Deliverable Number: D5.11
Deliverable Title: Survey Network Meeting report 3
Work Package: 5 ‘Training and dissemination’

Deliverable type: Report
Dissemination status: Public

Submitted by: SHARE ERIC (MEA)
Authors: Elena Sommer (SHARE, MEA), Hervé L'Hours (CESSDA, UKDA), Tom Emery (GGP), Linn-Merethe Rød (CESSDA, NSD), Marianne Høgetveit Myhren (CESSDA, NSD), Straume Øyvind, (CESSDA, NSD)

Date submitted: September 2018

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 654221.
SERISS (Synergies for Europe’s Research Infrastructures in the Social Sciences) aims to exploit synergies, foster collaboration and develop shared standards between Europe’s social science infrastructures in order to better equip these infrastructures to play a major role in addressing Europe’s grand societal challenges and ensure that European policymaking is built on a solid base of the highest-quality socio-economic evidence.

The four year project (2015-19) is a collaboration between the three leading European Research Infrastructures in the social sciences – the European Social Survey (ESS ERIC), the Survey for Health Aging and Retirement in Europe (SHARE ERIC) and the Consortium of European Social Science Data Archives (CESSDA AS) – and organisations representing the Generations and Gender Programme (GGP), European Values Study (EVS) and the WageIndicator Survey.

Work focuses on three key areas: Addressing key challenges for cross-national data collection, breaking down barriers between social science infrastructures and embracing the future of the social sciences.

Please cite this deliverable as: Sommer, E., L'Hours, H., Emery, T., Rød, L.M., Høgetveit Myhren, M., Straume, Ø., (2018) Survey Network Meeting report 3. Deliverable 5.11 of the SERISS project funded under the European Union’s Horizon 2020 research and innovation programme GA No: 654221. Available at: www.seriss.eu/resources/deliverables
## Content

1. Introduction

2. SERISS work on ‘New forms of data – legal, ethical and quality issues’ & Perspective on the GDPR implementation

3. Legal and ethical issues of combining survey data with new forms of data

4. Linking survey data with social media data in social research – Twitter example

5. Legally and ethically using, re-using, and preserving social media research data harvested from an API

6. Linking survey data with administrative data in practice

7. Consent-related challenges and strategies

8. Implementing consent for record linkage in a cross-national survey: a practical example from SHARE

9. Discussion session and next steps

Appendix 1: Workshop agenda

Appendix 2: List of participants

Appendix 3: European Social Survey (ESS) Round 9 Respondent Information Sheet
1 Introduction

The ‘SERISS Survey Experts Network’ is a series of workshops thematically based around SERISS work packages. The aim of the workshops is to bring together survey practitioners and researchers (e.g. representatives from national statistics institutes, cross-national European surveys, survey agencies and survey methodologists) in order to facilitate a productive exchange of knowledge and practices in state-of-art survey research, to initiate a discussion on how to tackle specific challenges in survey methodology and data management, and to encourage future cooperation between different organisations.

A number of large organisations involved in cross-national survey research agreed to join the forum including: Eurostat, Eurofound, PIAAC (OECD), ISSP, European Union Agency for Fundamental Rights (FRA), Gallup World Poll, RTI, Kantar Public and Ipsos MORI. These organisations are invited to send representatives to all workshops. Further, external participants with expertise in particular topics discussed during the workshops (e.g. representatives from national statistics institutes, academics) are also invited to each workshop.

There will be in total four annual workshops over the lifetime of the SERISS project. The first workshop ‘Representing the population in surveys’ dealing with sampling challenges took place on 8th December 2016 and was hosted by Munich Center for the Economics of Aging (MEA). The second workshop ‘Harmonised coding for socio-economic survey questions’ was organised around the outputs of the SERISS Work Package 8 ‘A coding module for socio-economic survey questions’ and took place on 4th-5th September 2017 in Amsterdam. The third workshop ‘Legal and ethical issues of combining survey data with new forms of data’ took place on 19th June 2018 at City, University of London. The fourth and the last workshop on survey fieldwork management and costs will take place in spring 2019. The workshop reports are available on the SERISS website.

This report is a summary of the third SERISS Survey Experts Network workshop ‘Legal and ethical issues of combining survey data with new forms of data’. The main purpose of the workshop was to bring together researchers, survey practitioners and other stakeholders (e.g. representatives from data archives, national statistics institutes, social media platforms) involved in data collection, data preservation and data linkage to discuss challenges and solutions for linking survey data with new forms of data in light of the new General Data Protection Regulation (GDPR).

In total 32 participants attended the third workshop (see Appendix 1 for the workshop agenda and Appendix 2 for the participants list). To enable exchange between participants, the workshop had an interactive format with presentations followed by discussion slots. The report summarises the presentations given during the workshop, the discussion outcomes, and planned future steps.
2 SERISS work on ‘New forms of data – legal, ethical and quality issues’ & Perspective on the GDPR implementation

Presentation by Marianne Høgetveit Myhren, Norwegian Centre for Research Data (NSD)

The goal of the SERISS Work Package 6 “New forms of data: legal, ethical and quality issues” is to address the major legal and ethical challenges of accessing large-scale data on an individual level with a particular focus on social surveys and linking social survey data with new types of data such as biomarker, social media data and administrative data. Further, this work package aims at monitoring the implementation of the new European General Data Protection Regulation (GDPR) in the social surveys context and producing guidelines for researchers and survey practitioners reflecting the new GDPR requirements.

The specific objectives of the work package are to:

• Determine and provide guidance related to the legal and ethical issues when using traditional and new forms of data in the social sciences

• Ensure social scientists can work effectively when handling personal data within an evolving legal and ethical framework whilst retaining the confidence of the public

• Increase the range and depth of the CESSDA infrastructure to ensure that new forms of data can be shared and reused

The overview of tasks and submitted deliverables is available on https://seriss.eu/resources/deliverables/

Legal and ethical aspects affect all stages of data-based research including data collection, data processing, data analyses, data preservation, data access, and data re-use. A good balance between research and data privacy interests is crucial to secure personal data protection in research. A legal framework that explicitly allows data producers and data controllers to disseminate data within and across national borders is essential to ensure sustainable and harmonised access conditions in Europe. The GDPR released on 25. May 2018 to protect personal data applies to any controller or processor in the EU who processes personal data regardless of whether the processing takes place in the EU or not as well as to any controller or processor outside the EU if they offer goods/services or monitor behaviour of EU citizens. The key goal of the GDPR is to harmonise rules across Europe making Europe more fit for the digital age and facilitating cross border data flow. Its further purpose is to ensure a high level of data protection in order to provide legal certainty and trust as well as to provide more control of their data to citizens. The GDPR will be supplemented by national laws.

Although the GDPR is associated with more continuity than change for the research sector/community, there are some implications for the research community. Institutions will be assigned increased responsibility for the data they hold and process (“accountability”) and data protection officers will be mandatory for many institutions. Individuals, on the other hand, will get more rights to control their data, e.g. right to data portability. There will be new requirements for consent and information to be provided to data subjects. However, broad
consent to certain areas of scientific research is possible when recognised ethical standards for scientific research are kept and there is a broad definition of scientific research. In general, the GDPR is research-friendly and safeguards the interests and the needs of scientific research institutions. Despite limited flexibility there is room for national supplementary provisions and derogations, especially in the field of research thereby providing room for distinct national adaptions.

The GDPR applies to personal data which is defined as any information that can be used to identify a person ('data subject'), directly or indirectly (see Figure 1 for examples of personal data). The current definition of personal data is implemented across Europe with various degrees of strictness. A standardised practice and implementation of the definition can therefore improve the possibilities for data access, data sharing and in turn improve conditions for cross-national research. It also implies that more research projects will fall within the scope of the GDPR than before in some European countries. Sensitive personal data are special categories of personal data that are subject to additional protection. In general, organisations require stronger grounds to process sensitive personal data than they require to process "regular" personal data.

**Figure 1: Types of personal data**


Anonymised data implies a process undertaken on personal data and refers to information that does not relate to an identified or identifiable natural person or to personal data when it is presented in such a way that the data subject is not or no longer identifiable. Anonymisation of data should be irrevocable but should still be checked at regular intervals in light of new technologies. The GDPR does not apply to anonymised data (EU GDPR Recital 26).

The GDPR lists six principles of how personal data should be treated (Article 5). Personal data must be:

- Processed lawfully, fairly and in a transparent manner *(lawfulness, fairness and transparency)*
- Collected for specific purposes and not processed further for incompatible purposes *(purpose limitation)* – exemption for research/archiving purposes in accordance with EU GDPR Article 89(1) – further processing not incompatible with original purpose
- Adequate, relevant and limited to what is necessary *(data minimisation)*
- Accurate and where necessary up-to-date *(accuracy)*
- Kept in identifiable form no longer than necessary *(storage limitation)* - exemption for research/archiving purposes in line with EU GDPR Article 89(1)
- Processed with appropriate security *(integrity and confidentiality)*

Processing of personal data must be met with one of the legal bases in Article 6 (1) in order for it to be lawful. The relevant legal bases for the processing of personal data for research purposes are:

- Consent from data subject for one or more specific purposes
- Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller

Processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party

Data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data, data concerning health or data concerning a natural person’s sex life or sexual orientation are considered as special categories of data that are subject to additional protection (Article 9)

Processing of special categories of data is in principle prohibited, but can be processed if based on:

- Explicit consent for one or more purposes (Article 9 (a))
- If it relates to personal data which are manifestly made public by the data subject (i.e. social media data) (e)
- Processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject (j).
The duty to inform data subjects is emphasized in the new legislation, but GDPR does give options for research on a legal basis without informing the data subjects, article 14, 5b states that the duty to inform does not apply when:

the provision of such information proves impossible or would involve a disproportionate effort, in particular for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, subject to the conditions and safeguards referred to in Article 69(1) or in so far as the obligation referred to in paragraph 1 of this Article is likely to render impossible or seriously impair the achievement of the objectives of that processing.

In survey research, informed consent is a typical legal basis for processing of personal data. Consent is defined as **freely given** (must be a genuine choice with an option to refuse/withdraw without consequences and not be a dependent relationship), **specific** (clear information on extent and consequences), **informed** (content and form requirements, should be easily understood and easily accessible using clear and simple language) and **unambiguous** ("opt in" action is required; silence, pre-ticked boxes and inactivity are not valid (Recital 32)) indication from a person that affirms that his/her personal data may be processed (Article 4).

The controller must be able to demonstrate that consent has been given and it should be as easy to withdraw consent as to give it. There should be explicit consent when processing special categories of data for one or more specified purposes. Consent can be used as the legal basis to transfer data outside of the EU (Article 49). National legislation may impose more requirements for consent than required by the GDPR. Broad consent for certain areas of scientific research is applicable when recognised ethical standards for scientific research are kept (Recital 33). Figure 2 presents the content to be covered in consent according to the GDPR.

Four essential principles to retain trust are covered by the GDPR: transparency; user control; privacy by design and default; and accountability. Further, to ensure the rights and freedom of the data subject, appropriate technical and organisational measures should be in place. Furthermore, the GDPR strengthens the data protection officer arrangement and in practice makes it mandatory for most of the research sector (Article 37). The data protection officer will be a main element in the system for regulating, controlling and documenting the processing of personal data for different purposes.

When producing informed consent for survey participants, the following considerations should be taken account of:

- Will personal data be processed?
- Will special categories of data be processed?
- What is the legal basis for processing personal data (consent, public interest or legitimate interest)?
- Has necessary information for participants to exercise their rights been provided?
- Are appropriate safeguards in place (i.e. appropriate technical and organisational measures to ensure principle of data minimisation and pseudonymisation/anonymization/encryption)?
- Is it necessary to derogate from any of the data subject’s rights?
- Are there issues related to ownership/terms of use/ other legal issues?
- Are there plans for archiving/re-use of data?
Within the framework of SERRIS Work Package 6 ‘New forms of data – legal, ethical, and quality issues’, a GDPR compliant participant information sheet was designed for the European Social Survey (ESS) Round 9 to serve as the basis for informed consent for ESS respondents. ESS is a cross-national survey that processes both personal data and special categories of personal data and collects data from both respondents and non-respondents. Documentation of given consent is a challenging task, especially in cross-national context. Although the proposed information sheet has been specifically designed for the ESS, it can be used as a template for other survey infrastructures in Europe. The full draft of the information sheet is available in the Appendix 3.
3 Legal and ethical issues of combining survey data with new forms of data

Presentation by Andrew Charlesworth, Bristol University

Recent technological development has contributed to the emergence of new forms of data such as data ‘born digital’ and old forms of data that have been digitalized (e.g. government and commercial transactions, or government and other registration records), internet usage, tracking data, as well as satellite and aerial imagery. The legal framework regulating the use of new forms of data for research purposes includes the following aspects:

- Data protection, privacy, confidentiality
  - E.g. GDPR/DPA 2018, protection of private information
- Records-specific rules
  - E.g. re-use of Public Sector Information Regulations 2015, Access to Health Records Act 1990
- Intellectual property
  - E.g. copyright
- Other ownership/contractual claims
  - E.g. Twitter Terms of Use

The use of new forms of data for research is associated with certain limitations and potential data privacy issues. Figure 3 presents some of these limitations.

**Figure 3: New forms of data vs. data privacy**

<table>
<thead>
<tr>
<th>Principle</th>
<th>Automated analysis of datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection Limitation</td>
<td>The larger the data collection, the better the potential for identifying interesting correlations.</td>
</tr>
<tr>
<td>Data Quality</td>
<td>‘Messy data’ is fine, it’s not clear what is relevant until its analysed, and even inaccurate or incomplete data can be useful.</td>
</tr>
<tr>
<td>Purpose Specification</td>
<td>Data may have been collected for a particular purpose, but analysis may indicate further unrelated and previously unknown, but valuable, purposes. Data as collected may not be obviously PII, but analysis of it may identify individuals.</td>
</tr>
<tr>
<td>Use Limitation</td>
<td>There may be value in sharing and aggregating data that may not be apparent at the time of collection</td>
</tr>
<tr>
<td>Security Safeguards</td>
<td>It may be unclear what security issues if any arise from a particular collection of data, or its analysis.</td>
</tr>
<tr>
<td>Openness</td>
<td>Where data is collected and analysed, it may not be obvious that it is PII, and even in circumstances where it is, the researcher may have no way of informing the data subject of its use.</td>
</tr>
<tr>
<td>Individual Participation</td>
<td>Data which is anonymous may still be utilised in ways that can cause risk/harm to an individual.</td>
</tr>
<tr>
<td>Accountability</td>
<td>How and when might a researcher to be held accountable and for what?</td>
</tr>
</tbody>
</table>

Source: Presentation by Andrew Charlesworth, 19.06.2018, London.

[www.seriss.eu](http://www.seriss.eu) GA No 654221
Ethical standards in research have the three following main aims:

- to protect the rights and legitimate interests of research participants (e.g. by normally requiring informed consent)
- to create trust in both researchers and research outcomes (e.g. by endorsing or rejecting practices affecting the actual or perceived reliability or trustworthiness of research data)
- to preserve the research environment (e.g. by seeking to prevent practices which might restrict future access to research participants or reduce the effectiveness of future research)

What is possible by law in terms of data collection and data analysis is sometimes questionable from the ethics point of view and there are some grey zones, e.g. data collection in the public domain without consent. Also asking for consent for GDPR reasons might conflict with a consent for ethical reasons and consent can mean different things in different contexts. Providing granular assurances to data subjects that data processing would be limited, and then calling on public task as a blanket derogation to justify further or broader use could be seen as actively misleading resulting in reducing the confidence of the individual data subject, of potential future data subjects and of the general public. Therefore, it is important to differentiate and communicate the difference between ethical consent and consent for legal GDPR purposes.

In terms of the impact of the GDPR on research, the existing framework for UK research compliance remains broadly the same but there is scope for divergence in exemptions and derogations between EU Member States, with implications for cross-border research collaborations and certain areas of research are likely to see greater scrutiny than others (e.g. Facebook/Cambridge Analytics case).

The key issues to ensure ethical use of personal data for research are:

- Ensuring training and education of researchers
- Building data protection into data management plans
- Awareness of research involving personal data
- Input into, and awareness of, institutional policy
- Accountability and research governance (documentation, audit trail, responsibilities, outcomes)
- Understanding the sometimes fuzzy line between ‘ethical research’ and ‘lawful research’ - consent as both a legal and ethical requirement, and the alternatives to consent

Research ethics committees (REC) are faced with the following challenges:

- Administration vs Academics – two-way communication/co-operation
- Management perceptions of risk and value - REC seniority/workload
- ‘Cut and paste’ ethics applications
- Differences in discipline cultures and in scale
- External projects
- Expertise gap in RECs
Different academic disciplines face varying challenges to existing practices making ‘one-size-fits-all’ solutions and ‘quick fixes’ inappropriate. RECs already address data protection to varying degrees and can identify discipline-specific risks, good practice and problems with suggested ‘solutions’. As RECs have varying expertise in data protection, it is important that they are not simply used as ‘gatekeepers’, but rather trained and developed as ‘facilitators’ consulting researchers on ethic aspects.

4 Linking survey data with social media data in social research – Twitter example

Presentation by Curtis Jessop, NatCen Social Research

Linking survey participants’ answers to publicly available information from their Twitter accounts allows survey data to benefit from real-time, ‘natural’ behavioural and attitudinal data. A further benefit of such a linkage it that it adds the ‘who’ to Twitter data by providing information from the sample frame and allowing for the analysis of different groups. In terms of ethics, there are three particular areas of consideration: consent to link the data; processes for linking the data; as well as access to, archiving, and sharing the data.

In the case of consent, survey participants can be asked for consent to access their Twitter data and to link the Twitter data to their survey answers directly during the survey interview or by contacting them at a different time point (if contact details are provided as, for example, in panel surveys). Further consent is needed for archiving these data. The major challenge related to consent is balancing the amount of information that is provided to the survey participants. On the one hand, all the necessary information with enough detail needs to be provided so that consent is informed. On the other hand, the amount of information and the form in which it is presented should not overwhelm the respondents as this could potentially result in increased occasions when respondents do not read the consent and/or have a negative impact on respondents’ willingness to provide the consent.

An example of consent to access Twitter data that was used in a study by Al Baghal et al. (forthcoming) is presented in Figure 4. In addition, respondents received further information via a Q&A list focusing on the following aspects:

- What information will you collect from my Twitter account?
- What will the information be used for?
- Who will be able to access the information?
- What will you do to keep my information safe?
- What if I change my mind?
In terms of data processing, a traditional ethical requirement for survey data is that data are anonymised. In the case of social media data, the data are inherently identifiable in their raw format, but this is also when they are most useful. If social media data (e.g. Twitter) are linked to the survey data, the survey data are also no longer anonymous and the linked Twitter data can no longer be ‘public’.

Based on existing processes for access to identifiable survey data, quasi-anonymisation (e.g. by removing the Twitter handle, although the removal of handles does not ensure full anonymity as a google search for text strings will provide information about the account owner) and the use of cut-down of survey data and variables derived from the raw Twitter data as suggested in Figure 5 could be implemented to reduce risks of data being identifiable.
In order to maintain the usefulness of the data, the focus should, however, move away from anonymisation towards secure access methods while still guaranteeing ethical use of the data and providing thorough justification for the research. Access to raw data in a secure environment would involve, among other measures, training for researchers, documentation of access, offline access and restrictions of data transfer outside this environment.

Further challenges of linking survey data and social media data are related to archiving and sharing of data which is important to enable replication of results and to maximise value of data. In particular, there are still uncertainties in terms of responsibilities for maintaining the data and legal issues of sharing Twitter datasets, as well as dealing with deleted Tweets and/or withdrawn consent after the dataset has already been archived.

There was discussion about the fact that, for information which includes personal data, a secure access location is a useful but expensive approach to re-use when researcher travel is factored in but that the technical elements are making secure remote access increasingly viable. Further, it’s important to differentiate between what happens to the data within a project and what data are going to be provided to the community as there is greater value in open dissemination where possible over more closed data. Open datasets with fewer identification risks and derived variables but potentially lower research potential could exist in archives alongside more complete versions accessible securely on a specific project request and with additional conditions (permissions, training, statistical disclosure mitigation measures).

The UK Data Service at the UK Data Archive provides the secure remote access system Secure Lab\(^1\) whereby the deployment of the access route depends on a mixture of the data owner preference and the recommendation of the archive, both based on data sensitivity. Although the hardware and software problem providing secure access has been largely

---

\(^1\) [https://www.ukdataservice.ac.uk/get-data/how-to-access/accesssecurelab](https://www.ukdataservice.ac.uk/get-data/how-to-access/accesssecurelab)
solved, the rapid development of big data requires clear legal and governance measures around secure remote access. Actors including UKDA, IAB and GESIS are working on a bottom-up approach to widening these services as part of a submitted European Open Science Cloud (EOSC) proposal.

5 Legally and ethically using, re-using, and preserving social media research data harvested from an API

Presentation by Sara Day Thomson, Digital Preservation Coalition (DPC)

The Digital Preservation Coalition (DPC) is a not-for-profit membership organisation aiming at raising awareness of the importance of the preservation of digital material. It serves as a knowledge and experience exchange network for its members and as an advocate and catalyst for digital preservation ensuring long-term access to digital content and services. As demonstrated in the DPC Technology Watch Report Preserving Social Media (Thomson 2016), there is a spectrum of ethical issues that interact with public or individual opinion on how social media data can be used, re-used and preserved for research. The choice of an appropriate approach depends on the following aspects:

- **What data is used and how?**

This includes the form of data - platform that collects the data (e.g. Twitter, Facebook, etc.); content and metadata (e.g. text, embedded URLs, geolocation, number of followers, etc.) - and type of data analysis (e.g. qualitative, quantitative, mixed-methods) that is applied to it. Currently a case by case evaluation is required for legal compliance and ethical practices.

- **What are the legal & regulatory restrictions?**

This includes privacy and data protection; intellectual property, and open data policies (see Figure 6).

**Figure 6: Legal and regulatory restrictions of social media use, re-use, & preservation**

• **What are the platform policy restrictions?**

Different platforms have different types of policies and different restrictions in terms of how their data can be used for research. It is also important to keep in mind that different platforms have different functions and purposes, sharing different forms of content that come with different types and levels of risk for users. Facebook, for example, contains a lot of private data and personal interactions, generating more issues with Data Protection and privacy. Twitter, by comparison, offers a more public environment, but still creates potential issues with violation of privacy as well as issues of user awareness. YouTube, compared with both Facebook and Twitter, is filled with original audio-visual material, which carries additional copyright restrictions.

Common platform policies include: Terms of Service; User Agreement; Privacy Policy; Terms of Use; and Developer Agreement or Developer Policy (regulates use of API). Twitter, for instance, does not allow sharing or storing deleted tweets or content from a deleted account. However, the User Agreement for Twitter states that data in the public domain may be re-used by Twitter and shared by Twitter through a license agreement. Thus, a licence to use data from one of Twitter’s APIs does not imply a licence to deposit that data with a repository. Therefore, even the storage of tweets harvested via a Twitter API by a third party is considered a violation of the developer agreement. While repositories can comply with this restriction by sharing the tweet IDs of a dataset, these IDs are unlikely to rehydrate back to the original dataset, because they will not include tweets that have been deleted by users.

There are also many conflicts between Twitter’s other Terms of Service (ToS) and research methods for using social media data. For example, Twitter has a requirement, likely created for use in journalism, that tweets be displayed with the user handle and all original text and metadata present on the original tweet. However, researchers often prefer to (or are required to) protect the identity of individual users but may still need to quote text or metadata.

• **What are the ethical implications?**

Apart from institutional ethical requirements (Figure 7), user perspectives should be taken into consideration when making ethics-related decisions. A research data management plan and tools should be developed with legal issues, platform restrictions, and ethics (user perspectives) in mind. To quote Kelsey Beninger (2017: 592): “Research needs to be supported by an inductive and flexible approach to ethical thinking; what principles need to be considered in the context of your study and how can you think about these to ensure the actions you decide to take support an ethical study”.

Based on the finding from the study *Research Using Social Media: Users’ Views* conducted by NatCen Social Research (Beninger et al. 2014³), ethical practices in research can be improved at different research stages:

- Sampling & recruitment stage (e.g. by stating credentials; forming a diverse and inclusive pool of users; clearly articulating purpose and aims of the study)

---


Collecting and generating data (e.g. protecting participants by taking into consideration the privacy risks created by the platform used; improving representativeness of findings)

Reporting results (e.g. protecting identities and maintaining trust;setting an example of transparency & honesty in reporting)

**Figure 7: Ethical requirements for social media data use for research**

![Ethical Requirements Diagram](Image)

- **How to preserve social media?**

Social media is vulnerable to loss. There is no legal or regulatory requirement for platforms to preserve data, or even to store it for the long term. To quote LinkedIn’s User Agreement: „LinkedIn is not a storage service. You agree that we have no obligation to store, maintain or provide you a copy of any content or information that you or others provide”, or Twitter Terms of Service: “The Twitter Entities shall not be liable for ... any data loss”. Preservation of social media is also dependent on changes in platform policy and ownership. Because historical data is less commercially valuable that the up-to-date data for platforms, these data will only become more difficult and expensive to capture and preserve in future. Although immediate sharing of preserved social media is currently not possible, preserving the bits, metadata, and documentation is still critical to making social media content accessible in future. Clear guidance provided by research support and librarians on how to preserve social media data would promote its use in research.

One of the main benefits of social media research is that it enables understanding of how social media can facilitate and mirror real-world events and can, for instance, help researchers make predictions about future events. Without preserving historical social media data, researchers might lose the data they need to trace patterns. Projects like Documenting the Now (DocNow) have made great progress in aligning social media ethical preservation...
with social movements played out on social media. The preservation of social media should be also aligned with initiatives for transparent government and for open data in public institutions. In order to have a complete picture of a government’s interactions with its citizens, all modes of communication used should be preserved.

During the discussion it was mentioned that while ongoing EC copyright legislation would probably provide exemptions for research involving data mining of text, the e-privacy regulation in contrast may require additional and/or more granular consent for the use of social media metadata. The Social Data Science Lab at Cardiff was referenced as an actor with strong relationships with the social media platforms and it was suggested that large collections, especially those with a national remit for social media data collection, might have one-to-one understandings with the platforms, but that there is a need to move beyond such ad hoc arrangements and build a cooperative consensus. The main issues with social media data lie less in using the data for initial analyses but in preserving and preparing such data for re-use and the impact regulations have on reproducibility of research when a study can only be published with no links to the underlying (meta)data.

6 Linking survey data with administrative data in practice

Presentation by Peter Jones, UK Office for National Statistics (ONS)

Following the aspiration of UK government to deliver future Censuses for England and Wales using alternative sources of data rather than collecting the data solely via questionnaires, there has been an ongoing work at the Office for National Statistics (ONS) on the development of a census combining administrative data and surveys. The unprecedented linkage potentials in terms of scale enable development of a so-called “person spine” from administrative records using different sources of administrative data whereby a statistical ID could be assigned across the different datasets and maintained longitudinally. In addition, data collection via supporting surveys alongside administrative data would facilitate data collection on topics unavailable in administrative data sources as well as allow to measure and adjust for the quality and reliability of administrative data. To assess the feasibility of such a combined approach, the ONS developed a plan to test and implement a Population Coverage Survey (PCS) with approximately 350,000 households sampled annually that could be linked to the person spine data via statistical IDs or personal identifiers.

ONS is permitted to link survey records to admin data under the provisions of the Statistics and Registration Services Act 2007 and will continue to do so under the Digital Economy Act 2017. While ONS are in the process of re-wording some of the information provided in survey documentation to reflect on GDPR requirements, explicit consent for data linkage is not required on the basis that data is being processed in the “performance of a task carried out in the exercise of official authority”.

ONS has developed privacy preserving approach to record linkage which is based on pseudonymisation and separated functions model. The main challenge of linking admin records and survey data is that person spine data contains information that is often outdated (e.g. for cases of emigration or deceased persons) or not accurate (e.g. duplicate
registrations or registrations at wrong address) as the quality of data partly depends on individuals informing the service providers about changes. Therefore, a coverage survey for admin data is needed to assess and adjust for the quality of the admin data sources.

As part of the Coverage Survey Questionnaire Development, a series of focus groups and cognitive testing around the proposed Population Coverage Survey (2013) were undertaken across England and Wales focusing on the following aspects: address histories and concurrent addresses; administrative source-specific information; quality of data on date of birth and full names; mode of data collection; proxy data collection; survey purpose, data confidentiality and security; whether the survey should be compulsory. In addition, some test data were collected to run simulation studies (test data, 2011 Census and admin data).

Coverage adjustment simulations based on an independent Population Coverage Survey showed up to 25% difference from Census estimates in particular age groups. Whereas overcoverage on admin record is relatively common, the Coverage Survey targets at measuring non-response in the Census as there are various reasons for why households and individuals are missed from the Census enumeration (e.g. missing addresses in the address frame, completing a single form for multiple families, non-response of a whole household or of some household members).

To capture over-coverage, implementation of so-called dependent sampling/interviewing procedure has been discussed as a potential approach. In contrast to independent sampling procedures such as simple random sampling or area-based stratified sampling, dependent sampling involves address sampling based on household characteristics and sampling of multiple addresses for individuals registered at different locations. Dependent interviewing (see Figure 8) involves sampling of individuals in admin records and checking if they are resident at address.

ONS are in the early stages of discussion around what a dependent sampling/interviewing approach might look like. This includes public acceptability research, more detailed understanding of the legal basis for using data in this way, and ethical considerations.

While dependent sampling / interviewing is an established method for measuring over-coverage, the ONS is also researching the potential for model-based adjustments that would not rely on a separate survey for over-coverage being collected. We continue to work closely with colleagues in other National Statistics Institutes (NSIs) that have similar programmes that aim towards a future administrative data census, in particular New Zealand and CSO Ireland.
7 Consent-related challenges and strategies

Presentation by Joseph Sakshaug, Institute for Employment Research (IAB)

Linking of data from different sources of administrative data (Figure 9) is a common practice at the Institute for Employment Research (IAB) in Nuremberg and IAB has conducted several studies to explore strategies to increase survey participants’ willingness to provide consent for linking their survey data with administrative data.
According to the Federal Data Protection Act (2013, Part I, Section 4; Code of Social Law X, 2013, Section 75), informed consent by the respondent is usually required prior to linkage. The purpose of the consent process is to ensure respondents are informed about:

- Which data sources will be linked
- Intended uses of the linked data
- Possible benefits (and risks, if any)
- Responsibility of ensuring data confidentiality
- Voluntary nature of request

The consent-related research conducted at IAB has demonstrated that:

1. Consent rates vary, but not in ways intended
2. Differential consent can lead to bias in estimates
3. Consent rates can be easily manipulated
4. Respondents don’t (always) understand request (see example in Figure 10)
Figure 10: Example of consent understanding among survey participants

Consent Understanding: IAB Study

<table>
<thead>
<tr>
<th>Percent correct by linkage consent</th>
<th>Consenters % correct</th>
<th>N</th>
<th>Non-consenters % correct</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answers will be passed along to IAB</td>
<td>88.3</td>
<td>977</td>
<td>57.8</td>
<td>142</td>
</tr>
<tr>
<td>Survey data merged with IAB</td>
<td>93.3</td>
<td>982</td>
<td>36.7</td>
<td>147</td>
</tr>
<tr>
<td>Name/Adress saved with linked data</td>
<td>68.3</td>
<td>981</td>
<td>38.8</td>
<td>147</td>
</tr>
</tbody>
</table>

Kreuter et al., 2015


Linkage consent rates can strongly vary across studies but studies from the US show that they are generally decreasing (e.g. from 85% to 50% for the National Health Interview Survey between 1993 and 2005; from 88% to 65% for the Survey of Income and Program Participation between 1996 and 2004; from 90% to 76% for the Current Population Survey between 1994 and 2003).

There are several different reasons why survey respondents agree or disagree to linkage with administrative data. According to Sala, Knies, and Burton (2014; IJSRM), common reasons for agreement are:

- “Being helpful with the research”
- “Trust in survey/fieldwork agency”
- “Having a clear understanding of why”
- “Other (positive) considerations”

Among the common reasons for linkage refusal are (Sala, Knies & Burton, 2014):

- “Concerns about sharing of confidential data”
- “Other (negative) considerations”

Non-consent with linkage reduces analytic sample size and can result in increased variance of linked-data estimates. It is also associated with certain bias as people who consent may be systematically different from those who do not consent based on key survey and administrative variables (see Figure 11) which may adversely impact linked-data estimates.
Figure 11: Non-consent bias

Bias in Administrative Estimates

IAB PASS Study (welfare recipient sample)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Nonresponse Bias</th>
<th>Measurement Bias</th>
<th>Linkage Consent bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.1</td>
<td>0.03</td>
<td>-0.3*</td>
</tr>
<tr>
<td>Foreign citizen (%)</td>
<td>-5.6*</td>
<td>-2.5*</td>
<td>-0.9*</td>
</tr>
<tr>
<td>Welfare receipt (%)</td>
<td>3.2*</td>
<td>-7.1*</td>
<td>-0.3</td>
</tr>
<tr>
<td>Disability (%)</td>
<td>0.4</td>
<td>6.0*</td>
<td>0.01</td>
</tr>
<tr>
<td>Employed (%)</td>
<td>1.0</td>
<td>-0.6</td>
<td>0.3</td>
</tr>
<tr>
<td>Income (30 days)</td>
<td>-71.4*</td>
<td>394.5*</td>
<td>1.7</td>
</tr>
</tbody>
</table>

Non-consent bias is present, but relatively small compared to other error sources
Sakshaug and Kreuter, 2012


Several empirical studies have demonstrated the effect of consent wording and consent placement in the survey questionnaire on willingness of participants to agree to a linkage request. Evidence from behavioral economics and psychology, for example, suggests a strong effect of a default option in active vs. passive opt-in and opt-out wording consent (Thaler and Sunstein 2008). Although there is a strong effect also in linkage requests (Bates 2005; Pascale 2011), it is not always clear if behaviour matches intentions. Ellikson and Hawes (1989), for example, found some evidence in the late 1980s that most respondents who signified their consent to have their children participate in school research by failing to mail back a card indicating their refusal (opt-out consent) did, indeed, intend to allow their children to participate in the research, whereas most of those who failed to mail back a card indicating their consent (opt-in consent) did not intend to refuse (Ellikson and Hawes 1989; cited in Singer 1993).

In addition, there are concerns about possible bias. Das and Couper (2013) tested variations of an opt-out linkage consent request in a probability-based online panel in the Netherlands and found out that those who objected to linkage (opted out) achieved lower scores in the follow-up knowledge questions.

Further, according to Kahneman and Tversky (1979), faced with risky choices, decision-making is influenced by framing in terms of gains or losses. Therefore, the wording of the consent can be designed to emphasise benefits of linkage and/or losses in case of non-consent (see Figure 9 for an example of wording). “Loss framing” usually results in higher consent rates.
Figure 12: Placement and framing of a linkage consent

Placement and Framing

Gain frame

Front: The information that you will give us in the course of the interview will be more useful if you agree to link with the data of the Federal Agency. Are you consenting to the transmission of the information?

Back: The information that you have already given us in the course of the interview are more useful if you agree to link with the data of the Federal Agency. Are you consenting to the transmission of the information?

Loss frame

Front: Unfortunately, the information you will give us in the course of the interview will be less useful if you disagree to link with the data of the Federal Agency. Are you consenting to the transmission of the information?

Back: Unfortunately, the information that you have already given us in the course of the interview are less useful if you disagree to link with the data of the Federal Agency. Are you consenting to the transmission of the information?


In addition, the placement of the linkage consent in the survey can have an effect on consent rates. Back placement is most common in surveys (Sakshaug, Tutz, Kreuter, 2013) as it has been assumed that rapport established between the researchers and the survey participants during the survey completion would have a positive effect on willingness of participants to provide a linkage consent. However, experiments suggest higher rates when asked upfront or in the context of related survey items (Sala, Knies, and Burton, 2014; Sakshaug, Tutz, and Kreuter, 2013).

For example, in a study conducted at IAB (see Figure 12 for framing of the consent), the back placement resulted in higher non-consent rates, especially in the web mode (see Figure 13). While there was almost no difference on consent provision with regard to the framing (gain vs. loss) if the consent request was placed at the front, the “loss framing” worked better if the consent request was placed at the back of the survey as compared to the “gain framing” (Figure 12). However, the manipulation gain-loss is not as strong as front-back, and therefore the findings suggest that it is recommendable to place the consent request up front. A possible explanation is that respondents get exhausted at the end of the survey and are less willing to commit to further requests.

Figure 13: Consent to linkage by framing and mode

<table>
<thead>
<tr>
<th></th>
<th>Phone</th>
<th>Front</th>
<th>Back</th>
<th>Total n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain</td>
<td>90.8</td>
<td>78.7</td>
<td>538</td>
<td></td>
</tr>
<tr>
<td>Loss</td>
<td>90.5</td>
<td>81.2</td>
<td>610</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>613</td>
<td>595</td>
<td>1,228</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Web</th>
<th>Front</th>
<th>Back</th>
<th>Total n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain</td>
<td>82.6</td>
<td>62.4</td>
<td>520</td>
<td></td>
</tr>
<tr>
<td>Loss</td>
<td>86.3</td>
<td>75.4</td>
<td>439</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>511</td>
<td>498</td>
<td>1,029</td>
<td></td>
</tr>
</tbody>
</table>

References


Implementing consent for record linkage in a cross-national survey: a practical example from SHARE

Presentation by Daniel Schmidutz, SHARE-ERIC

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a multidisciplinary and cross-national panel study conducted every two years in 27 European countries and Israel. The first wave of data collection took place in 2004. In 2018/2019 the eighth wave of data collection is carried out. SHARE collects micro data on health, socio-economic status, and social and family networks of individuals aged 50 or older via face-to-face computer assisted personal interviews (CAPI). SHARE data are made available free of charge to registered users for scientific research purposes.

Figure 14 shows the (simplified) basic organisational structure for the data collection with regard to general questionnaire modules that are fielded in all participating countries in a particular SHARE wave. On an international level, the study is centrally coordinated by SHARE Central in Munich, Germany. The implementation of SHARE on the national level is coordinated by Country Teams located at SHARE’s national scientific partner institutions. In each country the interviews are conducted by a survey agency commissioned by SHARE.

Figure 14: Organisational structure with regard to the general data collection


In case of record linkage (the combination of SHARE survey data with administrative record data from official sources), additionally providers of administrative data are involved on a national level as presented in Figure 15. In contrast to the general SHARE questionnaire modules, the implementation of the record linkage module is optional and country specific in SHARE. In SHARE Wave 8 record linkage is implemented in 8 SHARE countries.
Figure 15: Organisational structure if a record linkage module is implemented

![Organisational Structure Diagram]


The entire SHARE data collection is subject to the GDPR (European General Data Protection Regulation) as personal data is processed. The legal basis for data processing in SHARE is consent. In order to ensure compliance with the GDPR in SHARE after May 2018 the consent procedure for wave 8 was developed together with the SHARE-ERIC Data Protection Officer (DPO). Consent for participation in SHARE is provided orally by respondents and documented by the interviewers in the CAPI software. Prior to the interview, respondents are provided with a standardised data protection statement. This data protection statement has been produced centrally by SHARE-ERIC in cooperation with its DPO in the English language and translated into national languages by the Country Teams. It includes required information for data subjects in accordance with the requirements for consent of the GDPR and fulfills the information duties for the collection of personal data from data subjects. The entire procedure as well as the data protection statement have been reviewed and approved by an ethics committee.

For the purpose of record linkage additional explicit consent is needed. Regarding this, SHARE decided to take a different approach: In contrast to the procedure developed for the production of consent documents for the general SHARE data collection, with regard to record linkage no standardised template has been used in order to develop the national consent documents. Several reasons supported this decision, such as the fact that different sources and types of administrative data are used in the national record linkage projects and that the linkage procedure and access to the data varies in accordance with local requirements (see Figure 16). Instead of a standardised template, the Country Teams were provided with guidelines that include information, instructions and a checklist for the production of the record linkage related consent documents.
In terms of preparation of consent for record linkage, the SHARE Country Teams are responsible for ensuring compliance with national legal, ethical, and administrative requirements as well as for the production of the national consent documents and their translation into English. SHARE Central supports the Country Teams by providing them with supporting information and clauses regarding the rights of the data subjects to be included in the consent documents as well as the wording regarding the period of storage of personal data. SHARE Central is also responsible for the central documentation of the implemented procedure. The supporting information includes legal definitions and explanations (e.g. that consent needs to be freely given, specific, informed and unambiguous or with regard to the conditions for withdrawal) and clarifies consent options (i.e. written or documented oral consent in the CAPI software). The instructions provided by SHARE Central to the Country Teams include aspects such as, for example, when to ask for consent (separately during the linkage module and not as part of general consent for participation) and how to present the consent information to respondents (in a written form, in a concise, transparent and easily accessible form, using clear and plain language).

The checklist for Country Teams used for the production of the national record linkage related consent documents covers the following points:

- Mentions that consent is voluntary (and independent from participation in the study)
- Identity and contact details of data controllers and the DPO
- If special categories of personal data are linked to the survey data this has to be explicitly mentioned
- Purpose of data processing (e.g. research projects in the area of the linked data)
9 Discussion session and next steps

Chaired by Elizabeth Lea Bishop, GESIS Leibniz Institute for the Social Sciences

The discussion was divided into three main themes with participants forming three discussion groups according to their main interest, one to discuss each of the following themes:

- Consent-related challenges of linking survey data with social media data
- Challenges of linking survey data with administrative data
- Content of informed consent for survey participation – respondent information sheet

Participants were asked to share their experience and ideas on how specific challenges could be tackled and to discuss forms of potential collaboration that could contribute towards improvement.

The first interest group ‘Consent-related challenges of linking survey data with social media data’ noted that there are still uncertainties regarding whether the “public task” argumentation can be applied to social media data. Apart from legal regulations, ethical issues of liking survey data with social media should be addressed when designing consent to protect respondents’ data. The broad consent should go beyond mentioning data usage but also provide information on data storage. The consent should be as specific as needed but the same time not overdo the level of details if these details are not really required. The consent should include what is necessary to be included for legal purposes as well as what is important for respondents to know for ethical reasons. Apart from potentially resulting in low willingness to provide consent among survey participants, consents that are more specific than necessary can unnecessarily restrict the feasibility of additional research questions or collaborations that emerge after the data has been collected. Group participants were interested in collaboration to design consent templates and checklists that could be made publicly available as well as in sharing consent wording that has already been used in implemented projects with other organisations running similar projects. SERISS
Workpackage 6 is already working on such templates and the outputs will be made publicly available on the SERISS website.

The second group ‘Challenges of linking survey data with administrative data’ noted that ownership of administrative data is very complex, even for admin data providers (e.g. in case of cross-linking data from different sources of admin data) and it matters whether the linkage is done onwards from admin data sources or by a research organisation. Granular consent is problematic because researchers sometimes do not know for sure in advance which variables they will need for their analysis. Achieving a balance between access to all variables (which can be problematic for ethical reasons) and getting access to too few variables to be useful for analysis is among the major challenges of linking survey data with admin data. In addition, there are possible quality issues with using register data, especially as a replacement for primary survey data collection. However, despite some potential quality issues of register data several European countries (e.g. Slovenia and Estonia) are moving towards a register-based census.

The third group ‘Content of informed consent for survey participation – respondent information sheet’ used the draft of the European Social Survey (ESS) Round 9 respondent information sheet (see Appendix 3) as an example to discuss the key points that should be included in informed consent for survey participants. In the case of the ESS, it is not necessary to collect signed consent from survey participants as “public task” rather than “consent” is used as the basis for processing. Ensuring that participants are able to give informed consent by providing them with reference material in written form is sufficient. The information sheet can be sent to respondents with advance letters announcing the upcoming survey or provided by the interviewer at the doorstep. However, the latter case could possibly conflict with the voluntariness principle if there is not enough time for the participants to read the information thoroughly or to make an informed decision about their participation. Therefore, the information sheet should not be too long but at the same time include all required information. The amount and level of detail of provided information (e.g. information about collection of auxiliary data for non-respondents, performing back checks, or storing of indirectly identifiable data) is a subject for discussion in the specific project as there are so far not standardised guidelines. As it is extremely difficult to guarantee absolute anonymity of published survey data, the ESS respondent information sheet avoids using the word “anonymous” and uses instead the following statement “We will make every effort to ensure that no participant is identifiable in the results of the survey or in any publications”. Several participants mentioned that it might be difficult for participants to understand the difference between “publication of anonymous data” and “storing of potentially indirectly identifiable data” and some participants expressed their concerns that providing this information might result in lower response rates. At the same time, several other participants argued that such transparency reflects a serious approach towards handling of personal data and might increase trust in researchers resulting in higher response rates.

In general, there is uncertainty regarding the effect that GDPR will have on response rates in surveys and the willingness of survey participants to provide consent allowing the linking of their survey data with data from other sources. As this is a new field, workshop participants and their organisations could benefit from collaboration aimed at designing and sharing consent templates and content checklists for respondent information sheets as well as sharing experience and training infrastructures. Despite a growing offer on GDPR training addressed towards commercial companies and commercial projects, GDPR training for social researchers is still relatively rare. Therefore, organising a shared training for research projects facing similar issues could benefit several projects at relatively low costs per project.
As a next step SERISS Work Package 6 ‘New forms of data: legal, ethical, and quality issues’ will produce several publicly available deliverables addressing the GDPR and its implications for social research. SERISS Workpackage 5 is also developing training on GDPR and legal and ethical issues pertaining to data collection, aimed at social researchers. To enable future communication and information exchange between the workshop participants, a digital network has been set up on Basecamp. This platform will be used to share materials, feedback and ideas for collaboration.
Appendix 1: Workshop agenda

| PART 1: LEGAL AND ETHICAL ISSUES OF COMBINING SURVEY DATA WITH NEW FORMS OF DATA |
| Chair – Tom Emery, Generations and Gender Programme |
| 10:00 – 10:15 | Welcome and introduction of participants |
| 10:15 – 10:45 | SERISS work on ‘New forms of data – legal, ethical and quality issues’ & Perspective on the GDPR implementation |
| | Marianne Haqetveit Myhren, Norwegian Centre for Research Data (NSD) |
| 10:45 – 11:15 | Legal and ethical issues of combining survey data with new forms of data |
| | Andrew Charlesworth, Bristol University |
| 11:15 – 11:30 | COFFEE BREAK |

| PART 2: LINKING SURVEY DATA WITH SOCIAL MEDIA DATA |
| Chair - Janež Štebe, Slovenian Social Science Data Archives (ADP) |
| 11:30 – 12:00 | Linking survey data with social media data in social research |
| | Curtis Jessop, NatCen Social Research & Tarek Al Baghal, University of Essex |
| 12:00 – 12:30 | Legally and ethically using, re-using, and preserving social media research data harvested from an API |
| | Sara Day Thomson, Digital Preservation Coalition (DPC) |
| 12:30 – 13:15 | LUNCH BREAK |

| PART 3: LINKING SURVEY DATA WITH ADMINISTRATIVE DATA |
| Chair – Eric Schulte Nordholt, Statistics Netherlands (CBS) |
| 13:15 – 13:40 | Linking survey data with administrative data in practice |
| | Peter Jones, UK Office for National Statistics (ONS) |
| 13:40 – 14:05 | Consent-related challenges and strategies |
| | Joseph Sakshaug, Institute for Employment Research (IAB) |
| 14:05 – 14:30 | Implementing consent for record linkage in a cross-national survey |
| | Daniel Schmidutz, Survey of Health, Ageing and Retirement in Europe (SHARE ERIC) |
| 14:30 – 14:45 | COFFEE BREAK |

| PART 4: DISCUSSION SESSION |
| Chair – Elizabeth Lea Bishop, GESIS Leibniz Institute for the Social Sciences |
| 14:45 – 15:45 | Group discussions and exchange of ideas |
| 15:45 – 16:00 | Future collaboration suggestions and closure of the workshop |
## Appendix 2: List of participants

<table>
<thead>
<tr>
<th>Participant</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tarek Al Baghal</td>
<td>Institute for Economic and Social Research (ISER), University of Essex</td>
</tr>
<tr>
<td>Elizabeth Lea Bishop</td>
<td>GESIS Leibniz Institute for the Social Sciences</td>
</tr>
<tr>
<td>Aleksandra Bujnowska</td>
<td>Eurostat</td>
</tr>
<tr>
<td>Andrew Charlesworth</td>
<td>University of Bristol Law School</td>
</tr>
<tr>
<td>Sara Day Thomson</td>
<td>Digital Preservation Coalition (DPC)</td>
</tr>
<tr>
<td>Tom Emery</td>
<td>Generations and Gender Programme (GGP)</td>
</tr>
<tr>
<td>Hayk Gazulyan</td>
<td>Kantar Public</td>
</tr>
<tr>
<td>Mark Handel</td>
<td>Facebook</td>
</tr>
<tr>
<td>Oshrat Hochman</td>
<td>GESIS Leibniz Institute for the Social Sciences</td>
</tr>
<tr>
<td>Marianne Høgetveit Myhren</td>
<td>Norwegian Centre for Research Data (NSD)</td>
</tr>
<tr>
<td>Sally Horton</td>
<td>IPSOS MORI</td>
</tr>
<tr>
<td>Curtis Jessop</td>
<td>NatCen Social Research</td>
</tr>
<tr>
<td>Peter Jones</td>
<td>UK Office for National Statistics (ONS)</td>
</tr>
<tr>
<td>Tomas Cizek</td>
<td>Czech Social Science Data Archive</td>
</tr>
<tr>
<td>Hervé L'Hours</td>
<td>UK Data Archive, UK Data Service</td>
</tr>
<tr>
<td>Ruud Luijkx</td>
<td>European Values Study (EVS)</td>
</tr>
<tr>
<td>Sophia MacGoris</td>
<td>Eurofound</td>
</tr>
<tr>
<td>Marco Paccagnella</td>
<td>Programme for the International Assessment of Adult Competences (PIAAC, OECD)</td>
</tr>
<tr>
<td>David Reichel</td>
<td>European Union Agency for Fundamental Rights (FRA)</td>
</tr>
<tr>
<td>Linn-Merethe Rød</td>
<td>Norwegian Centre for Research Data (NSD)</td>
</tr>
<tr>
<td>Lorna Ryan</td>
<td>European Social Survey (ESS ERIC), CITY, University of London</td>
</tr>
<tr>
<td>Joseph Sakshaug</td>
<td>Institute for Employment Research (IAB)</td>
</tr>
<tr>
<td>Daniel Schmidutz</td>
<td>Survey of Health, Ageing and Retirement in Europe (SHARE ERIC)</td>
</tr>
<tr>
<td>Eric Schulte Nordholt</td>
<td>Statistics Netherlands (CBS)</td>
</tr>
<tr>
<td>Kim Smouter-Umans</td>
<td>European Society for Opinion and Market Research (ESOMAR)</td>
</tr>
<tr>
<td>Elena Sommer</td>
<td>Survey of Health, Ageing and Retirement in Europe (SHARE ERIC)</td>
</tr>
<tr>
<td>Janez Štebe</td>
<td>Slovenian Social Science Data Archives (ADP)</td>
</tr>
<tr>
<td>Øyvind Straume</td>
<td>Norwegian Centre for Research Data (NSD)</td>
</tr>
<tr>
<td>William Thorn</td>
<td>Programme for the International Assessment of Adult Competences (PIAAC, OECD)</td>
</tr>
<tr>
<td>Jessica Trixa</td>
<td>GESIS Leibniz Institute for the Social Sciences</td>
</tr>
<tr>
<td>Martin Vavra</td>
<td>Czech Social Science Data Archive</td>
</tr>
<tr>
<td>Ana Villar</td>
<td>Facebook</td>
</tr>
</tbody>
</table>
Appendix 3: European Social Survey (ESS) Round 9 Respondent Information Sheet – Draft, June 2018

The European Social Survey is a survey of social attitudes, beliefs and behaviours of individuals in European countries

Protecting your personal information is important to us. This leaflet outlines what we do with the information we collect and store.

Why are we contacting you?
[ESS has asked National statistics agencies to produce representative lists of possible participants, based on random sampling information about [your age, gender, address] or household].

If you are under 16 years of age at the time of the interview, a parent or guardian will be asked for permission for you to participate. As a parent/guardian you can request to see the questionnaire in advance.

Your participation in the ESS
During the ESS interview, you will be asked a number of questions about your attitudes, life events, as well as more factual questions. We will ask about, politics, religion, trade union membership and about your attitudes to social issues. Your answers from the interview will be recorded electronically. We also gather information about the area you live in.

Your participation in the survey is voluntary. You are free to choose which questions you want to answer, and you may stop the interview at any time without giving any explanation.

If you decide not to participate
If you decide not to take part in the ESS, for whatever reason, we will ask and record the reason(s) why, for example if this is due to a language barrier or illness/sickness. This information helps us to check that the people who participate in the survey are representative of the overall population in your country. For information about your privacy and rights, see below.

What happens with the collected information?
- We will treat all the information about you with strict confidentiality and in accordance with EUs General Data Protection Regulation (GDPR) and national data protection laws.
- Only the [FWO], that collects data, will have access to your contact (eg name and address) information.
- When the survey is finished, the [FWO] will send the ESS-data, without your name or contact details to the ESS Archive (NSD - Norwegian Centre for Research Data, Bergen, Norway).
- Your name and contact information will be deleted when the data is published and no later than June 2020.
- The rest of the collected ESS-data will be securely stored for an indefinite period. They are made available for use in scientific studies by researchers, students and others interested in Europeans’ social attitudes.
There is a possibility but it is unlikely that some information (such as citizenship, age, country of birth, occupation, ancestry and region) may combine in a way that would identify you. This is a possibility only for the survey responses which are not published. Access to these data will only be given to researchers after approved applications and confidentiality agreements are in place.

The results of the survey will be published on the ESS website within May 2020.

We will make every effort to ensure that no participant is identifiable in the results of the survey or in any publications based on the study.

Your rights

As long as we can identify you in the ESS results (the survey responses), you have the right to object to the processing of your personal data, to access, rectify and erase any information about you, and to ask us what information we hold about you. Once details such as your name and address are removed, then it will no longer be possible to delete the information you provided.

You also have the right to lodge a complaint to your national supervisory authority or with the UK Information Commission’s Office.

Who is responsible?
The European Social Survey European Research Infrastructure Consortium (ESS ERIC) undertakes the survey and is the Data Controller. Its headquarters are in London, UK.

What organisations are involved?
In [country], [Name of National Coordinator] at [Institution] and [the fieldwork agency] is responsible for the recruitment and carrying out the survey. See [link to national websites] for an overview of the key members of this team.

What gives us the right to collect your personal data?
The lawful bases used in ESS are that it is undertaken as a task in the public interest and necessary for research and archiving purposes, in accordance with the General Data Protection Regulation and national laws.

Where can I find out more?
To find out more about the study, visit [national ESS web page] and www.europesocialsurvey.org

Contacts:
FWO/National team [enter phone number] [email]

Data Protection Officer, ESS ERIC (Data Controller): dpo.esseric@europesocialsurvey.org

National Data Protection Authority: [enter contact details]
UK Information Commission’s Office
By Post: Wycliffe House
Water Lane
Wilmslow Cheshire
SK9 5AF
By Telephone 01625 545700
https://ico.org.uk/global/contact-us/email/

Full privacy notice at: www.europeansocialsurvey.org/<sitepage for data protection>