UK Data Service

Biomedical resources at UK Data Service and Data Management Planning

Soc-B Workshop

Cristina Magder & Hina Zahid

11 January 2022

Economic and Social Research Council

Copyright © (2021) UK Data Service,, created by (UK Data Service) University of Essex.

Can you hear us?

Housekeeping

- Check your speaker/headset is plugged in / volume is on.
- Click on audio to change to listening via phone
- Questions can be submitted via the chat box anytime
- Please be on mute throughout the presentations

Overview

Part 1: Biomedical resources at UK Data Service and Data Management Planning

- Data sharing and its' benefits
- Overview of UK Data Service and UK Data Archive
- Examples of Biomedical data available at UK Data Service
- Overview of Data Management Planning (DMP) and its' importance
- Depositing with UKDS

Part 2: Ethical and legal considerations for data sharing

Data lifespan and data sharing

- data have a longer lifespan than the (research) project that creates them
- data can be (re) used for various purposes when well managed, preserved and shared
- who is benefiting from data sharing? researchers, public, funding bodies, scholarly community and research participants

Exercise: Benefits of Data Sharing

How does data sharing benefit each group?

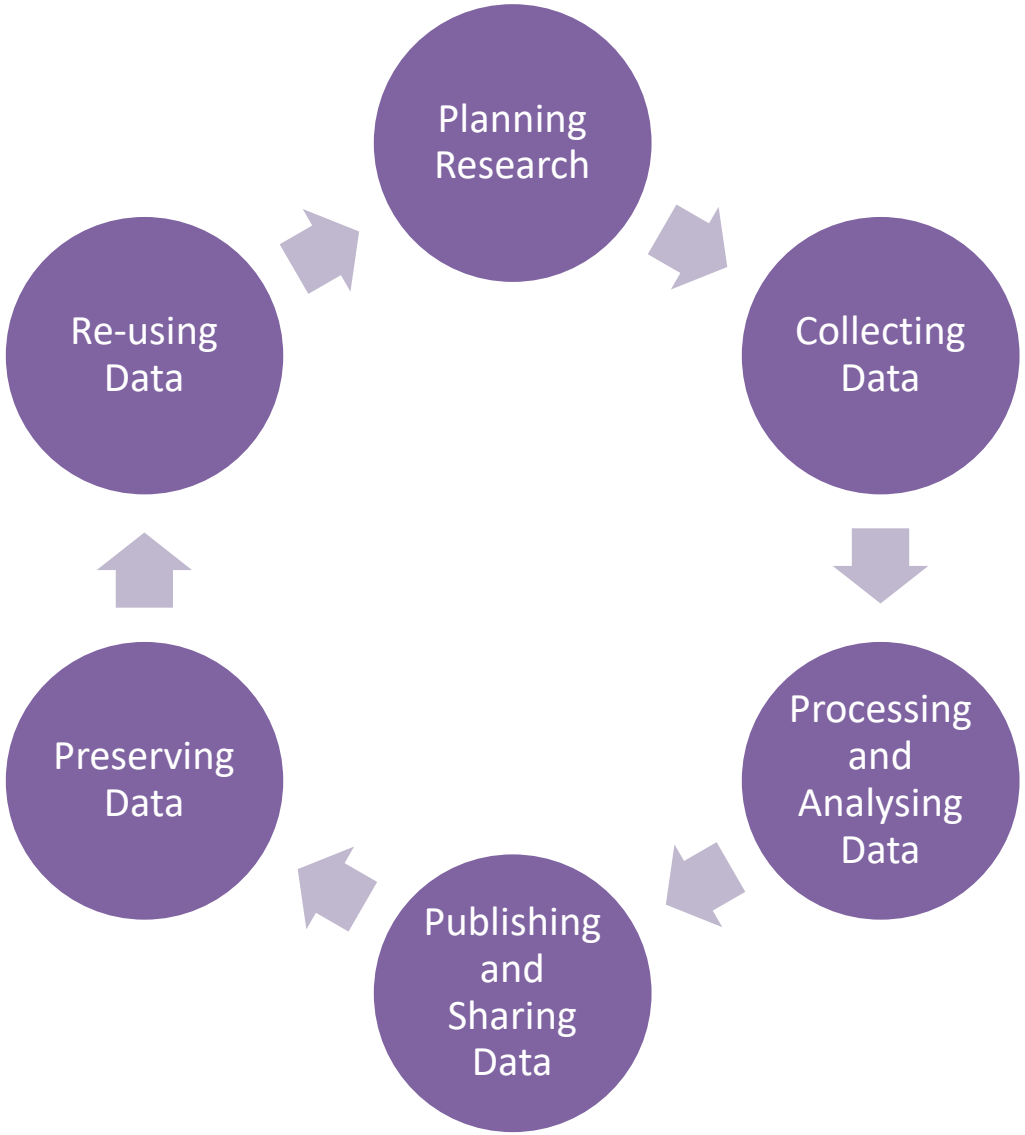
- **Researchers/Data Creators**
- **Funding Bodies**
- **Public**
- **Research Participants**

Please include ideas (with examples if possible) on the Padlet available at https://padlet.com/dcmagd/88h7esaysnjpev90

Benefits of data sharing

Researchers	Funding Bodies	Public	Research Participants
increases visibility of scholarly work and increase citations (DOI)	make optimal use of publicly funded research	advance science to the benefit of society	allow maximum use of their contributed data/information
archiving provides long-term safe storage for data	avoid duplication of data collection	adoption of emerging norms – ‘open access’ publishing	minimise data collection on the hard-to-reach (e.g. ill, elites)
assists in implementing publishers’ data retention policies and enables collaborations	maximise return for investment	compliance with laws and regulations (transparency, openness, accountability)	enable participants’ experiences to be understood as widely as ethically possible

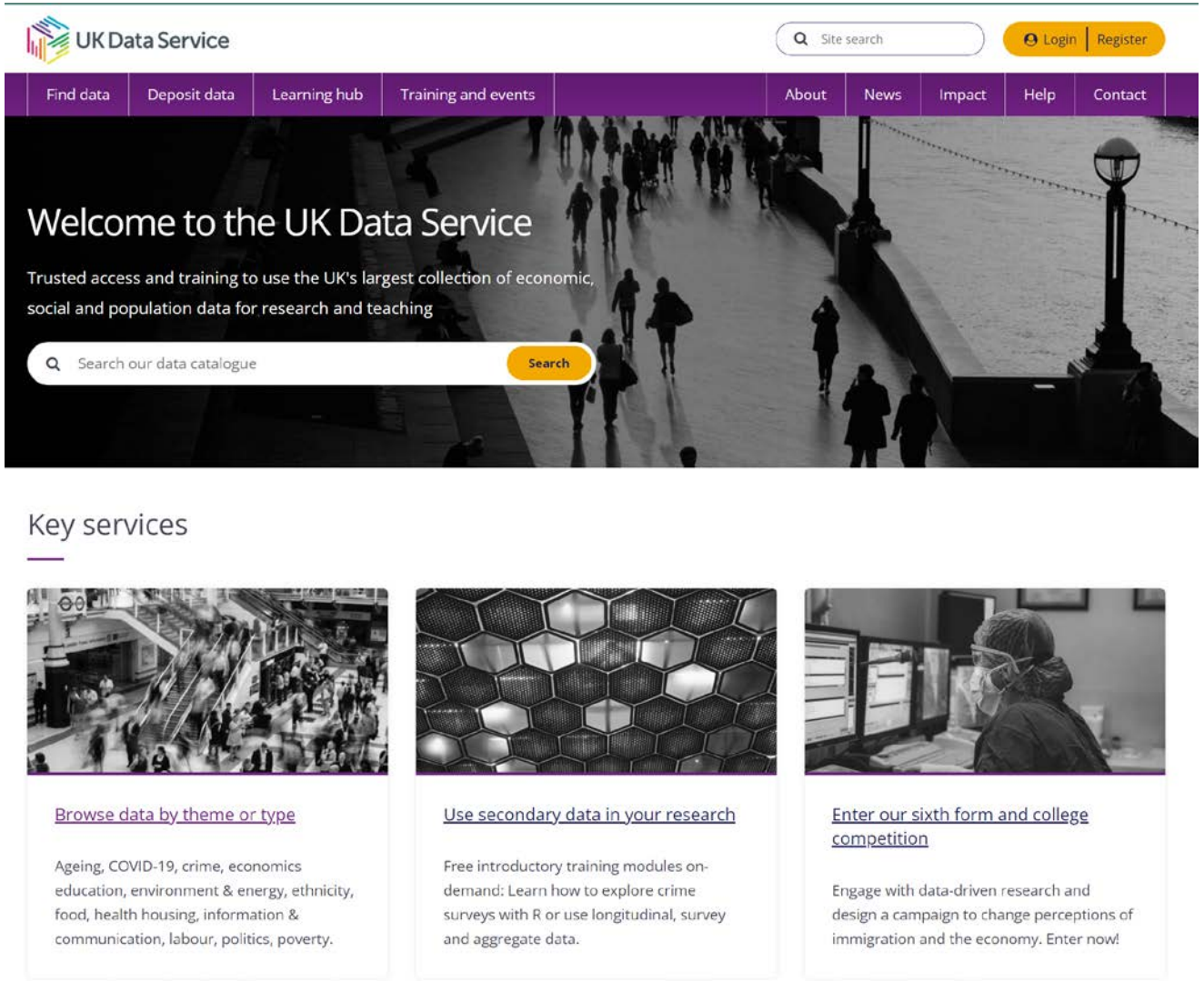
Research Data Lifecycle



UK Data Service

- UK's largest collection of UK and international social, economic and population data
- funded by the Economic and Social Research Council (ESRC)
- provide users with access, support, guidance and training to facilitate high quality social and economic research and education
- support the development of best practices for data preservation and sharing standards

UK Data Service



- UK Data Service Data Catalogue
- UK Data Service Training Events
- UK Data Service Research Data Management

UK Data Archive

- the lead partner of the UK Data Service
- centre of excellence in acquiring, curating and providing access to the largest collection of social science and population data for over 50 years
- provide a Trusted Digital Repository
- sharing best practice, we help to lead international best practice in data management
- based at the University of Essex

UKDS User Community

- Academic researchers and students
- Government analysts
- Charities and foundations
- Business consultants & data analysists
- Independent research centres & think tanks

Some stats on UKDS

- **8,500** datasets in the collection
- **250** new datasets and new editions added each year
- **50,000+** registered users
- **200,000+** downloads worldwide per year
- **50+** applications for restricted data per month
- **4000+** user support queries per year

UKDS: Sources of Data

- National statistical authorities
- UK government departments
- Intergovernmental organisations
- Research institutes
- Individual researchers

UKDS: Data Portfolio

UK Surveys

Large-scale government funded surveys

Longitudinal

Major UK surveys following individuals over time

International

Multi-nation aggregate databanks and survey data

Census

Census data 1971 – 2011

Business

Microdata and administrative data

Qualitative

Range of multimedia qualitative data sources

Licensing and access framework at UKDS

Open	• No real disclosure risk. Under open licences; almost no restrictions on reuse
Safeguarded	• Zero to low real disclosure risk. Requires authentication and authorisation e.g. registered user and End User Licence Agreement
Controlled	• Real disclosure risk. Requires project approval, user vetting and training; access via a safe setting; output checking

5 Safes Framework

A set of principles the UK Data Service developed with the ONS and HMRC in 2017 which enable data services to provide safe research access to data

- **Safe data:** data is treated to protect any confidentiality concerns
- **Safe projects:** research projects are approved by data owners for the public good
- **Safe people:** researchers are trained and authorised to use data safely
- **Safe settings:** a SecureLab environment prevents unauthorised use
- **Safe outputs:** screened and approved outputs that are non-disclosive

FAIR Principles

Findable

Accessible

Interoperable

Re-usable

Force 11 FAIR Principles

Live Demo using the UK Data Service Data Catalogue

UK Data Service Data Catalogue

Available here

Search for studies or series by typing in a search term (words, a study number, year) in the search box

Use the Humanities and Social Science Electronic Thesaurus (HASSET)

Browse by theme or data type

Biomedical data: UK Surveys – HSE

Health Survey for England (HSE)

- designed to monitor trends in the nation's health

The aims of the HSE series are:

- to provide annual data about the nation’s health
- to estimate the prevalence of certain risk factors associated with these conditions
- to examine differences between population subgroups in their likelihood of having specific conditions or risk factors;
- to assess the frequency with which particular combinations of risk factors are found, and which groups these combinations most commonly occur;
- to monitor progress towards selected health targets
- since 1995, to measure the height of children at different ages, replacing the National Study of Health and Growth;
- since 1995, monitor the prevalence of overweight and obesity in children

Biomedical data: UK Surveys – HSE

Core topics

- General health and longstanding illness
- Smoking and average weekly alcohol consumption
- Socio-economic information: sex, age, income, education, employment etc
- Prescribed medications (nurse)

Measurements

- Height and weight
- Blood pressure (nurse)
- Waist and hip circumference (nurse)
- Blood sample for cholesterol, glycated haemoglobin (nurse)
- Saliva sample (nurse)

Biomedical data: UK Surveys – HSE

Health Survey for England

Abstract FAQ's Resources Access data

Access data

GN 33261
Health Survey for England, 1991-

GN 33543
Health Survey for England, 1994- : Special Licence Access

Click on a series name or an arrow to see the datasets.

Add to account

HSE Pgp3 and HSV2 Serum Antibodies

Aim

- investigate trends in exposure to chlamydia among adults in England, and to identify factors associated with having antibodies to these infections

Main Topics

- national prevalence of antibodies to Chlamydia trachomatis and Herpes Simplex Virus (Type 2)
- sexual behaviour
- demographics

Biomedical data: UK Surveys – NDNS

National Diet and Nutrition Survey

- designed to assess the diet, nutrient intake and nutritional status of the general population aged 1.5 years and over living in private households in the UK
- core topics include food shopping, preparation and eating habits, smoking and drinking, physical activity, general health

Measurements

- height and weight
- spot urine sample
- blood pressure measurements
- blood sample

Biomedical data: Long Data – UKHLS

Understanding Society study, or the United Kingdom Household Longitudinal Study (UKHLS)

- began in 2009
- builds on the long running British Household Panel Survey (BHPS) established in 1991

Tracks change at individual and household level by examining:

- what people are experiencing (circumstances and key events in their lives as they happen)
- attitudes, identity and subjective well-being (expressions)
- how individuals and households respond to policies and key events (behaviours)
- changes in societal conditions (outcomes)

Biomedical data: Long Data – UKHLS

Measure	Applications
Height and weight, Waist circumference, Percent body-fat	BMI and assessment of excess body fat: obesity and risk factor for range of major chronic conditions and social outcomes
Respiratory function (Spirometry) (FVC,FEV_1, PF, FEV_1/FVC)	To detect both obstructive and restrictive respiratory diseases including COPD.
Diastolic and systolic blood pressure, resting pulse rate	Risk factor for stroke and heart conditions Risk cardio-vascular disease
Grip strength	Indicator muscle strength. Functional limitations and disability in older ages.
Blood samples (non-fasting), 19.8ml)	For the extraction of analytes and DNA
Short questionnaire on health on day of measurement, medications etc	Factors that may need to be considered in analysing physical measures and bloods

Biomedical data: Long Data – CLS Cohorts

The Centre for Longitudinal Studies (CLS) is an Economic and Social Research Council resource centre based at University College London

CLS is responsible for running four of Britain's internationally-renowned longitudinal studies:

- 1958 National Child Development Study (NCDS)
- 1970 British Cohort Study (BCS70)
- Millennium Cohort Study (MCS)
- Next Steps (previously the Longitudinal Study of Young People in England (LSYPE1))

Biomedical data: Long Data – MCS

The Millennium Cohort Study is providing information about babies born at the beginning of the 21st century, their progress through life, and the families who are bringing them up, for the four countries of the United Kingdom

Measurements

- physical measurements since the age of 3 (height, weight, waist measurement and body fat)
- at ages 7 and 14 physical activity and sedentary behaviour monitored via activity monitors
- at age 14 cohort members and parents provided saliva samples for DNA extraction and subsequent genotyping - the first time a triad of DNA samples has been collected from 2 biological parents and their child in a large scale nationally representative study

Biomedical data: Long Data – MCS

GN 33445

Millennium Cohort Study - Linked Administrative Data

SN	Study description	Explore online	Select
8785	Millennium Cohort Study: Age 5, Sweep 3, 2006: Foundation Stage Profile and Teacher Survey		Login
8714	Millennium Cohort Study: Linked Health Administrative Data (Scottish Medical Records), Outpatient Attendance, 2001-2015: Secure Access		Login
8713	Millennium Cohort Study: Linked Health Administrative Data (Scottish Medical Records), Inpatient and Day Care Attendance, 2000-2015: Secure Access		Login
8712	Millennium Cohort Study: Linked Health Administrative Data (Scottish Medical Records), Scottish Birth Records, 2000-2002: Secure Access		Login
8711	Millennium Cohort Study: Linked Health Administrative Data (Scottish Medical Records), Scottish Immunisation and Recall System, 2000-2015: Secure Access		Login
8710	Millennium Cohort Study: Linked Health Administrative Data (Scottish Medical Records), Prescribing Information System, 2009-2015: Secure Access		Login
8709	Millennium Cohort Study: Linked Health Administrative Data (Scottish Medical Records), Child Health Reviews, 2000-2015: Secure Access		Login

Health administrative data linked to the survey data (Scottish Medical Records, ICD-10)

COVID-19 Antibody Testing Cohort Studies

- National Child Development Study
- 1970 British Cohort Study
- Next Steps
- Millennium Cohort Study, 2021

+

- MRC National Survey of Health and Development

University College London, UCL Institute of Education, Centre for Longitudinal Studies. (2021). *COVID-19 Antibody Testing in the National Child Development Study, 1970 British Cohort Study, Next Steps and Millennium Cohort Study, 2021*. [data collection]. UK Data Service. SN: 8823, DOI: 10.5255/UKDA-SN-8823-1

University College London, MRC Unit for Lifelong Health and Ageing. (2021). *COVID-19 Survey in Five National Longitudinal Cohort Studies: MRC National Survey of Health and Development, 2020-2021: Special Licence Access*. [data collection]. *3rd Edition.* UK Data Service. SN: 8732, DOI: 10.5255/UKDA-SN-8732-3

Exercise: To share or not to share

Consider your research team has conducted a detailed survey on heart health conditions also gathering DNA, saliva and blood samples

Can you share the data? What are some challenges and opportunities?

Share the opportunities and challenges identified in sharing these data on the Padlet available at https://padlet.com/dcmagd/awni80k6de9ulnl3

Challenges

- Personal, confidential or sensitive information
- Linkage of data in multi-disciplinary projects
  - social with biomarkers or geo-located data
  - difficult to conceal identity of participants / fieldwork locations
- Lack of trust in others mis-using data
- Efforts and funding compete with ‘the science’

Opportunities

- Not burden over-researched, vulnerable groups
- Make best use of hard-to-obtain data, e.g. elites, socially excluded, over-researched
- Extend voices of participants
- Provide greater research transparency

Data Management Planning

A data management and sharing plan helps researchers consider: when research is being designed and planned, how data will be managed during the research process and shared afterwards with the wider research community

Benefits:

- think what to do with research data, how collect, how look after
- keep track of research data (e.g. staff leaving)
- identify support, resources, services needed
- plan storage, short & long-term
- plan security, ethical aspects
- be prepared for data requests (FoI, funder)

ESRC Data Management Plan (DMP)

- Assessment of existing data
- Information on new data
- Quality assurance of data
- Backup and security of data
- Expected difficulties in data sharing
- Copyright / Intellectual Property Right
- Responsibilities
- Preparation of data for sharing and archiving

ESRC DMP guidance

Key planning considerations

- know your legal, ethical and other obligations towards research participants, colleagues, research funders and institutions
- know your institution’s policies and services: storage and backup strategy, research integrity framework, IPR policy, institutional data repository
- assign roles and responsibilities to relevant parties
- incorporate data management into research cycle
- implement and review management of data during project meetings and review

DMP tools and templates

- Funder template for DMP
  - ESRC DMP requirements in data policy and DMP guidance
  - MRC DMP guidance and template
  - AHRC DMP requirements
  - NERC DMP guidance and template

- DCC's DMPonline tool

Benefits of depositing data with the UKDS

- no costs associated with depositing
- an online platform for creating, uploading, and managing deposits
- a three-tier licensing framework facilitating a robust access policy
- the Five Safes framework enabling controlled access to sensitive or confidential data
- data management guidance and support ensuring legal sharing and re-use
- persistent identifiers (DOIs) supporting resource discovery and citation
- promotion of deposits via newsletters and long-standing connections with partner Archives and organisations (e.g. CESSDA)
- de-identified study level metrics for funding proposals or extensions

UKDS appraisal process

We have an appraisal process

Assessment according to our Collections Development Selection and Appraisal Criteria, informed by our Collections Development Policy

Our decision will be one of the following:

- Accept the data in the curated repository,
- Accept the data in the self-deposit repository ReShare.
- Advise you to use an institutional repository or alternative place of deposit.

How to deposit via UKDS

- If you wish to deposit data with the UK Data Service, you first need to register and create an account.

- If you are depositing data from large-scale government departments or larger research organisations you can deposit data in our curated repository.

- If you are a researcher you can deposit your data via the UK Data Service self-deposit repository ReShare.

- ReShare is also suitable for the direct deposit of datasets described in a data paper or data descriptor in a peer-reviewed journal, such as Scientific Data and the Research Data Journal.

Any questions?

Thank you.

UK Data Service

Email: datasharing@essex.ac.uk

Twitter: UKDataService