DARE UK (Data and Analytics Research Environments UK)

DARE UK Interest Group (IG) Charter Template

Name of proposed Interest Group: **Improving Transparency Around Linkage Outputs (ITALO) Interest Group**

The WHY

<u>Introduction:</u> [A brief articulation of what broad-based challenge(s) or issue(s) the IG will address, how this IG is aligned with the DARE UK mission, and how this IG would be a value-added contribution to the sensitive data research community]

There has long been strong consensus between researchers and government data providers on what constitutes an informative data linkage quality report to improve transparency, reproducibility, accuracy of linkage process, and the validity of analyses and interpretation (Gilbert et al, 2018, GUILD). However, in the UK, data linkage quality reporting has yet to match the desired standards. Marginalised communities such as ethnic minorities, migrants or people living in more deprived areas are more prone to have linkage errors. As linked administrative data are used increasingly for policy making, there is an imminent statistical and ethical case for researchers, data owners and government bodies to ensure the evidence generation processes are proactively transparent (as reflected in DARE UK phase 1 public engagement activity) and equitable for all.

This Interest Group (IG) will address the critical challenge of ensuring high-quality data linkage in sensitive data research, focusing on the adoption of clear, consistent data linkage quality reporting standards. By convening researchers, data linkers, and data owners, the IG aims to (1) identify training needs around adapting existing linkage quality standards, such as GUILD, (2) pinpoint barriers preventing widespread implementation, and (3) share pockets of good practices and strategies across four nations to overcome these challenges. The conversation will also incorporate other factors affecting linkage quality—such as national opt-outs in health records; and data missingness —to foster a more holistic approach.

This IG will help deliver on Phase 2 priorities by instilling a culture of transparency and confidence for analysts to work with sensitive data. As a value-added contribution to the sensitive data research community, this IG will serve as a dedicated forum for exchanging best practices, developing robust training curricula, and co-creating guidance on linkage quality. The outcomes—shared frameworks, streamlined processes, and a strengthened culture of collaboration—will improve the reproducibility, efficiency, and trustworthiness of sensitive data research across the UK.

The WHAT

Research teams frequently face significant difficulties in evaluating linkage quality—particularly by sociodemographic factors—as well as identifying potential linkage patterns when only limited output is available from data owners. The back-and-forth required to ask for further clarifications is time-consuming, yet it often yields insufficient insights.

Additionally, there is a gap between developments in linkage methods and how linked data are ultimately utilized by researchers. For instance, while the ONS Demographic Index offers sophisticated techniques—

relying on clusters and confidence intervals rather than a single "spine"—these innovations are not widely adopted in standard practice.

Finally, current linkage processes do not easily allow researchers to assess the sensitivity of different linkage strategies or how these strategies might shape a study cohort. Researchers may have varying priorities—some favouring higher sensitivity for certain analyses, others higher specificity. Yet existing approaches make it difficult to tailor linkage methods accordingly.

Collectively, these challenges highlight the need for a dedicated forum—an Interest Group—that brings together researchers, data linkers, and data owners to share knowledge, develop training resources, and collaborate on common standards for assessing and reporting linkage quality.

<u>Objectives:</u> [Set of focus areas for discussion linked to the driving user scenario(s)/use case(s) above, articulate how this group is different from other such groups inside or outside of the DARE UK programme, which specific strategic theme within the [DARE UK programmes recommendations](#) (see page 19) is this IG primarily aligned with. *]

We compared our IG to the ONS Data Linkage Champions Network.

- **Broader Stakeholder Engagement and Trustworthiness**: While the ONS Data Linkage Champions Network typically focuses on government-led projects and methods, and has a closed membership to only include people working in the government; this Interest Group explicitly seeks to include researchers, data owners, and data linkers from diverse institutional settings, spanning multiple data types and research domains, and at different parts of the linkage pipeline.

- **Attention to Data Quality Reporting and Proactive Transparency**: A key goal of this IG is to co-develop and promote unified reporting standards (e.g., how to document linkage performance by sociodemographic groups; or by types of activities which the data is intended to capture). This level of systematic reporting and transparency broadens beyond single-agency contexts.

- **Focus on Training and Practical Integration – Capability and Capacity**: This IG not only discusses linkage methods but also emphasizes how to incorporate these approaches into regular research practices. Training needs and capacity-building are central to its remit, acknowledging that different users (e.g., academic researchers, NHS data owners, government analysts) have distinct needs. This IG hopes to bridge these gaps by working closely with larger collaborations with capacity to deliver such training needs, such as ADR UK and HDR UK.

<u>Outcomes:</u> [What does this IG intend to achieve including outputs that may lead to WG topics, any initial ideas on outputs that could be delivered through WGs spawned under this IG]

Intended Achievements of the IG and Potential WG Outputs

1. Practical Standards and Guidance for Linkage Quality
   a. The IG will coordinate efforts to develop, refine, and socialize robust linkage quality reporting standards, with a view to making them accessible and adaptable across diverse research domains (e.g., health, social sciences).

b. Through discussions and collaborations, the IG will document best practices, highlight common pitfalls, and create user-friendly guidance materials for stakeholders at varying levels of expertise.

2. Prototype Implementations and Technical Development
   a. A notable outcome could be the formation of a Working Group (WG) to develop an open-source codebase for implementing data linkage quality reports, designed to be easily adopted in diverse sectors (health, social research, policy evaluations, etc.).
   b. One concrete example is to build a prototype leveraging Splink—an open-source package developed by the Ministry of Justice—within an existing collaboration with NHS England. This prototype would (a) demonstrate how standardized linkage quality reporting can be embedded into real-world workflows, and (b) provide a repeatable template for others to adapt.

3. Evolving Ecosystem of Collaboration
   a. By convening a broad network of stakeholders, the IG will open avenues for additional topic-specific WGs—for instance, addressing nuances in linkage for particular data types, dealing with missingness, or tackling the ethical aspects of national opt-outs.
   b. These WGs may produce targeted outputs such as policy briefs, technical standards, or specialized software modules that integrate with the broader linkage quality framework.

Overall, the IG's impact will be demonstrated through its collective outputs—actionable standards, a shared codebase, and a stronger community of practice that together advance the field of high-integrity data linkage for sensitive data research.

The HOW

Participation/Collaboration: [Which communities will be involved, and what relevant skills/knowledge/experience should they have, are there other adjacent or relevant IGs/WGs this group will coordinate and/or collaborate with]

Communities Involved

1. Researchers and Methodologists
   a. Encompassing epidemiologists, data scientists, statisticians, data linkage engineers and analysts who routinely work with linked datasets. They bring expertise in study design, advanced analytical methods, and practical data handling challenges.
      i. Joseph Lam and Tony Stone contribute to the National Centre for Research Methods – Introduction to Data Linkage Course, which is ran 2-3 times per year with around 30 attendees each, from academic institutes, government and other organisations across UK.
      ii. Ludivine Garside has experience of re-scoping a linked data research project to reflect emerging data quality issues, as well as experience of working with data owners to link sensitive datasets when no dedicated third-party linkage team is available.
   b. These individuals contribute valuable insights on how linkage quality affects research outcomes, as well as the training and resources needed to raise standards.

2. Data Owners and Custodians

a. Spanning health systems (e.g., NHS England, NHS Scotland, etc.), social care agencies, and government departments.
    i. Giulia Mantovani is on the organising committee for the ONS Data Linkage Champions Network. This gives us the reach to government-based linkage architects and data custodians Giulia also leads the NHS England Data Linkage Hub, which is developing a [quality assurance framework for data linkage](#).

b. We work closely with the lead authors of GUILD, such as Ruth Gilbert and Katie Harron. Their technical and governance expertise is crucial for understanding limitations on data sharing and output disclosure, as well as processes for managing national opt-outs or data missingness. We have contacted methodologists and experts in the field such as Rosie Cornish (Bristol) and Richard Silverwood (UCL).

c. ADR UK is interested as this aligns with ongoing work - they are planning about data quality and relevant training (such as the data quality training course led by Peter Christen).

d. We received letter of support from key partners, such as the Head of Research Strategy and Commissioning for ADR England Strategic Hub, Karen Powell; lead author for the GUILD framework, Ruth Gilbert; and director of UK Longitudinal Linkage Collaboration, Andy Boyd.

3. **Developers and Technical Stakeholders**
    a. Those working on linkage technologies and open-source solutions (e.g., Splink, Ministry of Justice). We have contacted Robin Linacre (founding developer for Splink) and Stephen Jobling at Ministry of Justice, and they are interested to join as members. Their input can help ensure the linkage quality standards are not only theoretically sound but also practical to implement in real-world workflows.

    b. Software engineers, data platform developers, and IT professionals form the backbone of operationalizing these standards at scale.

4. **Public Engagement**
    a. Given DARE UK's emphasis on proactive transparency, we plan to engage the ADR UK Public Engagement Panel to establish public opinion, and when we have funding, on-going engagement of the public in our IG. This will help ensure that the IG's work aligns with public expectations regarding trust, transparency, and ethical data use.

    b. We have spoken with members of PEDRI (Public Engagement in Data-Driven Research & Innovation, ADR UK, also a DARE UK interest group), to help us engage with most relevant populations.

Mechanism: [Describe how often the group will meet and sustain progress between meetings]

Once commissioned, the Interest Group will convene **quarterly online meetings** to discuss emerging challenges, share updates on ongoing projects, and plan collective activities. Between these sessions, members will engage in local discussions and pilot self-evaluation exercises—for instance, using GUILD to assess data linkage quality of existing data. These interim activities allow for iterative feedback and refinement of standards, gaps, or best practices before the next quarterly gathering. Progress will be

tracked and shared via an online collaboration platform (on a Slack channel), enabling asynchronous communication and ensuring continuous momentum. By documenting key insights and outcomes from each local exercise, the group will maintain a clear record of learnings, facilitate peer support, and foster a collaborative atmosphere that sustains progress throughout the year.

We are also planning to lead a workshop at the 2025 ADR UK Conference on data linkage quality. Longer-term planning will subject to available funding to support in-person meetings from Jan 2026. We have made preliminary contacts with several other academic research networks to explore opportunity for jointly organised events in 2026.

<u>Potential members:</u> [Including a minimum of two proposed chairs and all members who have expressed interest]

FIRST NAME	LAST NAME	EMAIL	(Co-)Chair / Member
Joseph	Lam	Joseph.lam.18@ucl.ac.uk	Co-chair
Ludivine	Garside	LGarside@bristol.ac.uk	Co-chair
Giulia	Mantovani	Giulia.mantovani1@nhs.net	Co-chair
Tony	Stone	Tony.stone@sheffield.ac.uk	Co-chair
Peter	Christen	Peter.Christen@anu.edu.au	Member
Katie	Harron	k.harron@ucl.ac.uk	Member
Ruth	Blackburn	r.blackburn@ucl.ac.uk	Member
Robin	Linacre	Robin.linacre@justice.gov.uk	Member
Rosie	Cornish	Rosie.Cornish@bristol.ac.uk	Member
Andy	Boyd	A.W.Boyd@bristol.ac.uk	Member
Karen	Powell	karen.powell@esrc.ukri.org	Member

* 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.