DARE UK (Data and Analytics Research Environments UK)

DARE UK Working Group (WG) Charter Template

Name of Proposed WG:

SAFETEXT: community-led protocols for the safe and responsible use of de-identified and synthetic healthcare text for AI development

Affiliated DARE UK Interest Group (if applicable): *none*

The WHY

Introduction: [A brief articulation of what tractable challenge(s) or issue(s) the WG will address, how this WG is aligned with the DARE UK mission, and what this WG would deliver as a value-adding output or deliverable to the sensitive data research community]

(1) What is the research case? Why will the output(s) be useful and for whom?

AI applications that generate or handle healthcare free-text data (e.g. ambient AI, clinical documentation management, patient communication support, clinical training, etc.) are increasingly used across the UK, with a huge potential, promise and expectations to transform healthcare. Successful AI applications rely on good-quality, real-world or realistic data. However, clinical text often contains sensitive and personal health information, which brings risks not only to privacy of patients but also to third parties who may be mentioned in clinical narrative (e.g. '*patient is a full-time carer of her autistic son*'). This necessitates stringent governance and privacy measures to comply with legal and ethical standards, not only during the AI development but also once models are deployed in real-world setting. The absence of unified protocols risks inconsistent data handling practices, potential privacy breaches, and barriers to data and model access, which collectively hinder innovation and the equitable deployment of natural language processing (NLP) technologies.

Researchers, public, policymakers, regulators, and industry need to collaborate to define and implement agreed frameworks that promote transparency, safeguard patient privacy, and foster responsible use of healthcare text data for AI development. Such frameworks are critical to build trust, facilitate cross-institutional collaboration, and support the development of effective and ethical NLP solutions that can improve healthcare outcomes.

This working group (WG) aims to bring together the stakeholders to **understand the current practices and barriers, and develop a set of national guidelines for best practices in handling highly sensitive healthcare free-text data as part of the AI and NLP development lifecycle**. The key output of the WG will include a set of protocol templates that should be followed when providing access to such data to ensure trust, transparency, governance, audit and reproducibility. Trust in particular is a fundamental underpinning for all secondary use of health data: compromising that is hugely problematic and is often seen as a key barrier when it comes to using healthcare free-text data within the AI development lifecycle: from using free-text data for development, to validation and regular updates.

The outcomes of this WG will benefit

- AI researchers, both in academia and industry, by providing responsible, legally and ethically aligned access protocols to highly sensitive data;
- public, including patients and clinicians, with transparent and harmonised protocols focused on enhancing privacy and maintaining trust in secondary use of free-text data;
- TRE and NHS governance professionals, by offering with clear guidelines, protocols and best practice that would streamline and support routine sensitive data access;
- regulators with a specific framework that can be monitored, audited and translated to other domains; and eventually
- clinicians and patients, by providing realistic data to train and validate safer AI solutions.

We note that excluding free-text data from AI development brings significant risks to patient safety and further exacerbate inequities in healthcare as patients with certain diseases (e.g. mental health) or within certain settings (e.g. outpatients) may be less likely to benefit from AI applications as their data is mostly available in the free-text form.

(2) Which specific DARE UK programmes recommendations is this WG's output(s) supporting to address and how so? If none, please explain why this is the case. *

The lack of harmonised protocols for accessing realistic, local free-text data access has been recognised as a key barrier for further healthcare AI and NLP development, its deployment within the NHS and patients' trust in generative healthcare AI. This WG is established to contribute to the DARE UK's main mission to deliver "*a coordinated and trustworthy national data research infrastructure to support research at scale for public good*" by suggesting national guidelines to deal with highly sensitive healthcare free-text data, and aligning those with existing DARE UK frameworks (e.g. SATRE). Specifically, the WG will support the following DARE UK recommendations:

- trustworthiness and transparency ("Standardise, centralise and unify processes enabling access to sensitive data for research across the UK"),
- data and discovery ("Develop guidelines on privacy enhancing technologies (PETs) for use by TREs"; "Enhance the data lifecycle to support effective cross-domain sensitive data research"; "Explore the implications of new data types on approaches to making these data available for research"; develop meta-data standards),
- capability and capacity ("Use automation to ensure data research infrastructure services are reliably secure, auditable and reproducible", while "supporting data collectors and data guardians to routinely make their data accessible for research in the public benefit").

The WG will also contribute to the DARE UK's workstream on community building, engagement and standards, and "public information campaign to raise general awareness of how and why sensitive data is made accessible for research". Finally, the outputs of the WG will facilitate the "process for accredited researchers to request access to sensitive data from data guardians whilst maintaining appropriate levels of data privacy and security".

The WHAT

(3) A reasonably detailed description of the intended output(s) that will be delivered through this WG.

This WG will first assess the current practices and barriers to providing free-text data for healthcare AI development, through discussions and participation of our partners and collaborators. We will conceptualise and set out the ethical positions (i.e. healthcare text viewed as highly sensitive and

private) from the legal aspects (free-text is personal data that is likely to be special category data concerning health), as well as risk mitigation approaches (text de-identification, synthetic text generation) and the usage within trusted research (and development) environments (e.g. for model training or for validation). We will clarify the processes and legal positions for de-identified, pseudo-anonymised and anonymised free-text data, and review the current protocols and risk estimation practiced used with the UK. The other part that will be discussed is the embedding and propagation of privacy risks by the potential for models to retain and potentially replicate identifying information from the training set. Following the community discussions, the key outcomes will include:

- Protocols for responsible text de-identification and pseudonymisation (i.e. removal of direct identifiers) and generation of healthcare synthetic text, aligned with the current legal and governance frameworks;
- Procedures for validation of both clinical validity and re-identification risks to ensure reliable text data without compromising data integrity.
- Meta-data templates and standards to describe the provenance and details of the free-text data (e.g. de-identification or generations protocols used) to ensure transparency, trust and access.

The guidelines will be designed to follow the standard frameworks set by Five Safes and the SATRE project. We envisage the guidelines to specify detailed protocols for preparing de-identified and synthetic text that will include: (1) the objectives and justification for the AI task (Safe projects), (2) description of text deidentification/synthetic data generation processes (Safe data), (3) the evaluation procedures to be carried out by data controller (e.g. sampling strategy, metrics with acceptable thresholds) (Safe data, Safe people), (4) the data usage protocol, including access and processing environments (Safe setting) and the procedures to be taken if any identifiers encountered (Safe data), (5) the public domain output resulting from healthcare text use, in particular whether any quotation of original text will be required and the vetting procedure for this (Safe outputs). The WG will also propose meta-data schemas (e.g. an extension of RO-Crate or OMOP) to describe the resulting data to support provenance, transparency and reuse (e.g. within the HDRUK Gateway). We will look across different clinical narrative types (GP notes, mental health records, outpatient letters, ambulance notes etc.).

The WHO

(1) Why is this WG and its proposed members the right group to tackle this?

The proposed WG consists of academics, NHS practitioners, governance and legal experts, patients, regulators and industry, including senior academic leaders, early carer AI researchers and experienced industry collaborators. We will bring together a number of partners with experience in real-world healthcare text processing, including DataLoch that manages health and social care data for South-East Scotland, including GP narratives; SLaM CRIS – a local database of de-identified mental narrative records; the world's largest de-identified dataset of mental health clinical notes (900M notes across the UK); large, digitally mature NHS trusts such as Northern Care Alliance and Manchester University NHS trust; specialised tertiary-care providers such as The Christie NHS trust - the largest single-site cancer centre in Europe; an ambulance service (NWAS), etc.

The WG will include several AI researchers and software developers who have worked in automated, large-scale de-identification of free-text data in the UK, including development and deployment of both generic (e.g. CogStack's AnonCAT, Edinburgh's SARA project, Manchester/NCA MASK) and bespoke tools (e.g. software developed in Akrivia, SLaM CRIS and Evergreen Life).

Our WG also includes a strong legal, ethics and governance team, with experience in discussing social licences to process the data and working on large data initiatives (such as Born in Scotland and UK BioBank). Our members have pioneered engagements with public and patients, co-producing research in many research and innovation projects. For example, we have organised the first Citizens' jury in the UK to discuss access to free-text data (2018, http://healtex.org/jury/). This work has identified challenges and set the research road map for research in this area. The team has also established the national network for healthcare text analytics and has been recognised as multi-organisational hub for this domain.

Our initial members also involve colleagues from executive and governmental agencies (MHRA, NICE and UK Health Security Agency) and various national TREs (Scottish Data Safe Haven Network, TREvolution), bring in both regulatory and practical experience of running national services (such as CPRD or Born in Scotland).

(2) Which communities will be involved and what relevant skills/knowledge/experience do they have?

We have a strong multi-disciplinary community within the UK network on healthcare text analytics (http://healtex.org/), which we will build on. With its ~300 members, the network will be used as the key community for input, specification and feedback on the proposed protocols. Healtex collaborates closely with the wider health informatics and data science communities via HDR UK – we will present the proposed guidelines at the HDR UK annual 2026 conference. Some of our members are part of the NHS AI Ambassador network, and we will seek their feedback on our proposals. We have also made initial contacts with the Smart Data Research UK to discuss their experience in secure handling of data and data donation in particular (Smart Data Donation service), as this could be an alternative approach to providing data access.

The HOW and WHEN

(3) What related initiatives/groups/work – not represented through this WG or related IG membership – will the group engage with? Are there other adjacent or relevant WGs this group will coordinate and/or collaborate with?

Within the DARE UK, we have already started discussions with the UK Synthetic Data Community Group. This group has focused on structured data and is keen to work with us to establish protocols that refer to unstructured data. We will build on their experience and work to align the protocols between these different data modalities. We will also aim to establish links with the AI Risk Evaluation Community Group to work on modelling of re-identification risk evaluation. We will also engage with the Information Commissioner Office – initial contacts have been made through the partner organisations (MHRA). We will make a further open call for involvement to NIHR Health Informatics Collaborative and members of the Secure Data Environment Network in England and Scottland.

(4) Describe how often the group will meet and sustain progress between meetings?

The Working group will kick off by a remote meeting in October 2025, where we will set the terms of engagement and agree on the practicalities (e.g. documentation). We will confirm the lead roles for each of the activities we plan to do (e.g. leads for each of the workshops; coordination for writing-up; lead for public engagement; lead for meta-data development) and establish the PAG. We will organise monthly remote meetings of the core team to review the progress and discuss the planning for the key events. Face-to-face meetings will be held in each of the events the group is organising before and at the end of the meeting to agree next steps.

(5) A high-level work plan describing intermediate milestones and output(s) that will be developed during the WG's work, and the overall timeframe (max. 12 months) for the WG to deliver the proposed output(s).

The work will be organised in these five phases:

Workshop 1 (November 2025, Edinburgh): Current practice for accessing free-text for AI development. Exploration of the diversity of experiences and perspectives around access to data and approaches to preserving confidentiality in healthcare free-text, including evaluation practices (for both privacy risks and clinical fidelity). Activities will include overviews from data controllers, AI providers and public contributors, followed by discussions on the current work to mitigate privacy risks in free text (via automated de-identification and synthetic data generation). Participants will discuss and consider the methods presented, what makes good practices, challenges and any gaps that they see. Co-design with HDR UK, UK synthetic data network and PAG.

Output (December 2025): Up-to-date summary of methods currently in practice for privacy-risk mitigation, potential challenges and gaps.

Workshop 2 (February 2026, Manchester): Information Governance (IG) for accessing health free-text for AI development. IG specialists and TRE providers will introduce the legal landscape and current practices on how free-text data are made available for research and innovation, along with example use cases. Exploration of current protocols around free-text availability for research, as well as the barriers and limitations and the reasoning for these. Activities will include focused discussions on trust, transparency and risks, and potential mitigations, any regulatory gaps and overlaps. Co-design with ADR UK, TREvolution, Scottish Safe Havens Network and PAG.

Output (March 2026): Discussion summary capturing legal and IG perspectives on trust, regulation, risks, and mitigations to offer an outline framework to enable free-text data to be made available to AI applications.

Draft guidelines development (April-June 2026). Co-production of draft guidelines by co-chairs and PAG. Suggest key steps (as indicated above), possible metrics and alignment with the existing frameworks (e.g. Five Safes, the SATRE guidelines). Dry-test with current partners. Propose meta-data extensions to describe de-identified and generated synthetic data. Outline a sustainable approach for structure feedback.

Output (June 2026): Draft guidelines for preparing safe access to free-text data.

Workshop 3 (June 2026, Wrexham, HealTAC 2026): Guidelines for accessing free-text healthcare data for AI applications. This workshop will bring together the community to discuss draft guidelines and to consolidate all the learnings and insights to collaboratively create guidelines for healthcare free-text access. Activities will include a presentation and discussion of the outputs from the previous workshops, presentation of the draft proposal for the guidelines based on outputs from earlier workshops and examples of implementations. In the second part of the workshop, we will work in a small group mix to focus on areas to develop further, suggesting revised guidelines.

Output (June 2026): Draft guidelines for healthcare free-text access for AI development.

Guidelines refinement and final presentations (July-October 2026). The core team will refine the draft guidelines and meta-data frameworks following the feedback and any outstanding areas for

development. PAG will ensure the guidelines are sufficiently transparent and clear for non-specialist audiences. Prepare iconographic and documentations. Public presentations and launch events will be organised in September /October 2026 (e.g. during the HDR UK annual conference). Prepare news item for the DARE UK website and partner websites.

(6) Do the output(s) have potential for adoption by the target beneficiaries articulated in (1)? What is the adoption or implementation plan for WG member organisations? Such adoption or implementation should start before the WG timeframe is complete.

We will design the WG to have the guidance adoption as a key principle and a measurable target. Partners and members of the WG include representatives of all potential beneficiaries, including large data service providers (see above). We are expecting that the majority of them will test the guidelines by mapping their current practices to the community-proposed protocols, identifying the gaps and issues that will be addressed in the Guidelines review phase (July-September 2026). We acknowledge that implementation in some organisations may be slower as the guidelines may require re-validation of the outcomes (e.g. evaluation of the automated processes for identification of direct identifiers). We will discuss with partner organisations as how adoption and transitions can be done most effectively – for example, some organisations may decide to implement the guidelines for new projects only. As part of adoption practice, we will highlight the resources needed for transition. We will design a questionnaire to collect the adoption feedback from partner, and use it as a structured feedback process to ensure the guidelines and issues remain current.

Potential members:

FIRST NAME	LAST NAME	EMAIL	(Co-)Chair/Member
Goran	Nenadic	gnenadic@manchester.ac.uk	Chair
Arlene	Casey	arlene.casey@ed.ac.uk	Co-chair (de-id lead)
Angus	Roberts	angus.roberts@kcl.ac.uk	Co-chair (synthetic data)
Sarah	Markham	sarah.markham@kcl.ac.uk	Co-chair (public rep)
Elizabeth	Ford	E.M.Ford@bsms.ac.uk	Co-chair (PIE lead)
Jessica	Bell	Jessica.Bell@warwick.ac.uk	Co-chair (legal lead)
Robert	Stewart	robert.stewart@kcl.ac.uk	Co-chair (clinical lead)
Jaya	Chaturvedi	jaya.chaturvedi@kcl.ac.uk	Member (AI, NLP)
Alex	Bea	b.alex@ed.ac.uk	Member (AI, NLP)
Honghan	Wu	honghan.wu@glasgow.ac.uk	Member (AI, NLP)
Raquel	Iniesta	raquel.iniesta@kcl.ac.uk	Member (AI)
Anil	Bharath	a.bharath@imperial.ac.uk	Member (AI, PET)
Richard	Dobson	richard.j.dobson@kcl.ac.uk	Member (AI, NLP)
Viktor	Schlegel	v.schlegel@imperial.ac.uk	Member (NLP, PET)
Gail	Davidge	gail.davidge@manchester.ac.uk	Member (PIE)
Lewis	Hotchkiss	lewis.hotchkiss@chi.swan.ac.uk	Member (community)
Darren	Lunn	Darren.Lunn@mhra.gov.uk	Member (regulators)
Shaun	Rowark	Shaun.Rowark@nice.org.uk	Member (regulators)
Joshua	Harris	Joshua.Harris@ukhsa.gov.uk	Member (regulators)
Thomas	Searle	tom@cogstack.org	Member (industry)
Molly	Farrow	Molly.Farrow@akriviahealth.com	Member (industry)
Stuart	Christian	Stuart.Christian@akriviahealth.com	Member (industry)
Noura	Al-Moubayed	noura.al-moubayed@evergreen-life.co.uk	Member (industry)
Elaine	Mackey	elaine.mackey@manchester.ac.uk	Member (IG)
Anoop	Shah	a.shah@ucl.ac.uk	Member (clinical)
Brian	McMillan	brian.mcmillan@manchester.ac.uk	Member (clinical)
Stian	Soiland-Reyes	soiland-reyes@manchester.ac.uk	Member (TREs)
Gareth	Price	Gareth.Price@manchester.ac.uk	Member (TREs)
Dan	Schofield	daniel.schofield1@nhs.net	Member (NHSE)
Louise	Carter	louise.carter24@nhs.net	Member (NHS)
Lukas	Hughes-Noehrer	lukas.hughesnoehrer@mft.nhs.uk	Member (NHS)
Lauren	Scanlon	l.scanlon@nhs.net	Member (NHS)
Karim	Webb	Karim.Webb@nca.nhs.uk	Member (NHS)
Rhiannon	Williams	Rhiannon.Williams@nwas.nhs.uk	Member (NHS)

PET = privacy enhancing technologies; NLP = natural language processing; IG = information governance.

** Note, please do not hesitate to point out gaps in the current DARE UK set of strategic themes and/or recommendations that the programme should consider as it continues to evolve these. Community feedback and input is welcomed.*