

The data journey, from collection to use in research

Diabetes and education data research

DARE UK Public Dialogue workshops, January 2022

I'm now going to show you an example of what the data journey looks like from its original collection to its use in research. This is a hypothetical example, but the journey reflects that taken for real life projects.

1. Public services are used



- A child living with diabetes goes to an NHS hospital for an annual care check, where the doctor asks questions and records information about how they are feeling, such as how tired they are and how their appetite is. They also make a record of what medication they are receiving, and measure the child's weight, blood sugar levels and blood pressure, amongst other things.
- After the appointment, the child goes to school, where it's test day.

2. Records are made



- Following the hospital appointment, the doctor makes a record in the hospital's computer system, recording all of the details they collected about the patient's symptoms, their blood measurements and any medication they are taking. These notes are attached to the patient's personal health record, which includes personal details such as their name and address. The record is important for monitoring the patient's personal care.
- At school, the teacher makes a record of the child's test result in the school system. The result is attached to the child's personal school record, which includes personal details such as name and address. The record is important for monitoring the child's personal educational attainment.

3. Records are transferred



Llywodraeth Cymru
Welsh Government

- The place where the records are then transferred differs for each UK nation, though the overarching process I will present here is similar for each. For the purpose of this example, we will use Wales.
- The child's hospital appointment record then gets transferred from the hospital database to NHS Digital (the body responsible for keeping healthcare data safe and using it to improve understanding of health problems), where it is stored in the National Diabetes Audit. The audit was developed to help improve the quality of diabetes care in England and Wales, so the transfer of data is important for this purpose.
- The child's school record gets transferred to the Welsh Government, which it collects as an important part of monitoring and improving school performance.

4. A research team wants to access the records



- After going through a competitive funding application and successfully receiving funding from the Economic & Social Research Council to undertake a project exploring the relationship between diabetes and education in Wales, a team of academic researchers now needs to apply to access the data needed for the project.
- For their project, they need to link together the National Diabetes Audit, collected by NHS Digital, and Welsh education data collected by Welsh Government.
- This part of the process begins with an initial discussion with the data custodians (the people responsible for keeping the data safe and secure) about the details of the project, and involves working out how the data can be accessed and linked within the appropriate legal and ethical framework.

The screenshot shows the NHS Digital website. At the top right, there are logos for HDRUK (Health Data Research UK), ADRUK, and UK Research and Innovation. The main heading is "5. Applications are made to access the data". Below this is the NHS Digital navigation menu with links for Coronavirus, Services, Data, Cyber, Developer, News, and About. The breadcrumb trail reads "NHS Digital > Services > Data Access Request Service (DARS)".

The main content area is titled "Data Access Request Service (DARS)" and includes the text: "The Data Access Request Service (DARS) can offer clinicians, researchers and commissioners the data required to help improve NHS services." Below this is an embedded SAIL Databank page. The SAIL Databank page has a teal header with the logo and navigation links: Home, SAIL Data, Application Process, About Us, News, FAQ, Contact. It also has language selection buttons for English and Cymraeg. The main heading on the SAIL Databank page is "Our Two Stage Application Process". Below this, it states: "SAIL Databank has established a convenient and effective two stage application process to be followed by anyone who would like to access data via SAIL Databank for high quality research projects." The two stages are:

- Stage 1**: Initial application and scoping document (to be completed by a SAIL Databank analyst following a discussion about a potential project)
- Stage 2**: Application to gain Information Governance Review Panel (IGRP) approval

- The group of researchers therefore makes an application to NHS Digital to access the National Diabetes Audit.
- At the same time, they make an application to the SAIL Databank, which is the trusted research environment (TRE) where a copy of Welsh education data is kept for research.
- In their applications, they must explain the work they want to do – including linking the diabetes and education data together – and why this is important for improving the lives of children living with diabetes.

6. The applications are assessed



- NHS Digital and the SAIL Databank then assess the applications they receive, to see whether they match the criteria for accessing the data requested.
- The applications are also assessed by independent ethics committees at the same time, to ensure the project is in the public interest.

7. Researchers undertake training to access the TRE



- At the same time as applying to access the two different data sources, the researchers must apply to become 'approved researchers' to use the SAIL Databank trusted research environment (TRE), where the researchers will be able to access the data in a de-identified form.
- To become 'approved researchers', all members of the research team that want to see and analyse the data must undertake training and complete a test. They are also assessed on their academic qualifications and research abilities, and must be part of a recognised academic institution, such as a university.

8. Applications are approved and training is complete



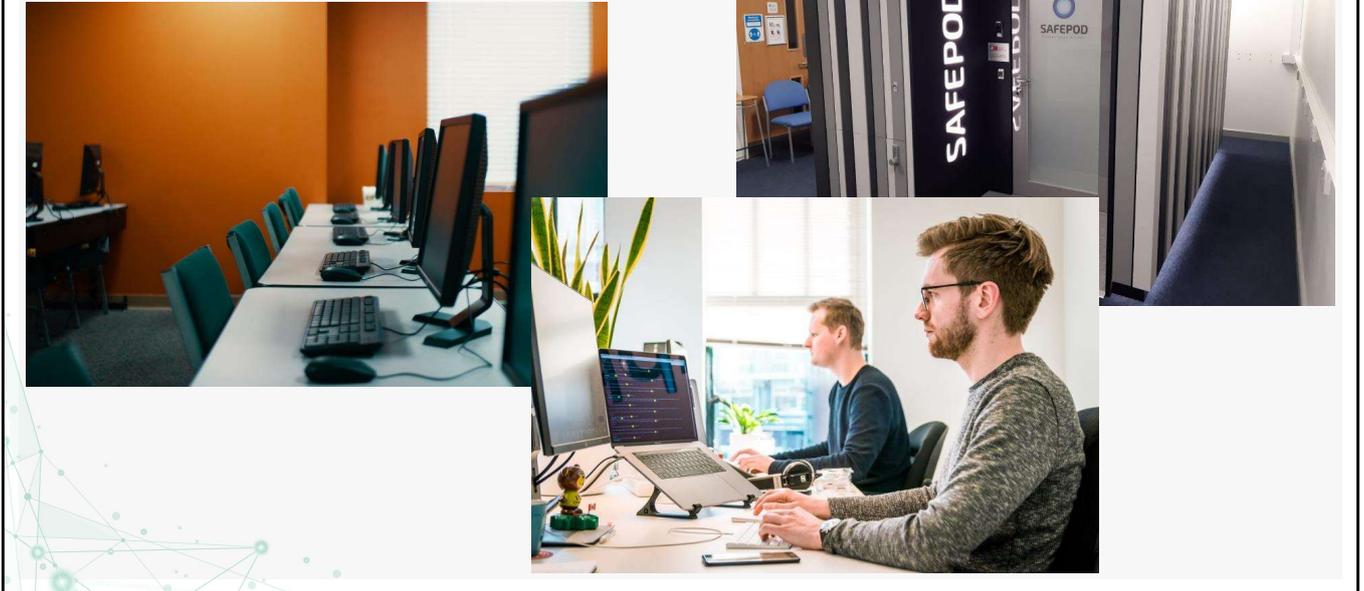
- The research team have been approved by both NHS Digital and SAIL Databank, who have assessed the project as being in the public interest and having a lawful basis for accessing the data requested in a de-identified form.
- Meanwhile, all researchers involved on the project have successfully completed their SAIL Databank training, and passed the test to access the trusted research environment (TRE).

9. The data is transferred to the TRE



- Welsh education data is already kept in the SAIL Databank trusted research environment (TRE) in a de-identified form.
- However, the National Diabetes Audit is held by NHS Digital, so must now be transferred to the SAIL Databank for the researchers to access it and link to the education data.
- Before it is transferred, any identifying information – including names and addresses – is removed so that the research team cannot find out the identity of the people the data is about.

10. Data is accessed



- Now that all the data requested is in the research team's digital 'work space' in the trusted research environment, and they are approved to access it, they can access the data to start linking it together and analysing it.
- The data they have access to includes the information about patient symptoms, medication, blood sugar levels and so on, and their school test results. This allows them to explore the relationships between these different factors in children's lives.
- The data the researchers access is de-identified – it does not contain any personal information such as names, addresses or identifying numbers.
- They access the data either in a secure, monitored room in the trusted research environment's own buildings, or via a secure connection to the trusted research environment.
- The length of the process – from the initial data access request to being able to access and analyse the data – can differ greatly, but can range from a few months to several years depending on the specifics of the project and the data requested.

11. Research outputs are checked



- When the researchers have finished analysing the data, before their findings can leave the trusted research environment, staff at the SAIL Databank must first check the findings to make sure they maintain the confidentiality of the people the data is about, and that it is not possible to identify anybody from the data.
- If it is possible to identify any individuals, this data is removed before the researcher is allowed to take the findings out of the trusted research environment (TRE).

12. Findings are published



When staff at the SAIL Databank are happy that the research findings maintain the confidentiality of the people the data is about, they are released, and the researchers can publish their findings – for example in academic journals, magazines and in the media – for other researchers, charities, policymakers and the public to learn about.