20Paper.pdf

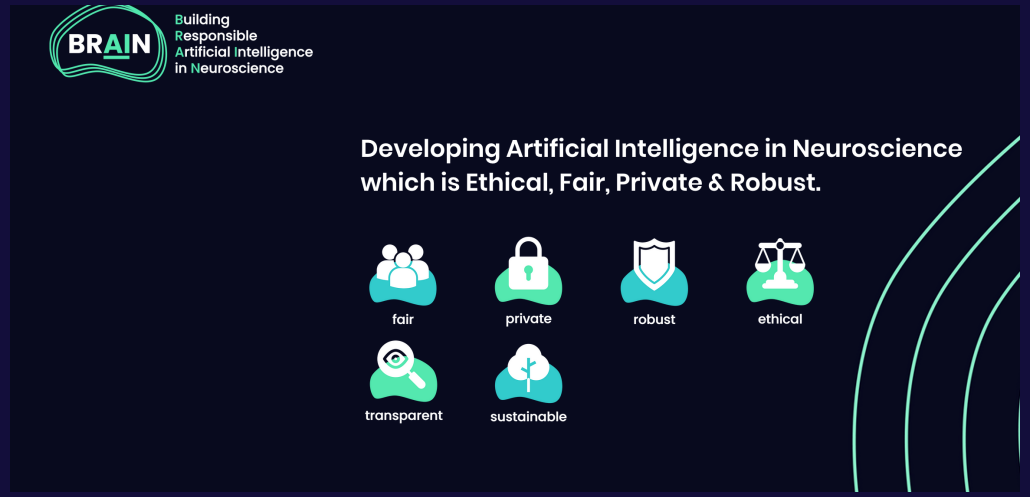
European Commission: Guidelines for Trustworthy Artificial Intelligence





https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Building Responsible AI in Neuroscience



Robust Models:

AI models in neuroscience are often trained on highly protected data collections which are accessed either via hospitals, universities or Trusted Research Environments (TRE's)

When models are released from these environments, then they could be susceptible to attack and therefore potentially revealing disclosive patient information

This is why it is important that we ensure peoples data is kept private within AI.

Fair and Unbiased AI:

Diversity in training data is crucial to ensuring that AI models are developed which do not discriminate. AI models developed in the UK mainly focus on data collections from white populations in Europe, leaving behind those ethnicities which are underrepresented

This is why we need to ensure that data is available from multiple ethnicities to build better representative AI

https://www.brain-research.co.uk/

Building Responsible AI in Neuroscience

Sustainability:

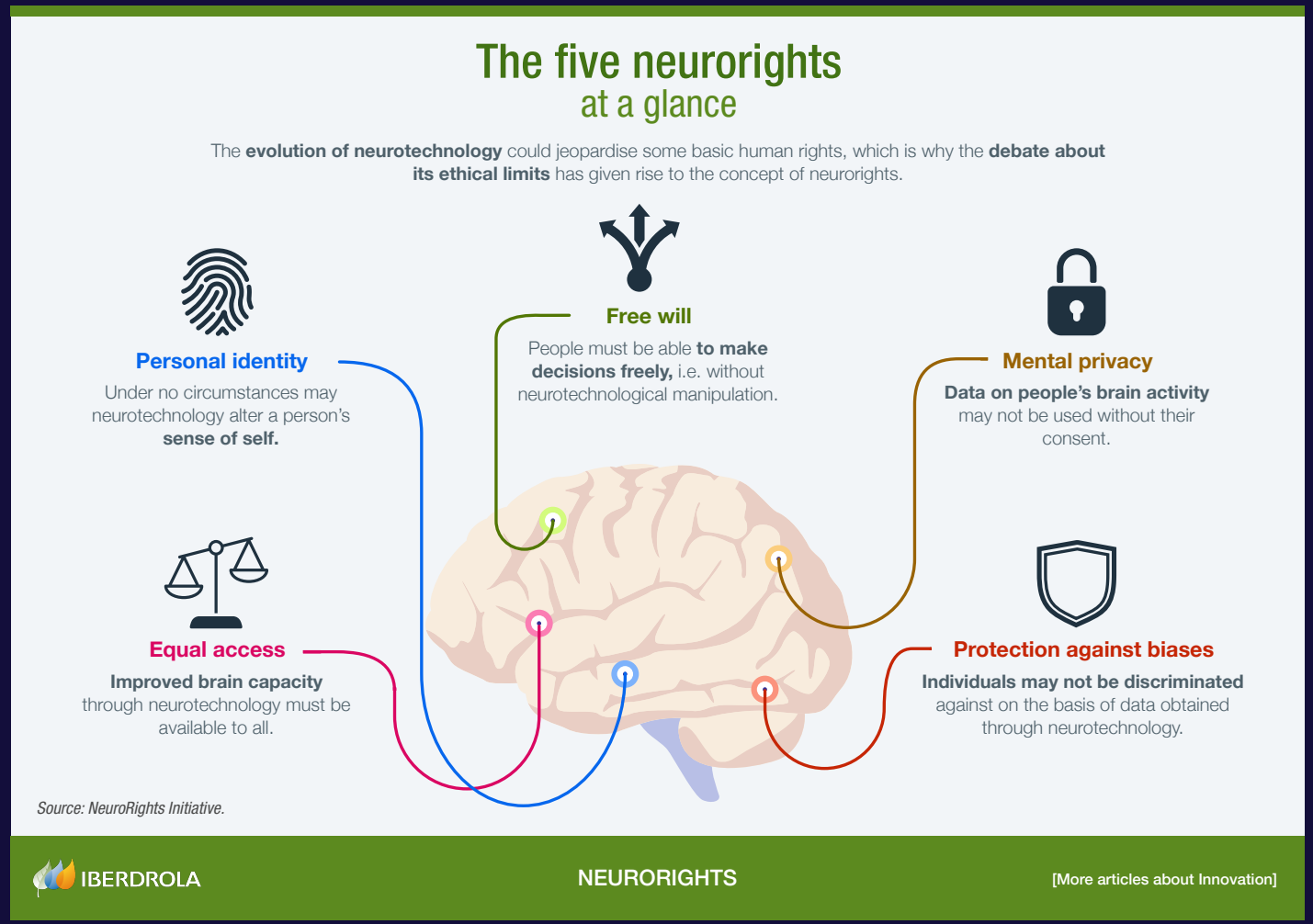
Processing large complex data such as neuroimaging and genomics can require lots of computational resources to generate the processed data

Not only this but the training and development of AI often requires the use of HPC clusters which can have a comparatively high environmental impact

Incorporating sustainability in neuroscience research can reduce this impact

https://www.brain-research.co.uk/

Neurorights



https://www.iberdrola.com/documents/20125/42616/Infographic_Five_Neurorights.pdf

Sample of NI AI Initiatives Led by CEBE

- Responsible AI Adoption within **AICC (AI Collaboration Centre)** – (CEBE – LoO)
  - Talent Pipeline and Future Workforce
  - Research and Innovation Services
  - Community and Collaboration
  - Data Accessibility, Governance and Ethics - dedicated role for AI Ethics Policy Advisor
  - Co-creation of a Responsible Governance & Ethics module for industry (SCEIS & Legal Innovation Centre)

- **engage tool** (AIRC & ISRC): electronic Town Meeting 'engagement', collecting information on the discussion topics electronically enabling the stakeholders to participate in debates and express themselves individually on key topics/issues
  - the engage e-participation service solution and methodological approach is a proven effective platform for gaining strategic consensus on a given topic using real-time digital technologies
  - cocreation of AI user stakeholder needs in collaborative R&D projects as well as in strategy development including AI for NI with DEL, Matrix & HSC

engage

Ulster University

AIM4Health
Artificial Intelligence approaches to addressing Mental Health inequalities in Ireland

NORTH SOUTH RESEARCH PROGRAMME
HEA HIGHER EDUCATION AUTHORITY

Comhairle Contae Fhine Gall
Fingal County Council

Project Description & Investigators

Artificial Intelligence approaches to addressing Mental Health inequalities in Ireland through improved diet and lifestyle: an interdisciplinary North-South investigation of the TUDA cohort integrating nutrition, environmental science and data analytics.

- Prof Michaela Black, Dr Debbie Rankin, Prof Jonathan Wallace, Prof Adrian Moore, Prof Helene McNulty, Dr Catherine Hughes, Dr Leane Hoey, Dr Anne Molly, Dr Mimi Zhang, Dr James Ng

Problem and Solution

Depression, often accompanied by anxiety, affects an estimated 10% of the Irish population and the cost of treatment is estimated at €4.7 billion annually in Ireland and £7.5 billion in the UK. Mental disorders pose serious health, economic and societal challenges, therefore addressing inequalities in older age is an urgent public health priority.

A multidisciplinary team from North and South, incorporating experts in Computing, Nutrition, Health, and GIS technologies to analyse multidisciplinary health data from North-South Trinity-Ulster-Department of Agriculture (TUDA) study with cohort of over 5000+ older persons. The health data include *clinical, nutritional, blood biomarker, demographic, and geographic* parameters. The project aims to use sophisticated AI and GeoAI techniques for analytics methodologies to explore trends and features within the cohort database amalgamating this multifaceted data and produce coordinated outcomes that address mental health issues in older persons.

Project Partners including Public Service Stakeholders

- Ulster University (Computing, Geography & NICHE), Trinity College Dublin (Mathematics & Medicine), Clinicians from St James Dublin

AI used within Project

- AI and GeoAI techniques for analytics

Impact to Public Service

The outcomes will identify key risk factors and novel solutions for incorporating into public health strategies aimed at promoting better mental health in older people. The risk factors identified from this project will inform dietary and lifestyle interventions to promote health and wellness in our ageing population, North and South. Engagement with PPI in North and South.

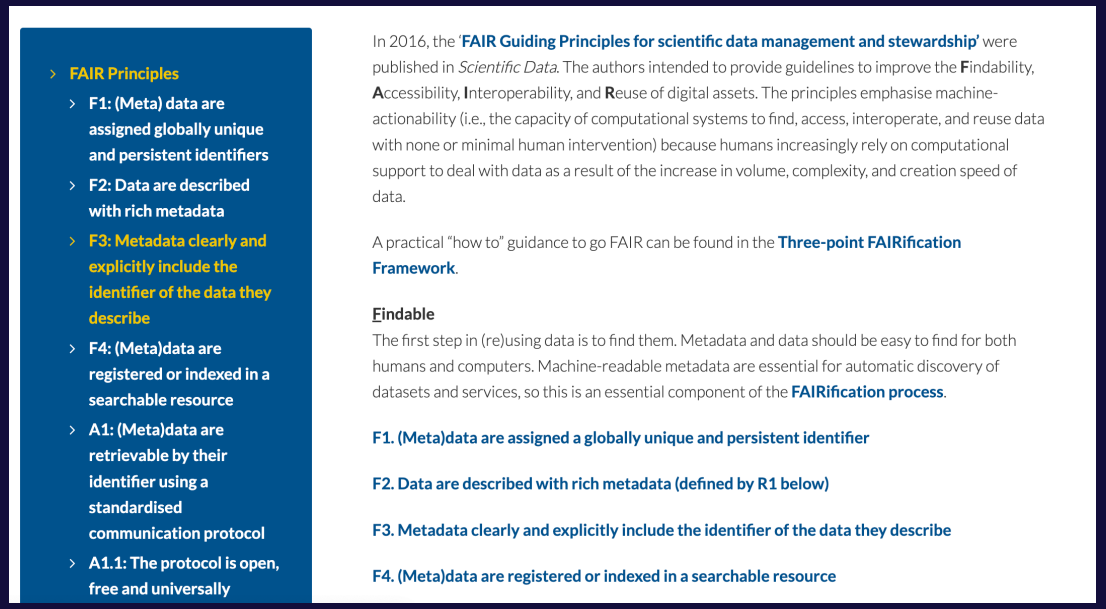
Responsible & Trustworthy AI Governance

- Policy Statement on Ensuring Research Integrity in Ireland Developed by National Research Integrity Forum

- FAIR (Findable, Accessible, Interoperable and Reusable) data principles in Ireland

https://www.iua.ie/wp-content/uploads/2019/08/IUA_Research_Integrity_in_Ireland_Report_2019.pdf

https://www.go-fair.org/fair-principles/



Standardised Architecture for Trusted Research Environments – SATRE

Project Description including investigators

- SATRE will compare openly available UK **trusted research environments (TREs)** hosting health, manufacturing, commercial, science and humanities data and bring them into alignment with a standardised TRE reference architecture
- Dr Dermot Kerr, Prof Sonya Coleman, Dr Justin Quinn

Problem and Solution

- The need for **TREs** is clear: Personal or sensitive data which have been collected for operational, commercial or governmental reasons need to be managed securely and safely for research use in an environment that encourages best practices. TREs are designed to enable only authorised projects and researchers access to sensitive data whilst minimising risk of data release or exposure.

Project Partners including Public Service Stakeholders

- University of Dundee, The Turing Institute, Ulster, UCL, Health Data Research UK (HDRUK), Research Data Scotland (RDS)

AI used within Project

- The TRE is used for secure hosting of data and secure processing using AI techniques

Engagement with Public Services

- Online focus groups have already been carried out and this is planned at various intervals throughout the project

Impact to Public Service

- SATRE outputs will be an informed reference TRE technical specification, a collection of educational media and detailed reports; all supporting DARE's aim of a national research data infrastructure supporting public services

Post-Brain Injury Virtual Driving Performance Assessment

- Spatial Computing and Neurotechnology Innovation Hub (**SCAN-iHub**), ISRC
- Higher Education Research Council (HERC), DfE NI
- **Traumatic Brain Injury (TBI)**:10 million people affected annually; an increasing cause of death and disability globally, particularly in lower-middle income countries.; commonest cause of death and disability in under 40s in high-income countries; global costs of $400 billion a year (0.5% of annual global output)
- Development of an effective **fitness-to-drive assessment** enabling clinicians to **support** individuals to **return to driving** following brain injury when the individual poses the necessary skills.
- ***AI*** *used to develop a* ***performance metric*** *to obtain ground truth from healthy users and flag at-risk behaviours in secure environment.*
- *Impact to Public Service: Assessment enables:*
  - *1. post-TBI rehab*
  - *2. healthy ageing*
  - *3. potential alternative driving assessment for young/healthy/high risk drivers*

Post-Brain Injury Virtual Driving Performance Assessment

Responsible & Trustworthy AI

Transparency: through **Documentation** re. development, purpose, training data, etc.

Robustness & Safety: through **Monitoring**: in real-world (virtual) applications checking for unexpected behaviours.

Fairness: through **Diverse Training Data:** Ensuring diverse set of inputs.

Privacy & Data Protection: through **Data Anonymisation: Data Encryption**

Regulations: through GDPR compliance.

Accountability: through **Traceability**: Logs to track system's decisions/actions

Responsibility: Responsibility for the system's actions.

Human Oversight: **Human-in-the-loop:** Ensuring a provision for human intervention.

Ethical Considerations: through **Stakeholder Involvement:** Clinicians now – patients later

Planning to Provide:

Ongoing Assessment: through **Regular Audits:** Regular reviews/audits of systems - ensure trustworthiness. **Feedback Loops:** Create mechanisms for users to provide feedback on the AI's performance and behaviour.

Openness: through **Open Source:** so wider community can inspect/verify/contribute.

Collaboration: Engage with the wider community to share best practices and learnings.

Regulations & Standards: through **Compliance**: Plans to ensure system complies with existing regulations/standards, e.g. UK driving test regulations.

LUCIA

LUng Cancer-related risk factors and their Impact Assessment

HORIZON-MISS-2021-CANCER-02

LUCIA (Horizon Europe)

Funded by the European Union

UK Research and Innovation

Ulster University

European Commission

LUCIA

LUng Cancer-related risk factors and their Impact Assessment

- €13.5M Project
- 4 years (Kicked off Jan 2023)
- 20 partners

YAGHMA B.V. (YAG)

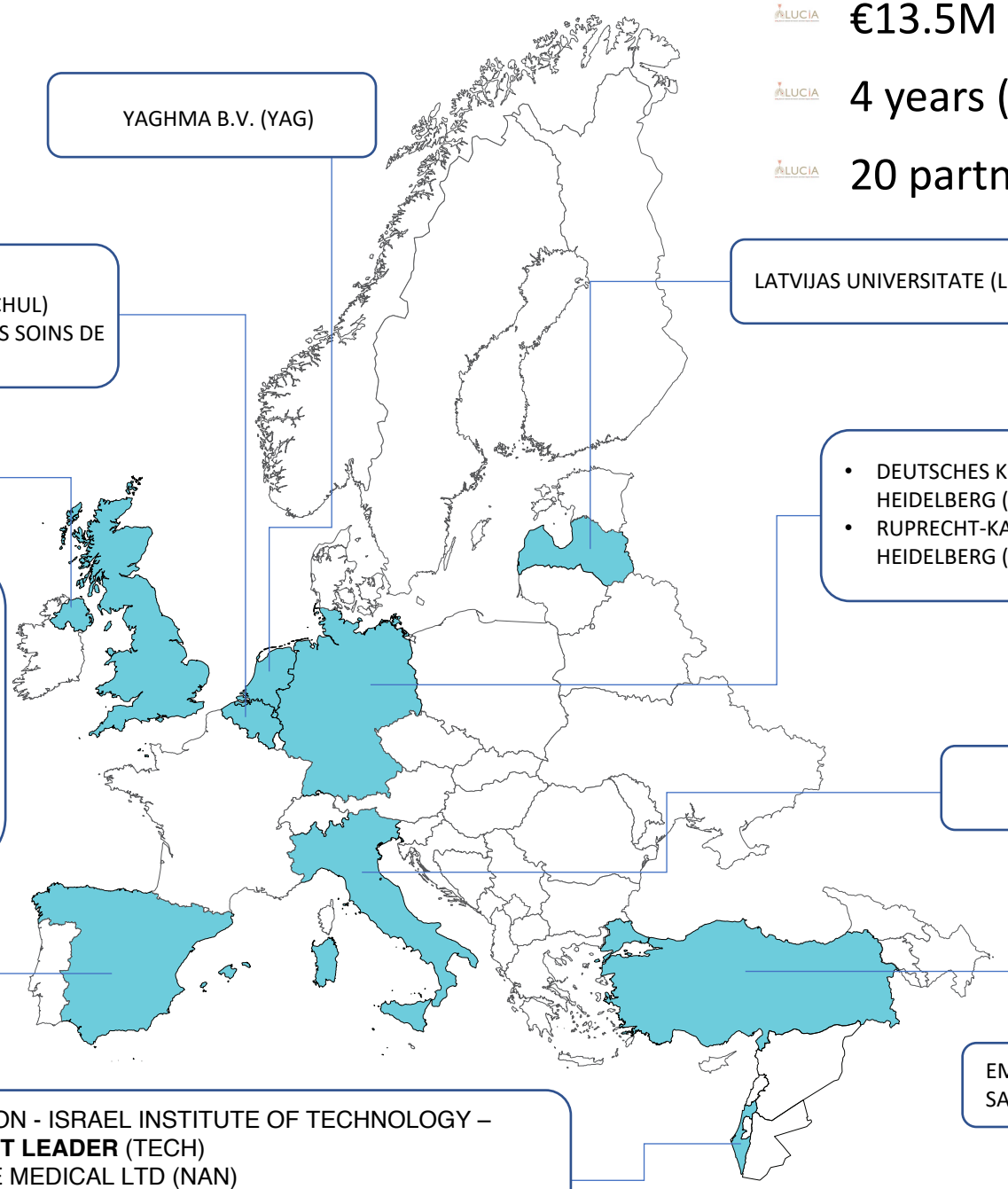
- TIMELEX BV (TLX)
- CENTRE HOSPITALIER UNIVERSITAIRE DE LIEGE (CHUL)
- FEDERATION EUROPEENNE DES HOPITAUX ET DES SOINS DE SANTE (HOPE)

LATVIJAS UNIVERSITATE (LU)

ULSTER UNIVERSITY (ULSTER)

- DEUTSCHES KREBSFORSCHUNGSZENTRUM HEIDELBERG (DKFZ))
- RUPRECHT-KARLS-UNIVERSITAET HEIDELBERG (UHEI)

- FUNDACIO INSTITUT DE RECERCA BIOMEDICA (IRB BARCELONA)
- FUNDACIO CENTRE DE REGULACIO GENOMICA (CRG)
- FUNDACION CENTRO DE TECNOLOGIAS DE INTERACCION VISUAL Y COMUNICACIONES VICOMTECH (VICOM)
- UNIVERSIDAD POLITECNICA DE MADRID (UPM)
- BILBOMATICA SA (BILB)
- SERVICIO ANDALUZ DE SALUD (SAS)
- ASOCIACION INSTITUTO DE INVESTIGACION SANITARIA BIOCRUCES BIZKAIA (BCB)



DEXAI - Etica Artificiale (DEX)

EMODA YAZILIM VE DANISMANLIK SANAYITICARET LIMITED SIRKETI (EMO)

- TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY – **PROJECT LEADER** (TECH)
- NANOSE MEDICAL LTD (NAN)
- PRONAT INDUTRIES LTD (PRON)

Funded by the European Union

UK Research and Innovation

Map courtesy of https://freevectormaps.com/

LUCIA Overview

- Lung Cancer (LC) is the biggest cancer killer worldwide, with five-year survival following diagnosis varying between 5% to 25%
- Though tobacco smoking has long been recognized as the major risk factor for LC, many cases (incl. LC patients that are non-smokers) cannot be explained by this
- LUCIA aims to establish a novel toolbox for discovering and understanding new risk factors that contribute to LC development
- Deliver a toolbox encompassing the analysis of three aspects:
  i. personal risk factors, which include a person's exposure to chemical pollutants as well as behavioural and lifestyle factors;
  ii. external risk factors, such as urban, built and transport environments, social aspects and climate; and
  iii. biological responses to these personal and external risk factors, including changes in genetics, epigenetics, metabolism and aging

LUCIA Overview

The impact of the identified personal and external risk factors and the associated biological responses will be then validated in three clinical use cases:

- general population risk assessment and screening
- precision screening of high-risk populations
- digital diagnostics

The resulting evidence within LUCIA will be translated into policymaking recommendations, with the aim to implement them in screening programs for LC

LUCIA
LUng Cancer-related risk factors and their Impact Assessment

Ethics & Governance to Ensure Value-sensitive design of LUCIA sociotechnical systems – Lead by AIRC & ISRC

- **Ethical assessment framework building**: (1) scientific literature + (2) applicable ethical frameworks for LUCIA technologies (EC Guidelines for Trustworthy AI, AI Act)

- **Stakeholder group mapping**: (a) developers and digital innovators, (b) physicians, clinical researchers and patients, (c) researchers in SHH and legal, (d) standardisation and regulatory bodies.

- **Draft of the interviews to kick off value-sensitive design**: (a) focus groups (live project meeting), (b) semi-structured interviews (online)

- **E-workshops on ethics-by-design and value-sensitive-design (engage tool):**
  - *ethics-by-design workshop:* ethics-by-design and value-sensitive design frameworks explained (AI Act, Ethics Guidelines for Trustworthy AI)
  - *ethics-by-design workshop* assessment of LUCIA technologies and initial requirement elicitation

- **Digital research data generated in line with FAIR principles**

- **Awaiting final EU Artificial Intelligence Act - expected Nov 2023 to ensure adherence**

- ***Challenge of using powerful AI state of art solutions and ensuring explainability to build trust in AI***

Funded by the European Union

midas

Meaningful Integration of Data, Analytics and Services

Grant Agreement No. 727721

2016 – 2020 Project Overview

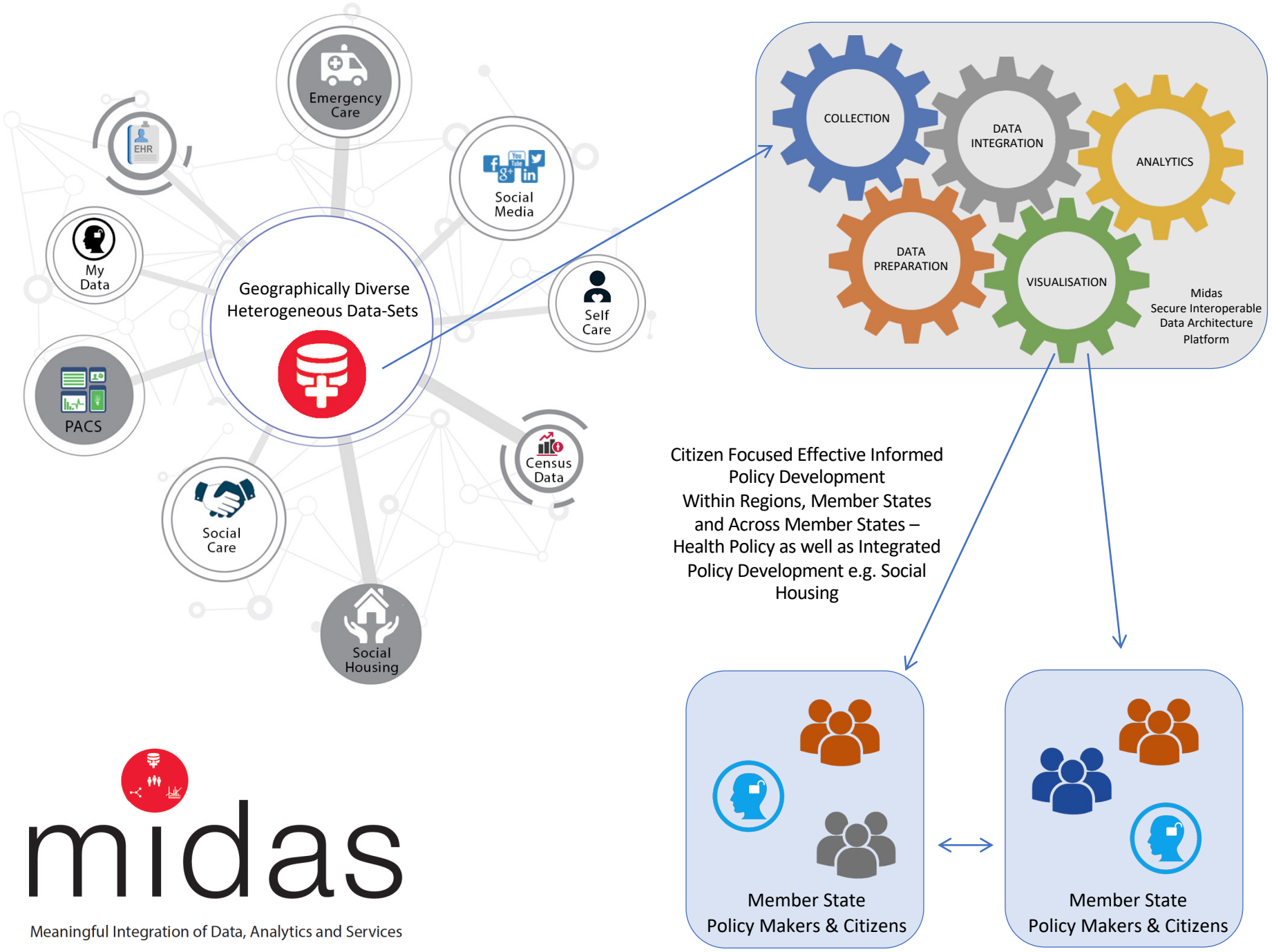
Prof Michaela Black

This project is funded by
the European Union

H2020-SC1-2016-CNECT
SC1-PM-18-2016 - Big Data Supporting Public Health Policies

http://www.midasproject.eu

@MIDASProjectEU



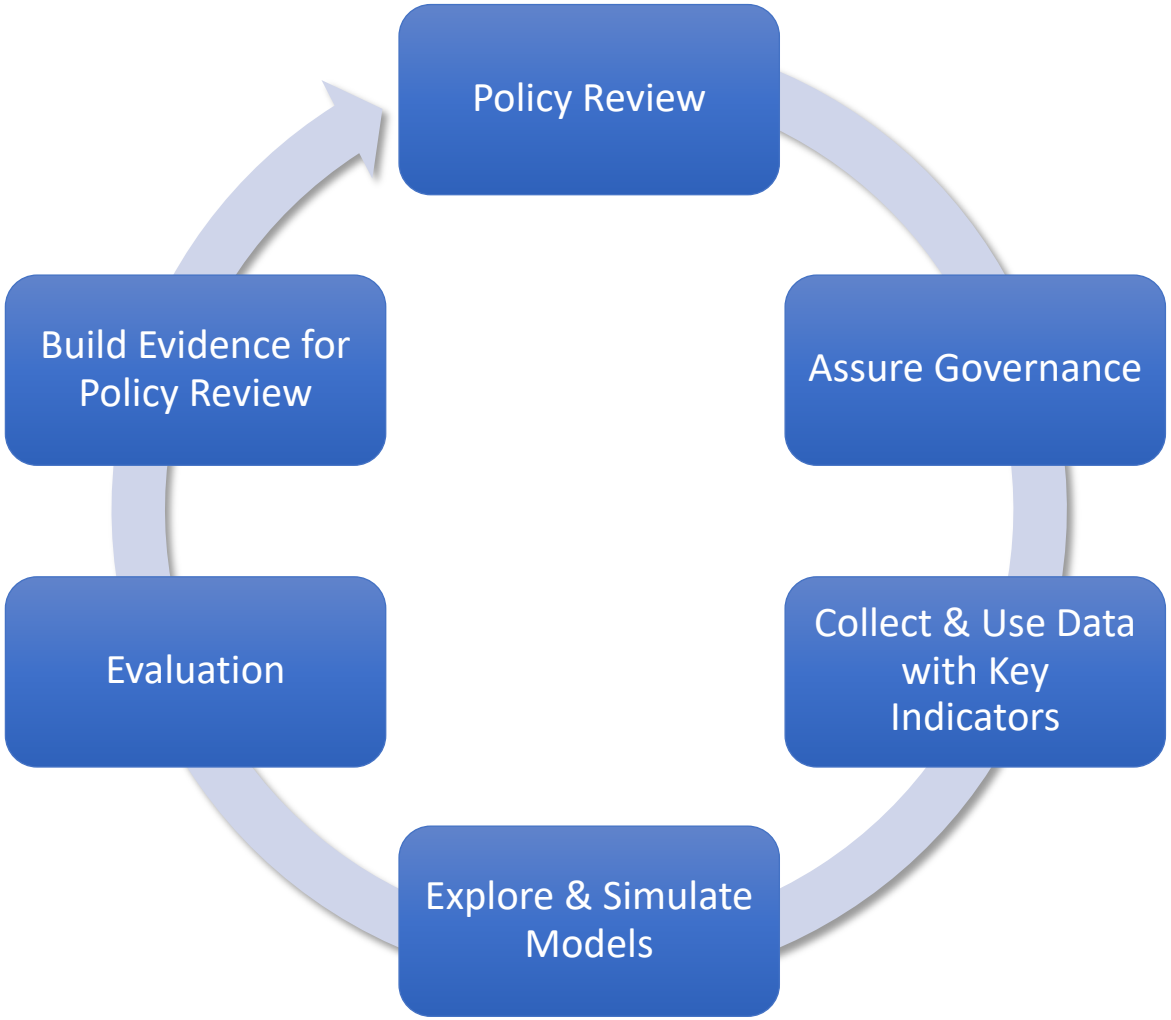
midas

Meaningful Integration of Data, Analytics and Services

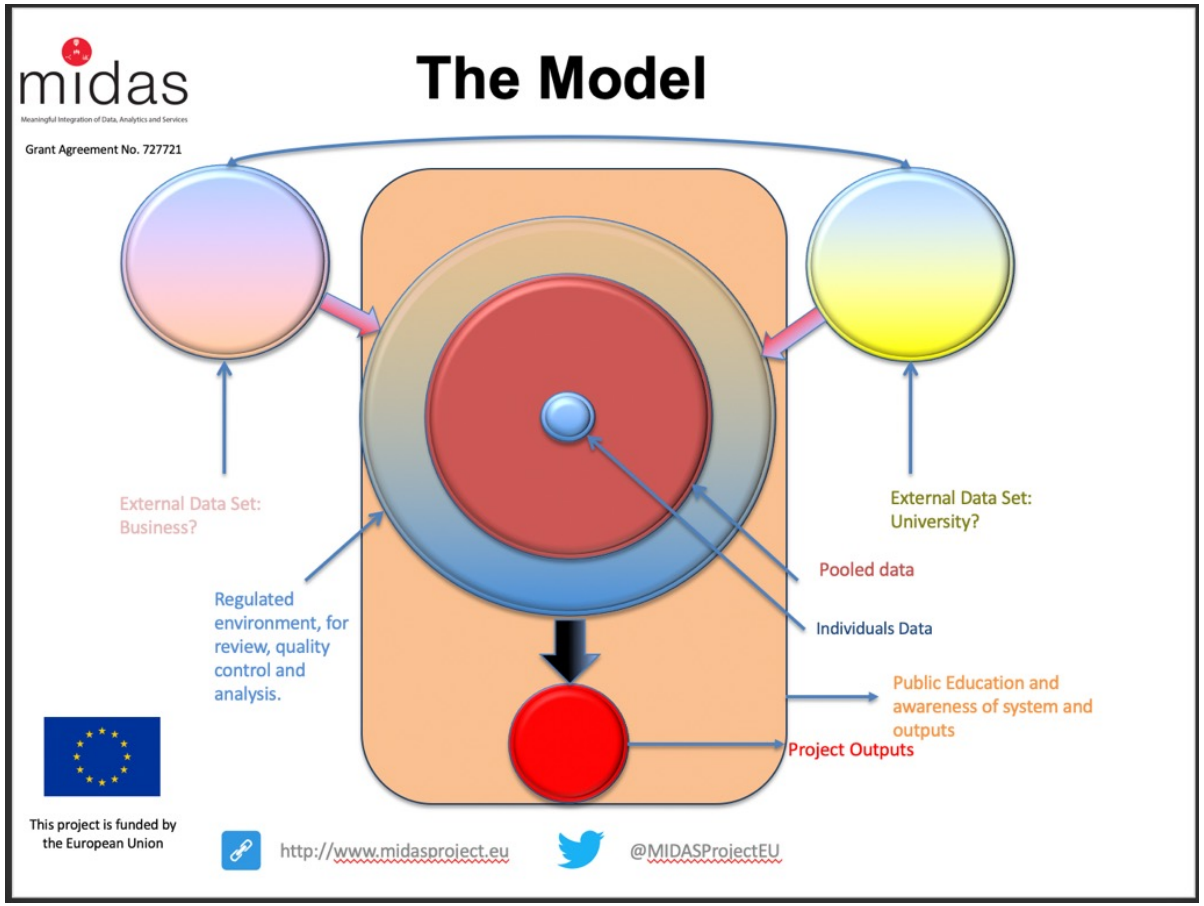
MIDAS Stakeholder-Centred Co-creation Approach

Scientific/Technical Team	Policy Board
University of Ulster (Coordinator)	South Eastern Health And Social Care Trust NHS (SET)
Dublin City University (DCU)	Health and Social Care Board (HSCB) Business Services Organisation (BSO) NI
City of Oulu University	Health Service Executive (HSE) Ireland
VTT: Technical Research Centre of Finland	Finland's National Institute for Health and Welfare (THL)
VICOMTECH (Spain)	BIOEF Public Foundation of the Health Department of the Basque Government
KU Leuven, Department of Electrical Engineering-ESAT (Belgium)	Department of Health Public Health England (DH PHE)
Industry: IBM Ireland	Arizona State University/ Mayo Clinic
SME: Analytics Engine (UK)	
SME: Quintelligence (Slovenia)	

Improved Policy Process



Trustworthy AI & Responsible AI: Honest Broker Model (HBM)



- Methodical and accessible model to assure this ethical and quality oversight
- Operational in Northern Ireland (HSC), and considered for adoption by the Basque region, with Finland proposing a similar governance system
- Healthcare data is viewed by the public as “special”, and as such provision should exist when using large data sets drawn from health service records, be that patient records through Electronic Health Records, Imaging repositories
- MIDAS research proved HBS can only exist within an environment where relationships are clear and the public understand the scope of the service, the reasons for the service and the potential benefits of the service

Synthetic Data
Data Access with Data Privacy

- A large amount of health and well-being data is collected daily, but little of it reaches its research potential because personal data privacy needs to be protected as an individual's right, as reflected in the data protection regulations
- Moreover, data that does reach the public domain will typically have under-gone anonymisation, a process that can result in a loss of information and, consequently, research potential
- Lately, **synthetic data generation**, which mimics the statistics and patterns of the original, real data on which it is based, has been presented as an alternative to data anonymisation
- MIDAS carried out initial review of this work on some Policy sites

Useful Resources

- AlgorithmWatch is a non-profit research and advocacy organization that is committed to watch, unpack and analyze automated decision-making (ADM) systems and their impact on society
  https://algorithmwatch.org/en/

- High-Level Expert Group on AI presented Ethics Guidelines for Trustworthy Artificial Intelligence 7 key requirements that AI systems should meet
  https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-aience

- AI for good
  https://aiforgood.itu.int/how-can-ai-improve-mental-health-for-100-million-people/

Ulster University

Thank you for listening

Questions?

#1 NI UNIVERSITY
GUARDIAN UNIVERSITY GUIDE (2023)

97% ENVIRONMENT FOR WORLD-LEADING OR INTERNATIONALLY EXCELLENT RESEARCH
ULSTER UNIVERSITY (REF 2021)

TOP REGION TO VISIT IN 2018
LONELY PLANET (2017)