

DPIA for *Understanding Society*

Step 1: Identify the need for a DPIA

Explain broadly what project aims to achieve and what type of processing it involves. You may find it helpful to refer or link to other documents, such as a project proposal. Summarise why you identified the need for a DPIA.

As part of *Understanding Society*: The UK Household Longitudinal Study (UKHLS) we have reviewed our processes which involve the processing and use of personal information and have highlighted the need to have a data privacy impact assessment. This is because of the large volume of data, collected from a large number of individuals. The data collected includes personal information, including special category data.

The UKHLS is a long-running longitudinal study. It started in 2008 with a sample of addresses. Individuals in households that participated in the original interview (Wave 1) became sample members. They are regularly invited take part in another wave of data collection for the survey, and reminded that the survey is completely voluntary. The UKHLS has incorporated the British Household Panel Survey (BHPS), a longitudinal survey which was started in 1991, with regular interviewing thereafter. The UKHLS includes an Innovation Panel which is used by researchers as a test-bed for innovative ways of collecting data and for developing new areas of research.

As a longitudinal survey, UKHLS – and the BHPS before it – aims to interview the same people each time, to enable researchers to look at individual-level changes. This requires us to collect the names and contact details of sample members, so that they can be re-contacted at the next wave and invited to participate. To aid in contacting sample members, we also ask them to provide the contact details of a 'stable contact' who would know where they move to if they change address. The collection and use of the stable contact information has been the subject of a separate DPIA. In addition to these personal contact details, the survey collects information about all aspects of their lives from respondents. This includes information about health, religious beliefs, sexual orientation, political opinions, and trade union membership. Therefore, we collect special category data.

ISER is the data controller for the UKHLS, the fieldwork agencies are the data processors, and are sub-contracted by ISER to manage the data collection, either

in-person (by interviewers visiting the sample members), by telephone (by interviewers calling the sample members), or online.

The sample information required to re-contact sample members is sent from ISER to the fieldwork agencies using a secure, encrypted online portal ("the portal"). The access to this portal is password controlled and can only be accessed by specific IP addresses. Access to the Portal is regularly reviewed to ensure that it is up to date. The sample information is processed by the fieldwork agencies and made available to interviewers. It is also used to send the invitations and reminders to the sample members to participate in the next wave of the study, these contacts are by post, email, or SMS. In addition, as part of the sample file, ISER sends the fieldwork agencies some data collected in previous waves of the study where this is required to route the respondent to the correct set of questions and to identify where a change of circumstances has occurred.

The sample details are stored securely in the fieldwork agency servers, and during face-to-face interviewing, the interviewers only have the sample information for their allocation loaded onto their encrypted, password-protected lap-tops. When the interviewers synchronise with the fieldwork agency servers – which is a task they are instructed to do daily – the data are transmitted to the servers and removed from the lap-tops.

The fieldwork agency process the data from the interviews, clean it and check it for errors. They restructure the data to an agreed specification, and then they transfer it to ISER using the encrypted online portal.

At ISER, the data are stored on secure servers. The sample information containing the names, addresses, and other contact details (email addresses, phone numbers) are separated from the survey data and stored securely in a database which has very limited access rights. Only those who need the database in order to update it, or to use it to generate the sample file for the next wave are able to access it. The survey (questionnaire) data are stored on protected secure servers and are processed by the UKHLS data team. This team checks, cleans and restructures the data. They also created added-value components to the data, such as derived variables, imputations and weights. The survey data are stored either in a protected area (equivalent to End User License data from the UK Data Archive) or a restricted area (equivalent to Special License or Secure Access data). The access rights to these areas are strictly controlled and reviewed regularly. The data are deposited with the UK Data Archive. The released data do not contain any names, addresses, or other contact details.

UKHLS data are also processed as part of the tracing process. Where a sample member moves and cannot be traced by the interviewer, the fieldwork agency return the information to ISER using a secure ftp site. This file is automatically picked up and processed. The file is used to generate another file which can be uploaded to the tracing service we use (currently GB Group) and is automatically matched against their databases to find updated contact details. Where new

details are found, the traced mover file is uploaded to the portal so that the interviewer is able to contact them to invite them to take part.

UKHLS data are also processed as part of the keep-in-touch policy for the study. Sample members are sent a short report of results from the study, including information about who had used the data and media coverage of the study. An email "Participant Update" is also sent out periodically during the year. For these mailings, the database is used to generate a mailing list of the current active participants.

UKHLS data are processed to support the "Associated Studies" component of the study. This is where external researchers can request to use the study as part of their research. The Associated Studies are typically small-scale qualitative research projects, which are funded separately and require separate ethical approval. For Associated Studies, sample members are written to by *Understanding Society* and asked if they would like to participate in this separate study. If they do want to take part, they return a form to ISER to consent for us to pass on their contact details to the researcher.

Step 2: Describe the processing

Describe the nature of the processing: how will you collect, use, store and delete data? What is the source of the data? Will you be sharing data with anyone? You might find it useful to refer to a flow diagram or other way of describing data flows. What types of processing identified as likely high risk are involved?

At the first wave of the survey, we contacted a sample of addresses by letter. The letter invited them to take part in the survey, and included information about the study. An interviewer then visited the households to see if they were willing to participate, and if so to arrange a suitable time to carry out the interview. Those households where an adult participated in this first wave became the core sample for the study. During the first interview, the interviewer collected names, dates of birth, and other contact details from the participants, including contact details for a stable contact, if the participant chose to provide one. Thus, all the personal data we hold has been given to us by the participant.

During the interview the data are used to route the participant through the correct questions. Data from a previous wave are used to validate the identity of the individual and to indicate whether there have been any changes in circumstances since the previous interview.

Data are used after the interview to contact the sample members, to provide information about how the survey has been used, and the research that has been published, and to invite them to participate at the next wave.

Data, once checked, cleaned and pseudo-anonymised, are deposited with the UK Data Archive and made available for secondary analysis by registered researchers.

During some interviews, we have asked for the consent from participants to allow us to link administrative data held about them to their survey responses. The consent question and information leaflet are first approved by the data-holder and then approved by the University Ethics Committee before being used. Before any data are transferred to enable the linkage, Data Sharing Agreements are in place between ISER and the data-holder, to ensure that the data linkage is done in a secure way.

All data are transferred using secure and encrypted procedures and we do not identify any high risk data processing.

Describe the scope of the processing: what is the nature of the data, and does it include special category or criminal offence data? How much data will you be collecting and using? How often? How long will you keep it? How many individuals are affected? What geographical area does it cover?

The data include the contact details of the sample member (name, address, telephone numbers, and email address), and the stable contact (if the participant chose to provide one), as well as the questionnaire responses that they have chosen to give. The data contains special category data – the questionnaire asks the participant to tell us about their ethnicity, political opinions, religious beliefs, trade union membership, sexual orientation. At the start of every interview the participant is reminded that the interview is completely voluntary and if they do not want to answer any question, they can just move on. In addition, the more sensitive questions are part of a 'self-completion' section where the participant can choose not to answer any question without revealing this to the interviewer. As more of the participants complete their survey online, they are able to go past questions without answering them if they do not want to. Of course, participation in the study is optional.

The data are collected as part of an ongoing longitudinal study, the strength of which is that it interviews the same people each time, to enable researchers to look at individual-level changes. The data are collected as a social science resource, and the study is funded in this way. The data are made available to researchers who are registered with the UK Data Archive. As a resource for secondary analysis, the survey is not collected for one particular project, or to answer a specific research question. This is why the questionnaire covers a wide range of topics, and different researchers will use different parts of the study. Time is a constraint on the amount of questions that can be asked, and so we try to make sure that all the questions in the survey are relevant and appropriate for research. A longitudinal study becomes more useful the longer it lasts, with researchers being able to analyse present outcomes by past experiences. It is therefore expected that the survey data, once deposited with the UK Data Archive, would remain in perpetuity. The sample data (names, addresses, contact details) are kept unless the sample member requests that we delete them.

The questionnaire is around 45 minutes long for each adult who chooses to participate. In addition, one adult will be asked about the household composition and a household-level questionnaire about household-level characteristics (such as number of rooms, rent/mortgage payments, utility bills). The household composition and household questionnaire take around 15 minutes to answer. Children aged 10-15 are also invited to complete a questionnaire (currently on paper). This is completely voluntary and the children do not have to take part if they don't want to. The parent/guardian of the child is first asked if they are willing for their child to take part, and if so, the child is then asked. The parent/guardian can request to see a blank copy of the paper questionnaire, to guide their judgement. However, once the questionnaire has been completed, the

same level of confidentiality is given to the child as with the adult, and the parent cannot see the completed questionnaire.

The main survey for UKHLS is annual, with sample members being invited to participate throughout the year. For other parts of the study, we invite the sample members to participate between the waves of the main survey – for example, with the monthly Coronavirus survey, where sample members are invited monthly. These are related to relatively short-term research projects (e.g., 12 months). These additional projects will be subject to a separate DPIA, and will often be supported by a project-specific privacy notice.

At the last completed wave of data collection, the Innovation Panel (IP11) consisted of 4,387 issued households, comprising 5,753 individuals in productive households, with 2,990 adult participants. In the main sample (Wave 9), there were 25,951 issued households, with 52,691 individuals in productive households, and 36,055 adult participants. The survey covers the UK, including Scotland north of the Caledonian canal, and Northern Ireland.

At each wave of data collection, we issue households where there was at least one adult participant at the last wave. We also issue households where we could not contact anyone at the last wave, and those where the participant gave a 'soft' refusal (for example, they were too busy). We remove from the issued sample all those who give an adamant refusal. We also remove from the issued sample households who had given a soft refusal in the last two years, who had not been contacted in the last two years, and those who had given a soft refusal two years ago and were not contacted last year. These households become 'dormant' and are not issued to interviewers or invited to participate online. However, if they contact ISER and request to be included in the survey again, they can be invited.

Describe the context of the processing: what is the nature of your relationship with the individuals? How much control will they have? Would they expect you to use their data in this way? Do they include children or other vulnerable groups? Are there prior concerns over this type of processing or security flaws? Is it novel in any way? What is the current state of technology in this area? Are there any current issues of public concern that you should factor in? Are you signed up to any approved code of conduct or certification scheme (once any have been approved)?

The UKHLS sample members have been part of the study for some time. During the year we send them a couple of short reports of recent findings from the study to let them know how the survey is being used. Each year, we write to them in advance of their annual interview and given them information about the upcoming survey. For those who are invited to participate online, this letter contains their unique log-in details. Where we have an email address or mobile phone number, we also send the invitation to complete online using these methods. The letter contains all the contact details to allow sample members to contact ISER – Freephone, email, Freepost, website with contact form. We know that some may not want to complete online, so we let them know that if they do not want to take part online, an interviewer will call them. For those adults who were issued directly to an interviewer, the letter lets them know that the interviewer will soon call. The initial call to continuing households is usually by telephone, so the interviewer can answer any questions the sample members may have, and then invite them to take part in the survey this year. If the sample member is willing, the interviewer can arrange an appointment for a suitable time. Individuals in the sample therefore have a great deal of control and they know – and are reminded each year – that the survey is completely voluntary and they do not have to take part. Furthermore, if they contact us after their interview and ask us to delete the data, we will do that up until the point it is deposited in the UK Data Archive.

Being part of the study for so long, sample members know how we use their data and are aware that we write to them each year in advance of their interview, using the contact details that they have given us. If they change address between interviews and we cannot contact them, they know that we will write or call their nominated stable contact, if they have chosen to give us one. The only information we ask the stable contact is whether they have a new address for the sample member, we do not share any other information with them. As part of the interview, some of the information that they had given at the last interview is fed back to them to see whether circumstances have changed (for example, “last year you told us you were a plumber, is that still your job?”). We also send sample members updates during the year with information on how the data have been used.

The sample does include children, with those aged 10-15 invited to complete a paper questionnaire. Verbal consent is given by the parent/guardian, and then the young person is asked if they would like to complete the questionnaire. They

are given an envelope so that when they have finished the questionnaire, they can put it in the envelope, seal it, and give it back to the interviewer. If the interviewer is not present (e.g., the parent/guardian completed their interview online), the questionnaire is sent to the household with a return envelope to send the completed questionnaire back.

As we try to interview the same individuals each year, there is a chance that someone who was able to give an interview in the previous year, is no longer physically or mentally capable. Interviewers are experienced and trained to recognise this, and so will not carry out an interview if they judge that the sample member is not able to give informed consent.

To recognise the fact that the interview may be a burden, especially with the invitation to take part each year, we do give the sample members a token of appreciation for their efforts. In line with many large surveys of this type, we include a gift-card with the advance letter, or offer a gift-card after the interview.

The way in which data are processed for the UKHLS is common to other longitudinal surveys, where details of the sample are required to be kept and used to issue the survey to the same people at the next wave of data collection. With the transfer of personal details between organisations, we recognise that are concerns about data security. ISER are ISO-27001 accredited, and it is a requirement of the funding for the study that sub-contracting fieldwork agencies are also ISO-27001 accredited. This is the international standard for information security management. This ensures that we maintain the highest standard for data security and that this covers not only the security of electronic data, but also other information such as letters, papers etc. Any electronic data that is transferred between ISER and the fieldwork agency is done via an online secure, encrypted Portal. Access to this is password-protected and is restricted to certain IP addresses.

In addition to the regular ISO-27001 audits and Cyber-Essential checks, the design, content, and processes used in the survey are regularly reviewed by the University Ethics Committee.

Describe the purposes of the processing: what do you want to achieve? What is the intended effect on individuals? What are the benefits of the processing – for you, and more broadly?

The purpose of the processing is to ensure that we produce a high-quality study that is representative of the population, and is a valuable resource to researchers in the UK, and all over the world. The study is a key investment for the ESRC and UKRI, and is downloaded by thousands of researchers from a large number of different disciplines.

We do not intend for the study to have any significant effect on the individual sample members. They receive some information during the year, and once a year they are invited to take part in a survey. If they decide to take part, they can take 45-60 minutes answering the survey. The survey is not designed as an intervention. ISER does receive a small amount of feedback to the interview, and to the information updates during the year, and these are largely positive. We hope that the sample members feel that they are part of an important social survey, and that they feel positive about their participation. They may also be interested in the ways in which the study is being used, and the research that has come from it.

ISER benefits from the study because it gives the institution and the university a positive reputation. Researchers at ISER who work on the study are internationally recognised as experts in survey methods and advise other studies. Social science in general benefits from the study because it is one of the largest studies of its kind in the world, and is a leading-edge innovative study. It is a valuable resource for social and medical scientists, who are able to access the data free of charge for their research.

Step 3: Consultation process

Consider how to consult with relevant stakeholders: describe when and how you will seek individuals' views – or justify why it's not appropriate to do so. Who else do you need to involve within your organisation? Do you need to ask your processors to assist? Do you plan to consult information security experts, or any other experts?

Sample members can contact the UKHLS team in a number of ways to comment on the study. Other stakeholders, such as the ESRC and government co-funders, are part of the management structure and are consulted regularly. The study also has a number of "Topic Champions" who are leading academics in their fields and who are regularly consulted about the study as a whole, and individually consulted to advise on the content of the survey in their field.

We ensure that our data processing systems are secure, and that access to personal information is strictly limited to those who need it. We have consulted extensively with our information security expert; Ray Ware, the ISER Information Technology and Security Manager and have had this DPIA reviewed by the University's Information Assurance Manager. We have consulted with the ICO regarding Stable Contacts' data processing, and have followed their advice and guidance which includes publishing enhanced user awareness resources and a dedicated DPIA.

Step 4: Assess necessity and proportionality

Describe compliance and proportionality measures, in particular: what is your lawful basis for processing? Does the processing actually achieve your purpose? Is there another way to achieve the same outcome? How will you prevent function creep? How will you ensure data quality and data minimisation? What information will you give individuals? How will you help to support their rights? What measures do you take to ensure processors comply? How do you safeguard any international transfers?

The lawful basis for processing personal data is Public Task.

The processing does achieve the purpose, and there is no other way to achieve the same outcome. Other surveys are not designed to capture the long-term life course in the same way as *Understanding Society*. Other longitudinal studies in the UK tend to be of specific ages (birth cohorts) or life-stages (e.g., English Longitudinal Study of Ageing). Only *Understanding Society* is a broad population representative study. The data collected as part of the study cannot be replicated by other means, e.g. use of administrative data.

We include information on the *Understanding Society* website about GDPR, and we have a Privacy Notice. ISER and Kantar Public (the current fieldwork organisation) are both certified to ISO 27001 and we ensure that the processing of personal data conforms to the requirements of this certification.

We do not transfer personal data internationally, and within ISER access to personal data is restricted to a small group of ISER personnel on a strict business need to access basis.

Step 5: Identify and assess risks

Describe source of risk and nature of potential impact on individuals. Include associated compliance and corporate risks as necessary.	Likelihood of harm	Severity of harm	Overall risk
	Remote, possible or probable	Minimal, significant or severe	Low, medium or high

Sample data, with personal details, are lost or accessed by unauthorised individuals, leading to unwanted contact from third parties or even disclosure of confidential information	Remote	Significant	High
Survey data, with information about special category data or financial information linked to individuals, are lost or accessed by unauthorised individuals, leading to embarrassment and potential blackmail of data subjects.	Remote	Severe	High
Sample members do not know what the study is about, or why they are being contacted.	Remote	Significant	Medium
Sample individuals do not want to take part and do not like being contacted, and report the invite as an intrusion to the ICO.	Remote	Significant	High
Kantar/NatCen could use the respondents contact details and / or survey responses for other research purposes.	Remote	Significant	High

Step 6: Identify measures to reduce risk

Identify additional measures you could take to reduce or eliminate risks identified as medium or high risk in step 5				
Risk	Options to reduce or eliminate risk	Effect on risk	Residual risk	Measure approved
		Eliminated reduced accepted	Low medium high	Yes/no
Sample data, with personal details, are lost or accessed by unauthorised individuals, leading to unwanted contact from	Sample data can only be transferred by known and authorised individuals and require them to use a known and authorised email. The data are encrypted on transfer.	Reduced	Low	Yes

third parties or even disclosure of confidential information				
Survey data, with information about special category data or financial information linked to individuals, are lost or accessed by unauthorised individuals, leading to embarrassment and potential blackmail of data subjects.	Survey data can only be transferred by known and authorised individuals who access the Portal from known IP addresses. For on-line responses, Individual respondent ID is validated by a requirement to enter date of birth. All data are encrypted on transfer. On receipt at ISER, sample details (name, address, etc) are removed before the survey data are copied to a 'protect' area, and only accessed by named researchers working on the project. The access to the researchers is granted only by the PI.	Reduced	Low	Yes
Sample members do not know what the study is about, or why they are being contacted.	Only sample members that have been part of the study are contacted. Advance letters are sent, giving more information about the study, and providing links to more information. All communication for the study includes contact details (Freepost, Freephone, email, online form on website).	Reduced	Low	Yes
Sample individuals do not want to take part and do not like	Sample members are part of an on-going longitudinal study and have been invited to participate in studies before. All	Reduced	Medium	Yes

being contacted, and report the invite as an intrusion to the ICO.	communications to sample members includes contact details to allow them to contact us to remove themselves from this particular study, or the larger study. Participation is completely voluntary and sample members are not under any coercion to take part.			
Kantar/NatCen could use the respondents contact details and / or survey responses for other research purposes.	Prohibited within the formal contract, which defines their role as Data Processor. Breach of confidence would subject the data processor to financial penalties and potentially catastrophic damage to reputation.	Reduced	Low	Yes

Step 7: Sign off and record outcomes

Item	Name/date	Notes
Measures approved by:	Dr. Jonathan Burton, Associate Director (Surveys)	Integrate actions back into project plan, with date and responsibility for completion
Residual risks approved by:	Ray Ware, IT and Security Manager. 03/09/2020	If accepting any residual high risk, consult the ICO before going ahead
DPO advice provided:	Sara Stock, Information Assurance Manager (DPO), University of Essex 04/09/2020.	DPO should advise on compliance, step 6 measures and whether processing can proceed

<p>Summary of DPO advice:</p> <p>This is really detailed and I'd be inclined to add at the top of section one a very brief reason for the DPIA being carried out – which is related to the volume of data, the number of individuals involved, and the collection including special category data.</p> <p>Other than that I don't have any comments.</p>		
DPO advice accepted or overruled by:	Accepted, by Jonathan Burton	If overruled, you must explain your reasons
<p>Comments:</p> <p>Updated the first paragraph under section one as advised.</p>		
Consultation responses reviewed by:	Ray Ware, IT and Security Manager. 18/09/2020	If your decision departs from individuals' views, you must explain your reasons
<p>Comments: NFA required.</p>		
This DPIA will kept under review by:	<p>Dr. Jonathan Burton, Associate Director (Surveys) in consultation with Ray Ware, IT and Security Manager.</p> <p>To be reviewed by Sara Stock, Information Assurance Manager (DPO), University of Essex on substantive change.</p>	The DPO should also review ongoing compliance with DPIA