

Training and Capacity Building Strategy

2021-2026

Published November 2021

1. Purpose

To deliver on Administrative Data Research UK (ADR UK)'s strategic objective to build capacity and capability of researchers trained to analyse complex linked administrative datasets, staff from across the Strategic Hub (SHub) have been given responsibility to explore options and work collaboratively with others in the partnership. This draft strategy proposes how we will deliver training specific to Administrative Data Research England (ADR England) linked datasets, as well as provide a level of coordination of the wider investment to identify the most efficient and effective general training for identified learner audiences.

2. Background and context

The approved ADR UK Full Business Case (FBC) set out the importance of training and capacity building for the long-term investment phase of the programme. This has since been embedded into the proposals from partners on what they will deliver as part of this phase, and has also shaped the ADR England strategy, which also highlights the need to collaborate across the partnership in this area.

The economic case within the FBC also makes clear how many researchers will need to be reached and trained to use complex, linked administrative datasets to ensure these are properly utilised, and so can deliver the required impact and return on investment (£2.3 million). Initially, it had been thought that ADR UK would fund a new Centre for Doctoral Training (CDT) to progress this work, aiming to fund 24 PhD students and reach a further 76 through Economic and Social Research Council (ESRC) studentships. However, exploring the feasibility and efficiency of that proposal resulted in a decision to focus instead on increasing the number of PhD students and other researchers at different levels of experience trained to use the more complex linked administrative datasets. This will have the benefit of firmly embedding awareness of ADR UK linked administrative datasets into the wider research community, instead of potentially just creating a silo of expertise with no spill-over into the wider research community.

To reach the desired return on investment of £4.45 on every £1, ADR UK has committed to reaching 1,307 'researcher years' over the five-year investment.

This is a term defined in the FBC as the ‘researcher complement’ multiplied by ‘duration’, on which different factors are then applied depending on whether a researcher is just attending a training course funded by ADR UK (via the £2.3 million training budget) or is fully funded by ADR UK to carry out research on linked datasets, through ADR England, Scotland, Wales or Northern Ireland activities via grants to core partners or ADR England grants.

This approach will enhance the ‘capacity building’ impact of training by reaching larger numbers of researchers who may benefit from being able to use these datasets in their research despite not being ADR UK-funded. The provision of information and training on our flagship linked datasets to these researchers will also help raise awareness across the social sciences about the ADR UK programme more broadly. There are already over 500 PhD students funded by ESRC each year alone, which is around 15% of the total new PhD entrants in the social sciences in the UK. Monitoring is now in place across the entire ADR UK partnership to ensure ‘researcher years’ can be routinely tracked throughout the current investment period.

This paper sets out how ADR England plans to deliver on this objective, interdependently with all ADR UK partners. Consultation on the ideas set out in this paper is taking place with partners, ESRC colleagues, a ‘sounding board’ of academics reached through a survey of CDTs and Doctoral Training Partnerships (DTPs), and other stakeholders including UK Data Service, Health Data Research UK (HDR UK) and the National Centre for Research Methods (NCRM).

Importantly, the ideas presented in this paper are also supported by responses to a recent survey run by ADR UK sent to researchers via our networks. It sought to understand how ADR UK and others can better support researchers to undertake impactful administrative data research and revealed several areas where training and capacity building could help.

3. Emerging themes

Themes emerging from stakeholder discussions fall into three categories:

3.1. *Identifying learners*

There is evidence to suggest that many students, researchers, and supervisors lack a basic understanding of what administrative data is and how it can be used in research. It is clear (anecdotally and from our recent survey) that we are not currently reaching many quantitative social and non-social scientists (including health and other data scientists, maths and computer scientists and others interested in ‘big data’), or qualitative or mixed methods researchers with an awareness of the potential value of administrative data. Even within the community that is already using administrative data, there is a demand for practical training on methods, on the use of specific linked datasets, as well as on elements such as understanding access, data ownership and approvals processes, legal gateways, researcher accreditation and funding available.

3.2. *Identifying advocates for administrative data*

As more government departments and others engage with us in linking administrative datasets and as access and processes become faster and smoother, we continue to carry the burden of past researcher experiences of delays and complications in using admin data for research.

We need to work on changing this reputation across academia by demonstrating that the situation is improving as part of the offer of enhanced training and information provision. To build the community of stakeholders and researchers, we need to identify more advocates who can demonstrate the progress we have made, champion our work and help us grow our networks.

3.3. Dependencies

With strong advocates and identified learners, we also need to acknowledge the dependencies on data owners, controllers and processors to ensure consistent, smooth and swift access to linked datasets; a continued 'flow' of new linked administrative datasets that are available to researchers; strong training delivery partners to provide high quality, reliable training; and the ongoing development of good practice, tools and resources across all stakeholders to support the effective use of administrative data.

4. Theory of change

Ultimately, within the current funding period, we want to reach a point where training and peer-led networks are leading to established relationships across different sectors which feed evidence-based decision-making. We have used the concept of Maturity Levels to describe the Theory of Change for this area of our work. They are:

- Building appetite for using admin data, awareness of training opportunities, needs and gaps.
- Commissioning training to fill the gaps.
- Ensuring the training is in place and being used.
- Training leads to peer-led networks to support further learning and outreach to decision makers.
- Training and peer-led learning to established cross-sectoral relationships to inform evidence-based decision making.

This sequence of progress sets out our opportunity to create a dynamic and flexible approach which drives the themes identified above. More work is needed over the duration of the strategy to shape and embellish these further. To initially explain this and present what needs to happen to deliver, we have described different aspects and levels of education on the use of administrative data with the intention that they can then be mapped across to the maturity levels.

5. Routes to delivery

5.1. Information provision

Across ADR UK, partners are required to “provide clear, easy to find information about the existence and use of datasets available for researchers to access”. Using websites, data catalogues, newsletters, published academic papers (where relevant) and other such engagement tools, we will develop existing and new sources of information on administrative data for a range of audiences including, but not limited to:

- **Agreed definitions** (such as for Research Ready Data and Synthetic data).
- How to **discover and access** linked administrative datasets.
- What linked administrative data can be **used for** and how they can be used in research, using examples of how this has already been done (research ideas -> access -> analysis -> impact).
- Who are the **main funders** of research related to the use of linked administrative datasets and any differences between them?

Activities to drive this work will include:

- Working closely with ESRC and other partners to identify events and opportunities to showcase linked administrative datasets and corresponding research.
- Developing regular contact and accessible materials for CDTs and DTPs to use to support engagement with administrative data research and building meaningful relationships to drive research opportunities at different levels.
- Working with academics to publish papers (and accompanying comms) which help to define and socialise definitions of terminology central to administrative data.
- Providing information/training across Government to increase awareness and understanding such as through POST and CSA networks.

5.2. *Signposting*

Given our limited resources, it is important that we maximise the use of our partnership and wider stakeholder relationships to share learning opportunities and reduce duplication. As shown through responses to our surveys, we need to identify how to build capability through training on:

- Producing good metadata, data descriptors and feedback loops (data owners and researchers).
- Developing good code and code sharing.
- Use of appropriate statistical methods for analysing large and complex datasets.
- Realising and accelerating impact.

We will work with partners and other stakeholders to:

- Map what linked administrative datasets are available and where information about them can be found across the ADR UK partnership (e.g. ONS catalogue, SAIL, Research Data Scotland and NISRA) and more widely (e.g. UKDS; HDR UK). This might also include mechanisms to increase understanding about the data itself, including its source and any examples of how it has already been used.

- Map what self-directed and instructor-led learning opportunities are available, who they are directed at and what the gaps are.
- Clarify what tools are available to support both data owners and users (such as BIT synthetic data generation code, MoJ SPLINK data linkage tool and others) and how to access them.

Activities to support this work will include:

- Meeting with partners and stakeholders in this space to undertake mapping and agree consistent signposting mechanisms including a possible ‘front door’ for researchers and also link into other mapping processes, such as NCRM.
- Development of a training page on both the ADR UK and International Population Data Linkage Network (IPDLN) websites ahead of the joint 2022 conference, including consideration of how best to develop the international training network. Ensure that it is cross-referenced with other sources of insight on this topic, such as NCRM, and that others signpost to ADR UK too.
- Working directly with BIT and MoJ on how others can best use their tools and what training needs to be provided.

5.3. *Opportunities within existing ESRC-funded programmes (e.g. research grants and SDAI)*

We will step up our engagement across existing ESRC-funded programmes to tap into learner audiences in order to grow awareness and opportunities for engagement with administrative data across the life course of researchers, including:

- Through Q-step.
- ESRC quantitative research PhD studentships.
- ESRC post-doc (ECR and more advanced) researcher networks.
- Supervisors.
- Qualitative researchers who could add value to their work by using admin data.

Activities to drive this engagement will include collaboration between ESRC and the Strategic Hub:

- Working with DTPs that have stronger quantitative research delivery to raise the profile of administrative data and testing mechanisms such as supervisor-led competition to drive applications for PhDs which use linked administrative data.
- Working with DTPs on the development of train the trainer activities.
- Working with ESRC during the ESRC Doctoral Training Network recommissioning process to provide the evidence base to justify studentships being steered towards using administrative data in both quantitative PhD studentships and with qualitative researchers.

- Developing relationships with government partners to support and co-fund collaborative studentships using admin data.
- Identifying if/where PhD studentships focussing on administrative data could usefully be developed and funded by ADR England, if there are funding gaps identified related to specific linked datasets that ADR UK is making available.

5.4. *Specific training opportunities (including instructor-led and self-directed courses and workshops)*

Across ADR UK partners are **required to** “Support the development and delivery of training opportunities for the more complex datasets, to widen access beyond those already trained in their use and those researchers already in receipt of ADR UK grant funding”.

For ADR England, this currently includes The Longitudinal Education Outcomes (LEO), Data First, MoJ-DfE, and GRading and Admissions Data for England (GRADE), with others also in development and close to being made more widely available such as Education and Child Health Insights from Linked Data (ECHILD) and Growing Up in England (GUIE). Courses and/or workshops to support the use of these datasets are a priority for ADR England in delivering this strategy, and this **will be a central pillar of work for counting ‘Researcher years’**, as defined in the ADR UK FBC economic case.

However, in addition to these, ADR England will also investigate the opportunities and benefits of providing training to build wider interest in these datasets using linked administrative data research for policy making and decision making. Over the period of the investment, we will:

- Identify different learner audiences for these datasets and understand how their learning needs differ (including non-academic audiences).
- Identify training delivery partners and commission training which can be tested for effectiveness (e.g. day courses and workshops, online courses, a possible ‘summer school’ approach) and looking at how best to build on and support existing tried and tested approaches, such as the Scottish Centre for Administrative Data Research (SCADR)’s Introduction to Administrative Data course.
- Work with training delivery partners to develop content and resources, such as synthetic data. This will also involve working closely with data owners.
- Develop ‘train the trainer’ and peer learning resources for researchers who are already using and are familiar with specific linked datasets.
- Establish sustainable training resources with inbuilt processes for review and updating material.
- Collaborate across the partnership to develop communities of practice.

Initial activities to support this work include:

- Testing contextual training for research fellows (such as planned training for existing DF fellows on the Courts system) and seeing if this can be packaged for different learners (such as ESRC-funded quantitative PhDs).
- Working with delivery partners already familiar with individual linked datasets
- Developing a package of work on synthetic data – drawing on the work of partners such as ADR NI and SCADR, and of other investments in BIT, NCRM, UKDS and current methodological development projects.
- Consult with research fellows on appetite to become trainers (or peer educators) on datasets, and facilitate testing of mechanisms for that to happen, such as the provision of a fellowship funding extension to provide training to new users.

ADR England is also keen to fund specialist PhD studentships which focus on the development of linked administrative data and on early pathfinder research specific to new data linkages. We will look for appropriate opportunities through our communications with academic applicants to schemes such as the Research Ready Data and Access funding call. More exploration of this kind of opportunity is needed.

5.5 *Experiential learning*

As ESRC develops new opportunities for policy fellowships, ADR England will feed into these and learn from their experience to scope experiential learning opportunities and measure their effectiveness. Identifying how our research fellows can champion these and be advocates for us will be an important theme of this work. As well as working with ESRC, ADR England will make use of its new public engagement work to increase partnerships across other sectors, such as Local Authorities and the Third Sector, to identify and test new ways of supporting ‘on the job’ learning.

6. Proposed ADR UK training and capacity building strategy

There are already significant training and capacity development activities happening across the partnership and it is central to this proposal that the SHub activity will build on what already works, and support mechanisms for increasing the value of existing opportunities. We aim to provide coordination to share learning opportunities, minimise duplication and re-invention, and to expand capacity through strong communication across the partnership on this theme.

We propose to set up a Task and Finish Group made up of representatives from across the partnership to refine this strategy so that it can steer our activities effectively and provide confidence that the strategic objective can be delivered.

7. Conclusions and Recommendations

ADR UK provides an important function supporting the development and use of linked administrative datasets, but we are increasingly aware from our pilot phase that if we don't train researchers to use the most complex datasets, which potentially have the highest research value, they won't get used. To mitigate this risk, we have outlined a stratified suite of activities to proactively develop and deliver this strategy and build resilience in this area.

We recommend that the activities outlined above are discussed in the context that this theme is an emerging one, particularly in ADR England. We will endeavour to learn throughout the investment period, build on successes from across the partnership, and adapt this strategy to focus on the most effective delivery mechanisms.

The facilitation role of ADR UK partners, supported by the SHub, will ensure that benefits of this workstream extend beyond those researchers that we directly fund to reach the wider research community, so that the value of the linked administrative datasets that we catalyse will be optimised across institutions, sectors and disciplines.