SCOTTISH CENTRE FOR
ADMINISTRATIVE
DATA RESEARCH

September 2023

Researcher handbook:

Data linkage and administrative data research in Scotland

Contents

Introduction and acknowledgements

This guide is a description of how data linkage and administrative data research projects work in Scotland as at 1 September 2023. This information will be updated as [Research Data Scotland (RDS)](#) services develop and all updates will be available on the RDS website. The expectation is that this handbook will be updated and eventually replaced by an interactive system within the [RDS website.](#)

This handbook is not intended to replace the guidance or protocols of individual data access channels, providers or systems. Rather, it is an informal guide for researchers to the context of data linkage and access in Scotland and what they can expect.

Content for this guide has been produced by collating information from a range of different sources, including guidance from Public Health Scotland, Scottish Government, Research Data Scotland and National Records Scotland. It has been collated by team members within the ADR Scotland team which is part of ADR UK, funded by ESRC.

This guide may be shared in full or in part, and ADR Scotland partner organisations may use any element of the content without attribution in their own web pages, handbooks, guides or other assets as required.

Please note: the processes described in this guide may be subject to change in future. The content is accurate to the best of the teams' knowledge as at 1 September 2023. Please contact relevant organisations directly if you have any queries.

The Data Access Process – to be updated by RDS by end 2023
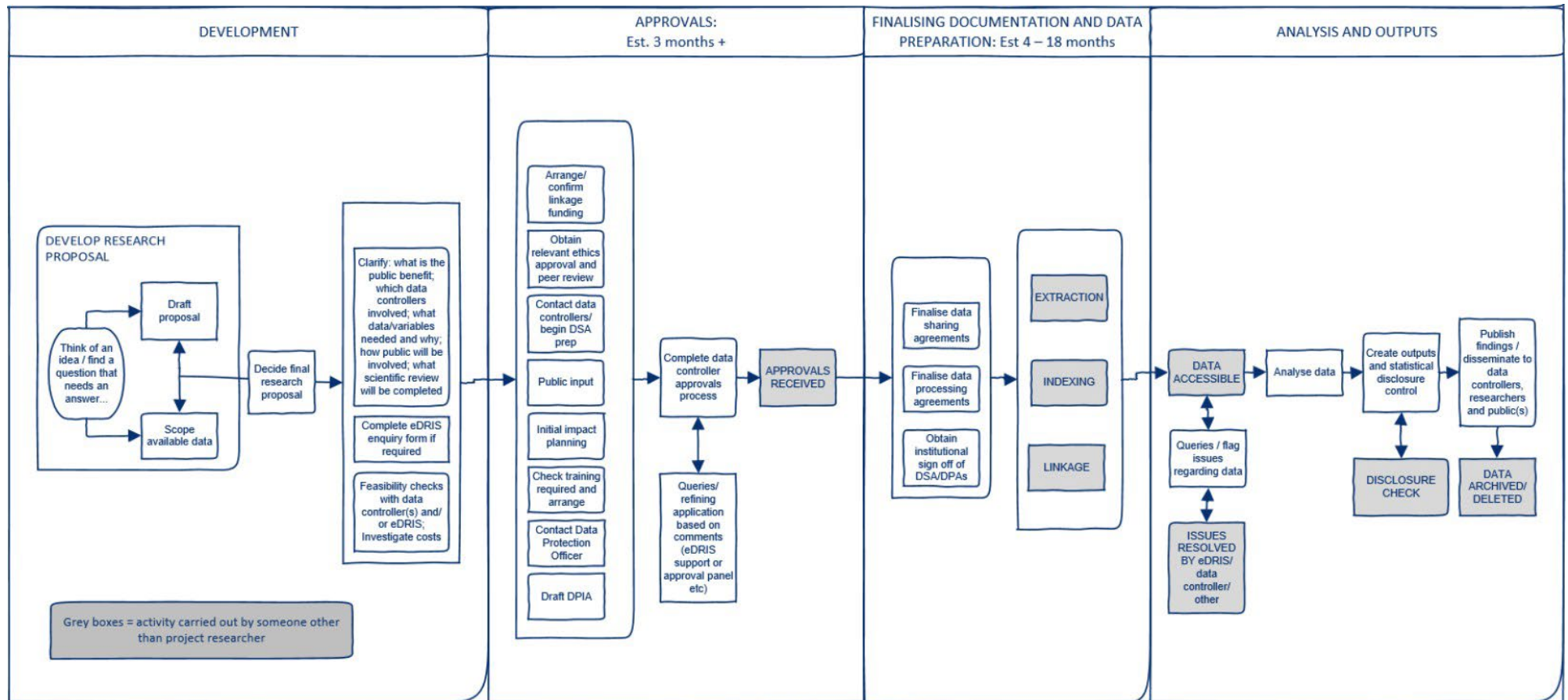
Key Foundations

Context and caveats

This guide is a description of how data linkage and administrative data research projects (using individual/household records) work in Scotland as at September 2023.

It is **written primarily for researchers** who:
- are interested in linkage;
- are new to this process;
- are familiar with what is required for a general research project;
- have a basic understanding of how data protection legislation affects data and research.

Processes in the rest of the UK are broadly similar to those in Scotland, but data controllers and points of application differ.

At the time of writing, Research Data Scotland (RDS) is an organisation working to unlock public sector data to make it quicker and simpler to do research and improve lives. Therefore there are several significant streams of work underway across the public sector to streamline and improve the process for both data controllers and researchers. These workstreams are likely to mean that some information in this guide may change in the future. Please do refer to RDS Accessing data webpage in conjunction with this handbook.

Administrative data is data collected in everyday life – examples of administrative data include hospitalisation records and educational attainment records. Due to the nature of administrative data, there are some things a researcher should keep in mind throughout the process of data linkage:

1. Administrative data is not captured or collected primarily for research purposes, so the data may be incomplete, or used and collected in a form not ideal for research. Specifically, data is unlikely to be clean, or easy to analyse without some manipulation and metadata is often unavailable.
2. Data providers are not necessarily used to sharing their data for research, and they have no obligation to share with researchers. Therefore much of the system is built on people who are supporting researchers alongside their main responsibilities. Data providers may be able to engage with a researcher's project for a range of reasons, and may not have:
   - A team available to support researchers
   - A clear process for sharing data
   - Knowledge of the research process or research methodology
   - A full understanding of what exactly happens to their data before researchers can access it

A positive attitude and willingness to support and accommodate data provider's requirements will go a long way.

1. Where to Start

Who can access data for research and what they are responsible for

Data linkage or administrative data projects can be led by academic researchers, public/third sector employees or approved industry partners.

The data linkage experience is often complex and requires [time and ongoing commitment](#) – projects can typically take from a few months to a few years. There are several services which can provide support with aspects of the process, but the lead researcher is responsible for driving the project.

The Research Leader will be expected to:

- Liaise with [data controllers/providers](#) about what data is required and whether their data can help;
- Complete all application forms for relevant data and approvals processes – many can be found in [RDS website](#);
- Learn about relevant information governance (IG) requirements in their own organisation (e.g. who is their Institution's Data Protection Officer, who can sign certain documentation, who controls the data involved);
- Respond to queries from data access panels/data controllers/ research coordinator/any support teams in a timely and professional manner; and
- Obtain relevant information governance training, for themselves and whoever will complete the analysis.

Training

Researchers will need to complete and provide proof of data protection/information governance training, which will remain valid for a set period of time, after which they will need to complete again.

The specific course may be dictated by the relevant data provider - for example, if a researcher is planning on using any data from ONS, they will need to complete the [ONS Safe Researcher Training](#).

Other providers, such as NHS Scotland, accept courses such as the

- [UKRI Using data about people in research](#)

- MRC '[Research, GDPR and Confidentiality Quiz](#) or the

- [10 bitesize modules on Research, GDPR and what you really need to know'](#) [online module.](#)

Some [data providers/controllers](#) include lists of approved courses in guidance notes or application forms. Researchers should check with their data provider to determine which course they will be expected to complete.

eDRIS and what they can provide

The electronic Data Research and Innovation Service (eDRIS) team is part of the Information Services Division (ISD) of Public Health Scotland (PHS). Their role is to facilitate and support the majority of research projects using health and administrative data in Scotland, for instance those projects funded under Health Data Research UK and the Scottish Centre for Administrative Data Research(SCADR).

Any researchers using the National Safe Haven or requiring health data from more than one NHS Scotland health board area will encounter eDRIS.

Projects that plan to hold or link non-health data in another location, for example within a university data safe haven may not need to interact with eDRIS. This handbook includes information for all types of projects in Scotland, but goes into further detail into the eDRIS process based on the large number of projects using the service.

Before going any further it is useful to refer to the Research Data Scotland webpages to view the eDRIS Pre-application checklist and the eDRIS initial enquiry form, as eDRIS deal with a large number of projects, and have a wealth of useful information about the feasibility and practicalities of doing research with administrative data – including the costs.

eDRIS will provide researchers with a dedicated research coordinator, who will support and advise on the various aspects of the research process, from applying for access to data; linkage; and producing outputs from research.

The level of eDRIS support required can include:

- Help with study design
- Provision of expert advice on coding, terminology, metadata and study feasibility
- Assistance with obtaining the required permissions for access to data
- Liaison between researcher and the data suppliers to secure data
- Support to access data within the National Safe Haven
- Provision of analyses, interpretation and intelligence about data (where required)
- Output checking and advice on statistical disclosure control to reduce the risk of identifiable data being published - note this check will always take place for some types of data

As mentioned above, an overview of the data access process for 2023 can be found on the RDS website, with further information on the eDRIS service is available on the Public Health Scotland (PHS) website, including a downloadable eDRIS Short Guide.

2. Developing an application

To begin research with administrative data, it is necessary to construct a well-defined research proposal which demonstrates the research is in the public benefit. Aligning with the interests of policy-makers can also often be desirable. This proposal should contain a set of research questions which can be feasibly answered by an included list of variables to be requested from specific data sources (along with the practicalities, legalities, and analytical methodologies of how this will be achieved).

Alongside the proposal, there are the various standard stages in any research process, such as providing a data protection impact assessment (DPIA) and completing an ethics application. A researcher's institution will have guidance on how to complete these, including what level of ethical approval required. On top of these, the researcher may have to facilitate data sharing and processing agreements between the relevant data controllers/processors. An eDRIS frequently asked questions guide is available online, which might help some researchers.

Establishing which datasets are available can be complex, depending on the nature of the project or study a researcher is planning. The process, and marrying available data with suitable research questions, can be time-consuming. Both must be complete before applying for access to data.

Establishing data availability

Establishing data availability can take time - particularly for those new to the administrative data environment in Scotland. eDRIS can provide assistance on public sector data, but the researcher will also need to do some work of their own to determine availability for any data that has not been previously used for research.

Recently ADR UK has created an ADR UK Data Catalogue, which makes it easier to discover the growing collection of administrative datasets that are available for public-good research. For each dataset there is a contact email to which researchers can direct further inquiries (although eDRIS can also help with many of the more general questions). Similar resources are available for Scottish Census data, data from census based longitudinal studies, and (to a lesser extent) for data relating to children and social care.

The NHS National Data Catalogue provides a full list of the datasets held by ISD, information about the datasets (population coverage, sometimes aspects of data quality, related publications), and a data dictionary of health and social care data definitions and standards.

Finding out about other sources of administrative data can be significantly more complex. Again eDRIS may be able to provide insight to researchers regarding datasets which have previously been involved in data linkage projects. Researchers will be required to establish:

A. what datasets relevant to their research aims are collected;
B. who to contact for information about specific datasets; and
C. if the available metadata is suitable enough to allow the development of a research proposal.

For datasets curated by ADR Scotland, the [dataset table includes links to metadata files](#). If the researcher cannot find metadata, it is possible to gather relevant information by discussions with the relevant contact who works regularly with the dataset.

In addition to the [ADR Scotland data records table](#) that is included within the RDS website, researchers may find the following respositories offer information on a variety of public sector datasets:

- NHS [National Data Catalogue (NDC)](#)
- [ONS Secure Research Service (SRS)](#)
- [Scottish Government Open Data Platform](#)
- [UK Data Service](#)

- [The Secure Anonymised Information Linkage (SAIL) Databank](#) is a safe haven for billions of person-based records on the population of Wales.
- [Northern Ireland Statistics and Research Agency](#).
- [UK census based longitudinal studies](#) (SLS, CeLSIUS, NILS)
- [Other resources](#)

Developing research questions

Alongside detailing the specific data requested, a research proposal must also define a set of explicit research questions which the researcher hopes to answer. Flexibility with respect to both research questions and data is usually required.

Researchers new to the data linkage world may initially feel faced with two options:
a) Develop a research question first, and then find the best dataset to answer it
b) Find a dataset and then develop research questions around it

Both of these approaches have challenges. Keeping firmly to a specific research question using existing data can be very challenging given the nature of administrative data (and if successful requires careful appreciation of the relative strengths/limitations of the data). The latter option of exploring a dataset to find an interesting question is not really possible, as approval is needed in order to access the data in the first place.

The more common approach is a fluid combination of the two. Researchers will tend to begin   with a reasonably broad scope of enquiry and perhaps a set of possible data sources (e.g. organisations that may collect relevant data). From this starting point, specific research questions often develop alongside the researcher's growing understanding of the available data and gaps.

The initial stage of constructing a research proposal can be an extended and iterative process, involving input and collaboration with eDRIS and other data users in some instances - largely due to the impact data availability has on the focus of research. Whether there is metadata available can add to the complexity of the process. It is important to remember that administrative data is not designed for research purposes, nor are data controllers likely to have research as a top priority. For researchers, the key here is to be prepared for extended periods of waiting or preparation, and to accumulate as much knowledge as possible about all and any potentially relevant data and access procedures.

Defining which data is required

What data can be requested

The aim of data linkage research is to allow researchers to do research using linked datasets without having access to any of the personal identifiable information (PII). However, the linking process still involves using and processing personal data, which carries privacy risks and requires careful handling. Even the use of pseudonymised data (in which identifiers are replaced by artificial pseudonyms) and de-identified data (from which direct and indirect identifiers have been removed) runs a potential risk of re-identification (see https://doi.org/10.1038/s41467-019-10933-3) and is therefore subject to data protection laws. These confidentiality risks, Data Protection laws and the fact that the majority of administrative data is not gathered with explicit consent, means researchers are only to be able to access the data they need to answer their questions and will need to be able to justify the need for each of the variables requested.

Additionally, data should be requested at the granularity necessary for the specific research purposes (e.g. researchers must justify the need for data-zone level vs local authority area level data). This is part of the data minimisation principle of the General Data Protection Regulation (GDPR), and researchers should bear this in mind when deciding what data is required - will differences between local authorities be compared? If so, why is datazone required? Similarly, if the researcher is interested in hospitalisations, what kind of conditions are they interested in? Researchers completing applications requesting 'all records' should be prepared to answer questions on why this is necessary.
For further information on data protection principles and guidance, please visit the RDS website.

Proxies/alternatives

When researchers face a lack of data pertaining to their research proposal, or potential issues with data quality, it is possible that there may be proxies for the variables they desire (which may be found either in the same dataset or elsewhere which could then be linked). For instance, data on prescribed drugs is at times used as a proxy for mental health diagnoses. Ideally mental health diagnosis would be obtained from primary care records, but at present these can be challenging and expensive to obtain. Prescribing makes a potential alternative as people typically receive medication that is relevant to their diagnosed condition, so it can be used to approximate diagnoses themselves. This approach carries with it certain caveats, which researchers must consider and decide if they are acceptable to their research proposal. Additionally, it is recommended to recruit a member of the research team with relevant knowledge in relation to the field - e.g. at what doses which drugs are likely to actually be prescribed for a mental health condition rather than for something else.
This attitude may also be required at the very broad level of the study design and links back to the fluid nature of developing research proposals using administrative data. For instance, if the researcher is unable to obtain data on violent crimes, it may be worth refocusing on hospital admissions for conditions caused by assault. Any adjustments such as these obviously entail careful consideration of how results are interpreted. These caveats and the limitations of the data must be acknowledged in any publications arising from the research.

Questions which must be answered during the application process

When considering the uses of datasets for an application, researchers are likely to find answers to specific questions relating to the datasets needed for the research proposal (e.g., sample sizes, timeframes, lists of variables), and specific metadata relevant to the variables required (e.g., variable completeness) - all of which are useful in the process. Further wider questions will also need to be considered, and answers documented within applications. It is recommended that the following questions are considered by the researcher early in the process:

- Is this an appropriate and sensible use of the data?
- Is the level of data detail appropriate?
- Where can the data be stored and accessed? Does the environment prevent the misuse of the data?
- Who will have access to the data/who is involved in the research project? Are they trustworthy? – How can they prove they understand what dealing with confidential data means (i.e. have they been trained?)
- Is there any confidentiality risk from publication? Is the research on such a small population that any publications will be identifiable?
- Who is/are the data controller(s)?
- How will the data controllers view this use of the data?
- Is there a process for applying for access? If so, what? If not, is it possible to set one up? For example, is there a data access panel to apply to, will a data sharing agreement (DSA) be required, if so between whom?
- What personally identifiable information (PII) is in the dataset? Is there any information that could be used to link to another dataset? Has it been linked before?
- Who is going to do the linkage?
- How is the data collected, and is it suitable for the research project - are there practical issues around those who input the data into the system? This point is pertinent for all research using administrative data (e.g. how do first responders code different types of incidents?).
- Are there likely to be inconsistencies in data collection (e.g., across different areas, time periods, or across both)?
- How might the data be influenced by targets/performance management in the organisations collecting it?
- How accurately are variables recorded? (e.g., a health practitioner might prioritise accurate recording of a patient's condition code over the recording of an external cause code)
- How long will the data need to be stored for? How long will publication take?

3. Practicalities

Where will the data be held?

Any data controller releasing their data for research will be concerned with where will the data be held. Linked data is often classified as personal data and therefore should be held in a suitably secure environment. Researchers should not expect to be able to download it to their work laptop.

It is possible that the data controller has a default expectation of where the data will be held - for example in one of the [Safe Havens/Trusted Research Environments (TREs)](#) that are set up around Scotland and the UK to hold confidential data for research (and are often the preferred option). However, safe havens or the data controllers may have certain restrictions - for example researchers may need to access the data from a specific location. This is important to be aware of if analysis must be performed in a limited time.

If a researcher is not using one of the safe havens (for example, when their organisation has an alternative [SafePod Network](#) which the data controller has indicated could be suitable), then they will be required in the various application procedures to provide specific details on the technological and security aspects of the system where the data will be held and accessed. These questions include:
- What are the access controls (physical and technological)?
- Does the system have any security accreditation (e.g. ISO27001)?
- How secure is the server? Is it within the EU?
- How/how often is data backed up?
- What security policies are followed?
- What audit measures are in place?
- How will data be deleted? How will backups be deleted?

Timescales involved

Administrative data research and data linkage projects are likely to take, at a minimum, several months of preparation/application by the researcher, followed by more months of extraction and data checking before the data can be accessed. Timescales will vary depending on factors such as the number of datasets involved, the number of data controllers, and how familiar the people involved are with data linkage. The process of refining the research questions based on discoveries about the data, gaining approval for use, and completing the relevant information governance paperwork will take coordination of several different teams who all have competing workloads. Often data controllers and processors operate queues to deal with this workload, so it may take time to assess and provide information on any particular project.

Linkage of data itself can be done fairly quickly, but there are a range of preparations, checks and processes the data needs to go through before it can be made available for analysis. Once a researcher gains access to the data, it is possible that further questions/issues may arise which may take further time to resolve.

Any outputs will need to be checked by a team for disclosure risk, again taking time. For this reason, it is recommended to be pragmatic as to what can be achieved in set timeframes - for example, if the researcher has limited time (such as PhD students), starting a data linkage project from scratch may not be the most appropriate route for their project.

Overall - it is important for researchers to build in enough time to complete all of the paperwork required, to be proactive in seeking out next steps, and to be patient.

What are the limitations of linkage?

While data linkage is a hugely useful resource, the benefits of which have been documented elsewhere, there are some limitations to it that researchers should consider.

For a dataset to be linked at the individual level, it will need to have some level of [personal identifiable information (PII)](#) within it, such as name, date of birth and location. Linkage rates are likely to be better for datasets with more identifiers: the best linkage is obtained when a dataset has reliable fields for first name, surname, date of birth, gender, home postcode. Records will also need a unique identifier, to be clear that records with the same identifier belong to the same person. Datasets with limited or no identifiers will have far less utility for linked data research.

Administrative datasets are likely to be large. Often cohorts will run into hundreds of thousands, and with repeated events (e.g. repeated years, repeated admissions) these will be multiplied to millions, with large numbers of variables. Overall it is possible that researchers will be working with millions of millions of records. Therefore processing will take time, both before data is accessed (during extraction and linkage) and during analysis. This also means that analytical methods utilised will need to be sufficiently nuanced to be able to account for the complexity and size of the data.

Finally, the [system of linkage](#) within Scotland necessarily requires no single party being able to see all of the linked data with the identifiers – meaning it is difficult to be absolutely certain at every step that the linkage is perfect. Administrative checks are completed before a researchers accesses the data to check the linkage, but it is vital the researcher also checks that the data is as expected.

4. Requesting Access to Data

The importance of visible public benefit

In order to gain access to administrative data in Scotland a research project must demonstrate a clear public benefit.

Projects should actively seek input from the public and maximise opportunities to identify the questions that society wants answered. One way to demonstrate this, especially if applying for data or funding from the Scottish Government (SG) is to outline how the research is in line with the Scottish Government's National Performance Framework and will support SG policies. The public engagement process could also involve public forums, policy makers, external experts (such as third sector organisations) or leaders of the public services. Demonstrating evidence of engagement with these non-researcher stakeholders will lend support to a researcher's application. In 2023, RDS joined SCADR's public panel, and created a joint public panel called Scotland Talks Data.

Researchers will be asked in all applications to demonstrate what public benefit the research will bring, and how the public have been involved in the development of the proposal. More information on how to involve the public can be found at
- Office for Statistics Regulation's review: Joining Up Data for Better Statistics, and
- UKRI 'Planning for Impact with Public Engagement and Beyond' guide.

Who are the main data controllers?

In order to access person level or household level data for research, researchers will need permission from the data controller. Data controller is a legal term under GDPR and may not be the same organisation as the one that provides the data (the data provider). For example, a local authority may commission a company to deliver a survey on attitudes to the local area. The company will collect and analyse the data, but it is most likely the local authority will be the data controller.

Researchers are likely to need to make an application to each data controller of the data they are requesting access to: i.e. if a researcher wishes to link data controlled by Scottish Government (SG) to health data that is controlled by NHS Scotland, they will need to make applications to both NHS Scotland and SG.

The landscape of data linkage is complicated, and it may not always be clear who the data controller is for the data which is of interest. It is possible that the variables of interest are controlled by one data controller (e.g. SG) but the variables that make the data possible to be linked (the PII) are controlled by another (e.g. local authorities). In order for a linkage study to be completed, approval and Information Governance (IG) documentation would be required for both data controllers.

Main Data Controllers in Scotland

This list gives some examples of the data held by these providers, but is not exhaustive. It should also be noted that some of these datasets may not be available for research or linkage.

NHS Public Health Scotland:
- Hospital activity statistics (in/outpatients, A&E, maternity, mental health)
- NHS24
- Prescribing information
- Audits/Registry (e.g., Cancer registry)
- Lifestyle (e.g., drugs misuse database)
- Community Health Index (register of unique identifiers of patients in NHS Scotland - used across health service data, enabling direct linkage)
- Social Care Survey (from 2018 onwards)
- Scottish Schools Adolescent Lifestyle and Substance Use Survey (SALSUS)

National Records of Scotland (NRS):
- Census data
- Marriages
- Births and Deaths

Scottish Government (SG)
Education Analytical Services (EAS) division:
- Additional Support Needs
- Scottish Survey of Literacy and Numeracy
- Educational Attainment
- Care of Looked after children
- Child protection

Justice Analytical Services (JAS) division:
- Scottish Crime & Justice Survey (some data available on UK Data Service)
- Court proceedings & Convictions (monthly snapshot from Police Scotland)
- Criminal Justice Social Work
- Liquor Licensing (Local Authority area level)
- Homicide (annual snapshot from Police Scotland)
- Recorded crime (quarterly snapshot from Police Scotland)
- Prison population (monthly snapshot, missing 2014/15)

Other SG Surveys – note some are available as standalone datasets (i.e. not for linkage) via the UK Data Service:
- Scottish household survey (SHS)
- Scottish health survey (SHeS)
- Growing up in Scotland Survey
- Social Care Survey (1998 - 2017)

Scottish Prisons Service:
- Prisons database
- Scottish Prisons survey

Police Scotland/Scottish Police Authority:
- Calls
- Incidents
- Crimes
- Custody
- Convictions

Local authorities
- Liquor Licensing
- Homelessness
- Social Work information

Other surveys/studies in Scotland and their controllers are:
- Scottish Longitudinal Study
- Scottish Mental Survey 1947/Lothian Birth Cohort - University of Edinburgh
- Aberdeen Children of the 1950s study (ACONF) - University of Aberdeen
- Edinburgh Study of Youth Transitions and Crime - University of Edinburgh
- Generation Scotland – University of Edinburgh

UK-wide data controllers - note at present not all of this data is available for research/linkage:
Department for Work and Pensions (DWP)
- Job Seekers Allowance
- Employment Support Allowance

Her Majesty's Revenue & Customs (HMRC) – some data may be available via the HMRC Datalab:
- Income Tax
- Tax Credits
- Child Benefit

Ministry of Defence:
- Armed forces and Service leavers personnel

Ministry of Justice – note some data now available through the ADR UK Data First project:
- Magistrates court appearances
- Crown court appearances

Access process

Some data providers will have a defined data access process and requirements for the use of their data. Below is a brief overview of the current main channels in Scotland.

NHS Public Health Service

For health data there is a central point of application for all research requesting data for more than one health board area in Scotland. This in the form of the [Public Benefit and Privacy Panel for Health and Social Care](#) (Health PBPP). This panel process ensures that the proposal is feasible, in the public benefit, and that the correct protocols are in place to ensure the privacy and security of the study data. Please note, requesting data from GP surgeries is not covered by this panel as each GP is a data controller in their own right.

The electronic Data Research and Innovation Service (eDRIS) team are part of Public Health Scotland. eDRIS Researcher Coordinators (RCs) work in collaboration with researchers, providing support and advice through the data access process.

All necessary data protection documentation should be attached with a Health PBPP application. This includes a [data protection impact assessment (DPIA](#)), any relevant data sharing agreements and data controller-processor contracts which are in place.

Applications to Health PBPP must go via eDRIS via the following steps – outlined in the [RDS website](#):

1. Researcher an [eDRIS enquiry form](#)
2. eDRIS will advise on feasibility and provide an estimate of costs
3. Researchers complete a [Public Benefit and Privacy Panel for Health and Social Care](#) (Health PBPP) application form. There is [guidance for applicants](#) providing information on how to fill out each section of the form, and eDRIS will assign a research coordinator to the project who can assist.
4. When all required documentation has been submitted and the panel have issued a full approval letter, eDRIS will coordinate the extractions from the relevant parties.
5. Linkage (where necessary) is done by trusted third party (National Records of Scotland, or a team within Public Health Scotland (PHS))
6. The eDRIS research coordinator will check and arrange access to the final linked datasets.
7. The linked data will made available in accordance with the permissions agreed with the data controllers for the project. Where applicable, eDRIS will provide a secure environment in the National Safe Haven for the analysis. In these cases, the linked data will be placed into a study area which can be accessed only by approved researchers in the research team.

Scottish Government

Following a similar procedure as NHS PHS, data access requests to the Scottish Government are made via the [Statistics Public Benefit and Privacy Panel (Statistics PBPP).](#) The Statistics PBPP is the access process for all data held by SG and the Scottish Census (controlled by NRS). Like the health counterpart, Statistics PBPP applications require providing data protection documentation such as a DPIA, and any [data sharing agreements](#) (where necessary).

The Scottish Government provides [guidance on requesting data](#). As of October 2018, the Education Data Access Panel is no longer active and has merged with Statistics PBPP.

NRS

National Records of Scotland control a variety of datasets with slightly different application procedures. Scottish Census data are requested through an application to the [Statistics Public Benefit and Privacy Panel (Statistics PBPP)](#). Both the panel and NRS will have to agree that your project has public benefit and the request is proportionate for your project to be approved.

Metadata for the Census 2011 can be found on [the Scottish Census website](#). For enquiries relating to accessing census data, contact dataaccess@nrscotland.gov.uk or an eDRIS research coordinator.

NRS Vital Events data (births, marriages, deaths) are managed via the eDRIS service and the [Public Benefit and Privacy Panel for Health and Social Care](#) - see above.

Police Scotland

Police Scotland are currently developing the processes for both the information governance and the practicalities involved in sharing data. This has been ongoing for several years due to the large numbers of systems involved. The academic research team in Police Scotland have recently been moved to a more central strategy position, which is hoped will support progress.

At the time of writing, accessing individual (person) level data from Police Scotland is unavailable, however, it may be possible to [request data at a geographic level](#) which may be suitable for research purposes. If researchers require access to Police data or you wish to arrange interviews with police personnel, contact Police Scotland's Academic Research Team at partnershipresearchethics@scotland.pnn.police.uk

Local Authorities

There are 32 local authorities in Scotland, each are a separate data controller, meaning that requests for data will need to be sent to all 32. Finding the person responsible (the information asset owner (IAO)) for each dataset is likely to be difficult and require a lot of persistence. However, it has been successfully completed previously, with a one-off SG-funded Health and Homelessness Data Linkage project. Steps are being made towards overarching arrangements which would make homelessness data more accessible for research without referring to each local authority individually, this is expected to be in place by end 2023.

SLS

The Scottish Longitudinal Study (SLS) is a large-scale linkage study on 5.3% of the Scottish population, created using data from administrative and statistical sources, and a standing dataset of linked records including: census data from 1991 onwards; vital events data (births, deaths, marriages); NHS Central Register data (gives information on migration into or out of Scotland); and education data (including Schools Census and SQA data). Timelines for accessing SLS data are likely to be considerably shorter than other projects, and researchers may find it worthwhile to consider whether a 5% sample of the population will meet their needs and accelerate their project.

An online guide informs researchers on how to go about doing research with the SLS. Researchers can find metadata, a helpdesk to discuss ideas, lists of outputs & projects, and guides for the research process.

5. Costs

Why there is a cost for linkage

Researchers are not charged for data in Scotland. However, the process of supporting researchers through the application, information governance and linkage process requires several teams with specialist knowledge, who often work on a cost recovery model. It is worth checking with data providers if there are any costs expected.

eDRIS, for example, charges for research co-ordination services. Charges are based on the institution leading the research and the complexity of the study. The pricing guide is available on the eDRIS website. After the initial scoping of the project is complete, eDRIS will provide researchers with a quote for the service. eDRIS request that researchers have funding secured before engaging their services beyond initial enquiries.

When to ask about costs

eDRIS and data providers will need enough information to understand the scale of a project before being able to give a reliable indication of costs. This may be after the researcher has completed a lot of work regarding data availability and defining the research questions, but is likely to be before a full application is drafted. Researchers should be aware that costs can change as the application process further defines a project.

Funding Options

A researcher may wish to build in data linkage costs into a funding application or apply for a small grant to cover the charges. Links to grant awarding bodies can be found on the eDRIS useful links page.

6. How does data linkage work?

For datasets to be linked at the person level (i.e. records for one person in both datasets), each dataset must have:

- some personal identifiable information (PII): e.g. name, gender, date of birth, postcode. The quality of the linkage will depend on how many of these fields are present, how many records are completed and how reliable it is
- a unique identifier (ID) for each entity in the dataset (e.g. each person)

As a result it is not possible to link an anonymised dataset at the individual level.
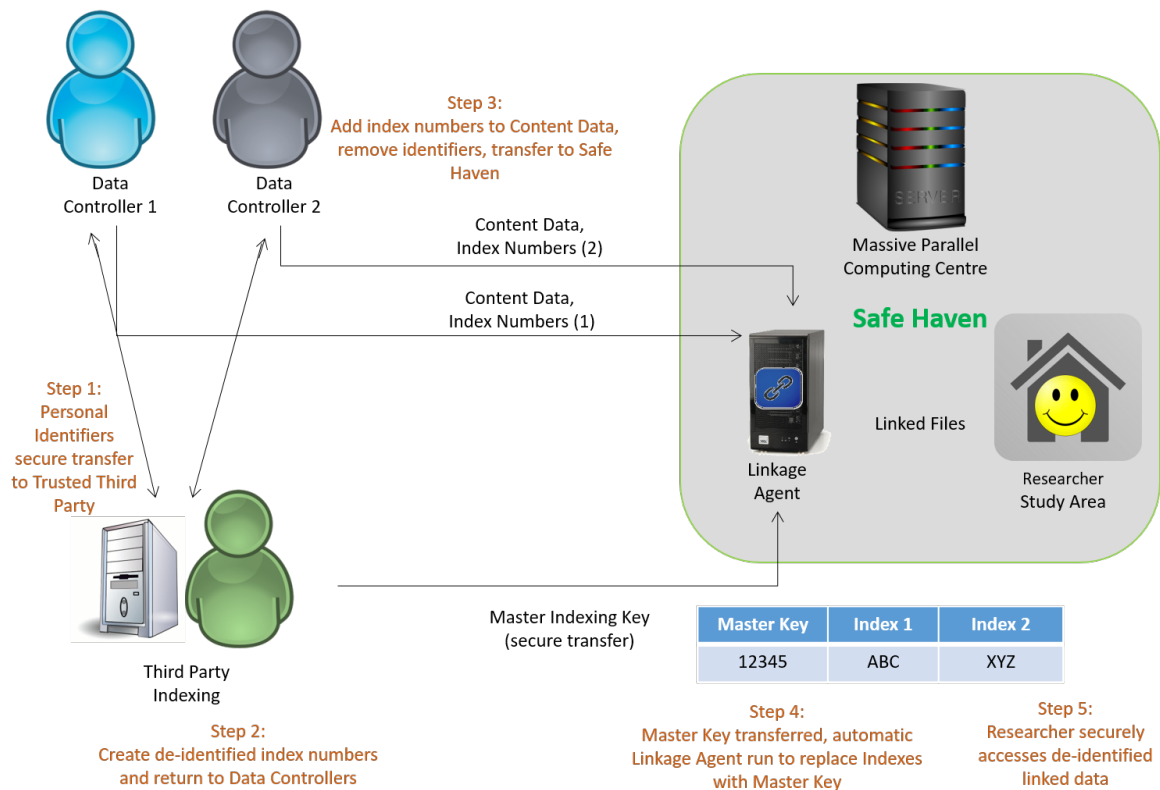
The steps describing the process of linking data using the eDRIS service are detailed below. This may vary for some data providers but provides an overview of one method of safe data linkage.

Data for a project will be linked using established probabilistic matching techniques based on the Howard Newcombe principles. Linkage will be undertaken by a trusted third party and a population spine used as an intermediary linkage tool. The population spine used by PHS contains the personal identifiers of all individuals in Scotland who have been in contact with NHS Scotland.

Process

1. Data providers will supply personal identifiers (only) plus their own person or record ID number to the indexing team.
2. The indexing team will probabilistically match the identifiers to the population spine using complex algorithms.
3. The data provider will receive a file back with their own person or record ID number and a unique person index ID number specific to that dataset. This is generated by the indexing team.
4. The data provider will attach the received index ID number to the remaining content of the dataset to be provided for linkage (also called content data or payload data) and send to eDRIS.
5. The eDRIS Research Coordinator will confirm that the agreed data has been received and send the file to the linkage agent. The linkage agent is an automated computer programme which carries out the linkage.
6. The linkage agent will receive all the datasets and their unique person ID numbers plus a master control file containing a master person ID and all the dataset unique person index ID numbers.
7. The linkage agent will then replace all the dataset-specific person ID numbers with the master person ID number on each of the content data files
8. The linked datasets are transferred securely to the approved analysis location for researchers to access. This allows a researching analysing the data to see all the records belonging to an individual across all the datasets without the need to see the personal identifiers.

A diagram depicting this process is below:

Step 3:
Add index numbers to Content Data, remove identifiers, transfer to Safe Haven

Content Data, Index Numbers (2)

Content Data, Index Numbers (1)

Data Controller 1

Data Controller 2

Step 1:
Personal Identifiers secure transfer to Trusted Third Party

Massive Parallel Computing Centre

Safe Haven

Linked Files

Linkage Agent

Researcher Study Area

Third Party Indexing

Step 2:
Create de-identified index numbers and return to Data Controllers

Master Indexing Key (secure transfer)

Master Key	Index 1	Index 2
12345	ABC	XYZ

Step 4:
Master Key transferred, automatic Linkage Agent run to replace Indexes with Master Key

Step 5:
Researcher securely accesses de-identified linked data

Data linkage for research in Scotland

The Scottish Government has been committed since 2020 to improving Scotland's future through the production of high quality data, analysis and evidence to inform policy and support improved service delivery. One approach they are taking to achieving this is through data linkage; Joined Up Data for Better Decisions outlines their strategy for better use of data that already exists in Scotland including the establishment of a Data Linkage Framework, with the Guiding Principles for Data Linkage supporting the creation of a culture where legal, ethical and secure data linkage is accepted and expected.

7. Information Governance (IG)

Key considerations

The process of linking administrative data is highly likely to involve processing of personal data. As a result, a range of safeguards must be put in place by law before a researcher may access the data. This is to protect the data subject's privacy, enables fair and legal access to data and is an essential step in allowing researchers to access and analyse such a large amount of data in a safe manner. Research Data Scotland have information on their website to help researchers navigate data security.

The Information Commissioner's Office (ICO) provides comprehensive information on GDPR, the documentation that is required, and support on how to complete it. Its website answers the following questions:
- What is personal information?
- Who does the GDPR apply to?
- Can we identify an individual directly from the information we have?

All information governance documentation is designed to demonstrate the degree to which each proposed project follows the Data Protection principles required under GDPR. These are covered more fully on the ICO website but briefly, the GDPR sets out seven key principles for using personal data:
- Lawfulness, fairness and transparency
- Purpose limitation
- Data minimisation
- Accuracy
- Storage limitation
- Integrity and confidentiality (security)
- Accountability

These principles should lie at the heart of a researcher's approach to using linked administrative data for research.

Contacting the relevant Data Protection Officer (DPO)

All organisations must have a Data Protection Officer by law. For academic researchers, the DPO of their institution will be able to advise on data protection requirements for research. It is recommended in all cases that researchers contact their DPO during the research proposal stage of their project for guidance and support.

DPIAs and why they are required

A data protection impact assessment (DPIA) is a process to help users of data identify and minimise the data protection risks of a project.
It is a legal requirement under GDPR to complete a DPIA in some circumstances, and it is recommended for all data linkage projects as good practice, indeed some data providers will not allow access without a completed DPIA.

A DPIA must:
- describe the nature, scope, context and purposes of the processing;
- assess necessity, proportionality and compliance measures;
- identify and assess risks to individuals; and
- identify any additional measures to mitigate those risks.

To [assess the level of risk](#), researchers must consider both the likelihood and the severity of any impact on individuals. High risk could result from either a high probability of some harm, (e.g. individual level routine dental data could easily be published online, causing some distress to a lot of people) or a lower possibility of serious harm (e.g. convictions of everyone in a small town could be sent to the residents, causing a lot of distress, for both victims and perpetrators).

Researchers can find additional support from their DPO, the [ICO website](#), and relevant experts (e.g. [eDRIS](#)).

Legal Gateways and Lawful Bases

In documentation (e.g. DPIAs and application forms), it is necessary to outline the legal gateway for accessing the data involved in a project. This is to comply with the GDPR Data Protection Principle 1 - that data should be processed fairly, lawfully, and in a transparent manner. The legal gateway may depend on the purpose of the research, or the researcher's institution's standing as a research institution. Researchers can seek support from their DPO in identifying the legal gateway for accessing the data.

In addition to a legal gateway, [the lawful basis](#) for processing the data for a project must be identified, along with a second lawful basis for processing any [special category data](#) i.e. anything to do with health, ethnicity, or any other sensitive topic. Lists of lawful bases for processing are available on [the ICO website](#).

Data controllers and data processors

Data providers will generally be the data controllers of their data (although this may not always be the case, or it may not be clear exactly who the data controller is, [see data protection terms](#)).

- The legal relationship a researcher has with the data provider will be defined by the data provider.

- The responsibilities of parties involved will depend on their status as a 'data controller' or 'data processor', and

- a researcher's institution will be required to sign an agreement that the researcher understands their status, and that they will abide by the responsibilities of their role under GDPR.

It is recommended that researchers find out who must sign these agreements for their institution as early as possible as it can take some time for the right department to process the information.

When the data is shared with the researcher, they may become a data controller of that data, either alone or jointly with the data provider. In this case, the researcher's institution will have a data sharing agreement (DSA) with the data provider - a researcher's DPO should be able to advise who can sign a DSA for their institution.

Data Controllers:
- Decide what the purpose or outcome of using the data
- Have autonomy as to how the personal data is processed
- Appoint processor(s) to process the personal data on their behalf

Alternatively, a researcher may be a data processor of the data. In that case their institution will have a data processing agreement (DPA), or controller/processor contract with the data provider. These are legal documents and should be sent to the researcher's institution's data protection or legal department to be signed.

Data Processors:
- Follows instructions from someone else regarding how the data is processed
- Do not decide what personal data should be collected from individuals. Processing of the data are implemented by a contract with someone else
- May make some decisions on how data is processed, but implement these decisions under a contract with someone else.

More information can be found on the ICO website.

In addition to the agreements between the researcher's institution and the data provider, it may be that further agreements are required with the institution or facility which will hold or link the data. For example, for non-health data using the National Safe Haven, a controller-processor contract with eDRIS is required as they are processing (i.e. holding/linking) the data on the researcher's behalf. eDRIS can provide their standard contract. If eDRIS are not involved, the researcher may consult their institution's research support office or data protection team to obtain a relevant controller-processor contract.

8. What to expect when accessing data

Safe Havens

A safe haven is a Trusted Research Environment (TRE) or Secure Data Environment (SDE) is a technical solution for the secure storage and handling of de-identified data. It provides a secure location to hold data, and an analytical platform for researchers to access the data they have permission to view and analyse for their project. A safe haven is not typically connected to the internet, and there are strict access controls on who can and cannot access, and the software available within it.

There are four [accredited safe havens](#) set up across Scotland. These were originally set up under the Farr Institute for health research, and so are mostly health focused. Local Safe Havens operate in the regional hubs:

- [Grampian Data Safe Haven (DaSH)](#) Aberdeen
- [Health Informatics Centre (HIC)](#) Dundee
- [DataLoch](#) Edinburgh and
- [West of Scotland Safe Haven](#) Glasgow

These are each individual safe havens with their own networks and security protocols. Data controllers may require researchers to use one or another of these, or researchers may be able to hold the data elsewhere. Most projects using data relating to all of Scotland, store their data in the National Safe Haven, which is managed by the [eDRIS](#) team.

Other data centres across the UK have their own secure locations to access data, such as the [ONS Secure Research Service](#), or [UK Data Service Secure Lab](#).

In 2023, over £2 million has been awarded by UK Research and Innovation – the UK's largest public funder of research – to fund a series of Driver Projects as part of the [DARE UK (Data and Analytics Research Environments UK)](#) programme. The projects will take a leading role in informing the design of what will be a UK-wide network of trusted research environments (TREs) – highly secure digital environments that provide access to sensitive data for analysis by accredited researchers.

Physical or Remote Access

Some safe havens providers allow remote access, where the researcher's institution's computer equipment can be used from the institution to access the data via a secure portal. However, there may be access controls put in place by the data provider, or based on the organisation accessing the data, which means physical access to a secure room (called a 'safe setting') is required.

The National Safe Haven allows remote and physical access. Physical access can be found at:

- [SCADR office within Bayes Centre](#), Edinburgh
- [Robertson Centre](#) - Glasgow
- [Health Informatics Building](#) building/University of Aberdeen - Aberdeen
- [HIC building](#), Ninewells Hospital - Dundee
- [SafePod](#) access through the [SafePod Network website](#) (In 2023, eDRIS is registering the Scottish NSH with the SafePod Network to increase the number of data access points)

What happens at a safe setting

If researchers are required to access a safe haven physically (via a 'safe setting'), they will be required to bring identification to verify that they are the researcher who has permission to access the data. Belongings will be kept in a locker on site as no personal belongings are allowed in the safe setting - this includes mobile phones and smart watches. The researcher will be on CCTV for information security purposes.

Terminals within safe settings only access the safe haven and the software within it. If extra software is needed, researchers should contact a member of staff who will be able to advise if this is possible. It is advisable to check this early on in the application process to ensure the analysis can be completed using the software available.

9. Outputs

Statistical disclosure control (SDC) - what it is and how it works

Researchers will sign several agreements throughout the course of the application and documentation process that say any outputs will not be able to identify an individual - therefore researchers should not attempt to publish any output that immediately identifies individuals or allows someone to learn something new about an individual. This is known as disclosure. One of the final steps for a researcher using any personal or individual level data is to undertake statistical disclosure control (SDC) on research outputs - this means using statistical techniques to make the outputs less identifiable.

Analysing linked data is likely to show that there are some small groups or unique individuals within the data. Any outputs of analysis that a researcher wishes to remove from the safe haven, for discussion amongst their project team, or publication, should not identify these. Even basic descriptive analysis - a table showing the maximum and minimum income within the data - is likely to have a risk of disclosure and therefore a) is unlikely to be released from the safe haven and b) should not be published.

Data controllers and processors have a responsibility to manage the risk of directly or indirectly reidentifying any individuals from linked data outputs. For outputs from a safe haven, the outputs will be checked by staff within the centre. For example, in the National Safe Haven, all outputs are checked by two members of the eDRIS team. This process is likely to take time, which researchers should consider when requesting output to be released (i.e. not requesting a release at 5pm for a conference the next day).

SDC protocols set out guidelines for assessing the risk of disclosure. Different data providers may have different requirements or protocols for researchers to follow, so it is advisable to ask what these are before attempting to release or publish anything. The emphasis in eDRIS is on having principles for researchers to follow, while allowing the flexibility for judgement to be made on a case-by-case basis.

Disclosure risk of an output will depend on:
- how sensitive the topic is;
- cell values/table design;
- populations, geographies and institutions involved;
- the likelihood of an attempt to identify an individual;
- the level of impact of any disclosure.

As general guidelines, researchers should avoid producing outputs with small cell numbers (e.g., <5 or <10, or models with df<10), or where individuals are represented graphically in any way (for instance as outliers, as minima/maxima, or in residual plots). Researchers should also think about any previously released outputs, as cross-tabulating between outputs might be disclosive. The best way to make outputs easily checkable and releasable is to do complicated analysis on large numbers of people, and to explain exactly what each output refers to and its purpose. Information contained in labels, titles, descriptions all make it easier to check for disclosure.

There are several guides and training on how a researcher can complete SDC. Researchers accessing any administrative data under the Digital Economy Act 2017, or accessing any data via ONS are required to attend ONS approved Safe Researcher Training, which includes guidance on disclosure control, and the 'five safes' of data research.

The guides below cover statistical disclosure control (SDC) and output requirements in more depth for specific data providers:
- eDRIS requesting outputs from SH and disclosure control
- Secure Data Access Professionals Group SDC Handbook (2019)
- Scotland Census 2001
- Disclosure control for tables produced from administrative sources
- Public Health Scotland

Appendix 1 – Glossary

Terminology	Acronym	Description
Administrative Data		Routinely gathered information, either electronic or paper, which is collected when registering people or carrying out transactions, or for record keeping – usually when delivering a service. Such data can include: social security payment records, educational attainment records, health records, court records, tax records, etc.
Administrative Data Research UK	ADR UK	A UK wide partnership transforming the way researchers access the UK's wealth of public sector data, to enable better informed policy decisions that improve people's lives.
Administrative Data Research Scotland	ADR Scotland	A partnership between the Scottish Centre for Administrative Data Research (SCADR) and Scottish Government (SG) with a vision to support researchers and data controllers to share data efficiently and effectively for the public good
Administrative Data Research England	ADR England	ADR England is a portfolio of data linking and research projects, delivered by a range of academic and government partners. ADR England projects provide policy-relevant insights using data held by UK Government departments and public bodies.
Administrative Data Research Ireland	ADR Northern Ireland	Administrative Data Research Northern Ireland (ADR NI) is a partnership between the Administrative Data Research Centre - Northern Ireland (ADRC NI) and the Northern Ireland Statistics and Research Agency (NISRA). The partnership is supported by the Health and Social Care Research and Development (HSCR&D) Unit.
Administrative Data Research Wales	ADR Wales	ADR Wales unites specialists in each field from Population Data Science at Swansea University and the Wales Institute of Social and Economic Research and Data (WISERD) at Cardiff University with statisticians, economists and social researchers from Welsh Government. The cutting-edge data analysis techniques and research excellence developed, along with the world-renowned SAIL Databank
Content Data		AKA payload data - the de-identified variables researchers receive for their research project, as specified in the research proposal and applications
Centre for Excellence for Children's Care and Protection (CELCIS)	CELCIS	CELCIS is a leading improvement and innovation centre in Scotland. They are passionate about improving children's lives by supporting people and organisations to drive long-lasting change in the services they need, and the practices used by people responsible for their care.
Centre for Longitudinal Study Information & User Support	CeLSIUS	Works closely with the ONS Longitudinal Studies (LS) Unit to provide custom analyses from the English Longitudinal study upon request by academic users. Analysis can only be run on ONS computers

Community Health Index	CHI	A unique number given to every individual who registers with the NHS in Scotland - often used to link Scottish health datasets together
Community Health Index Linkage and Indexing Team	CHILI	PHS internal team responsible for indexing health and new datasets to the CHI database
Data Access Panel	DAP	A generic term for a panel within an organisation that grants access to the organisation's data.
Data Controller	DC	A person or organisation who determines the purposes and means of processing personal data. Holds certain responsibilities in relation to the data and holds responsibility for sharing of the personal data
Digital Economy Act	DEA	Legislation enacted in 2017 to support the use of the digital economy in the UK, broadly designed to regulate electronic communications infrastructure and services. The Act includes provisions relating to (among other things) data sharing between Government departments to create a Digital Government - and is therefore a legal gateway for some Government data. At time of writing, the processes to access data under the DEA are in draft.
Data Linkage		A process of joining two or more datasets on a common variable at a particular level. Often the datasets are owned by different data controllers and the levels can be either individual (i.e. linking records about a single person in two or more datasets) or geographical (e.g. linking information about the number of people in a datazone to environmental information about that datazone).
Data Provider		A person or organisation providing data for a research project. May or may not be the data controller.
Data Processor		A person or organisation responsible for processing data on behalf of a data controller. Has specific legal responsibilities under GDPR.
Data Protection Act 2018	DPA 2018	The data protection law currently active in the UK, enacts several UK based parts of the GDPR.
Data Protection Impact Assessment	DPIA	A risk assessment document that identifies, analyses and outlines the process to minimise the data protection risks of a project or plan. Required for almost all data linkage projects in Scotland.

Data Protection Officer	DPO	A legally required role for public bodies and any organisation processing large scale processing of personal data, responsible for advising and monitoring compliance with all data protection laws and first point of contact for supervisory authorities on data protection.

Data Processing		The collection and manipulation of items of data to produce meaningful information. Includes collecting, cleaning, manipulating, analysing, storing and sharing
Data Sharing Agreement	DSA	A formal agreement used when sharing data that clearly documents what data are being shared and how the data can be used.
Education Analytical Services	EAS	A division within Scottish Government that is responsible for Education data.
Electronic Data Research and Innovation Service	eDRIS	A team of health data experts who support researchers wishing to request and make use of administrative datasets. Provide a single entry point and end-to-end support to researchers. Part of the Information Services Division (ISD) within NHS Scotland. Provides services to ADR Scotland researchers for the Scottish Centre for Administrative Data Research.
Edinburgh Parallel Computing Centre	EPCC	Part of the University of Edinburgh, providing novel and secure computing facilities and support to academia and industry.
Economic and Social Research Council	ESRC	Funding body supporting research on economic and social issues, part of UKRI. Funds ADR UK.
General Data Protection Regulation	GDPR	Regulation in EU law on data protection and privacy for all individual citizens of the European Union (EU) and the European Economic Area (EEA). Brought in May 2018. Enacted into UK law via DPA 2018.
Government Statistical Service	GSS	A community of all civil servants working in the collection, production and communication of official statistics.
Growing up in Scotland	GUS	A longitudinal research study, funded by Scottish Government, tracking the lives of around 14,000 children in Scotland born between 2002 and 2017.
Health Data Research UK	HDR UK	The UK national institute for health data science.
Indexing		The process of comparing if people in more than one dataset are the same person, and the data can be linked. In Scotland this is done via probabilistic matching the PII in the data to the Population Spine.
Information Asset Owner	IAO	A senior/responsible individual in an organisation with the responsibility to understand what information is held, what is added and what is removed, how information is moved, and who has access and why. This may be for a single dataset, or several data holdings within an organisation. As a result IAOs are able to understand and address risks to the information, and ensure that information is fully used within the law for the public good.

Information Asset Register	IAR	A list of information assets (can be data sets, databases, etc.) that an organisation holds.
Information Commissioner's Office	ICO	The UK's independent body set up to uphold information rights - including under GDPR and DPA 2018.
Information Governance	IG	The management of information at an organisation. Information governance processes and paperwork set the requirements and standards to ensure organisations handling data fulfil their obligations to handle information securely, efficiently and effectively.
Information Services Division	ISD	A division within NHS NSS which provides health information, health intelligence, statistical services and advice that support the NHS in progressing quality improvement in health and care and facilitates robust planning and decision making.
Justice Analytical Services	JAS	A division within SG that handles data relating to the criminal justice system.
Master Index Key/Master Control File		A lookup table containing the index numbers of all datasets required for the specific linkage project and how they relate, along with a master index number.
Metadata		Information that provides information about other data - in most cases in data linkage, metadata refers to information relating to the structure of the data, what variables are contained, what time period and population is covered, and what the linkage possibilities are.
Memorandum of Understanding	MOU	A type of agreement between two or more parties showing an intention of all parties on how to act. Not a legally binding document but often used instead of a data processing agreement between governmental bodies.
Medical Research Council	MRC	Part of the UKRI, a funding body which supports scientific research by funding projects to improve the health of people across the UK.
National Health Services - National Services Scotland	NHS NSS	A national health board of NHS Scotland responsible for shared services across Scotland and controls some national level health data.
National Health Services - Public Health Scotland	PHS	A national health board of NHS Scotland responsible for shared services across Scotland - also houses the eDRIS function.
National Records of Scotland	NRS	National Records of Scotland (NRS) was established on 1 April 2011, following the merger of the General Register Office for Scotland (GROS) and the National Archives of Scotland (NAS). Its purpose to collect, preserve and produce information about Scotland's people

		and history and make it available to inform current and future generations.
NHS Central Registry	NHS CR	A record of everyone who is known to the NHS in Scotland - forms the basis of the Population Spine.
Northern Ireland Longitudinal Study	NILS	A data linkage study of 28% of the Northern Ireland population including records from 1981 - date.
Northern Ireland Statistical Research Authority	NISRA	An executive agency within the Department of Finance (NI), the principal source of Official Statistics and social research in Northern Ireland, incorporates the General Register Office (GRO).
National Statistics Authority	NSA	An independent body in each country that have the objective of promoting and safeguarding the production and publication of official statistics that serve the public benefit.
National Statistician's Data Ethics Advisory Committee	NSDEC	A panel that provides an opinion on the ethical appropriateness of research proposals from the Office for National Statistics (ONS) and the Government Statistical Service (GSS).
Office for National Statistics	ONS	The national statistical institute of the UK and the UK's largest provider of Official Statistics.
Office for Statistical Regulation	OSR	The regulatory arm of the UK Statistics Authority (UKSA).
Official Statistics		Any set of statistics produced by an organisation named under secondary legislation and described by that organisation as an official statistic or part of a set of official statistics.
Payload Data		The de-identified variables researchers receive for their research project, as specified in the research proposal and applications.

Personally Identifiable Information	PII	Information that can directly identify an individual person without reference to any other information - for example, name, date of birth, postcode, genetic information.
Population Spine		A common data set against which PII is indexed when datasets are being linked. The Scottish population spine is based on the NHS CR merged with some other records, held securely by National Records of Scotland.
Public Benefit and Privacy Panel for Health	PBPP	A panel of information governance and data experts that provide approval for data to be used - by examining the public benefit vs the privacy risks.

Research Data Scotland (RDS)	RDS	RDS work with researchers, analysts and policymakers to unlock the potential of public sector data for the benefit of public good.
Scottish Centre of Administrative Data Research	SCADR	Research centre that analyses data from across the public sector, exploring what linking it in new ways can reveal. Alongside the Scottish Government, forms ADR Scotland, the Scottish element of the ADR UK (Administrative Data Research UK) partnership, working to improve access to administrative data across the country.
Scottish Government	SG	The devolved government for Scotland. Has a range of responsibilities that include: the economy, education, health, justice, rural affairs, housing, environment, equal opportunities, consumer advocacy and advice, transport and taxation.
Scottish Longitudinal Study	SLS	A study following 5.3% of the Scottish population from the 1991, 2001, 2011 Censuses, linked to Education and other data to answer research questions.
Statistics Public Benefit and Privacy Panel	SPBPP / Statistics PBPP	The single access point for anyone wanting to access data held by Scottish Government or the Scottish Census.
Trusted Third Party	TTP	The party responsible for indexing the personal identifiable information (PII) to a spine - meaning PII is kept separate from all other data.
Urban Big Data Centre	UBDC	Research centre promoting the use of big data and innovative research methods to improve social, economic and environmental well-being in cities. Jointly funded by the ESRC and the University of Glasgow.
UK Research & Innovation	UKRI	A collection of funding bodies who provide funding for research.
UK Statistics Authority	UKSA	The UK National Statistics Authority.

SCADR PUBLICATION

Original content created in Nov. 2019, with links and content updated by September 2023.

RESEARCHER HANDBOOK