How health record linkage works, how data is kept safe and secure, and what this means for participants.

This video shows how the **UK LLC links data:** https://www.youtube.com/watch?v=FXnHc8R3UW0

If you have any questions, please contact the study team at CSS Biobank: https://cssbiobank.com/contact-us

CSS BIOBANK
COVID SYMPTOM STUDY BIOBANK

CSS Biobank provides a secure file with CSS Biobank numbers and permission settings

The CSS Biobank study team will create **two sets of data**:

1 One set contains **personal details** (like name, date of birth, and NHS number) and **permission information**.

2 The other contains **CSS Biobank study data** (the data we want to link to health records such as questionnaire data) but no personal details.

Both sets of data include **unique CSS Biobank numbers** so they can be matched with health record data.

NHS

NHS uses personal details to find matching health records

The CSS Biobank number is encrypted

Identifying details are removed and securely stored

- The file with personal details is **securely sent to the NHS** to find health records. The CSS Biobank number is **encrypted** to protect identities.
- Once the NHS has found health records, **personal details** (name, NHS number, etc.) **are removed**. These identifying details are **stored securely** and **not shared**.

SeRP Gateway at Swansea University

De-identified health data is linked with CSS Biobank study data using CSS Biobank number

- The remaining health information (now de-identified) is **securely sent** to the Secure eResearch Platform (**SeRP**) Gateway at Swansea University, where it is **linked to the CSS Biobank study data** using the CSS Biobank number.

All data in the UK LLC is de-identified and subject to **Five Safe's governance principles**. The Five Safes is a well-established framework used to manage and govern access to **sensitive data** in a way that **protects privacy** while enabling **valuable research**. https://ukllc.ac.uk/confidentiality

UK LLC Trusted Research Environment

Only approved researchers can access the data and only in a safe, controlled way

Data can't be copied or removed

Once linked, the **de-identified dataset** is stored in a **Trusted Research Environment** (TRE) by the UK LLC. A TRE is a highly secure computer system where:

- Only **approved researchers** can access the data
- Data **cannot be copied or removed** from the TRE
- The data in the TRE **does not include** names, addresses, dates of birth, NHS numbers, or any other **personal identifiers** (like the names of hospitals).
- Researchers will **never be able to identify participants** from the linked data.