Data Protection & Ethics

Robin Rice
Data Librarian and Head, Research Data Support
Library & University Collections

Open Access Week, 2018
New University of Lisbon
Caparica, Portugal

Course objectives

- To engage participants in discussion of legal and ethical issues involved in data collection, data management, and data sharing with regard to human subject research.
- To introduce obligations on university researchers from the perspective of the General Data Protection Regulation (GDPR).
- To suggest strategies for working with personal and sensitive data that are both data protection-compliant and follow good practice in research integrity and research data management.
- To raise awareness of available research data services.

Overview

- Ethical and legal perspectives on research data
- Legal (GDPR) definitions, principles for research
- Strategies for
  - Data management plans (DMPs) and data protection impact assessments (DPIAs)
  - Data collection, consent, and transparency
  - Active data management and data security
  - Data sharing – anonymisation & controlling access
- References for further information

What do we mean by sensitive data?

Generally -

- *Data relating to people.*
- Data relating to rare or endangered species of plants or animals.
- Data generated or used under a restrictive commercial research funding agreement.
- Any data posing a threat to others or to national security.
- Any data likely to have significant negative public impact if released.

Research ethics

Ethical principles are in line with, or go beyond legal principals. Ethics cover:

- The purpose and nature of the research itself.
- The nature of consent obtained (e.g. opt-in versus opt-out participation).
- What data need to be safeguarded during analysis and destroyed after their use.

Ethics require a context-sensitive approach and a balanced risk assessment about likely harm to individual/community vs right to research/public to know

What research is subject to ethical review

- Using human participants or live animals
- Referencing individual subjects (people)
- Keeping identifiers for individuals

Ethics Committee by Nick Youngsor

What research is exempt from ethics review?

- Freely available in the public domain, or
- Obtained through pure observational studies of public behaviour:
  - are of human action that occurs in a forum open to the general public,
  - are non-invasive,
  - require no interaction with participants,
  - do not identify participants (so no camera may be used).
- Does this apply to internet/social media?
  - Not necessarily! Researchers need to consider the expectations of the providers of the information and seek consent where appropriate.
  - University guidance states that the Data Protection Officer should be contacted regarding research involving social media.

Research data and the Law:
General Data Protection Regulation, 2018

Songwriters: Sonny Curtis
I Fought the Law lyrics © Sony/ATV Music Publishing LLC

Disclaimer: I'm a librarian, not a lawyer -

THE UNIVERSITY of EDINBURGH

Schools & departments MyEd

Search

Home > Profile pages > Dr Rena Gertz (LLB, LLM, PC.dp)

Dr Rena Gertz (LLB, LLM, PC.dp)

Data Protection Officer

As the Data Protection Officer, I can provide advice and assistance to you. Please contact me, if:

- You have questions about filling in the privacy notice
- You would like to share personal data and are not sure if you can do so lawfully
- You are not sure if you need to carry out a privacy impact assessment
- You are not sure how to carry out the privacy impact assessment
- You have any other question about personal data
- There has been a serious data protection breach and we will need to notify the Information Commissioner

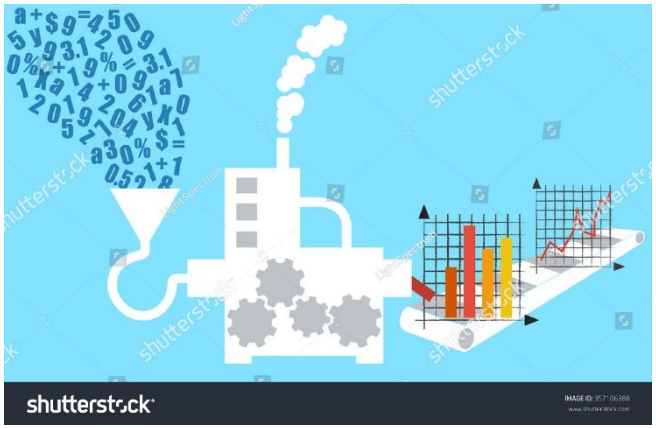
UKRI: "Why GDPR matters for Research"

- 1. Research is in the public interest.
- 2. Consent to take part in research is important.
- 3. GDPR recognises that research data is valuable, it can be kept long-term.
- 4. GDPR forces a record of historical decision-making.
- 5. GDPR safeguards reflect current research good practice.

From https://blog.esrc.ac.uk/2018/05/25/why-gdpr-matters-for-research

GDPR definitions: processing

- Data processing is any action taken with personal data. This includes the collection, use, disclosure, destruction and holding of data.



Note: The GDPR applies to those in the EU who collect personal data about human subjects anywhere in the world, or to anyone outside the EU who collects personal data on EU citizens.

GDPR definitions: data controller

- A data controller is a person or organisation that has full authority to decide how and why personal data is to be processed, and that has the overall responsibility for the data. This includes deciding on use, storage and deletion of the data.
- *Your university* is likely the data controller for research conducted by staff, and registered with your national authority (ICO in the UK).

GDPR definitions: data processor

- A data processor is a person or organisation that processes personal data on behalf of another organisation.

  - If the University passes personal data to a company or an organisation, but instructs this company or organisation on what should be done with that data and how to do this by means of a contract, then the receiving organisation is a data processor.

  - The University will only be legally responsible for any breaches of data protection legislation by a data processor if no contract is in place, and if the University has not satisfied itself that the data processor has adequate security provisions in place.

Activity: DC or DP?

- 1. A researcher makes a contract with a transcription service to transcribe interviews.
- 2. A researcher from the University of Edinburgh wins a funded grant in partnership with a researcher at another institution (co-PI). Both researchers/teams will analyse the data.
- 3. A Postgraduate student works on a funded University project. Both parties discuss and make decisions about the analysis of the data (purpose for processing).

GDPR definitions: personal data

- "'Personal data' means any information relating to an identified or identifiable natural person ('data subject)."
- An identifiable person is one who can be identified both directly and indirectly, in particular by reference to an identification number or to one or more factors specific to their physical, physiological, mental, economic, cultural or social identity.
  - Directly identifiable means identifiable from the information itself, for example, a name together with an address, age, telephone number.
  - Indirectly identifiable means not identifiable from the information itself, but from the information combined with data from another easily available source.

From: https://www.ed.ac.uk/records-management/guidance/data-protection/definitions/personal-data

Spectrum of identifiability

From: https://understandingpatientdata.org.uk/what-does-anonymised-mean

Activity: which of these might be personal data?

- **Biographical information or current living situation**, including dates of birth, Social Security numbers, phone numbers and email addresses.
- **Looks, appearance and behaviour**, including eye colour, weight and character traits.
- **Workplace data and information about education**, including salary, tax information and student [course enrolment].
- **Private and subjective data**, including religion, political opinions and geo-tracking data.
- **Health, sickness and genetics**, including medical history, genetic data and information about sick leave.

From: https://www.itgovernance.eu/blog/en/the-gdpr-what-exactly-is-personal-data

GDPR definitions: legal basis for processing

Whenever we use personal data we must have a legal basis for doing so.

Data protection legislation gives us a list of possible legal bases we can choose from. These are:

- necessary for contractual arrangements
- processing to comply with legal obligations
- processing to protect vital interests
- processing to perform a task in the public interest
- processing necessary for legitimate interest
- [explicit and freely given] consent

From: https://www.ed.ac.uk/records-management/guidance/checklist/legal-basis

University research and legal basis

- "In most circumstances the legal basis for using personal data for research will be 'public task'. This is [in] reference to the University's public research purpose as established by statute (e.g. the Universities Scotland Act 1966).
- By using 'public task' as the legal basis, we 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, and that you 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 (to participate in research) process."

P. 3, Research under the General Data Protection Regulation, UoE DPO

GDPR definitions: special categories data

Special category personal data is more sensitive, and so needs more protection.

In particular, this type of data could create more significant risks to a person's fundamental rights and freedoms. For example, by putting them at risk of unlawful discrimination.

(In the previous Data Protection Directive, this type of data was called 'sensitive, personal' data.)

- race
- ethnic origin
- politics
- religion
- trade union membership
- genetics
- biometrics (where used for ID purposes)
- health
- sex life
- sexual orientation

From: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/

Legal bases for research with special categories data (2 of 10 options)

- Explicit consent of the data subject
  - To rely on explicit consent for special categories of personal data, the same basic requirements as those for consenting to the processing of regular personal data apply. Even in written context, not all consent will be explicit.

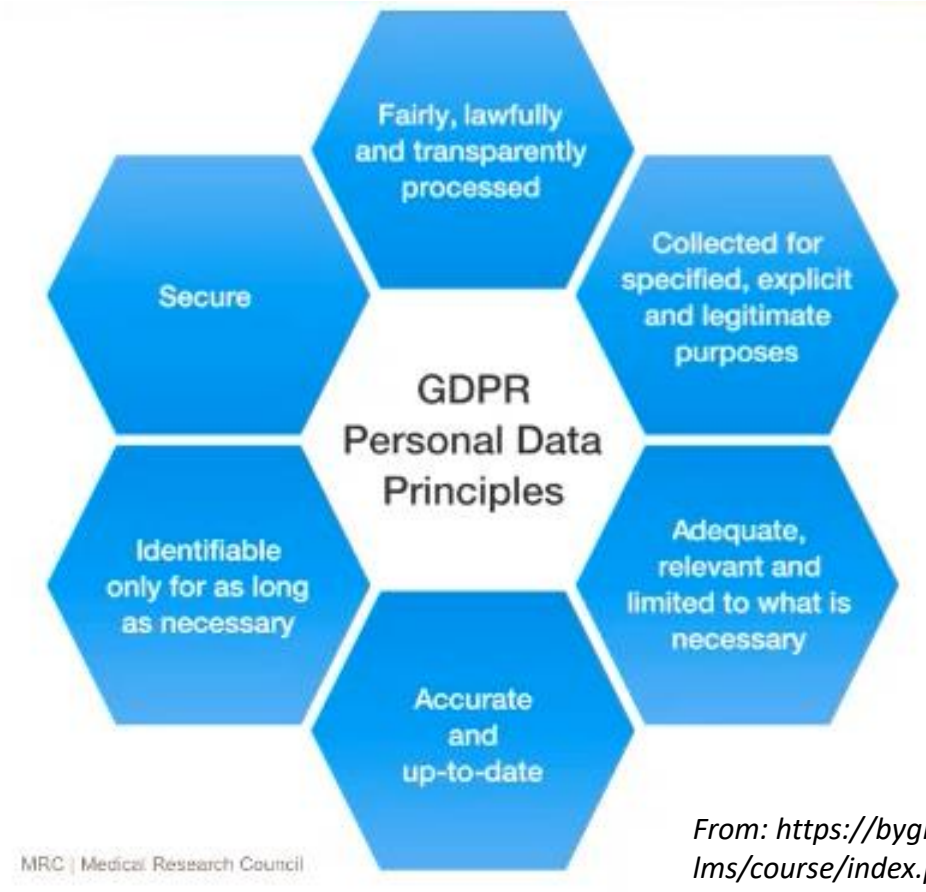
- Archive, statistical and research purposes
  - If at all possible, all personal data – both special categories and ordinary personal data – should be anonymised for archiving, research and statistics. If that is not possible, then data protection legislation allows the activities to be carried out under suitable safeguards.

From: https://www.ed.ac.uk/records-management/guidance/checklist/legal-basis/special-categories

The 6 Data Protection Principles

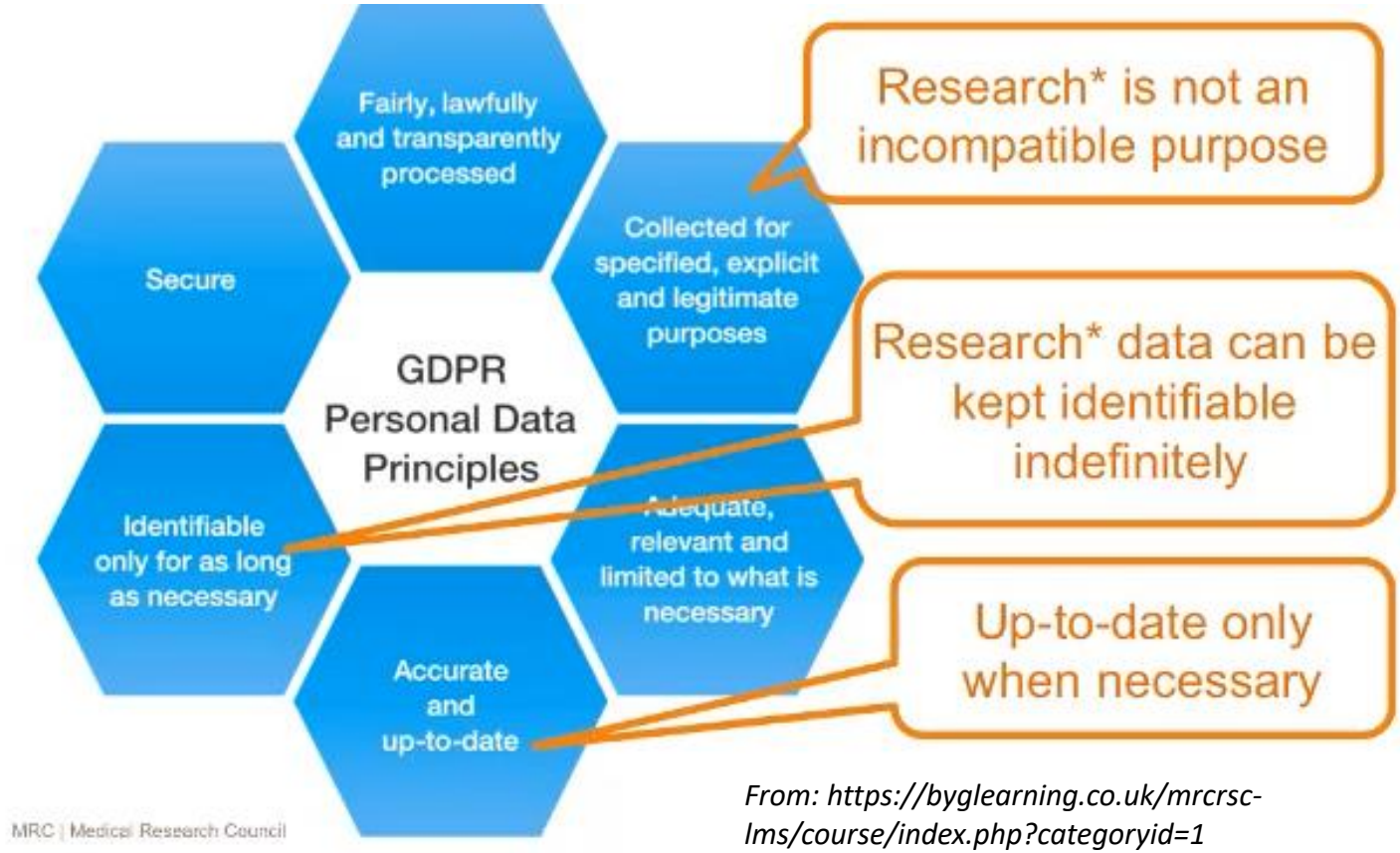
- 1. Lawfulness, fairness and transparency
- 2. Purpose limitation
  - Processing that's done for archiving purposes in the public interest or for scientific, historical or statistical purposes is given more freedom.
- 3. Data minimisation
- 4. Accuracy
  - Individuals have the right to request that inaccurate or incomplete data be erased or rectified within 30 days but such rights by research participants may be limited if it impairs the achievement of the research purpose AND if appropriate safeguards are in place.
- 5. Storage limitation
- 6. Integrity and confidentiality (security)

GDPR Principles Pictured



From: https://byglearning.co.uk/mrcrsc-lms/course/index.php?categoryid=1

GDPR Principles and Research



From: https://byglearning.co.uk/mrcrsc-lms/course/index.php?categoryid=1

GDPR Safeguards

SAFEGUARDS APPLIED IF HOLDING PERSONAL DATA TO SUPPORT RESEARCH

Pseudonymising

- Limit risk of disclosure
- Comes with own risk

Data Minimisation

- Adequate (sufficient to fulfill your purpose)
- Relevant to your research
- Limited to what you need:
  - Minimise number of participants
  - Minimise amount of data per participant
  - Minimise the degree of sensitivity

MRC | Medical Research Council

From: https://byglearning.co.uk/mrcrsc-lms/course/index.php?categoryid=1

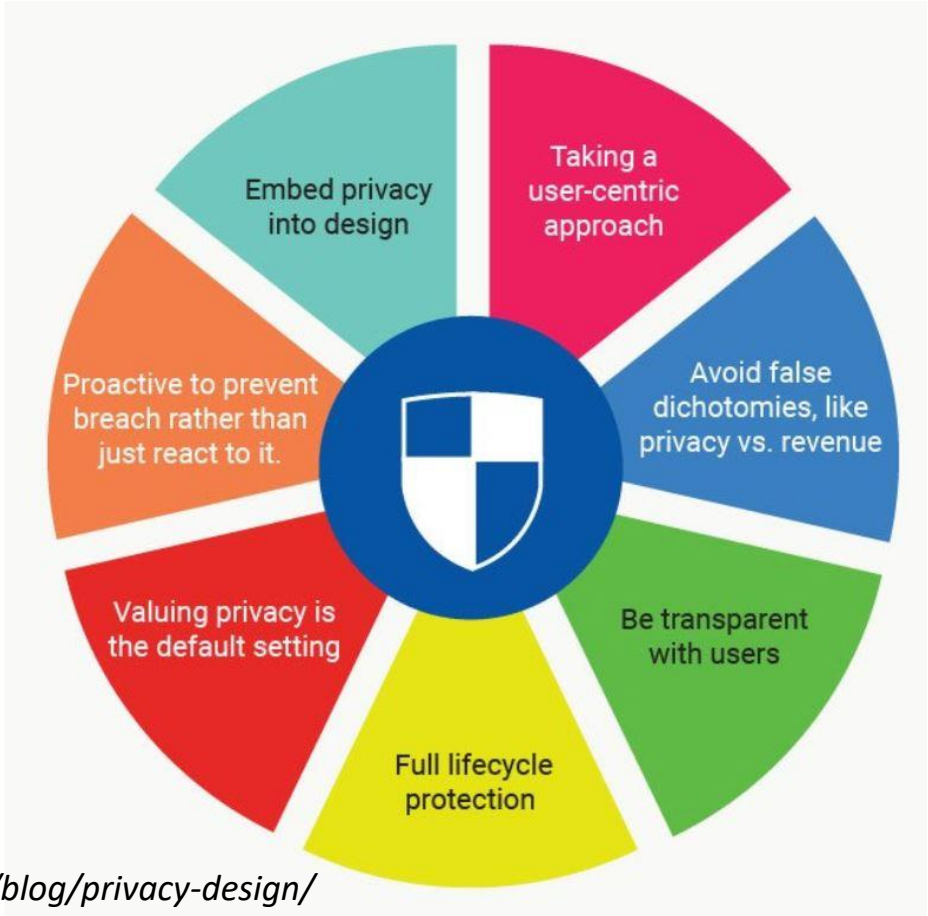
Safeguards for special categories data

- ✔ Implement additional technical and organisational measures, AND
- ✔ Cause no substantial damage or distress, AND
- ✔ No significant decisions being made about the participant, OR
  Research approved by an ethics committee

MRC | Medical Research Council

From: https://byglearning.co.uk/mrcrsc-lms/course/index.php?categoryid=1

GDPR: Privacy by design

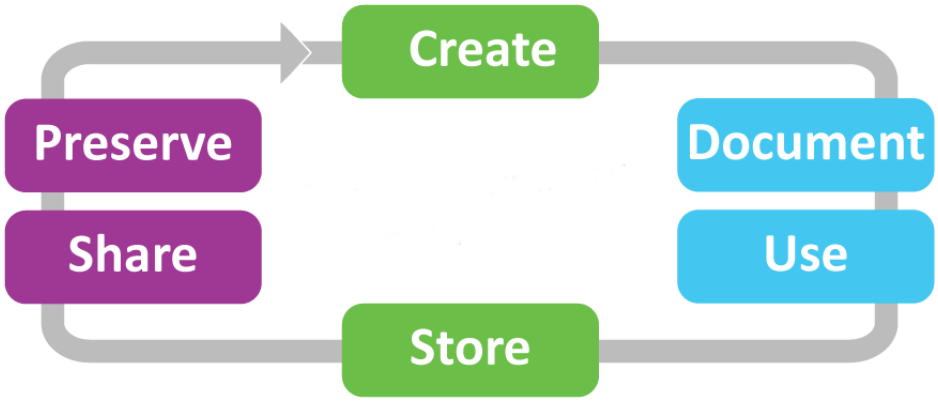


From: https://termsfeed.com/blog/privacy-design/

https://youtu.be/yhVqImna7cU (until 3:27)

VOX POP VIDEO: RESEARCH DATA MANAGEMENT

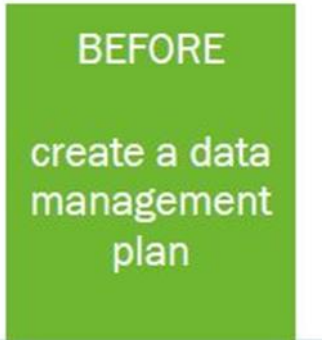
Research Data Lifecycle



MANTRA (handout)
FOSTER Open Science online courses
UK Data Archive 'Manage and Share Data'
Library
Your DPO
The CNPD – Comissão Nacional de Protecção de Dados
(the Portuguese Data Protection Authority)

TRAINING AND SUPPORT

RESEARCH DATA MANAGEMENT PLANNING



Before you start, plan: Data Management Plans

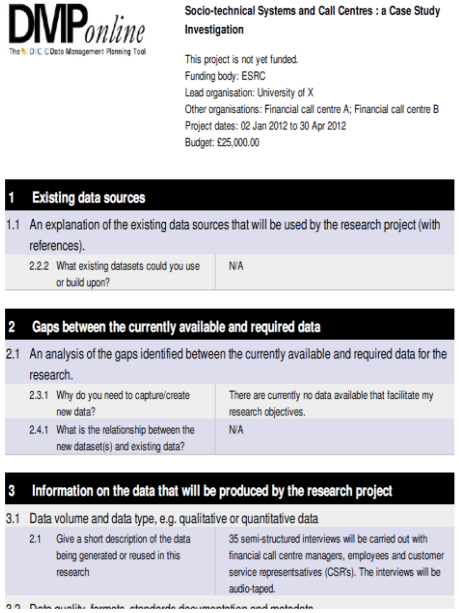
DMPs are written at the start of a project or postgraduate study:

to define how your research data will be:

- created & captured
- managed
- shared
- protected and preserved

They help researchers to properly manage their data for their use, meet funder needs & enable sharing.

The Digital Curation Centre offers a free service, DMPOnline: https://dmponline.dcc.ac.uk/



DMPonline
The DCC Data Management Planning Tool

Socio-technical Systems and Call Centres : a Case Study Investigation

This project is not yet funded.
Funding body: ESRC
Lead organisation: University of X
Other organisations: Financial call centre A; Financial call centre B
Project dates: 02 Jan 2012 to 30 Apr 2012
Budget: £25,000.00

1 Existing data sources	
1.1 An explanation of the existing data sources that will be used by the research project (with references).	
2.2.2 What existing datasets could you use or build upon?	N/A

2 Gaps between the currently available and required data	
2.1 An analysis of the gaps identified between the currently available and required data for the research.	
2.3.1 Why do you need to capture/create new data?	There are currently no data available that facilitate my research objectives.
2.4.1 What is the relationship between the new dataset(s) and existing data?	N/A

3 Information on the data that will be produced by the research project	
3.1 Data volume and data type, e.g. qualitative or quantitative data	
2.1 Give a short description of the data being generated or reused in this research	35 semi-structured interviews will be carried out with financial call centre managers, employees and customer service representsatives (CSR's). The interviews will be audio-taped.

Before you begin your research: DPIA

- Conduct a Data Protection Impact Assessment if your ethics committee requires it.
- A DPIA is a form of risk assessment in relation to data protection that will help ensure your compliance with the Law.
- It will assist you to evaluate protections and alternative processes to mitigate potential privacy risks.
  - UoE Guidance and template available: https://www.ed.ac.uk/records-management/guidance/research/data-protection-impact-assessment

DPIA: Mapping data flows

Understanding data flow

Where are you getting data from?

Are you linking it to other data?

Where are you keeping the data?

Are you sharing/transferring data?

Where / who is analysing data?

Where are you storing / archiving data?

etc…

MRC | Medical Research Council

WORKING WITH DATA

DURING

working with data

During your research project

- Informing research participants and gaining informed consent
- Storing your data securely
- Managing access and permissions to data
- Managing files and copies
- Using encryption as safeguard
- Reporting any breaches to the DPO

Consent documentation: what is included?

A written consent documentation includes:

- an information sheet and
- a consent form signed by the participant.

Information sheets

An information sheet should cover:

- Purpose of the research
- What's involved in participating
- Benefits and risks
- Terms for withdrawal
- Usage of the data
- Strategies for ethical use of data
- Details of the research

Consent forms

A consent form should cover the following points:

- The participant has read and understood information about the project.
- The participant has given the opportunity to ask questions.
- The participant voluntarily agrees to participate in the project.
- The participant understands that s/he can withdraw at any time without giving reasons without penalty.
- Procedures regarding confidentiality are explained.
- Separate terms for consent for interviews, audio, video or other forms of data collection.
- Separate terms of consent for using data in research, publications, sharing / archiving.
- Participant assigns the copyright I hold in any materials related to the project.
- Signatures and dates of signing for the participant and the researcher.

See UK Data Service Legal & ethical issues:: https://www.ukdataservice.ac.uk/manage-data/legal-ethical

Video

- Prof. Lynn Jamieson, a sociologist with the Centre for Research on Families and Relationships, talks about obtaining consent: http://edin.ac/1JrrkQ1

Activity: consent forms

Have a look at the consent forms you're given and answer the following questions:

- What are your initial impressions of the information sheet and each of the consent forms?
- Is there anything that's missing or anything that you feel is unnecessary?
- Is the information sufficient and appropriate for the processing likely to be done?
- What about data sharing and archiving?

Storage

- Use a centralised, managed storage option such as DataStore
- Understand how data are backed up
- Do not store data with personal identifiers on cloud systems, especially if data may be stored on servers outside the European Economic Area
- Security of data
  - Use appropriate safeguards for personal data

Security of data

- Storage
- Access
- Retention
- Disposal
- Breach

Security (safeguards)

- Pseudonymize/anonymize data before storing.
- Store identifiable data by stripping off identifiers.
- Store identifiers in a separate encrypted container.
- Encrypt identifiable data on portable devices, or encrypt the entire device.
- Give access to data only to authorized people.
- Keep identifiable data on a secure, backed-up central server and do not allow copies to proliferate.

Anonymisation & pseudonymisation

Anonymisation is the complete removal of any identifiers, which means irreversibly preventing the identification of the individual to whom the data relate. The individual will not even be identifiable anymore when linked with other information which is available or likely to be available.

Pseudonymisation is replacing any identifying characteristics of data with a pseudonym, such as replacing a name with something else – such as another name of the same culture for qualitative data, or a random case ID for other research. The difference to complete anonymisation is that there will be a key to re-identify the individuals, which will be kept separately.

Rena Gertz- UoE Data Protection Officer

Encryption

- Encryption is the process of converting data into an unreadable code. You must have access to a password or a secret encryption key to be able to read an encrypted file.
- Encryption comes in strengths. A higher key size takes exponentially longer to crack.
- For health data, NHS Information Governance Guidelines recommend using an encryption algorithm that supports a minimum key length of 256 bits, such as AES 256, 3DES, or Blowfish.

When should data be encrypted?

Personal data and special categories data in particular should be encrypted for the entire period it is held. This includes:

- **Data at rest:** Data held on storage media or computing devices should be protected to prevent unauthorized access. Protection measures should be applied to *ALL* copies of the data, including those held on back-up media.
- **Data in motion:** Data being transferred from location A to location B should be encrypted to prevent it being intercepted and accessed. This covers electronic transfer (e-mail, FTP, etc.) and physical transmission (e.g. sending a USB disk through the post).

Options for encrypting data

Several options are available to protect personal and sensitive data using encryption:

- **Encrypt a disk in its entirety:** Full Disk Encryption may be applied in order to protect all data held on the drive.
- **Encrypt one or more partitions on the disk:** Personal and sensitive data can be held on the encrypted partition, while the anonymized material can be held on the un-encrypted partition.
- **Create an encrypted container (archive):** A file that, when accessed using appropriate software, can be accessed and used in the same way as a physical drive.

There are software applications that offer encryption functionality at different degrees of granularity. These include Veracrypt, Microsoft Bitlocker, Apple FileVault, and 7-Zip, among others.

UoE Encryption advice:
https://www.ed.ac.uk/infosec/how-to-protect/encrypting

Encryption tutorials:
https://www.youtube.com/watch?v=y4Iosu-Yfsw&list=PLG87Imnep1SmnFGhAjFVHonQSVmMlpHkV

Accessing / transferring data

- Who is authorized to access data?
- Are you authorised to transfer data?
- Is the person who is receiving the data authorized to have it?
- Are you transferring the data securely?
  - Use secure channels e.g. SFTP.
  - Be cautious about cloud services e.g. Dropbox, Google Drive. If you must then use encryption.
  - Never send sensitive data over email unless encrypted.

Retention of data

Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes.

You must:

- Review the length of time you keep personal data.
- Consider the purpose or purposes you hold the information for in deciding whether (and for how long) to retain it.
- Securely delete information that is no longer needed for this purpose or these purposes.
- Update, archive or securely delete information if it goes out of date.
- *Remember*: You can keep non-personal data as long as you like (e.g. anonymised data), as this is outwith the data protection legislation.

Secure deletion

When you dispose of a computer or a laptop or any kind of device, you must ensure it is securely deleted.

Ordinary methods of deleting files,

- Moving to the wastebasket and then emptying the basket
- Using the DELETE command (on the command line)
- formatting the disk

DO NOT delete the data, they just mark the space where the files are as being "available for re-use“:

These methods work:

- Using a secure disk eraser like DBAN (Darik’s Boot and Nuke)
- Using whole disk encryption

See: http://www.ed.ac.uk/infosec/how-to-protect/secure-deletion

Breaches and incidents

Incidents are usually:

- A breach of one of the principles of the Data Protection Act and/or confidentiality law OR;
- Technology-related (cyber incidents).
- A breach can be caused by a cyber-incident but some cyber incidents may not involve a breach of information.

Examples of breaches and cyber incidents

Breaches	Cyber incidents
Identifiable data lost in transit	Phishing email
Lost or stolen hardware	Denial of service attack
Lost or stolen paperwork	Social media disclosure
Data disclosed in error	Website defacement
Data uploaded to website in error	Malicious damage to systems
Non-secure disposal – hardware	Cyber bullying
Non-secure disposal – paperwork	Other
Technical security failing	
Corruption or inability to recover data	
Unauthorised access or disclosure	
Other	

ACTIVITY: data protection breaches

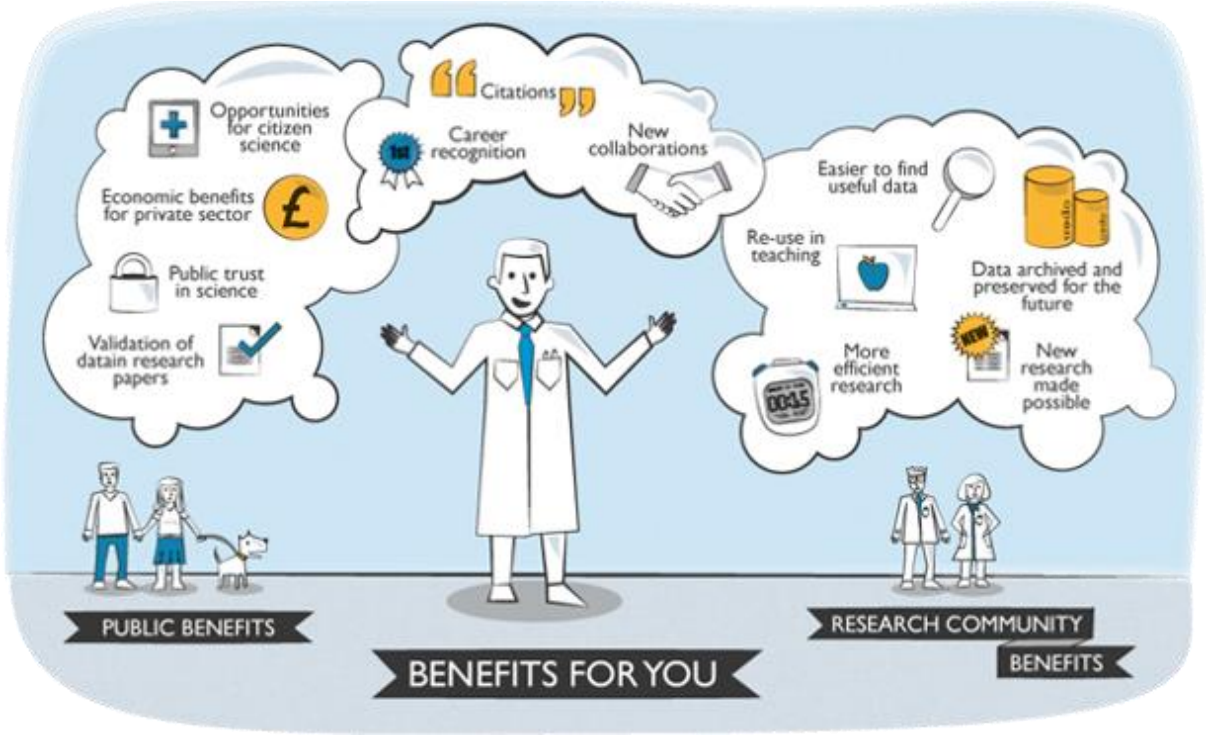
(see UKDA handout, Information Security handout)

AFTER

share and archive your data

SHARING AND PRESERVING DATA

Why share data?



From: Journal of Open Archaeology Data, CC-BY 3.0

https://youtu.be/yhVqImna7cU (from 3:27)

VOX POP VIDEO REDUX: RESEARCH DATA SHARING

How do GDPR and FAIR intersect?

"As open as possible, as closed as necessary"

- *Horizon 2020*

- Findable
- Accessible
- Interoperable
- Reusable

G

D

P

FAIR

Open data

- Data that contain no personal or disclosive information, e.g. anonymised, or attributed with explicit, freely given permission.
- Open data are usually licensed under an open licence such as a Creative Commons Licence and users do not need to register to access the data.
- By depositing such data with full documentation in an open access, trusted digital repository, it is more likely for researchers to meet the FAIR guidelines with their data.

Zenodo data repository

- Zenodo is OpenAire's open access multi-disciplinary data repository: https://zenodo.org/
- Assists researchers who want to share their data, get credit for data publication, and preserve their data for the long-term (DOI, licence, citation)
- It can help researchers comply with funder requirements to preserve and share their data and complies with the Horizon 2020 open by default requirements.
- Data should be fully anonymised before deposit unless subjects have given rights to publish their data.

What can a researcher do to be able to share?

- Plan for sharing (via a Data Management Plan).
- Don't collect personal information that's not needed (minimisation principle)
- Informed consent: get consent to share data or inform participants that their anonymised data will be shared and how.
- Attribute, anonymize, or aggregate individual's data.
- Log all your data processing (including documenting steps you take within your analysis package).
- Document all authorised data flows.

How to create an anonymised open dataset

Numeric data, eg. surveys	Qualitative data, eg. interviews
Remove names and identifiers	Share the edited transcript, not video or audio unless consented
Renumber and resort case ids	Agree a pseudonym with each subject
Group numbers into categories - banding	Remind subject not to disclose personal or sensitive information, eg. about family members
Top and bottom code numbers (age, salaries)	Replace proper nouns in text (names, placenames etc.) using square brackets, don't blank out
Use standard codes (eg. SOC, SIC) and geographic boundaries at appropriate levels; not fine-grained	Avoid over-anonymising or data will lose value
Check for low cell counts in cross-tabs	Keep a log of all replacements, generalisations or removals made; store separately from anonymised data

Keep an anonymisation log

example anonymisation log:

Interview and page number	Original	Changed to
Int1		
p1	Age 27	Age range 20-30
p1	Spain	European country
p3	Manchester	Northern metropolitan city or English provincial city
p2	20th June	June
p2	Amy (real name)	Moira (pseudonym)
Int2		
p1	Francis	my friend
p8	Station Road primary school	a primary school
p10	Head Buyer, Produce, Sainsburys	Senior Executive with leading supermarket chain

https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/qualitative

When open data access is not possible…

- Because potential harm to research subjects is too great.
  - Information that can be used to discriminate requires extra protection.
- Not permitted by the data producer, funder, health authority etc.
  - Sometimes precautions are required even for anonymized data.
- Because anonymization is either not feasible or would destroy value of dataset.
  - Population too small to be anonymous, e.g. those with genetic condition.

… then restrict access

- Restrict access by publishing a statement about Data Access. All of your raw data has a risk of disclosure so would need to be made available under safeguarded conditions.
- Users will have to agree to certain conditions, such as not to disseminate any identifying or confidential information on individuals, households or organisations, and not to use the data to attempt to obtain information relating specifically to an identifiable individual.
- Safeguarded data may have additional conditions such as requiring data owner permission or prohibiting commercial use.

Use ‘the 5 Safes’ to restrict access

- 5 Safes:
  - Safe data: public use dataset (fully anonymised)
  - Safe projects: approved research project; in public interest
  - Safe people: accredited training process
  - Safe settings: use of secure access facility or remote secure access
  - Safe outputs: outputs checked by responsible staff-person and data linkage may be offered as part of the service
- For example, the UK Data Service Secure Lab provides Approved Researchers with controlled access to sensitive or confidential data, enabling researchers to access and use datasets in a secure and responsible way: *https://www.ukdataservice.ac.uk/use-data/secure-lab/security-philosophy*

Resources for further information

- **Comissão Nacional de Protecção de Dados: https://www.cnpd.pt**
- Guide to the General Data Protection Regulation, ICO:
- https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr
- Research and the General Data Protection Regulation, University of Edinburgh DPO: https://www.ed.ac.uk/records-management/guidance/research/data-protection
- Association of Internet Researchers 2012 report, Ethical Decision-Making and Internet Research: https://aoir.org/ethics
- Consent for sharing, UK Data Service: https://www.ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing/consent-forms
- Recommended informed consent language for data sharing, ICPSR: https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/confidentiality/conf-language.html