



# CLOSER-UK LLC webinar: Data access in the Secure eResearch Environment (SeRP)

**Richard Thomas**  
Senior Data Manager  
April 2022

# Overview



- Available data
- Upcoming data
- Secure eResearch Platform (SeRP)
- Data in the SeRP
- Files in/out the SeRP

# Available data – studies cohort



20+ longitudinal studies with >200,000 participants

**ALSPAC:** Avon Longitudinal Study of Parents and Children

**BCS70** British Cohort Study

**Born in Bradford**

**ELSA:** English Longitudinal Study of Ageing

**EPIC-Norfolk:** The European Prospective Investigation into Cancer (EPIC) Norfolk Study

**EXCEED:** Extended Cohort for E-health, Environment & DNA

**The Fenland Study**

**Generation Scotland**

**GLAD:** Genetic Links to Anxiety and Depression Study

**MCS:** Millennium Cohort Study

**NCDS58:** 1958 National Child Development Study

## Next Steps

**NICOLA:** Northern Ireland Cohort for the Longitudinal Study of Ageing

**NIHR BioResource\_COPING:** National Institute of Health Research BioResource, COVID-19 Psychiatry and Neurological Genetics Study

**NSHD46:** MRC National Survey of Health and Development Cohort/1946 Birth Cohort

**SABRE:** Southall and Brent Revisited (non-linked data only)

**TRACK-COVID** Study

**TwinsUK**

**Understanding Society** - the UK Household Longitudinal Study

# Available data – linked NHS D



## NHS D datasets

- GP Data
- Vaccinations
- COVID-19 test data (Pillar 1-3)
- Hospital Episode Statistics (HES):
  - Accident and Emergency
  - Admitted Patient Care
  - Outpatient
  - Critical care
- CHES
- Cancer & mortality registers
- Community Mental Health (MHSDS, IAPT)
- Prescribing data

# Available data – NHS D COVID



## Covid-specific NHS datasets

- GP Data - GDPPR Data for Pandemic Planning and Research (COVID-19)
- Vaccinations:
  - CVS – Covid Vaccination Status
  - CVAR – Covid Vaccine Adverse Reactions
- COVID-19 testing data (Pillar 1-3)
  - SGSS – Second Generation Surveillance System (pillars 1 & 2)
  - NPEX – National Pathology Exchange (pillar 2)
  - IELISA – Non-hospital antibody testing – finger prick test (pillar 3)
- CHES – COVID-19 Hospitalisation in England Surveillance System
- *Work being done to derive covid status output – coordinating with LHW*

# Upcoming data – built environment



## Air Quality (NO<sub>2</sub>, PM<sub>2.5</sub>, O<sub>3</sub>)

- Land Use Regression models @25m resolution
- Annual averages 2010-2021
- National coverage
- New modelling to describe fluctuations in Air Quality during lockdown at higher temporal resolution (GIS dispersion hybrid model based on LUR and direct emission source contributions)

## Noise

- National (~95% coverage) model based on CNOSSOS (Common Noise Assessment Methods) methodology
- Applied to property/postcode (i.e., modelled to participant)
- Weighted, hourly, evening/daytime, 24 hour average from year 2000

## Greenspace and greenness

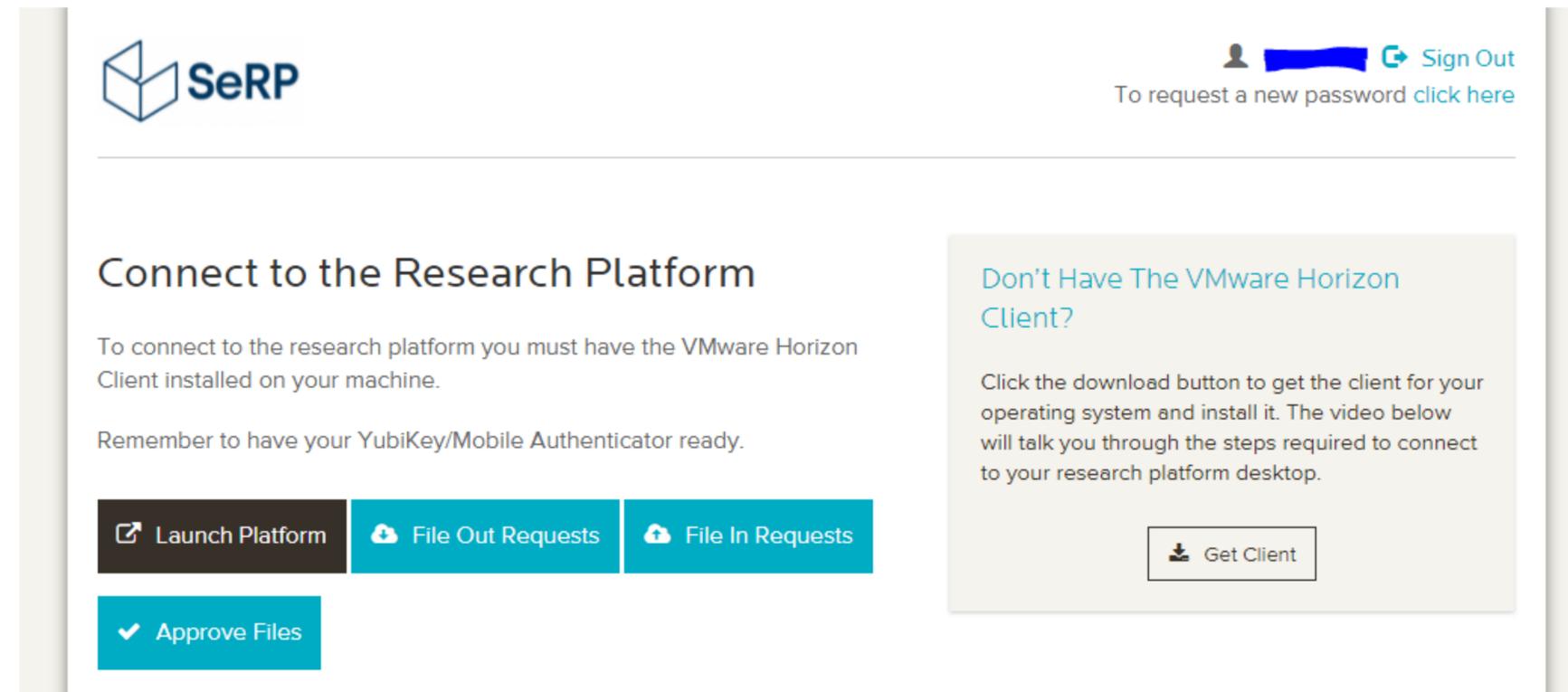
- National model based on NDVI
- Allocation of buffers around property/postcode (250m, 500m, 1000m and 2000m)
- Buffers based on circular geocoordinate and walkable (road) network

## What is the SeRP

- SeRP – Secure eResearch Platform – Trusted Research Environment (TRE)
- Windows 10 environment with standard software packages available SSMS, python (anaconda distribution), R (+Rstudio), Stata, SPSS
- Extra controls on data in and out the system
- Restricted internet connections

## Logging into the SeRP

- Username and Yubikey will be issued by LLC team
- You set password and account will be ready to access
- Download VM Horizon Client – allows connection to desktops
- Enter login details...



The screenshot shows the SeRP user interface. At the top left is the SeRP logo. At the top right, there is a user profile icon, a blue bar, and a 'Sign Out' button. Below this is a link: 'To request a new password [click here](#)'. The main content area is titled 'Connect to the Research Platform'. Below the title, it says: 'To connect to the research platform you must have the VMware Horizon Client installed on your machine.' and 'Remember to have your YubiKey/Mobile Authenticator ready.' There are four buttons: 'Launch Platform' (dark grey), 'File Out Requests' (teal), 'File In Requests' (teal), and 'Approve Files' (teal). On the right side, there is a light grey box titled 'Don't Have The VMware Horizon Client?' with the text: 'Click the download button to get the client for your operating system and install it. The video below will talk you through the steps required to connect to your research platform desktop.' and a 'Get Client' button.

# Getting started in the SeRP

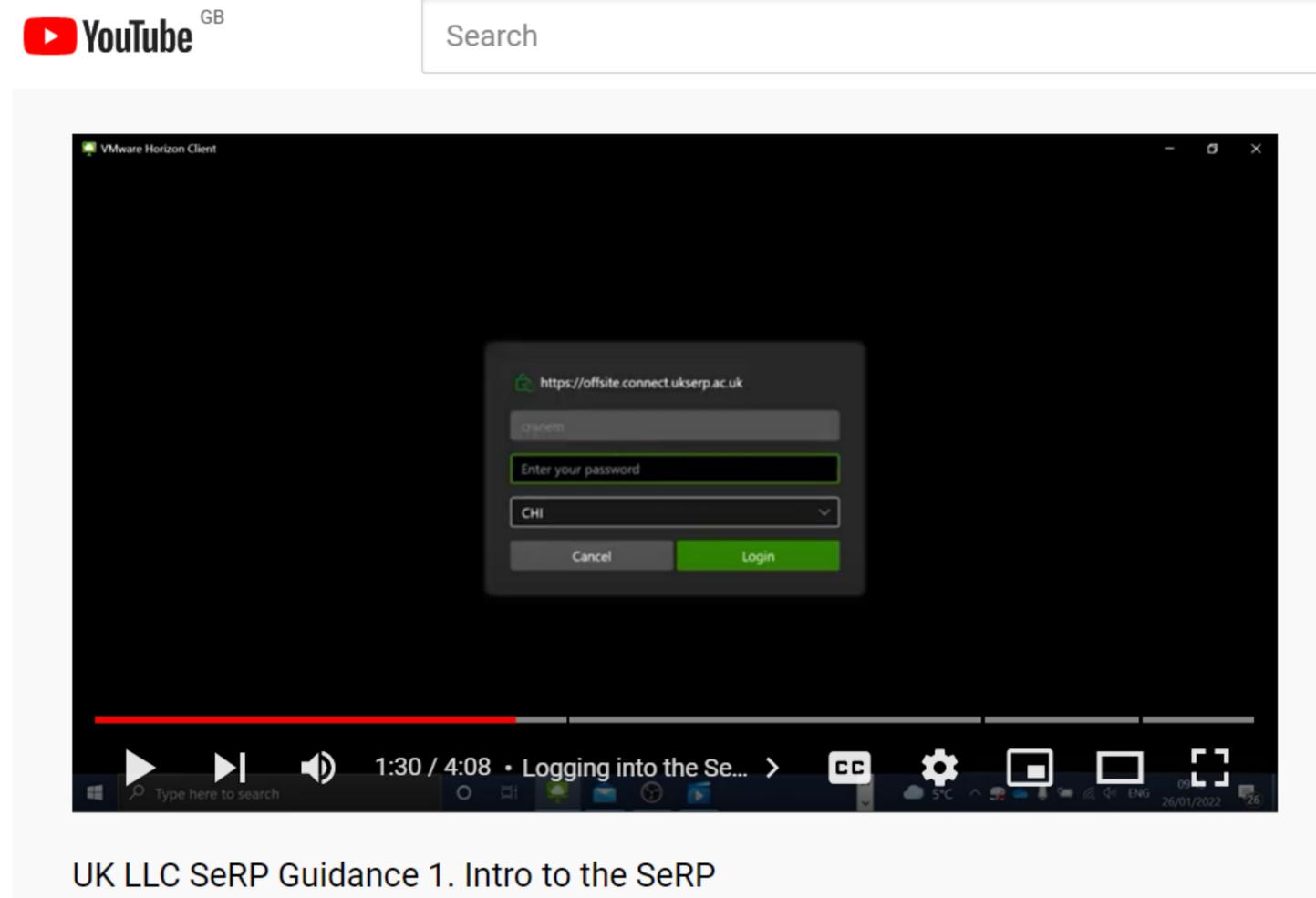


## Documentation

- S:\ReadAll folder – accessible to all
- Contains TRE user guide
- Video walk through – guiding through first time login

## Syntax

- Contains script to setup python environment
- Helper syntax for python, R and Stata:
  - Connects user to database
  - Pulls data views
  - Adds variable and value labelling
- Jupyter notebook with python/pandas available
- R helper now available
- Stata helper currently under development



# Your project area



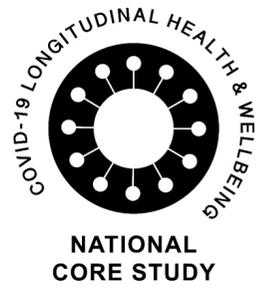
- Each research project will have folder in shared drive with rw permissions:

LLC_0001	15/02/2022 13:42	File folder
LLC_0005	27/01/2022 17:12	File folder
LLC_0006	10/02/2022 15:29	File folder
LLC_0007	10/02/2022 15:30	File folder

- All working should be kept in this area
- Folder will come with pre-defined folder structure
- Create sub-folders at will

Sub Folder	Purpose
available_views	Contains text file outputs from all data provisions with a list of SQL database views that were made available during the provision.
data	Contains researcher generated datafiles created during analyses
documentation	For storing documentation pertinent to the research. This can be generated with the TRE or sent in via <u>'file-in'</u> request.
llc_guidance	A sub-folder containing key UK LLC requirements documents for ease of reference (e.g. your Data Request Form)
outputs	Proposed publication-ready analytical outputs to be submitted through the 'file-out' review process
syntax	Contains researcher generated analytical syntax/scripts

# Navigating data in the SeRP



- Data stored in SQL database with project-specific views (based on request):
  - ⊕ LLC\_0001.BCS70\_COVID\_w1\_v0001\_20211101
  - ⊕ LLC\_0001.BCS70\_COVID\_w2\_v0001\_20211101
  - ⊕ LLC\_0001.BCS70\_COVID\_w3\_v0001\_20211101
  - ⊕ LLC\_0001.BCS70\_response\_1970\_2016\_v0001\_20211101
  - ⊕ LLC\_0001.BIB\_CV\_W1\_v0001\_20211101
  - ⊕ LLC\_0001.BIB\_SOCIODEMO\_v0001\_20211101
  - ⊕ LLC\_0001.ELSA\_elsa\_covid\_w1\_eul\_v0001\_20211101
  - ⊕ LLC\_0001.ELSA\_elsa\_covid\_w2\_eul\_v0001\_20211101
  - ⊕ LLC\_0001.ELSA\_wave\_9\_elsa\_data\_eul\_dhdi\_v0001\_20211101
- In form <project number>.<source>\_<descriptive name>\_<version number and date>
- E.g. LLC\_0001.BCS70\_COVID\_w1\_v0001\_20211101
- All metadata in accessible views – requires query and match to bring together
- Helper syntax will ease load/make available all available data in stats/software packages

# Jupyter notebook example



- Jupyter notebook – recommended way of working
- Example in working area to aid data familiarisation
- Notebook has underlying python “helper” class with functions to enable initial data overview:

```
In [14]: helper.short_print(views)
```

	TABLE_NAME	SCHEMA	number_of_variables	variable_names	number_of_rows
0	GENSCOTT_OCCUPATION_v0001_20211101	GENSCOTT	8	LLC_0007_stud_id, avail_from_dt, income, joble...	20943
1	GLAD_FILE2_v0001_20211101	GLAD	351	LLC_0007_stud_id, age, age_category, age_categ...	36922
2	MCS_COVID_w1_v0001_20211101	MCS	304	LLC_0007_stud_id, avail_from_dt, cw1_aldrpp, c...	5217
3	MCS_COVID_w2_v0001_20211101	MCS	781	LLC_0007_stud_id, avail_from_dt, cw2_acaper_1,...	8417
4	MCS_COVID_w3_v0001_20211101	MCS	828	LLC_0007_stud_id, avail_from_dt, cw3_acaper_1,...	9361
5	MCS_basic_demographic_cm_v0001_20211101	MCS	14	LLC_0007_stud_id, adc11e00, avail_from_dt, bdc...	9610
6	MCS_basic_demographic_parent_v0001_20211101	MCS	24	LLC_0007_stud_id, add05s00, add07s00, add13s00...	18127
7	MCS_mcs7_cm_interview_v0001_20211101	MCS	795	LLC_0007_stud_id, avail_from_dt, g_comp, g_com...	9451
8	NCDS58_COVID_w1_v0001_20211101	NCDS58	302	LLC_0007_stud_id, avail_from_dt, cw1_aldrpp, c...	4634
9	NCDS58_COVID_w2_v0001_20211101	NCDS58	779	LLC_0007_stud_id, avail_from_dt, cw2_acaper_1,...	5618

# Data structure



## Study/cohort data

- predominately one row per person:

study_id	gender	mood	datetime
100001	Male	1	01/01/2020
100002	Male	2	03/01/2020
100003	Female	1	10/01/2020

- Minimised to table/blocks of data

## NHS D data

- NHS D data – many rows per person (event based) with datetime stamps:

study_id	diag_3_01	diag_3_02	diag_3_03	diag_3_04	diag_3_05	datetime
100001	H25	M13				01/01/1995
100001	N40	K21	Z86			01/01/1997
100001	E78					01/01/2005
100001	E11					01/01/2011
100002	Y56	L12				01/01/2011
100002	Z50	C78	Z95	K59	Z76	01/01/2020
100003	Z95					01/01/1999
100003	H25	E66				01/01/2001
100003	X78					01/01/2004

- Minimised to medical code lists

# Medical codelists

- Data minimised to medical code lists for some NHS D datasets e.g. HES
- Standardised codelists available e.g. <https://www.caliberresearch.org/portal/codelists>

## Diseases of the blood

### Chronic anaemia

<a href="#">chronicanaemia_gprd</a>	Chronic anaemia (primary care)
<a href="#">chronicanaemia_hes</a>	Chronic anaemia (secondary care)

### Inflammatory markers

<a href="#">c_reactive_protein_gprd</a>	C reactive protein (CRP)
<a href="#">esr_gprd</a>	Erythrocyte sedimentation rate (ESR)

### Chronic anaemia

<a href="#">myelodysplasia_gprd</a>	Myelodysplastic syndrome (primary care)
<a href="#">myelodysplasia_hes</a>	Myelodysplastic syndrome (secondary care)

- Or custom (user defined)
- Currently working on a new model for codelist definitions – allows data exploration before definition

# File ins/outs



Inputs and outputs to the SeRP are controlled...

## Files in

- Documentation and syntax
- No data

## Files out

- Anonymous aggregated outputs only, examples:
  - Tables of results
  - Graphs
  - Statistical model outputs
- Rules of allowable outputs follow ONS Safe Researcher principles and a checklist will be available
- File out requests should be made when intending to publish paper, report, presentation etc



Thank you

