



UNDERSTANDING SOCIETY DATA ACCESS STRATEGY

1. OVERVIEW

- 1.1. *Understanding Society* is an innovative, world-leading longitudinal data resource designed to address the social issues facing 21st Century Britain. The study offers an unprecedented insight into UK households as they respond to regional, national and international change. *Understanding Society* is currently the largest study of its kind in the world and will make a major contribution to the UK's unique and internationally renowned portfolio of longitudinal studies.
- 1.2. All access to and use of *Understanding Society* data, and those of its predecessor the *British Household Panel Study* (BHPS), is governed by the procedures set out in the Data Access Strategy, which seek to be fair, open and transparent. Unless explicitly mentioned, the BHPS access strategy is automatically taken to be the same as that for *Understanding Society*.
- 1.3. The aim of the Data Access Strategy is to ensure that the data¹ produced by *Understanding Society* is made as widely available as possible to the research community (nationally and internationally), whilst ensuring that sensitive data and/or data which is or may be disclosive of the identities of participants is secure, maintaining the legal and moral responsibility to the study participants and complying with the undertakings given to them (See Annex 1 for more information on participant consents).
- 1.4. The Understanding Society Data Access Committee (DAC) was established in 2009 to administer and provide oversight for the operation of data access arrangements. The DAC was responsible for developing the Understanding Society Data Access Strategy.
- 1.5. The DAC was wound up in June 2015. Its responsibilities for regulating access to biological samples, genotype and other related data were handed to a new data access committee covering a range of longitudinal and cohort studies supported by ESRC and other funders, METADAC. Responsibilities for regulating access to other survey and linked data were passed to the Understanding Society Scientific Leadership Team (SLT). Oversight and governance of data access arrangements is provided by the Understanding Society Governing Board.
- 1.6. At the May 2017 meeting of the Understanding Society Governing Board it was agreed to:
 - i. Add the Office for National Statistics (ONS) Secure Research Service (SRS), (previously the Virtual Microdata Laboratory), as an alternative data storer for *Understanding Society* data.
 - ii. Allow international access to data categorised as 2b subject to review and approval from a special Understanding Society internal committee.
 - iii. Re-classify *Understanding Society* epigenetic and genotype data as category 2b with the outcome that it can be considered by METADAC for international access.

¹ Unless otherwise specified, the term 'data' when used covers: all information or material gathered during the Study, including blood; DNA, genotypes derived from these samples; and information arising from linkage to administrative data.

- 1.7. The Strategy identifies a series of mechanisms to provide access to the data collected by *Understanding Society*. These procedures apply to all data collected, not just under the main study commissioned by ESRC, but also any co-funded add on studies or data generated by research using *Understanding Society* data.
- 1.8. Rules and procedures for access to survey and linked data are set out in Annex 2. Rules and procedures for access to biological samples and genotype data are set out in Annex 3.
- 1.9. The Strategy builds on existing agreements developed by the ESRC and UK Data Service for accessing data collected by complex longitudinal surveys. It recognises the importance of developing procedures, protocols and standards to support ethical safeguards surrounding data access and the reuse of data for research purposes.

2. DATA ACCESS PRINCIPLES

- 2.1. The procedures and processes that have been applied to provide access to *Understanding Society* data derive from the key principles are set out below:
 - i. To maximise the impact of the study, the data produced by *Understanding Society* is made as widely available as possible to a range of users;
 - ii. All access to *Understanding Society* data is governed by the procedures set out in the Data Access Strategy, which aim to be fair, open and transparent. The controls applied are proportionate to potential risks and are designed not to create unnecessary barriers to research;
 - iii. *Understanding Society* data (including samples) will be available to all bona fide researchers (whether in universities, government departments, charities or commercial companies), for research that can demonstrate public interest;
 - iv. The *Understanding Society* data custodians are the University of Essex;
 - v. Access requests will be checked to ensure they have had relevant scientific and ethical approval;
 - vi. The cost of accessing *Understanding Society* data and samples will be minimised. All applicants will be expected to bear the cost of obtaining data and/or samples;
 - vii. Users who are not ordinarily resident in the UK are currently not eligible to apply for access to the most sensitive, confidential and/or disclosive data;
 - viii. Access to survey and related data are managed by the UK Data Service (UKDS) and the ONS SRS (see Annex 2); access to genotype or epigenetic only data are managed by the European Genome/Phenome Archive (EGA), and access to combined genotype or epigenetic and survey data are scrutinised by the METADAC (See Annex 3);
 - ix. All requests for access to the biological samples collected by *Understanding Society* are governed by the METADAC;
 - x. As a depletable resource, the use of the biological samples will be carefully controlled, in order to optimise the long-term value of the resource;
 - xi. Access to administrative data linked to survey data will be governed by the specific data sharing agreements agreed by the relevant data owner and ESRC. In principle, the goal is to hold such linked data in the UKDS Secure Lab and/or ONS SRS.

Data Custodianship

- 2.2. ESRC remains the owner of the property in the *Understanding Society* database, however as per the Human Tissue Act² the samples cannot be 'owned'. The Understanding Society SLT is the data custodian, responsible for the safekeeping of tissue samples and control of their use, and eventual disposal (if required), all in accordance with legislation and the terms of the consent given by the donor.
- 2.3. Where *Understanding Society* data have been linked to administrative data, the relevant data owner remains the owner of the linked administrative data.

Ethical Considerations

- 2.4. Access to the data is granted in line with the terms of consent (see Annex 1) agreed with the study participants. When assessing data access requests by whatever route consideration is given to whether the proposed research is consistent with undertakings given to Study participants when they gave informed consent.

Commercial Use

- 2.5. Commercial organisations can apply for access to *Understanding Society* data and are subject to the standard procedures. For some data the terms of consent may prohibit commercial access. As per any application commercial organisations will be required to demonstrate the public benefits that are likely to flow from the research use. Applicants must confirm that their use of the data is for research purposes and not for commercial exploitation. No organisation, commercial or otherwise should be allowed to gain control or ownership over the *Understanding Society* resource.

International Access

- 2.6. International access to *Understanding Society* data is important and unnecessary barriers should not get in the way of such research. However, in some instances, procedures for protecting participants' data from disclosure risk may not be effective outside the UK and access may need to be barred.

Data Security

- 2.7. ESRC has ultimate responsibility for data security. ESRC considers all issues relating to information security and data protection a high priority and has been working with the Department for Business, Energy & Industrial Strategy, the Information Commissioner's Office, ONS and the Cabinet Office to address these issues. The UK Data Archive, a service provider for the UK Data Service, is ISO27001 compliant and accredited (the information security standard set by the Government). The ONS SRS is accredited to government security standards to hold data up to OFFICIAL SENSITIVE and to make data available for statistical research projects up to OFFICIAL. Data security at ISER is managed by the Information Oversight Committee and at ONS the SRS is managed by the Secure Research Service Security Working Group.

² <http://www.hta.gov.uk/legislationpoliciesandcodesofpractice/legislation/humantissueact.cfm>

Violations of Access Conditions

- 2.8. Access arrangements will ensure that an appropriate set of penalties are applied should violations of access conditions take place. Penalties can be imposed on users and/or their institutions. Further details for survey and other linked data can be found in Annex 2.

ANNEX 1: UNDERSTANDING SOCIETY CONSENT INFORMATION

Summary of consents in *Understanding Society*

The overall mechanism of consent is oral. Communication with respondents takes place through advance letters, information leaflets and what is said by the interviewers. Through these means participants are informed of the purpose of the study, how they were selected, the sponsor, how the data will be used, which includes their protection from harm by maintenance of confidentiality of the data. Participants indicate consent by answering questions. Participants are also informed that each year they will be asked to participate.

For study components that go beyond the usual question and answer situation specific consent is sought. Information is conveyed in an information leaflet and/or by the interviewer. Consent is indicated orally or written.

The information leaflets are posted on the website under Fieldwork documents, e.g., <https://www.understandingsociety.ac.uk/documentation/mainstage/fieldwork-documents>

Information conveyed by interviewers can be found in the questionnaires for relevant waves: <https://www.understandingsociety.ac.uk/documentation/mainstage/questionnaires>

Each of these areas is organized by wave. Written consent forms are also posted under Fieldwork documents for the appropriate wave. Tables 1 and 2 summarise the consents for all current waves of the main study and the Innovation Panel (IP) respectively, although it should be noted some waves are not available to researchers at the time of the production of this document.

Table 1: Summary of consents for mainstage study

Summary of consents for mainstage study components wave 1 to 9			
<i>Wave</i>	<i>Study component</i>	<i>Communications</i>	<i>Consent</i>
All waves	Basic survey	Advance letters, Information leaflet, Participant handbook	Oral
W1	Link educational records of adults age 16 – 24 and children age 4-15	Information leaflet	written
W1	Link health records of adults age 16+ and children age 0 – 15	Information leaflet	written
W2	Nurse visit following interview for subset of adults in General Population Sample component	Information leaflet	oral (most procedures) written – blood samples for research and/or genetic analysis
W3	Nurse visit following interview for subset of adults in BHPS sample component	Information leaflet	oral (most procedures) written – blood samples for research and/or genetic analysis
W4	Link educational records of adults for those who turned 16 ^a since W1 consent and	Information leaflet	written

	those who did not consent before. Adults age 16 – 24 and children age 4 -15		
W4	Link health records of adults age 16+ and children age 0 – 15. Those who did not consent before and those who turned 16 ^a	Information leaflet	Written
W4	Link benefit records of adults age 16+	Information leaflet	Oral
W5	Link higher education statistics agency records. Original sample member or new entrant finished higher education in 1995 or after from a UK institution, continuing sample member received a higher educational qualification since Wave 1.	Information conveyed by interviewer	Oral
W5	Link HMRC records of adults age 16+	Information leaflet	Oral
W5	Link records from DVLA. Adults who are licensed drivers with access to car/van for personal use which is registered in UK	Information conveyed by interviewer	Oral
W6	Link educational records of adults for those who turned 16 ^a since W1 consent and those who did not consent before. Adults age 16 – 24 and children age 4-15	Information leaflet	Written
W6	Link health records of adults rising-16s	Information leaflet	Written
W7	Link to education records. Respondent is a rising 16 year old, or was born after 1978 and is either a new entrant since wave 4, has no consent information, was asked consent and rejected	Information leaflet	Oral

	only once, or gave consent but form is not present or is present but is not valid and went to school in the UK		
W7	Link to education records. Respondent is the responsible adult of any child aged 2-15 in the household where the child is either a new entrant since Wave 4 or at this wave, there is no education consent on record for that child, the education consent has only been asked once and rejected once for that child, or education consent has been given for that child but there is no form on record or the form is on record but is not valid	Information leaflet	Oral
W7	Link to education records. Respondent is currently living in England and is the responsible adult of any child in the household born in 2008 or later and NPD consent was given with a valid form on record for the sample member.	Information leaflet	Oral
W7	Link to benefits records. Adults who hadn't previously given consent (including new entrants)	Information leaflet	Oral
W8	Link to HMRC records. Respondent is eligible for HMRC consent questions (not given consent before) or is a new entrant	Information leaflet	Oral
W8	Link to energy consumption records. If property is in	Question	Oral

	England, Wales or Scotland and any household respondent is the home owner or renter of the property		
W9	Link to education records. Respondent is a rising 16 year old this wave and has never completed an adult interview or was a rising 16 year old last wave	Information leaflet	Oral
^a Rising 16's asked consents in each subsequent wave			

Table 2: Summary of consents for Innovation Panel (IP)

Summary of consents for Innovation Panel (IP) components wave 1 to 10			
<i>Wave</i>	<i>Study component</i>	<i>Communications</i>	<i>Consent</i>
All waves	Basic survey	Advance letters, Information leaflet, Participant handbook	Oral
W1	Link educational records of adults age 16 – 23 and children age 4-15	Information leaflet	written
W1	Link DWP records of adults aged 16+	Information leaflet	written
W2	Link health records of adults age 16+ and children age 0 – 15	Information leaflet	written
W2	Link educational records of adults for those who turned 16 since W1 consent and those who did not consent before. Adults age 16 – 24 and children age 4 -15	Information leaflet	written
W4	Link DWP records of adults aged 16+	Information leaflet	written
W8	Link DWP records of adults aged 16+	Information leaflet	oral
W9	Link records to Credit Reference Agency (CRA) records and also to be passed on to the FCA for adults aged 16+	Question	oral
W9	Link DWP records of adults aged 16+	Information leaflet	oral

W9	Link to information about employers of adults aged 16+ in paid employment	Question	oral
W10	Link to Electoral Register of adults aged 16+	Question	oral
W10	Link DWP records of adults aged 16+	Information leaflet	oral
W10	Link to information from Twitter account of adults aged 16+ who have a twitter account	Question	oral

ANNEX 2 DATA ACCESS RULES AND PROCEDURES: SURVEY AND RELATED DATA

1. OVERVIEW

- 1.1. This annex sets out the framework for providing access to all *Understanding Society* data other than biological samples, genotype, epigenetic and other related data, which are handled by the METADAC³ (see Annex 3). Applications to access survey and other related data are handled by the *Understanding Society* Scientific Leadership Team (SLT) through either the UK Data Service (UKDS) and/or the ONS SRS. These procedures operate within the overall framework of the *Understanding Society* Data Access Strategy.
- 1.2. There are three categories of *Understanding Society* data based on the likelihood and potential risk of disclosure. Each have defined access mechanisms:
- The majority of users access data via a standard licence known as an 'End User Licence' (EUL). Their application is authorised directly by the UKDS;
 - Access to more detailed data, which are potentially disclosive of the identities of individuals, households or organisations is provided via a special licence through the UKDS and/or ONS SRS;
 - For a small number of users access to sensitive, confidential and/or disclosive data is provided via 'Secure Access' through the UKDS⁴ and/or the ONS SRS. This type of access is only available to persons who are designated as 'ESRC Accredited Researchers' for UKDS or 'Approved Researcher' for ONS SRS⁵.

Data Categorisation

- 1.3. The *Understanding Society* DAC agreed a set of principles through which the *Understanding Society* data is categorised, reflecting an assessment of the likelihood and potential impact of disclosure. Data that risk the disclosure of information which could identify individuals, households or organisations with which they are associated will require the highest degree of security and management.
- 1.4. *Understanding Society* data falls into one of three categories, which are defined by the likelihood and potential impact of disclosure:
- **Impact Level '1'**
Low impact (non-disclosive data: e.g. most survey responses and digital reports of biological assays and direct physical measurements);
 - **Impact Level '2'**
Medium impact (potentially disclosive: e.g. medium level and coarse geographies); for non-UK applicants this category is subdivided into two groups, with the less disclosive subdivision (2a, also known as Conditional Access for the BHPS) available to international applicants available on the same terms as for UK applicants, and the more disclosive subdivision (2b) only available if additional scrutiny of application is undertaken;

³ <http://www.metadac.ac.uk/understanding-society/>

⁴ <http://ukdataservice.ac.uk/get-data/secure-access.aspx>

⁵ <https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme>

- **Impact Level '3'**

High impact (e.g. data with the most detailed geographies) and linked individual level administrative data.

Table 3 provides examples of data included in each category.

Table 3: Examples of data included at different impact levels

<i>Impact level</i>	<i>Data included</i>
Impact Level 1 (EUL)	Content of main interview survey, with some restriction on degree of disaggregation of categories, for example of occupation, industry, country of birth. Top coding of income. Region is lowest level of geography.
Impact Level 2a (Special Licence (or Conditional Access for BHPS), available to non-UK users)	Main interview survey data with detailed classifications - full occupation, industry, country of birth classifications, month as well as year of birth, no top coding of income. Main interview survey data with intermediate geographies e.g. local authority district, parliamentary constituency, travel to work area. Genotype or epigenetic data without linkage to survey data.
Impact Level 2b (Special Licence, available to non-UK users with additional scrutiny over 2a data)	Main interview survey data with neighbourhood spatial data; or with organisation identifiers, e.g. School identifier codes or linked genotype or epigenetic survey data.
Impact Level 3 (Secure Data Service)	Very fine geographies: postcodes and grid references. Administrative data linked at individual level (e.g. education data).

Table 4 below shows the how different combinations of data would be classified under this proposal.

Table 4: Classification of data combinations

	Main interview survey data with detailed classifications	Intermediate level spatial data	Neighbourhood spatial data
Main interview survey data with detailed classifications (2a alone)			
Intermediate level spatial data (2a alone)	Category 2b		
Neighbourhood spatial data (2b alone)	Category 2b	Category 2b	
Organisation identifier (e.g. school codes) (2b alone)	Category 3	Category 2b	Category 3

1.5. Note when combining genetic or epigenetic data with survey data, it is only possible to use variables in the EUL version of the survey data.

1.6. In general projects are limited to a maximum of three 'special licence' datasets due to the increased risk of disclosure through small intersections. It is generally not allowable to mix census level Special Licence neighbourhood geographies for different years, for example, LSOA 2001 and LSOA 2011. This is in order to prevent the risk of the identification of small geographical areas due to boundary changes. However, this is judged on a case by case basis, and if appropriate data combined like this are treated as level 3 data and can be used via the UKDS and/or ONS SRS secure access.

Commercial Use

1.7. Commercial organisations can apply for access to *Understanding Society* data and are subject to the standard procedures (outlined in section 1.2 of this Annex). As per any application commercial organisations will be required to demonstrate the public benefits that are likely to flow from the research use. Applicants must confirm that their use of the data is for research purposes and not for commercial exploitation. No organisation, commercial or otherwise, should be allowed to gain control or ownership over the *Understanding Society* resource.

Data Linkage

1.8. Where consent has been obtained, *Understanding Society* data may be linked to administrative data (e.g. school and hospital records) and to text, voice and video information. Data linkage will be carried out in secure conditions to ensure that the owners of the administrative data cannot link data collected in the Survey to the name and address of the person concerned. Therefore any application to link *Understanding Society* data will be classified as Impact level '3'.

1.9. Access procedures for researchers wishing to link the data at individual, organisation and area level will be governed by the conditions of access set out in this Strategy. Where the data owner is not ESRC, approval will be required from the data owner, before any decision is made by the SLT on whether to grant access. The ESRC recognises that Intellectual Property Rights in relation to some data linked to the Study from other sources, such as administrative data from a government department, will remain with the organisation which supplies it and cannot be transferred to the ESRC.

Violations of Access Conditions

1.10. Access arrangements will ensure that an appropriate set of penalties are applied should violations of access conditions take place. Penalties can be imposed on users and/or their institutions. Further details can be found in section 2.6 below.

2. DATA ACCESS PROCEDURES

2.1. *Understanding Society* data is made available for researchers to undertake their analysis by identifying and requesting data from the UKDS or the ONS SRS.

2.2. The UKDS and the ONS SRS provide the interface with researchers and review applications to ensure they comply with this strategy and their own specified data security requirements. If researchers cannot meet these their applications are rejected at this stage⁶.

2.3. Each Impact Level of data has its own access mechanisms. The majority of users will apply to use Impact level '1' data via a standard licence known as an 'End User Licence' (EUL). Their application is authorised directly by the UKDS. Access to more detailed data, classified as Impact level '2',

⁶ <http://www.data-archive.ac.uk/media/132701/UKDA171-SS-MicrodataHandling.pdf>

which are potentially disclosive of the identities of individuals, households or organisations is provided via a special licence. This level of access is authorised directly by the UKDS or ONS SRS after consultation with the SLT. For some users access to Impact level '3' sensitive, confidential and/or disclosive data is provided via 'Secure Access' through the UKDS and/or ONS SRS. Such applications are reviewed and approved by the SLT, and if appropriate the data owner of any linked administrative data.

- 2.4. *Understanding Society* biomedical data that are not distinct from the main survey data (measurements collected by the nurse e.g. blood pressure and analytes derived from the blood samples) are covered by standard access procedures. As indicated in paragraph 1.1 access to biological samples, epigenetic and genotype data is regulated by METADAC.
- 2.5. The arrangements for access to each category of data are outlined below. Penalties for misuse of data under each category are included. Full details of penalties attached to each possible type of misuse are detailed in the UKDS Breaches Penalties Policy⁷ or ONS Research Data Access: Compliance and Breaches Policy.

Impact Level '1': Standard Conditions

Access Agreements

- 2.6. Users are asked to register with the UKDS and agree to the terms of the 'End User Licence' (EUL).

Penalties

- 2.7. Penalties for misuse of the data depend on the severity of misuse but include:
- suspension of rights for the user to access UKDS services for a fixed period or permanently;
 - suspension of the user's institution's access to UKDS for a fixed period or permanently;
 - suspension of access from all ESRC data services (both individual and institution) for a fixed period or permanently;
 - sanction from future ESRC funding (individual and/or institution) for a fixed period; and
 - whatever penalties may exist in law under the Statistics and Registration Services Act 2007, the Data Protection Act, or any other relevant legislation.

Impact Level '2': Special Licence

Access Agreements

- 2.8. Potential users are required to register with the UKDS or ONS SRS and apply for 'Special Licence' datasets for a particular research purpose. Users can request data from the UKDS or ONS SRS in the usual way and will be directed to the 'Special Licence' forms for completion. The UKDS or ONS SRS will seek approval for the user and the research from the *Understanding Society* SLT, which will act as data custodian under delegated authority from the Governing Board, on behalf of the ESRC as the data owner. Where the data owner is not the ESRC, approval will be sought from the data owner before a decision is made.

⁷ http://ukdataservice.ac.uk/media/176861/UKDA142_SDS_SecurityBreaches_public.pdf

2.9. Access agreements for users outside of UK HE sector may be restricted for this level of data depending on the security and institutional arrangements of their host organisation. Decisions will be made on a case by case basis, depending upon information provided in the 'Special Licence' application.

Penalties

2.10. As per Section 2.7 above with all licenced data, penalties will depend on the severity of misuse.

Impact Level '3': Secure Access

Access Agreements

2.11. The Statistics and Registration Services Act 2007⁸ widens the potential for access to official statistics by individuals who are designated 'Approved Researchers' for a particular usage.

2.12. The ESRC has adopted and extended this model for access to detailed data distributed by the UKDS which is not explicitly covered by the Statistics and Registration Services Act 2007 (i.e. data created by academic research units rather than Government Departments), creating a parallel designation of 'ESRC Accredited Researcher'. For researchers applying through the ONS SRS the same mechanism has been approved by the ESRC for users with Approved Researcher status.

2.13. Users for this level of data access are required to be registered users of the UKDS or ONS SRS as appropriate. They will also need to complete the 'ESRC Accredited Researcher' approval process for UKDS applications or the 'Approved Researcher' process for ONS SRS applications, providing details of their intended usage which the SLT, on behalf of the data owners, will use to judge whether secure data are in fact required. Where data ownership is shared or does not sit with ESRC, approval will be sought from the data owner, before a decision is made on whether to grant access.

2.14. On gaining approval, users are required to complete a training session introducing them to the UKDS or ONS SRS as appropriate and their responsibilities in its use, including information about statistical disclosure control, handling of confidential data, their responsibilities in law and the potential penalties.

2.15. On successful completion of training, users of the UKDS are asked to sign an 'Agreement of Terms of Use of the Service' and will be allocated credentials to access the data securely. For access to data through the ONS Secure Research Service, users will need to sign an Approved Researcher Declaration and an SRS SysOps document.

Penalties

2.16. Policies developed by the UKDS or ONS SRS will be applied to the use of *Understanding Society* data for this level of data access. These include all of the measures listed above.

2.17. Applicants applying to access *Understanding Society* data who wish to determine whether the data is suitable for the planned research will be able to consult the online documentation or consult the *Understanding Society* support staff at the University of Essex <https://www.understandingsociety.ac.uk/documentation/help>

⁸ <http://www.legislation.gov.uk/ukpga/2007/18/contents>

3. APPLICATION REVIEW PROCESS

- 3.1. Requests to access data are made to the UKDS or the ONS SRS. The UKDS or ONS SRS check applications to ensure that they comply with this strategy and that the applicant(s) satisfy the security and best practice guidelines in force. If not, then they will liaise with the applicant(s) either to help modify their application as appropriate or if this is not possible to reject the application.
- 3.2. Applications for requesting access to Impact level '2' or '3' data will be forwarded to the SLT by the UKDS or ONS SRS. The SLT will determine whether the proposal meets the criteria established by the Committee (see 3.4 below). For international access to data classified as impact level 2b, or other applications considered sensitive, an internal SLT group will review the applications. The internal group, consisting of at least three experienced SLT members will independently assess the application against the rules laid down in this document. Where there is not a consensus the *Understanding Society* Director will make the final decision.
- 3.3. Applications for data are considered by the SLT within 10 working days. Applications are assessed using the following criteria, and in the light of the impact levels outlined above:
 - (1) Has the application been submitted by bona fide researchers who can demonstrate public interest?
 - (2) Does the application violate (or potentially violate) any of the ethical permissions granted to the study or any of the consent forms signed by the participants or their guardians?
 - (3) Does the application run the risk of producing information that may allow individual participants to be identified?
 - (4) Does the application run a significant risk of upsetting or alienating study members or of reducing their willingness to remain as active participants in *Understanding Society* based research?
 - (5) Does the application address topics that fall within the acknowledged remit of the *Understanding Society* project, as understood by participants?
 - (6) Does the application demonstrate that ESRC policy regarding deposit of data will be adhered to?
- 3.4. If access is granted, the decision is communicated to UKDS or ONS SRS and data will be made available through them under a Secure Access / Special Licence agreement as appropriate.
- 3.5. The SLT will review decisions in the case of appeals received from potential users. If the SLT is content with the decisions made it will refer the complaint to the published categorisation principles. A further appeal may be made to the *Understanding Society* Governing Board. A sub group of the Governing Board will be convened to discuss the appeal. Their decision will be reported to the next Governing Board meeting.
- 3.6. The Governing Board will also have oversight of the whole process. SLT will report to them at each meeting: numbers of applications in different categories considered, number approved and rejected, any issues arising which suggest the need for a review of the Data Access Strategy. Should the Governing Board consider a substantial review of the Data Access Strategy is required it may initiate a consultation with external experts, but the Board would be responsible for agreeing a revision.

4. DATA RETURN POLICY

- 4.1. It will be a requirement that any researcher who is approved for access to *Understanding Society* data and in the process of undertaking their research generate new data, the new data and associated metadata must be offered for deposit to a standard agreed with the UKDS and/or ONS SRS. Deposit of analysed data will be the responsibility of researchers and penalties may be incurred for failure to adhere to the agreed policy.

ANNEX 3 ACCESS TO GENOTYPE OR EPIGENETIC DATA

1. *Understanding Society* has a significant biological component, including direct measures like blood pressure and the collection of biological samples. The study has captured bio-medical assessments on approximately 20,000 adult survey participants and obtained approximately 13,500 biological samples.
2. This Annex describes the process of gaining access to two products of this data collection. The results of genotyping deoxyribonucleic acid (DNA) samples, including data from a genome wide scan using an Illumina human core exome array and the results of methylation analysis of the DNA samples using an Illumina methylation EPIC array. Further information about the Infinium human core exome array and the Infinium MethylationEPIC array is available at <https://www.illumina.com/products/by-type/microarray-kits/infinium-core-exome.html> and <https://www.illumina.com/products/by-type/microarray-kits/infinium-methylation-epic.html>, respectively. With consent, 10,500 DNA samples have been genotyped. Access to genotype and methylation only data are managed by European Genome/Phenome Archive (EGA)⁹, and access to genotype and methylation data linked with EUL survey data are scrutinised by the METADAC¹⁰.
3. The principles guiding access to *Understanding Society* genotype and methylation data are those of the overall study expressed in the main Data Access Strategy. There are a small number of variations set out in the paragraphs below.
4. The data should be available to researchers consistent with the terms of consent for the specific data requested. The consent form for blood and DNA collection is appended to this annex. Communications with study participants for the collection of biological samples from which this data derives can be found in the Wave 2 Nurse section of the website for fieldwork documents: <https://www.understandingsociety.ac.uk/documentation/health-assessment>.
5. Given the nature of the consents provided by participants, the data are not available for commercial research purposes. 'Commercial research' means any research that has, as its primary purpose, the development of products or procedures that may be sold or offered for profit or commercial advantage. Any application from a commercial organisation rather than a charity or not-for-profit would be subject to scrutiny by the METADAC to ensure that there was a demonstrable public benefit arising from the proposed research.
6. Genotype and methylation only data are considered Impact Level 2a for risk of disclosure, while genotype and methylation data combined with other data are considered Impact Level 2b. For genotype and methylation only data, Wellcome Trust Sanger Institute (WTSI) DAC would grant access to non-UK users providing they meet all the criteria and terms and conditions and provided that the disclosure risk can be managed in a manner equivalent to that for UK based researchers. Genotype-combined with survey data applications would be considered by the METADAC. Only survey data available via the End User Licence would be available to link to genetic data.
7. For disclosure risk and / or quality control reasons a number of cases have been removed from the genome wide scan data files before depositing. These include those cases where genetic and reported gender or parentage (where both child-parent have provided DNA) results do not match.
8. Where an application relates to genome wide scan data from the Illumina human core exome array, with the addition of an age band variable but with no other data from the survey, applications should be made to the WTSI DAC. The same would apply if the same were required for

⁹ <https://www.ebi.ac.uk/ega/studies/EGAS00001001232>

¹⁰ <http://www.metadac.ac.uk/understanding-society/>

methylation data from the EPIC array. The data at the EGA will include standard variables derived from the genome wide information, for example family correlations and the first ten principal components. The WTSI DAC has been given delegated responsibility for considering such applications within the framework agreed with the former Understanding Society DAC. If the WTSI has any concerns about an application, it will be referred to the METADAC. Once an application is approved, the WTSI DAC will authorise applicants so they can download the dataset from the EGA.

9. Applications for genome wide scan data and/or methylation data linked to EUL survey data should use the application form on the *Understanding Society* website. This includes the terms and conditions to which the applicant agrees. Applicants are asked to specify the nature of the proposed research and all data to be used in the project. The METADAC will evaluate the application on the basis of:
 - advice from the METADAC technical support team, which will provide advice on the disclosure risk associated with the application.
 - where the METADAC or the technical support team consider appropriate, advice from the relevant experts.
 - an assessment of whether the applications meets the criteria set out in section 3.4 of Annex 2 of the Data Access Strategy.
10. The evaluation of the proposal will be made by the METADAC within six weeks of its submission. The METADAC Secretariat will inform the applicant of the decision in writing.
11. Where the request involves a combination of genotype and methylation data with EUL survey data, the dataset will be supplied by ISER. This will include a small subset of survey data to minimise disclosure risk, and depending on the research question, applicants can request the full genome and / or methylation scans or list the *Single Nucleotide Polymorphism* (SNPs) required.
12. Once an application has been agreed, any future additions to the dataset (for example, if an additional linked phenotype variable is required) will have to be processed by the METADAC Technical Review Team and must comply with the original application. If the additional request is not deemed to be part of the original proposal, a new application will be required.
13. Each application will generate one data file which will be randomly sorted and given a unique identifier. The data set will not include identifiers shared with the survey data set available from the UK Data Service or the genome wide or methylation scan data at EGA.
14. Whether a researcher accesses the genome wide or methylation scan data alone via EGA or in combination with survey data, there will be a non-negotiable requirement to supply the *Understanding Society* Health Research manager with three outputs:
 - any new scientific data (e.g. genotypes, or new variables) created in the project. These will be shared with the wider scientific community increasing the value of the resource;
 - notification of all publications using the data;
 - any incidental information on genetic variants identified during the research project that could potentially meet three criteria*:
 - scientific validity (the genotyping is of adequate quality);
 - clinical significance (the disease or condition caused by the genetic variant is potentially serious) , and
 - potential benefit, i.e. a valid approach exists to prevent or cure the condition / disease of concern and that early knowledge of the genetic risk to which an individual is exposed could enhance the efficacy of that prevention/cure.

Please note researchers are not being asked to investigate clinically relevant findings, but to report them if they are identified during the course of their research.

15. *Understanding Society* has no current plans to share incidental clinically significant findings with respondents. Indeed the informed consent given by respondents, in keeping with most other studies, specifies that '*that no personal test results from my DNA will be given to me*'. However, there are discussions within the ethical and scientific communities about what constitutes best practice in this respect, and advice may change in the future making it a requirement to report genetic findings to study participants if they satisfy the above criteria. *Understanding Society*, in common with other studies, has therefore decided to collect information on potentially relevant findings from research groups using the DNA data. It should be emphasised, however, that under no circumstances should researchers attempt to contact participants about potentially clinically relevant findings, this would be done, if there was a change in policy, by the *Understanding Society* study team after retrospectively agreeing new consents with participants or alternative ethically approved processes are agreed.

*Knoppers BM, Joly Y, Simard J, Durocher F. The emergence of an ethical duty to disclose genetic research results: international perspectives. *Eur J Hum Genet* 2006;14(11):1170-8

APPENDIX 1 – TEXT FROM CONSENT FORM FOR BIOLOGICAL SAMPLES

I consent to a qualified nurse taking a sample of my blood on behalf of the Institute for Social and Economic Research/National Centre for Social Research.

I have read and understood the Information for Participants leaflet about the second stage of the survey. The nurse has explained the procedures, and I have had an opportunity to discuss these with him/her.

Initial:

I consent to my blood being taken, stored and used in scientific research. I understand that all blood test results and related information will be coded so I cannot be identified. For purposes of scientific analyses, links to my name will be held separately and securely from any data collected. The sample will not be tested for HIV. I also understand my right to withdraw consent for storing the blood sample.

Initial:

I give my consent for a sample of my DNA to be taken from my blood, stored and used in scientific research.

I understand that:

- *the DNA samples and related information will be coded to ensure that my personal identity is not revealed to researchers carrying out scientific analysis*
- *links to my name will be held separately and securely, for administering the study and data collection*
- *that no personal test results from my DNA will be given to me*
- *the data and samples will be owned by the Study and the ESRC. No samples or information will be sold.*
- *The DNA analyses will not be used for paternity analysis, life insurance, mortgage applications or police records.*

I also understand my right to withdraw consent for storing the blood sample.