VSC Research Data Management Policy

1. Introduction

1.1 The purpose of the policy, updated from 2016, is

- **a)** To make clear the responsibilities of the VSC and its researchers for managing research data well, in order to adhere to accepted good practice in data sharing;
- **b)** To set out VSC 's obligations as a Data Controller under the Data Protection Act 2018 and the General Data Protection Regulation 2018 ("UK GDPR") with respect to the research data processing. This legislation defines 's responsibility to ensure that there are appropriate organisational and technical measures in place to protect personal data.

2. Purpose

2.1. This policy defines the responsibilities at individual and institutional level which should guide the work of those involved in research data collection, curation, storage and maintenance. It sets out the basis on which the VSC's staff and postgraduate research students will process any personal data collected from data subjects in the course of conducting research, or that is provided to those conducting research by data subjects or other sources for the same.

2.2. This policy will ensure that research data produced by its staff and postgraduate research students will be managed to the highest standards throughout the research data lifecycle in line with relevant legislation. The VSC acknowledges the importance and significance of legislation in this area, inter alia, the Freedom of Information Act 2000, the Data Protection Act 2018 and UK GDPR as amended from time to time. [1]

2.3. The policy will also ensure that the research data the VSC generates through projects will be securely stored and available to staff and postgraduate student researchers for the duration of their research, in addition to being available for the long-term conduct of research, teaching and for wider exploitation for the public good, by individuals, government, business and other organisations, as a project develops and after research results have been published. This will also ensure that, when required, accurate and retrievable research data are available to verify and defend the process and outcomes of research.

2.4. This policy and any other documents referred to in it sets out the basis on which we will facilitate compliance with research funders' data-related policy statements or codes of practice and where applicable, grant terms and conditions.

Part 3. Scope and definitions

3.1 Compliance.
This policy sets out rules on data protection and the legal conditions that must be satisfied when we obtain, handle, process, transfer and store personal data when undertaking research. All of the VSC's staff and postgraduate research students are required to follow policies and procedures to ensure compliance with legal obligations. The VSC Data Protection Officer ("DPO") is responsible for ensuring compliance with the Act and with this policy. That post is currently held by Peter Thompson who can be contacted at Peter@voicestudycentre.com. Any questions about the operation of this policy or any concerns that the policy has not been followed should be referred in the first instance to the DPO.

3.2. FAIR principles for data sharing

Where it is lawful to do so, the VSC supports the broad global consensus that publicly funded research data should be made openly available as soon as possible with as few restrictions as necessary. Additionally, many UK and international funders have embraced FAIR principles for data sharing (making data findable, accessible, interoperable, and reusable), which are facilitated by the development of online research data discovery and sharing capabilities. UK Research and Innovation (UKRI), for example, have agreed a set of common principles for research data policies which are based on the expectations that [2]:

- Publicly funded research data are a public good and produced in the public interest.
- Data that has acknowledged long-term value should be preserved to remain accessible and usable for future research.
- Published results should always include information about how to access the supporting data.
- Such research data should be made openly available with as few restrictions as possible in a timely and responsible manner.

These obligations are further developed by the ten principles articulated in the Concordat on Open Research Data [3]. Both are referenced throughout.

3.3. Research and Development

The policy covers all research and experimental development and the associated research data generated by the VSC. Research and experimental development (R&D) is defined as per the Frascati manual, as 'creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications'. The term R&D covers three activities: basic research, applied research and experimental development.

- **Basic research** is experimental or theoretical work undertaken primarily to acquire new knowledge of the underlying foundation of phenomena and observable facts, without any particular application or use in view.
- **Applied research** is also original investigation undertaken in order to acquire new knowledge. It is, however, directed primarily towards a specific practical aim or objective.
- **Experimental development** is systematic work, drawing on existing knowledge gained from research and/or practical experience, which is directed to producing new materials, products or devices, to installing new processes, systems and services, or to improve substantially those already produced or installed. R&D covers both formal R&D in R&D units and informal or occasional R&D in other units.

The UK GDPR also assumes a "broad conception" of research, including technological development, fundamental and applied research. Research may then be taken as activity aimed at generating new knowledge and advancing the state of the art in a given field in accordance with relevant sector-related methodological and ethical standards, in conformity with good practice. [4]

---

[2] https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open/

[3] https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf

[4] The Royal Historical Society notes that UK GDPR provides for a privileged role for research, without making a distinction between research undertaken for scientific, arts and humanities, or statistical reasons. See *Data Protection and Historians*. https://files.royalhistsoc.org/wpcontent/uploads/2020/07/19092331/20200707_RHS_Data_Protection_Historians_WEB2.pdf

3.4. Research Data

Research data is defined as that which is collected, observed, or created for purposes of analysing to produce original research results. Such research data are the recorded information, regardless of the form or the media in which they may exist, necessary to support or validate a research project's observations, findings or outputs. In practice, the nature of research data can vary widely depending on discipline. It can be textual, numerical, qualitative, quantitative, final, preliminary, physical, digital or print. Research data comes in very many formats, including: word processed documents, PDFs, spread sheets, scanned lab books, online surveys, digital recordings, databases or computer software. It may include, but is not limited to:

- Instrument measurements
- Experimental observations
- Still images, video and audio
- Text documents, spreadsheets, databases
- Manuscripts, text corpus/corpora historical records and archive materials
- Quantitative data (e.g. household survey data)
- Survey results & interview transcripts
- Simulation data, models & software
- Slides, artefacts, specimens, samples
- Computer log files, emails, web pages and forum posts

Researchers are encouraged to refer to the JISC Research Data Management Toolkit for further consideration. [5]

3.5. Researchers

Researchers are defined as members of the VSC, including staff and postgraduate research students, honorary and visiting research fellows, and those who are not members of the VSC but who are conducting research on VSC premises or using VSC resources and facilities.

3.6. Definition of data protection terms:

- **Anonymised data** that cannot be linked to a living individual is not subject to the Data Protection Act, though there may still be ethical reasons for protecting this information. What counts as "anonymised" is measured by a "likely reasonably" test. The UK's Information Commissioner's Office states: "Anonymisation is the process of turning data into a form which does not identify individuals and where identification is not likely to take place." This means that if, on the balance of probabilities, third parties cross-referencing "anonymised" data with information or knowledge already available to the public cannot identify individuals then data is not personal and not subject to the Act.
- **Consent:** agreement which must be freely given, specific, informed and be an unambiguous indication of the Data Subject's wishes by which they, by a statement or by a clear positive action, signify agreement to the Processing of Personal Data relating to them.
- **Data** is information which is stored electronically, on a computer, or in certain paper based filing systems.
- **Data Breach**: any act or omission that compromises the security, confidentiality, integrity or availability of Personal Data or the physical, technical, administrative or organisational safeguards that we or our third-party service providers put in place to

[5] https://www.jisc.ac.uk/guides/rdm-toolkit

protect it. The loss, or unauthorised access, disclosure or acquisition, of Personal Data is a Personal Data Breach.

- **Data Controllers** are the people who or organisations which determine the purposes for which, and the manner in which, any personal data is processed. They are responsible for establishing practices and policies in line with the Act. The VSC DPO is the data controller of all personal data used in our business, including research and academic purposes.
- **Data Minimisation.** Research should be designed so that the data collected is sufficient to address the research question(s) and to achieve the research objectives. Only data that is needed to achieve these aims should be collected. Research teams must minimise how much data they collect (i.e. minimise the number of participants and the number of data items collected about each participant), as well as minimising the degree of sensitivity associated with the data. In other words, researchers must only collect the data they need.
- **Data Processors** include any person or organisation that is not a data user that processes personal data on our behalf and on our instructions. Employees of data controllers are excluded from this definition but it could include suppliers which handle personal data on the VSC's behalf.
- **Data Subjects** for the purpose of this policy include all living individuals about whom we hold personal data. A data subject need not be a UK national or resident. All data subjects have legal rights in relation to their personal information.
- **Data Users** are those of our employees whose work involves processing personal data. Data users must protect the data they handle in accordance with the VSC Data Protection Policy, the Policy in Relation to Special Categories of Personal Data, the Research Data Management Policy, and any applicable data security procedures at all times.
- **Identifiable natural person.** One who can be identified, directly or indirectly, by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person
- **Personal Data.** The UK GDPR defines personal data as any information relating to an identified or identifiable natural person. As this definition is very broad it can be analysed into four elements:

  i. **Any information**. Every piece of information, regardless of its form (digital or analogue text, sound, image or audio-visual material) and of its content (facts and opinions, true or false), can potentially constitute personal data;

  ii. **Relating to a person**. This information can relate to a person in three ways:
    - via the content: the information says something about the person;
    - via the purpose: the information can be used to evaluate or influence the status or behaviour of the person (e.g. a call log of a telephone can be used to evaluate the behaviour of its owner);
    - via the result: the information can have an impact on the person's rights and interests (i.e. the person may be treated differently from others; e.g. statistics about the person's performance at work).

  iii. **Identified or identifiable person**. A person is identified if he/she is singled out from a group. She is identifiable if she can be identified by any means "reasonably likely to be used" by the controller or by a third person. For example, pseudonymised data are still relating to an identifiable person. In contrast, data that do not concern an identifiable person ("anonymous data") is not personal data and therefore the UK GDPR does not apply to their processing.

iv. **Natural person**. Only the information relating to natural (i.e. living) persons is to be considered personal data. This means that the UK GDPR does not apply to the processing of data concerning deceased people or legal entities.

- **Processing** is any activity that involves use of the data. It includes obtaining, recording or holding the data, or carrying out any operation or set of operations on the data including organising, amending, retrieving, using, disclosing, erasing or destroying it. Processing also includes transferring personal data to third parties.
- **Pseudonymisation:** Pseudonymising requires the physical separation of 'real-world' identifiers from the rest of the research data. A link is maintained between research data and 'real-world' identifiers via a cipher or code. The cipher is kept secure and separate from the research data. Thereby limiting how many people in the research team have access to real-world identifiers, making it more difficult to identify individuals from research data, and so helping to guard against accidental disclosure. However, the ICO notes that while pseudonymised data can help reduce privacy risks by making it more difficult to identify individuals, it is still personal data.
- **Research.** The UK GDPR adopts a "broad" definition of research, encompassing the activities of public and private entities alike. The UK GDPR aims to encourage innovation, as long as organisations implement appropriate safeguards. It is important that staff collecting data for research purposes process the data in line with the UK GDPR and VSC guidance.
- **Special Category Data** includes information about a person's racial or ethnic origin, political opinions, religious or similar beliefs, trade union membership, physical or mental health or condition or sexual life, or about the commission of, or proceedings for, any offence committed or alleged to have been committed by that person, the disposal of such proceedings or the sentence of any court in such proceedings. Special Category Data can only be processed under strict conditions, including a condition requiring the express permission of the person concerned.

4. Policy Context

4.1. UKRI: Concordat on Open Research Data

The VSC has adopted the following set of principles, derived from the Concordat on Open Research Data [6] , which should be followed by those conducting research in order to ensure that research data are managed in accordance with relevant legislative, regulatory, contractual, ethical, and other obligations.

Principle 1.	Open access to research data is an enabler of high-quality research, a facilitator of innovation and safeguards good research practice.
Principle 2.	There are sound reasons why the openness of research data may need to be restricted but any restrictions must be justified and justifiable
Principle 3.	Open access to research data carries a significant cost, which should be respected by all parties.
Principle 4.	The right of the creators of research data to reasonable first use is recognised.
Principle 5.	Use of others' data should always conform to legal, ethical and regulatory frameworks including appropriate acknowledgement.
Principle 6.	Good data management is fundamental to all stages of the research process and should be established at the outset.

[6] https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatonOpenResearchData.pdf

Principle 7.	Data curation is vital to make data useful for others and for long-term preservation of data
Principle 8.	Data supporting publications should be accessible by the publication date and should be in a citeable form.
Principle 9.	Support for the development of appropriate data skills is recognised as a responsibility for all stakeholders.
Principle 10.	Regular reviews of progress towards open research data should be undertaken.

These ten principles are included in various technological and organisational measures that the VSC has implemented, by *default and design*. These include, ethics review, data protection impact assessment and research data management planning as detailed below.

4.2. UK GDPR: Data protection by design and by default

Data protection by design and by default is a key element of the UK GDPR's risk-based approach and its focus on accountability. It requires the VSC to put in place appropriate technical and organisational measures which are designed to **a)** implement the data protection principles effectively, and **b)** integrate safeguards into its processing to meet the UK GDPR's requirements and protect individual rights.

- **Data protection by design** is an approach that ensures that the VSC considers privacy and data protection issues at the design phase of any system, service, product or process and then throughout the lifecycle. For research purposes this includes both physical systems (e.g. IT infrastructure) and developing suitable organisational policies and processes.
- **Data protection by default** requires that the VSC will only process the data that is necessary to achieve its specific purpose. It links to the fundamental data protection principles of data minimisation and purpose limitation. This does not require a 'default off' position, but does mean the VSC's researchers need to specify the data required for the specific purposes before the processing starts, appropriately inform individuals and only process the data required for this purpose.

The VSC's obligations: The VSC, as a data controller, is obliged to implement data protection by design and by default. This must take into account:

- The state of the art,
- The cost of implementation,
- The nature, scope, context and purposes of processing,
- The risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing.

Such measures may include ethics review, minimising the processing of personal data, pseudonymising personal data as soon as possible, transparency with regard to the functions and processing of personal data, enabling the data subject to monitor the data processing, enabling the controller to create and improve security features. In respect of this requirement, the VSC will provide training, support and advice for the research data management and research data management plans through its staff development programme for researchers. The VSC will advise on and provide the necessary mechanisms and services for storage, backup, registration, deposit and retention of research data assets in support of current and future access, during and after completion of research projects.

Researchers' obligations: Researchers, faculties, support units and where appropriate external collaborators, are required to work in partnership to implement good practice and

meet relevant legislative, research funder and regulatory requirements. The UK GDPR requires researchers to integrate data protection concerns into every aspect of their processing activities. Those undertaking research are therefore expected to maintain awareness of current requirements and obligations set by the VSC, and where applicable, those of the research sponsor, research partners, the supplier of externally sourced data and any other relevant bodies. The VSC's researchers must also adopt practices that are appropriate and conform to best practice within the subject domain. These should include the application of appropriate measures to protect research participants throughout the research lifecycle. In general, these practices must ensure that research data and records are:

- Accurate, complete, authentic and reliable;
- Identifiable, retrievable, and accessible;
- Retained in a safe and secure manner;
- Processed in a manner that is compliant with legal obligations (e.g. UK GDPR, contractual conditions and the common law of confidentiality and, where applicable, the requirements of funding bodies and project-specific protocols approved by the VSC's Research Ethics Committee;
- Available to others in line with appropriate ethical, data sharing and open access principles;
- Commensurate with the legitimate interests and protection of human participants of research data.

Researchers should familiarise themselves with the VSC's legal obligations for processing data that includes personal information under the UK GDPR and Data Protection Act (2018), and participate in appropriate training and professional development. They should also and seek advice where necessary.

UK GDPR and research.
The legal obligations for data protection in research are set out in **Part 6** (below) but in summary these are that personal data shall be:

- **Lawfulness, fairness and transparency**: Processed lawfully, fairly and in a transparent manner in relation to the data subject;
- **Purpose Limitation**: Collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes;[7]
- **Data minimisation:** Adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed;
- **Accuracy:** Accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay;
- **Storage limitation:** Kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed [8]

[7] Further processing for scientific or historical research purposes in accordance with Article 89(1), not be considered to be incompatible with the initial purposes
[8] Personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical

- **Integrity and confidentiality:** Processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures.

Appropriate safeguards may include, but are not limited to:

- The implementation of a robust, publicly available Data Management Plan
- The approval of the processing of data for research purposes by an ethics committee;
- Data Protection Impact Assessment, even if not required by law (to provide for reinforced accountability);
- Pseudonymisation as expressly recognised by the UK GDPR;
- Functional separation (i.e. taking measures to ensure that data are not used for other purpose than research, and in particular that the data are not used to take decisions or actions with respect to individuals);
- Increased transparency (e.g. providing the data subjects with more information than actually required by law; also making the information publicly available);
- Opt-out mechanisms, allowing data subjects to request removal of their data (even if their right to object is limited and the processing is not based on consent that could be withdrawn) or other mechanisms allowing the data subject to monitor the processing;
- Maintaining a detailed record of processing activities, exceeding what is required by law (reinforced accountability and transparency);
- Reinforced security, including e.g. access restrictions (only certain members of the research team can actually consult the personal data) or storage on a computer with no Internet connection
- Encryption using state-of-the-art techniques
- The use of Privacy Enhancing Technologies;

Part 5. Organisation and Technical Measures

5.1 Ethics Review

The VSC Research Integrity and Ethics Policy sets out the ethical review and approval process for ensuring the VSC's research is carried out in the public interest and is lawful under the UK GDPR and Data Protection Act. All staff and student research that is within the scope of the Research Integrity and Ethics Policy must therefore seek ethical review, and data processing should in no circumstances commence until ethical consent has been given**.**

5.2. Data Protection Impact Assessment (DPIA)

The UK GDPR creates a legal obligation to think about and mitigate data protection issues and privacy concerns at the project planning stage, before any data is gathered (e.g. data protection by design and default). A DPIA is an assessment to help researchers identify any potential risks a project might have as regards intruding into participants' privacy. The DPIA then assists with implementing appropriate measures and controls to minimise and manage those risks. The legislation has made DPIAs mandatory for higher risk data processing to ensure that privacy and data protection are key considerations from the start of any project and then taken into account throughout the project's lifecycle. The European Commission

purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject

(DG Research and Innovation) recommends that a DPIA be carried out for data processing operations that may entail higher ethical risks [9], such as:

Types of personal data	• Racial or ethnic origin • Political opinions, religious or philosophical beliefs • Genetic, biometric or health data • Sex life or sexual orientation • Trade union membership
Data subjects	• Children • Vulnerable people • People who have not given their explicit consent to participate in the project
Scale or complexity of data processing	• Large-scale processing of personal data • Systematic monitoring of a publicly accessible area on a large scale • Involvement of multiple datasets and/or service providers, or the combination and analysis of different datasets (i.e. big data)
Data-collection or processing techniques	• Privacy-invasive methods or technologies (e.g., the covert observation, surveillance, tracking or deception of individuals) • Using camera systems to monitor behaviour or record sensitive information • Data mining (including data collected from social media networks), 'web crawling' or social network analysis • Profiling individuals or groups (particularly behavioural or psychological profiling) • Using artificial intelligence to analyse personal data • Using automated decision-making that has a significant impact on the data subject(s)
Involvement of non-EU countries	• Transfer of personal data to non-EU countries • Collection of personal data outside the EU

For most research studies, the DPIA screening is included in the Research Ethics Committee approval process. However, some funders will require a separate DPIA, which should be competed using the ICO template, or one provided by the funder if that is specified. In cases where a research project does not require ethical review, researchers must still complete a DPIA screening, and if necessary a full assessment, which must be authorised by the DPO.
https://ico.org.uk/media/2258461/dpia-template-v04-post-comms-review-20180308.pdf

5.3. Data Governance & Research Data Management Planning

Institutional data is key to data-driven decision making, identifying new opportunities, planning and risk management. The VSC Data Governance Policy exists to ensure relevant data is fit for the purposes of internal and external reporting, and is appropriately categorised for storage, retrieval, destruction, backup, and access as needed to ensure proper management and protection of Institutional data.

[9] Ethics and data protection (July 21) https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf

In addition to completing a DPIA all staff must comply with research funders' expectations to submit a Data Management Plan (DMP) as part of grant applications. When a DMP is not required, it is recommended that Lead Investigators and Postgraduate Research students nevertheless generate, execute and update one. This is also in some cases required in the ethics review process.

Completion of the data management plan and DPIA will allow researchers to understand data flows and access privacy risks. Through doing so, all staff and research students will:

- Develop and record appropriate procedures and processes to collect, store, use, reuse, access and retain research data associated with their research program;
- Establish and document agreements for managing research data when involved in a joint research project, collaborative research, or research undertaken in accordance with a contractual agreement;
- Include within research grant proposals appropriate consideration of the cost and time implications of data storage and management
- Ensure that the integrity and security of their data is maintained;
- Plan for the on-going custodial responsibilities for the research data at the conclusion of the research project or on departure from VSC;
- Include recommendations for the destruction of research data.

Research data management plans will, where appropriate, use research funder templates or the Digital Curation Centre's DMPonline tool https://dmponline.dcc.ac.uk/ . Once completed the research data management plans must be reviewed by INSPIRE and the Data Governance Group before data collection commences.

5.4. Research Data sharing arrangements

Where proposed research projects involve collaboration with third parties (e.g. another VSC, NHS Trust or other external partner) and the sharing of personal data, special category data, criminal convictions or offences data or pseudonymised data is anticipated, an appropriate contract or data sharing agreement must be put in place before any data is exchanged.

5.5. Research Data Storage

All research data must be stored in the VSC's managed environment that protects against a data breach (as defined by the UK GDPR) and more general research data loss and corruption, unauthorised access and modification, and complies with relevant legal, regulatory, contractual, and other obligations for the period that it needs to be kept. The VSC's Cybersecurity protocols must be observed in this respect. A personal data breach means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data. This includes breaches that are the result of both accidental and deliberate causes. It also means that a breach is more than just about losing personal data. To maintain data security, in all cases the following VSC policies on the storage of data and use of information technology and systems must be followed:

- Data Protection Policy
- Policy in Relation to Special Categories of Personal Data
- Information Technology and Systems: Acceptable Use Policy

In circumstances where institutional storage is not available (such as remote working or during fieldwork), the researcher should take steps to maintain the integrity of the data, and ensure that appropriate security measures are in place to prevent unauthorised access. The research data must be transferred into a managed storage environment at the earliest opportunity.

Any data breach must be reported to the DPO immediately the researcher becomes aware of the breach (please see **6.12** below).

5.6. Research Data Repository Deposit

Research data created or captured by VSC researchers through projects should be offered to an appropriate externally supported data repository, or one designated by the funder, except in circumstances that would breach Intellectual Property Rights (IPR), commercial considerations, ethical, confidentiality, or other obligations, including the UK GDPR, as detailed below (**Part 6**). UKRI provides support for such data infrastructures [11] and they should be used as a priority as:

- The costs of hosting are met by the infrastructures
- The visibility of the data is enhanced
- The infrastructures have very high levels of governance
- The infrastructures provide very high levels of support for data curation, access control and licencing

The deposit should take place upon completion of the research, or upon the publication of results, whichever may be sooner. The VSC will also provide a Research Data Repository that may be used to manage research data over time, in such cases that deposit in an external repository is not possible.

In all cases, research data offered to a repository will be supported by a Research Data Management Plan. The plan must detail the manner in which the research data will be curated (e.g. the criteria used for the selection of data to be deposited, as distinct from the entire data set), the manner in which the selected data will be preserved for the full period of retention, in conformance with the relevant legislative, regulatory, or contractual obligations. Where necessary, appropriate safeguards should be put in place to protect participants and ensure that access conditions are met (e.g. anonymisation, access passwords, inclusion of rights statement). The research data management plan should specify measures taken to comply with the UK GDPR and Data Protection Act (2018) and a named contact responsible for any queries over the data being deposited.

[11] Researchers can also consult the JISC OpenDOAR directory of open access repositories https://v2.sherpa.ac.uk/opendoar/about.html

VSC data repository: By default, data deposited in a VSC data repository should be open data that contain no personal or disclosive information. That is to say, such data should be fully anonymised at the point of deposit and therefore outside of the UK GDPR requirement. This in all instances, *without exception*, shall apply to data deposited in the VSC research data repository. It is the depositing researcher's responsibility to ensure data is anonymised, and any deposited datasets found to contain personal details will be rejected.

External data repositories: Data deposited in third party repositories, such as those hosted by the UK Data Service or UKRI offer additional access controls, for example through different tiers of access (e.g. for 'safeguarded data' that contain no personal information, but where it is considered to contain a residual risk of disclosure and 'controlled data', for data that may be disclosive). Under normal circumstances no VSC derived data falling into these or similar categories should be considered for deposit..

The ICO Data Anonymisation Code of Practice notes that:
Where an organisation converts personal data into an anonymised form and discloses it, this will not amount to a disclosure of personal data. This is the case even though the organisation disclosing the data still holds the other data that would allow re-identification to take place. This means that the DPA no longer applies to the disclosed data, therefore:

- There is an obvious incentive for organisations that want to publish data to do so in an anonymised form;
- It provides an incentive for researchers and others to use anonymised data as an alternative to personal data wherever this is possible; and
- Individuals' identities are protected.

5.7. Retention periods

Research data generated through projects, and any associated records should be retained for as long as they are of continuing value to the researcher and the wider research community, and as long as specified by research funder, patent law, legislative and other regulatory requirements. In general, as specified in the UKRI guidance on best practice in the management of research data [12] the UK Research Councils expect data that underpins findings in publications should be accessible for at least ten years after publication. Individual Research Councils' and other funders' data policies and good research practice guidance provide additional requirements and should be consulted and retention periods specified in each Research Data Management Plan. In many instances, researchers will resolve to retain research data and records for a longer period than the minimum requirement. Researchers should note that personal data can be kept in non-anonymised form for no longer than is necessary for the purposes for which the personal data are processed. This notwithstanding, longer storage is possible when the data are processed for research purposes with 'appropriate safeguards', such as pseudonymisation as detailed in **6** (below). In this regard, data submitted to the repository will be kept indefinitely by default unless a clear retention schedule is provided.

5.8. Data deletion and destruction

[12] https://www.ukri.org/manage-your-award/publishing-your-research-findings/making-your-research-data-open/

When research data and records are to be deleted or destroyed, this should be done so in accordance with particular concern for confidentiality and security. The VSC Record Management Policy should be observed in this respect. For additional information please refer to the Head of Special Collections and Archives (LLR). The UK Data Archive also recommends that the following guidelines for destroying data should be adhered to:

- Data must be deleted from the system on which it has been stored using a secure erasure programme, such as Eraser (http://www.heidi.ie/eraser/index.php) or similar - which repeatedly overwrites files a number of times, until such time as the original data could not be retrieved forensically;
- The recycle/trash bin must be emptied, preferably to be immediately followed by running a secure erasure programme;
- Portable media holding any data must be destroyed and disposed of in a secure manner;
- Backup tapes must either be completely overwritten and degaussed (demagnetised) before being reused or disposed of;
- Paper copies must be destroyed by shredding, preferably using a cross-cut shredder;
- Before the PC, laptop or other device used for data storage leaves the possession of the organisation or individual (for destruction or second-hand sale, etc.), the hard disk must be completely erased using a secure erasure programme;
- If research data are held in the cloud or by a third-party service provider, researchers should ensure that it has securely deleted the data together with any back-ups.
- If data have been shared with partners or transferred to third parties in the course of the project, researchers should ensure that they have deleted the data, unless they have a legitimate basis for retaining them.

5.9. Data Curation

Not all research data is suitable for long term preservation, or sharing in a data repository. As a general guideline, researchers should only store the data that underpins research publications, although there are likely to be exceptions for larger data sets which have enduring and wider value. The Digital Curation Centre has created a comprehensive guide to appraising and selecting data for preservation which outlines the key issues in this area. [13] Researchers should consider, for example the following three steps:

Step 1	**Identify purposes that the data could fulfil**
Verification:	Enable others to follow the process leading to published findings and potentially reproduce or verify these
Further analysis:	Increase opportunities for further analysis of the data e.g., using new methods, integration with other sources for meta-analysis, whether through new collaborations or third-party analyses
Building academic reputation:	Data that is discoverable has greater visibility, which can boost citation rates for the published findings
Community resource:	Development: publish a data resource of value to a known user group, e.g., a reference dataset, methods test-bed, or domain database
Further publications:	The publication of a data article will contribute to scholarly communication and debate about data management or reuse in your domain
Learning & teaching:	Embedding data in a learning/teaching or public engagement

[13] https://www.dcc.ac.uk/sites/default/files/documents/publications/Five%20Steps%20to%20decide%20what%20data%20to%20keep.pdf

	resource to enhance its interactivity, engage users in learning about or participating in the research
Private use:	Find the data more easily in years to come to exploit other potential uses
Step 2	**Identify data that must be kept**
Are there research data policy reasons to keep it?	Will the data underpin an article submitted to a journal that has policy requiring it to be available?
	Will data produced through UKRI or other funding underpin a published research output?
Do regulations require the data to be available?	Does the data need to be retained to comply with Freedom of Information or Environmental Information regulations?
	Are there disciplinary regulations that require data to be retained as part of the research record?
Are there other legal or contractual reasons?	Does the data provide information of commercial value, or is it used in a patent application?
	Do contractual terms and condition state or imply that the data must be retained?
	Is it reasonable to believe that the data may be used in public enquiries or police investigations, or in any report that could be legally challenged?
Does it contain personal data relevant to the reuse purpose?	Does the data contain details that directly identify an individual or can be used to infer their identity, either in isolation or through linking it to another data set?
	Does your institution's ethics approval allow the data to be retained for further research?
	Does the consent agreement allow data to be reused for the purpose that you are now envisaging?
	Did the data subjects give their informed consent to its archiving?
	If so, is it feasible to adhere to any conditions of their consent e.g., any commitment to anonymise the data?
	Can the data be securely stored and actively managed to recognised information security standards?
Step 3	**Identify data that should be kept**
Description:	Is there enough information, e.g., from an up-to-date Data Management Plan, about what the data is, how and why it was collected, and how it has been processed, to assess its quality and usefulness for the aims you identified?
Quality:	Is the data quality good enough in terms of completeness, sample size, accuracy, validity, reliability, representativeness or any other criteria relevant in the domain?
Known users:	Are there users waiting for this data, or is there past evidence of a demand e.g., will this add value to an established resource or series?
Recommendation	Does the funder, or a learned/professional society or equivalent body in the research field recommend sharing data of this type or on this research theme?
Integration potential:	Does the data describe things that fit standardised terms or vocabularies in other research domains, such as geographic locations and time periods?
Reputation:	Was the data produced by a research group or project that is highly rated on the originality, significance and rigour of previous research outputs? Will making the data available be likely to significantly enhance a group or project's reputation?

Appeal:	Could the data have broad appeal e.g., as it relates to a landmark discovery, a significant new research process, or international policy and social concerns?
Non-replicable:	Would reproducing the data be difficult/costly (or impossible as in the case of unrepeatable observations)?
Cleared:	Is the data classified according to its sensitivity and free from privacy/ethical, contractual, license or copyright terms and conditions that restrict public access and reuse? Are any restrictions normal for the study domain?
Open format:	Is the data in a format that does not require license fees or proprietary software/ hardware to reuse?
Independent:	If any specialist software/ hardware is needed to use data, is that widely used in the field of study and readily available?

Metadata. Data curation is vital to make data useful for others and for long-term preservation of data. Therefore, all research data should be accompanied by high-quality documentation and metadata to provide secondary users with essential information to independently understand the data, enable discovery, and allow for scientific re-use. Documentation should describe at least the origin of data, fieldwork and data collection methods, processing and/or the researcher's management of the data. Individual data items such as variables or transcripts should be clearly labelled and described. All deposited data should follow the advice on data privacy as set out in this policy.

Publications. Research publications underpinned by funded research should include a short statement describing how and on what terms any supporting research data may be accessed, ideally via a formal citation. Researchers should therefore only publish research data with publishers (or responsible digital repositories) that provide persistent information for data, in the form of a persistent identifier. The normal standard to ensure persistent identification is the assignment of a Digital Object Identifier (DOI) for the dataset.

Non-digital data. Research data that is not generated in digital format will be stored in a manner to facilitate it being shared in the event of a valid request for access to the data being received. The expectation in this regard is that originators convert and store such data in digital format in a timely manner. The process required for such access, and if necessary, conversion, will be detailed in the research data management plan. In cases where conversion to a digital format is not possible, a copy of the VSC data-record should be printed and stored with the physical dataset. UKRI best practice recommends that data underpinning publications should be retained for at least 10 years after publication but any enquiries on retention periods should be referred to the Head of Special Collections at VSC.

5.10. Cost recovery

Where permitted, the management and sharing of research data should be supported through external research funding. Such funding should seek to recover the costs of any need for long-term curation and dissemination that requires the provision of staff and system resources. Information Technology and Systems will provide a schedule of associated costs. Researchers should liaise with ITS, LLR and INSPIRE regarding such costings when developing and updating research data management plans.

5.11. Privileged Use and Access Rights
In order to ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake funded work may, where provided for in the terms of the grant agreement, be entitled to a limited period of 'privileged use' of the data they have collected to enable them to publish the results of their research. This period of privileged use shall not preclude the publication of metadata at the earliest opportunity. Where a delay in dissemination of deposited data is needed to allow grant holders to publish their research findings, an embargo period can be applied to the data. This embargo period should generally be no longer than 12 months from the end of the grant but may be longer depending on circumstances. At the end of the exclusivity period the data must be made available on an open access basis. In cases where National Data Centres are used, these arrangements will be made directly with their officers. Where data is deposited in the VSC data repository the embargo period will be defined upon ingestion.

5.12. Data Rights.
Without prejudice to the rights of individuals in so far as their rights as data subjects are concerned, rights assigned to funded research data should not unnecessarily restrict its management, sharing or use. The ability to store, manage, share and use research data is dependent upon intellectual property rights being understood and allocated from the outset of the research process. When a VSC researcher is collaborating with any external partner, they should agree between them the rights and responsibilities of each party with respect to data collected, including key decisions about data storage, backup and security, registration, access, transfer, retention, destruction or archiving and licensing. Rights information indicating ownership and permitted use of research data should be clear and unambiguous and documented in consent forms, contracts, and partnership or collaboration agreements. Where possible and appropriate, researchers are encouraged to apply a licence that enables research data to be accessed and analysed and reused by many parties, such as CC-BY licence. All rights should be in accordance with the VSC's Intellectual Property Policy.

Staff research and IP. Subject to third party interests, and in accordance with the VSC's Intellectual Property Policy, research data which supports a scholarly work produced by a member of VSC staff, shall be owned by the VSC and retained for an appropriate period of time in accordance with Point **5.7.** This period may extend beyond the period of employment of the staff member with the VSC.

Student research and IP. Subject to third party interests and in accordance with the VSC's Intellectual Property Policy, research data which supports a scholarly work produced by a student of the VSC, will remain the intellectual property of the student and be retained for an appropriate period of time in accordance with Point **5.7.**

Where research is supported by a contract with, or a grant to the VSC that includes specific provisions regarding ownership, retention of and access to data, the provisions of that agreement will take precedence

Part 6. Obligations arising through the UK General Data Protection Regulation (UK GDPR) and the Data Protection Act 2018 (DPA 2018)

Data Protection legislation covers how personal data should be processed. Personal data is any information that identifies a living individual, including opinions about that individual and/or any intentions a data controller has towards that individual. Personal data collected and used for research is covered by the UK GDPR and DPA 2018. Under the legislation the

VSC is defined as a "public authority" [14]. The important sections of data protection legislation for research are:

i. **Six data protection principles that govern how personal data should be processed**
   - Lawfulness, fairness and transparency
   - Purpose limitations
   - Data minimisation
   - Accuracy
   - Storage limitations
   - Integrity and confidentiality

ii. **The lawful bases under which research data can be processed**
   - Public Task
   - Contract Research

iii. **The conditions for processing special category data**
   - Research
   - Explicit Consent
   - Academic purposes in the public interest

iv. **Rights relating to making requests on:**
   - a right to be informed
   - a right of access
   - a right to rectification
   - a right to erasure
   - a right to restrict processing
   - a right to data portability
   - a right to object
   - rights in relation to automated decision making and profiling

6.1. UK GDPR implications for open data

The VSC supports the principles in the Concordat on open Research Data that recognise that research data should wherever possible be made available for use by others in a manner consistent with relevant legal, ethical and disciplinary frameworks and norms. The UK GDPR does not prevent research data from being archived and shared for research use by others, as long as the data protection principles are met. The VSC, like all research organisations, must meet all legal requirements relevant to the processing activity (e.g. common law of confidentiality) and specify a lawful basis for data processing for their activities. This means that if researchers are processing personal data for research purposes, they should know the lawful basis they are relying on because they may be asked to specify it.

6.2. Lawful basis for processing personal data in research

Public Task. There must always be a lawful basis for processing personal data and, in the context of research activities, VSC's lawful basis is "the performance of a task carried out in the public interest or in the exercise of official authority vested the controller", known as 'public task'. The VSC can demonstrate that it meets the requirements to use this lawful basis under our Royal Charter, as the objects of the VSC are to "advance education

[14] Public authorities (e.g. universities, NHS, research council institutes) are funded by the public purse in order to conduct tasks that are in the public interest. Public authorities are defined in the DPA 2018, as those bodies that are subject to freedom of information legislation across the UK.

and disseminate knowledge by teaching, scholarship and research for the public benefit". [15] By using 'public task' as the legal basis, the VSC can ensure that as a publicly-funded organisation, it is always one of our official public tasks when we use personal data from people who have agreed to take part in research. It also reassures data subjects that the researchers are part of a reputable organisation that has a genuine reason to hold and use personal data. This is in addition to the control given to participants through the research ethics consent process.

Contract Research. VSC researchers may also undertake research projects requested and funded directly by an organisation(s), for this contract research or consultancy the lawful basis for processing personal data will be 'the performance of a contract'.

Consent. Under the UK GDPR, the VSC ***will not*** rely on consent as the legal basis for processing research data[16]. This means that although researchers usually need to obtain the consent of participants to take part in research projects to meet ethical requirements, they do not need to have their consent to process their personal data. This is because the VSC's lawful basis for processing personal data in research activities is public task.[17]

It is therefore important to distinguish consent for the processing of personal data from other consent processes or requirements. One way to achieve this in practice is for researchers to indicate clearly in a consent form, where the participant's consent is being asked for processing their personal data and where consent is being asked for taking part in the research, for use of the collected information, etc. If the form suggests that consent is required to process personal data, then that consent can also be withdrawn and researchers will not be able to use the data in their research, which can be very difficult at later stages or when close to publication (also see **6.10**. Participant rights and exemptions).

6.3. Special Category Data

The UK GDPR defines certain data as being special category data, and this requires ***additional*** conditions for processing. The special category data are as follows, and are very likely to be of particular relevance to research carried out in the humanities, arts and social sciences within the VSC:

- Personal data revealing **racial or ethnic origin**
- Personal data revealing **political opinions**;
- Personal data revealing **religious or philosophical beliefs**;
- Personal data revealing **trade union membership**;
- **Genetic data**;
- **Biometric data** (where used for identification purposes);
- Data concerning **health**
- Data concerning a person's **sex life**; and
- Data concerning a person's **sexual orientation**.

---

[15] https://www.VSC.ac.uk/media/VSC-website/content-assets/documents/governance/ordinances/charter-statutes-1.pdf

[16] The exception to this may in some cases be where 'special category' data are processed which requires an additional protection, as detailed in 6.3 below.

[17] UKRI / MRC guidance notes that "the law does provide the lawful basis of 'consent' to process personal data; and 'explicit consent' as a condition for special category personal data. However, we envisage that research organisations will not need to rely on these to support their research activities where an alternative lawful basis such as public task. The types of organisation that may need to rely on consent are those involved, for example, in marketing, who have traditionally used pre-ticked boxes to indicate agreement to share personal data widely". *GDPR: Lawful basis, research consent and confidentiality,* May 2018.

The additional conditions which must be met for processing special category data in the context of the VSC's research is most likely to be 'Scientific and Historical Research Purposes', although 'Explicit Consent' and 'Public Interest and Academic Purposes' could also be considered in some circumstances.[18]. These are detailed in **6.4, 6.5** and **6.6.**

6.4. Additional Conditions: *Scientific or Historical Research Purposes.*

In most cases, when processing special category data, researchers are advised to use the Article 9(2)(j) provisions where 'processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes' (providing that using the data is proportionate for the research and the fundamental rights and interests of the data subject are safeguarded). The relevant basis in UK law is set out in the DPA 2018, Schedule 1, part 1, condition 4. This condition requires researchers to:

- Demonstrate that the processing is necessary for archiving, research or statistical purposes. It must be a reasonable and proportionate way of achieving one of these purposes, and researchers must not have more data than they need;
- Comply with the safeguards and restrictions set out in Article 89(1) of the UK GDPR and section 19 of the DPA 2018 (see below); and
- Demonstrate that the processing is in the public interest.

Researchers must also note that the public interest basis of the research has to be demonstrated. As the VSC relies on its status as a public authority and its Royal Charter for the legal basis of processing as a public task (e.g*. carrying out a specific task in the public interest which is laid down by law*) the public interest condition does not need a further test. However, the VSC must still be able to demonstrate that the data processing is necessary for the public interest purpose. 'Necessary' means that the processing must be a targeted and proportionate way of achieving the research purpose. For example, researchers do not have a lawful basis for processing if there is another reasonable and less intrusive way to achieve the same result. As such all research that processes special category data must have ethics committee approval. These are best considered as part of the safeguards as follows.

Safeguards: When processing special categories of data for research purposes, researchers must meet an additional condition, known as the **Article 89(1) safeguards**. These require that "suitable and specific measures to safeguard the fundamental rights and the interests of the data subject" are in place. These safeguards consist of technical and organisational measures and provide research participants with assurance that:

1. The data processing is necessary to support research;
2. The data processing will only be used to support legitimate research activities that are considered to be in the public interest;
3. Data subjects' interests are safeguarded/protected;
4. Demonstrate that the processing is not likely to cause substantial damage or distress to individuals;
5. Not use the data to take any action or make decisions in relation to the individuals concerned [19]

[18] This does not include personal data about criminal allegations, proceedings or convictions, as separate rules apply. If research includes such data, researchers must contact the Head of Data Collection for advice.
https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/criminal-offence-data/

[19] Unless carrying out approved medical research as defined in section 19(4) of the DPA 2018

Technical and organisational measures are:

6. The minimisation principle: that researchers only gather the minimum amount of personal data necessary for the specified research purpose (e.g. if we don't need to collect information about ethnicity, we don't ask for it);
7. Demonstrate why the research cannot use anonymised data (that data should be anonymised wherever possible, either at point of capture or once collated);
8. Consider whether the research could use pseudonymisation to make it more difficult to link the personal data back to specific individuals (where data cannot be anonymised, it is, wherever possible, pseudonymised)
9. Consider other appropriate safeguards and security measures (store the data securely and that appropriate technical and organisational measures are in place to protect personal data).

Public interest test
Besides having these technical and organisational measures in place, researchers must be able to prove that the research is in the public interest, and as discussed above, necessary for achieving that task. The types of evidence for proving that research is in the public interest will be familiar to researchers and very likely already in place. At VSC, this will best be determined by the researcher in conjunction with the Ethics Committee, who will ensure that approval for research is only given if:

9. The research is proportionate;
10. The methodology is appropriate to the aims and intended outcomes of the research and confirms to field-specific standards and best practice,
11. The research and methodology meets appropriate professional standards or codes of conduct and / or guidance;
12. Is undertaken within an appropriate governance framework.

Where research has been through peer review (for example by a grant funding body such as UKRI or the EU) public interest is also likely to be demonstrated.

6.5. Additional Conditions: *Explicit Consent*.
While researchers are advised to use scientific or historical research purposes carried out in the public interest as the additional condition for processing special category data **(6.4)**, in cases where this does not apply, explicit consent may be considered. In this case the data subject must give explicit consent to the processing of those personal data for one or more specified purposes. 'Explicit consent' is not defined in the UK GDPR, but must meet the usual UK GDPR standard for consent. In particular, it must be freely given, specific, affirmative (opt-in) and unambiguous, and able to be withdrawn at any time. It should also be a genuine choice. In practice, the extra requirements for consent to be 'explicit' are likely to be:

- explicit consent must be confirmed in a clear statement (whether oral or written), rather than by any other type of affirmative action
- it must specify the nature of the special category data;
- it should be separate from any other consents researchers are seeking.

6.6. Additional Conditions: *Public Interest and Academic purposes.*
If gaining explicit consent is not possible (e.g. it would interfere with the validity of the research) and there is doubt regarding the status of the research, researchers may also consider using 'public interest' as the legal basis for processing special category data under Article 9(2)(g). The substantial public interest condition in this case would be condition 13: 'Journalism, **academia**, art and literature' (DPA Schedule 1 Part 2).

Consent Form Wizard
The Ethics and Legality in Digital Arts and Humanities (ELDAH) working group of the Digital Research Infrastructure for the Arts and Humanities (DARIAH-EU) has created the Consent Form Wizard to support EU humanities researchers in UK GDPR compliance. The DARIAH ELDAH Consent Form Wizard (CFW) is a tool which supports humanities researchers in obtaining valid consent for data processing in the context of their specific professional activity. Based on an online questionnaire, the system generates a consent form template, which observes the Articles of the UK GDPR. The CFW does not provide formal legal advice though.

The Consent Form Wizard is available at the following website: https://consent.dariah.eu/

6.7. Common law – confidentiality.

The law around information about people is further complicated in the UK as researchers must also comply with the common law of confidentiality. At the time of publication of this policy, information is considered confidential in common law if:

i. It can be related to an identifiable individual (similar definition of identifiable as used for personal data, but personal data can only relate to a living person, confidential information can relate to the living or deceased), ***and***
ii. It is not in the public domain (no such limit is placed on the definition of personal data), ***and***
iii. It is given with the expectation that it will be kept confidential. Individuals do not have to be explicit about their expectations, when entrusting others with their information this expectation is often implicit.

When an individual entrusts a researcher or research team with confidential information, the team must handle this in line with 'reasonable expectations'. In other words, confidential information should only normally be shared when there would be 'no surprises' for the individuals concerned. Where participants would not expect the researcher to be sharing their confidential information with others, researchers can manage their expectations by informing them of their intentions (e.g. in project materials or during discussions about participation) and asking them if they are happy with these plans. They should understand what is being proposed and what this might mean for them, before they decide whether the researcher can share their confidential information with others. Researchers should also always consider if they could limit the sharing of information to robustly anonymised information only [20]. Robustly anonymised information can be shared without having to consider reasonable expectations as information has to be identifiable to be subject to the common law of confidentiality. If in doubt as to the current status of the common law in relation to "confidential information" please refer to the DPO.

6.8. The six data protection principles and research

Principle 1: Data must be processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency').

Personal data must be collected and used in accordance with Data Protection legislation. This principle means that individuals should know who is collecting the research data, where

[20] See the ICO anonymisation code for further guidance.
https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf

it will be kept and what will be done with it. If researchers are collecting the data, they should have a Participant Information Sheet (e.g. privacy notice) on the form or associated with the collection so that people are aware of what the researchers will be doing with their data.

UKRI guidance [21] explains that being fair with research participants includes respecting their rights and ensuring that personal data is used in line with their expectations. Transparency is therefore intrinsically linked to fairness. The fairness and transparency requirements give control to participants: they have greater awareness of how their data is being used and can object if they wish. The following should apply:

- Transparency information must be concise, easy to understand and easy to find;
- Transparency information is best provided at both the corporate and research project levels (a layered approach). Researchers should work with the DPO to ensure that the information provided to participants is coordinated, relevant and understandable, and explains how data is used to support research.
- Good research transparency should help participants understand that data is commonly linked with other data sources, kept for a long time, reused to address important research questions and how their interests are protected;
- Organisations should display corporate privacy information about research where people will notice it, for example, links on website homepages. Researchers should help their participants to notice privacy information using communication methods appropriate for the study population, for example, links from participant information sheets. Researchers can provide further detail in departmental or project materials;
- Where data was not collected from participants, but from other sources, there are exemptions to transparency requirements if the provision of information is 'impossible' or involves 'disproportionate effort'. In these circumstances UK GDPR transparency information must be publicly accessible as a minimum, further efforts to help people notice it are not required. If researchers think this exemption might apply, they should discuss this with the DPO

Principle 2: Personal data must be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes ('purpose limitation').

This information must be detailed in the Participant Information Sheet (PIS). However, further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89 (1), not be considered to be incompatible with the initial purposes. However, if researchers are being required by a research funder or intend to deposit the data in an open repository in anonymised or any other form so it can be used by other researchers, they must tell research subjects that they will be doing so when collecting the data.

Principle 3: Personal data collected must be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation').

Researchers should not collect any personal data that is not strictly necessary for the specified purpose. If researchers are obtaining or holding any special category data, they must take special care to properly consider its necessity.

[21] https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-GDPR-FAQs.pdf

Principle 4: Personal data collected must be accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy').

Personal data collected for research is likely to be a snapshot of a moment in time. As such, it is unlikely to need updating. Long term collections will be keeping to this principle anyway for the purposes of future research.

Principle 5: Personal data must be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed ('storage limitation').

Personal data may be stored for longer periods insofar as the personal data will be processed solely for scientific or historical research purposes in accordance with UK GDPR Article 89(1). This is subject to the implementation of the appropriate technical and organisational measures in order to safeguard the rights and freedoms of the data subject. This means that research data is exempt from Principle 5 as long as data subject rights and freedoms are safeguarded. This doesn't mean that researchers shouldn't consider how long to keep data. Researchers should consider destruction of the original dataset once an anonymised or other form is in an open digital repository, for instance.

Principle 6: Personal data must be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

Personal data should be kept securely so that no unauthorised access can occur.

Paper/hard media should be:

- Kept in lockable cabinets/cupboards when not in use,
- Kept in lockable offices if possible,
- Not left unattended on desks while processing and only processed in locations with access control (e.g. secure entry systems).

Electronic data should:

- Not be displayed where third parties can inadvertently see it,
- Not be forwarded to third parties accidently,
- Be encrypted or password protected in transit,
- Be kept on secure network drives or password protected/encrypted removable media. Users should log-off when not attending the equipment (e.g. when away from their desk).

Participant Information Sheet

To comply with fairness and transparency, researchers will need to provide a Participant Information Sheet ('PIS' which for non-research data collection is called the 'privacy notice') for any research project that processes personal data. The PIS needs to include the following:

- Who is the data controller (the organisation with the overall responsibility - this will be the VSC represented by the lead researcher);
- Enough information, in lay language, for the participant to understand what the project is about and what is required of them;
- Any significant risks to the participants involved;
- Safeguards put in place to limit risks;
- Consent to participate in the research (general);
- What the legal basis is to make the processing of personal research data lawful (see below "Legal bases for conducting research using personal data");
- Consent to process personal data
- Who participants can contact for more information (lead researcher's contact details), a complaints contact and the contact details of this organisation's Data Protection Officer Peter@voicestudycentre.com
- Details of how people can exercise their rights (see below "Research Participants Rights")
- Assurances that their data will be held securely
- For special categories of personal data, compliance with the common law duty of confidentiality
- A note that if the research project changes in any way, the amended PIS will be shown on the project's, Research Centre's, Institute's or School's website.

Research Privacy Notice

Please see our template available on our website.

6.9. Anonymisation and Pseudonymisation of research data

EU ethics guidelines on data protection issues [22] explain that one of the best ways to mitigate the ethical concerns arising from the use of personal data is to anonymise them so that they no longer relate to identifiable persons. Data that no longer relate to identifiable persons, such as aggregate and statistical data, or data that have otherwise been rendered anonymous so that the data subject cannot be re-identified, are not personal data and are therefore outside the scope of data protection law. Where it is necessary to retain a link between the research subjects and their personal data, researchers should, wherever possible, pseudonymise the data in order to protect the data subject's privacy and minimise the risk to their fundamental rights in the event of unauthorised access. Pseudonymisation and anonymisation are not the same thing and it is important that researchers are aware of

[22] https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-and-data-protection_he_en.pdf

the difference between them, as the UK GDPR (Article 89) requires researchers to use them wherever possible or feasible.

Anonymisation. Data in a form that does not identify individuals and where identification through its combination with other data is not likely to take place. While anonymised data are no longer considered personal data, anonymisation processes are challenging, particularly where large datasets containing a wide range of personal data are concerned. If researchers intend to anonymise the data collected for use in a research project, the timing of the anonymisation process is also paramount. Researchers are collecting 'anonymised' data only if the anonymisation happens at the point and time at which the data are collected from the research subject, so that no personal data are actually processed. If anonymisation takes place at a later stage, e.g. it is the intention to remove personally identifiable information during the transcription of audio recordings or at the point at which survey data are fed into a database, the raw data are still personal data and the research must include provisions for its protection up until the point at which they are deleted or rendered anonymous. The Information Commissioner's Office have produced a 'Anonymisation Code of Practice' [23] which includes advice on good practice for anonymising data, as well as discussions on the legal considerations of the DPA 2018 . Researchers should refer to this for essential guidance and apply it to the recognised best practice within their research disciplines.

Pseudonymisation. Processing the personal data in such a manner that it can no longer be attributed to a specific data subject without the use of additional information, which needs to be kept separately and subject to technical and organisational measures. For example, if the research method de-identifies individuals in a survey by giving each respondent a numeric identifier, the data will technically remain personal and under the UK GDPR. Researchers should also note that it will still be classified as pseudonymised data if they have another file which links that numeric information to the real names or other personal information. If researchers destroy the linkage key between the identifiers and the personally identifying information, then it classifies as anonymised data and no longer falls under the requirements of the UK GDPR.

Re-identification is the process of turning pseudonymised or anonymised data back into personal data by means of data matching or similar techniques.

6.10. Participant rights and exemptions

Under the UK GDPR, data subjects have the following general rights:

- **a right to be informed** (i.e. to be told about the collection and use of their personal data)
- **a right of access** (i.e. a right to obtain a copy of personal data together with other supplementary information)
- **a right to rectification** (i.e. a right to have inaccurate personal data corrected or, if incomplete, completed)
- **a right to erasure** (i.e. a right, in certain circumstances, to have personal data deleted)
- **a right to restrict processing** (i.e. a right, in certain circumstances, to limit the way that personal data can be used)

[23] Introduction. https://ico.org.uk/media/about-the-ico/consultations/2619862/anonymisation-intro-and-first-chapter.pdf
Chapter 1. https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf

- **a right to data portability** (i.e. a right, in certain circumstances, for individuals to obtain personal data in a commonly used and machine-readable format for reuse by another service)
- **a right to object** (i.e. a right, in certain circumstances, to request that processing of personal data stops)
- **rights in relation to automated decision making and profiling** (i.e. a right, in certain circumstances, to restrict the use of automated decision making or profiling)

Some exemptions and allowances from data protection regulations can apply when data processing is undertaken for scientific or historical research purposes. If you require further information on this, please contact the DPO. The specific safeguard relating to exemptions to data subjects' rights is that applying the right would prevent or seriously impair the achievement of the research purpose. The following data subject rights may be limited for research data sets:

- the right of access;
- the right to rectification;
- the right to restrict processing; and
- the right to object.

The UK GDPR also provides exceptions from its provisions on the right to be informed (for indirectly collected data) and the right to erasure.

The exemption and the exceptions only apply:

i. To the extent that complying with the provisions above would prevent or seriously impair the achievement of the purposes for processing;
ii. If the processing is subject to appropriate safeguards for individuals' rights and freedoms (see Article 89(1) of the UK GDPR – among other things, researchers must implement data minimisation measures);
iii. If the processing is not likely to cause substantial damage or substantial distress to an individual;
iv. If the processing is not used for measures or decisions about particular individuals, except for approved medical research; and
v. As regards the right of access, the research results are not made available in a way that identifies individuals.

Where any participant seeks to use one of the above rights, researchers should immediately seek advice from the DPO. Special care must be taken when considering an exemption or exception, and the DPO alone will be responsible for making and justifying this decision.

In considering an exemption or exception, consideration will be given to research integrity and reproducibility, other legal requirements, resource implications and the impact on scientific validity. Issues of practicality are also relevant, as while a research subject can ask for their data to be removed from a dataset, it will be very difficult to do so once the research has been published. These exemptions will be balanced with what is fair to participants, so, before applying an exemption, the DPO will need to be aware of what research participants have been told about their rights. What has been said about withdrawal from the study is important here. Also important is whether the research has substantially evolved from what participants understand. If a researcher has evidence that the information they hold is correct, the research data does not necessarily need to change. However, a note should be added to the effect that the research subject has challenged the accuracy of the data.

6.11 Transfer outside the EU and working with collaborators

The UK GDPR does not allow transfer of personal data outside the EU or a country without adequate protection without meeting a condition like explicit consent or a binding contract. Researchers must always therefore obtain explicit consent or anonymise data before sending data outside the European Economic Area (the EU plus Norway, Iceland and Liechtenstein). Some countries listed by the EU have received adequacy decisions (as of Jan 2021 these are Andorra, Argentina, Canada, Faroe Islands, Guernsey, Israel, Isle of Man, Japan, Jersey, New Zealand, Republic of Korea, Switzerland, United Kingdom, Uruguay but the list may change with time). [24] The UK is currently considered adequate for data protection purposes until 2025. In all cases researchers should use password protection or encryption where possible in transit, but bear in mind that some countries will require the use of encryption keys if personal data are transferred into them.

6.12. Reporting a Data Breach

The UK GDPR requires Data Controllers to notify any Personal Data Breach to the Information Commissioner and, in certain instances, the Data Subject. The VSC has put in place procedures to deal with any suspected Personal Data Breach and will notify Data Subjects and/or the Information Commissioner where we are legally required to do so. All breaches should be reported immediately to the DPO at foi@VSC.ac.uk. Researchers should preserve all evidence relating to the potential Personal Data Breach to enable the DPO to carry out an investigation and, as required, report to the Information Commissioner and/or Senior Management of the VSC Group.

7. Take-down Policy

The VSC will make every reasonable effort to ensure that all data published in the VSC research data repository complies with the Law of England and Wales (English Law). However, should any employee, student or third party feel that certain content in some way infringes English Law or rights the VSC will, upon notification, review the legal status of the relevant material and remove it if the complaint is found to be valid. Valid grounds for the removal of content from the repository include (but this is not an exhaustive list):

- Violation of intellectual property rights, including copyright;
- Breach of data protection law (e.g. Data Protection Act 2018, UK GDPR), including, but not limited to, a data breach;
- Breach of moral or other rights protected by law (for example, confidentiality, derogatory treatment of work, libel, privacy);
- Ethical issues including plagiarism, falsified research, and the failure to adhere to ethical guidelines;
- Issues of national security.

To register a complaint, the complainant should contact the DPO
using the email address Peter@voicestudycentre.com The following information should be included:

- Contact details;
- Details of the item (title, author, URL, etc.);
- A description of the grounds for the complaint including any evidence or proof.

On receipt of the complaint, the DPO will:

- Acknowledge that the complaint has been received;
- Make an initial assessment of the validity of the complaint;

- Temporarily remove the item from public view if further investigation is required.

Following the initial assessment, if a complaint is judged to be invalid, no further action will be taken and the complainant will be informed of this decision; valid complaints will be investigated by the Chair of the Research Committee, the DPO, and the author of the item in question. All attempts will be made to resolve the issue swiftly and to the satisfaction of both the complainant and the author.

If, as a result of investigation, the item concerned is judged to have infringed an aspect of English law, it will be permanently withdrawn from the repository. A historical metadata record of the withdrawn item accompanied by a notice detailing reasons for withdrawal will remain in the repository but will not be accessible by the public. If a data breach has occurred, this will be reported to the Information Commissioners Office (see 6).