Responsible Data Management: A Trainer's Toolkit

irct
International Rehabilitation Council for Torture Victims

Ministry of Foreign Affairs of the Netherlands

Contents

About this toolkit

This document has been developed as a part of the IRCT's Global Anti-Torture Evidence (GATE) Project, generously funded by the Ministry of Foreign Affairs of the Netherlands. This toolkit is intended to be used as a reference for a professional working in a torture rehabilitation centre to train others in responsible data management. It provides an introduction to, and practical exercises in, the area of responsible and ethical data management in an anti-torture and human rights context. It can be used as a stand-alone tool, or in conjunction with other resources on responsible data management.

This document was authored by Carrie Gaston.

1. Introduction

Information is a core part of any organisation and one of its most valuable assets. Information governance and responsible data handling techniques provide a framework for the handling of that information. In particular, the handling of ***person-identifiable*** and ***confidential*** information in a secure, confidential and *mindful* manner.

Everyone who works for or on behalf of an organisation should be made aware of:

- The importance of the information held that may be confidential or sensitive and relate to the users of your services, your staff, volunteers, donors/funders or anyone else associated with your organisation.
- The relevant legislation in the countries in which you operate, as well as relevant guidance and best practices for looking after such important information.
- Why YOU must take responsibility for how you obtain, record, use, keep and share information.
- The impact responsible data management has on business continuity and the ability to continue to provide a safe and reliable service to those you support.

Responsible data management is everyone's responsibility!

2. Data: Identifying different types of data

In this section, you will teach your participants different types and categories of data, how to identify these, and the risks and safeguarding associated with each.

Learning objectives – by the end of this section, your participants will:

- *Be able to identify different categories of data.*
- *Understand the possible risks associated with the different categories of data.*
- *Think about how to apply safeguards to different types of data in their own contexts.*
- *Have an understanding of anonymization of personal data.*

Types of data

In every organisational context, but more particularly in any type of health and social care organisation, we come into contact with various types of personal information about people.

It is important to be able to identify these different types of information so that they can be appropriately protected when they are used and shared.

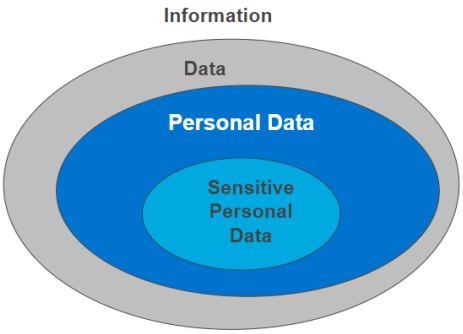
Information is any set of facts provided or learned about something.

Data is a set of values of qualitative or quantitative variables.

Personal data (pd) is data related to a living individual who can be identified:

- from that data,
- from that data and other information, which is in the possession of the person or organisation ("data controller" in data protection terminology).

Sensitive personal data (spd) is a special category of personal data that relates to racial or ethnic origin, political opinions, religious or similar beliefs, trade union membership, physical or mental health condition, sexual life, commission or alleged commission of any offence.



Confidential information is information where there are sensitivities around the handling and disclosure of the information. They may be either personal or organisational in nature. *Sensitive personal data* is usually always to be treated as confidential, however not all confidential information is of a sensitive or personal nature. For example, business critical information may fall into this category.

As well as thinking about how you and your organisation can *protect* data from a breach external to your organisation, it is also worth thinking about how to implement good practice in the confidential handling of data both externally but also *within* your organisation. This may mean that alongside this training, you think about implementing organisational policies and procedures that give your staff guidance on the good practice handling of anything that may be deemed confidential information. This could mean developing a Confidentiality Policy or Code of Conduct for staff.

Ask your participants if they can think of any types of information in their own contexts which may fall into this category. Ask the participants to list this information on a piece of paper and to think about the safeguards they may have in place in order to both identify and protect this category of information. Once they have finished, discuss the examples your participants wrote down.

The following is an exercise in data categorisation. The exercise should spark a discussion on personal and sensitive datasets, so it solidifies your participants' understanding of the various definitions. Wherever possible, repeat the exercise with actual physical versions of various datasets that are in use in the local context. By the end of the exercise your participants should be able to:

-differentiate between personal data, sensitive personal data and neither personal nor sensitive personal data.

Exercise 1: Understanding different types of data

Here are some examples of different sets of data and information. Ask your participants to place them in the appropriate category of sensitive, personal, neither. The given list itself is intentionally vague so that your participants are encouraged to ask appropriate questions about the given data, so they feel they have all the information required to label the item personal or sensitive. Tell them to think about whether the items that are neither personal or sensitive should still be treated as 'confidential' in nature. Also have your participants think about whether they might need any further information in order to determine the appropriate category and make a note of your questions.

Sensitive personal data	**Personal data**	**Data (neither pd or spd)**

Credit card details of a list of recent donors	a list containing the mental health information of clinic patients	names and addresses of customers

document containing the top 10 languages of your customer base, and the number of speakers of each language	list of clients and their political affiliation, using no names but identifier numbers	aggregate demographics data of all clients who attended the centre in the last year
a list of all 350 clients who attended in the last year, and their ethnicity and sexual orientation	anonymous survey results	a list of email addresses of clients who attend a Friday group
outcomes information on a population of clients e.g. their scores and changes in scores on a standard mental health measure	photocopies of individuals passports	information on the named police stations and detention centres that clients have told you about where they have been held
statistics on the top 5 different methods of torture and inhumane treatment your client base has told you about since your centre opened 5 years ago	a list of all of the religions your clients affiliate with	an internal report that contains sensitive business intelligence

Why is protecting personal information important? It is important to comply with legislation and best practices to protect *personal information*, because personal and sensitive information is valuable. Poor data handling and protection can cause personal, social and reputational damage. In our own context of torture rehabilitation, the risks may be even greater and involve the personal safety of the individuals who access our services.

Common ways information is lost:

- Losing information (including paper records) over the phone, via faxes, loss of computers or mobile devices.
- Theft of information, including via phishing attacks (see Glossary).
- Insecure storage and disposal of information leading to loss or theft.

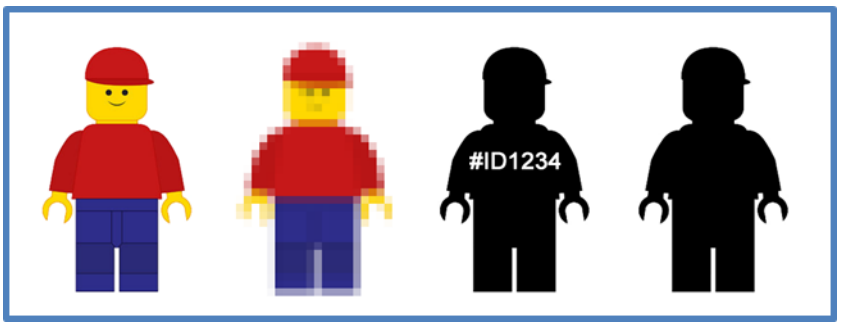
Human Error more damaging than cyber attacks

From October to December 2017, human error accounted for almost two-thirds of the incidents reported to the UK's Information Commissioner's Office (ICO) – the independent body set up to uphold information rights. Human error caused more loss or damage than insecure webpages and hacking, which only stands at 9% combined. Despite this, market attention and resources continue to focus on external threats, in particular cyber-attacks and hackers.

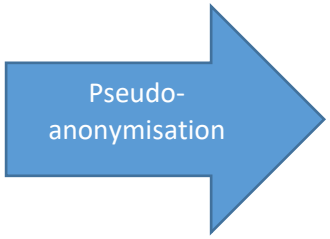
- Categorisation by the ICO of the types of breaches caused by human error reveals the major causes as:
  - data emailed to the wrong recipient (15.8%).
  - loss and theft of paperwork (13.1%).
  - data posted or faxed to the wrong recipient (13.0%).

- Other causes included insecure disposal of hardware and paperwork, loss or theft of unencrypted devices, and failure to redact data.

Taking the 'personal' out of 'personal data': Anonymisation & Pseudo-anonymisation



Pseudo-anonymisation is a procedure by which the most identifying fields within a data record are replaced by one or more artificial identifiers, or pseudonyms. There can be a single pseudonym for a collection of replaced fields or a pseudonym per replaced field. This helps to conceal an individuals' real-world identity ***but*** is not true anonymisation because the identity can be easily discovered given the key for the coding used.

Name: Alisha Santos Date of birth: 21 April 1980 Education: B.A. Psychology, University of Guelph Place of birth: Niagara Falls, ON Canada Current employment: Professor of Mathematics at Cambridge University Marital Status: married to Dean O'Donnell Number of children: 3		ID#357986 Date of birth: 1980-04-21 Education (highest level): 4 Place of birth: Niagara Falls, ON Canada Current employment: Professor Marital status: Married Children: y

Anonymization is the process of turning data into a form which does not identify individuals and where identification is not likely to take place. This allows for a much wider use of the information.

"We use the term 'Anonymised data' to refer to data that does not itself identify any individual, and that is unlikely to allow any individual to be identified through its combination with other data" (ICO anonymization code of practice, pg. 6)

Name: Alisha Santos Date of birth: 21 April 1980 Education: B.A. Psychology, University of Guelph Place of birth: Niagara Falls, ON Canada Current employment: Professor of mathematics at Cambridge University Marital Status: married to Dean O'Donnell Number of children: 3		ID#: 357986 Age group: 30-40 Education (highest level): Degree Place of birth: Southern Ontario Canada Current employment: teacher Marital status: married Children: yes

A quick word about aggregation….

Data aggregation is any process in which information is gathered and expressed in a summary form, for purposes such as statistical analysis. As long as your data is not ***personally-identifiable***, it is no longer subject to the same legislative protections, although it may well continue to be business sensitive, and therefore still in need of safeguarding (also see *confidential information* above).

The following exercise will help your participants think about how their own data in their own contexts relates to the data categories discussed above. Part (b) of the exercise will help your participants think about how they may safeguard certain categories of data, for example by using anonymisation techniques. By the end of the exercise your participants should be able to:

-list datasets in their own context.

-identify those datasets as personal data or not.

-have a discussion around which data their organisation may hold that is not personal data but may still need to be treated as confidential.

-start to think about applying safeguards to different types of data.

-think about how to apply anonymization to personal data prior to sharing.

-possibly start to think about data sharing arrangements.

Exercise 2: Understanding different types of data in your own context

(a) List the common data sets that you process (collect, handle, report, store, etc.) and try and categorise them into the different categories discussed above (personal data (pd), sensitive personal data (spd), neither). Are any of those data sets potentially 'confidential', but do not fit into the 'personal' category? Are the risks associated with this data different? Would you apply the same safeguards to this type of data?

(b) Think about a dataset you work with (for example, a client list) that may need to be shared. Can you think of how you might safeguard this information as much as possible before sharing it?

3. Data Protection in Legislation (EU)

In this section, you will introduce the data protection legislative framework. This newly updated legislation is only applicable across the European Union but is helpful to learn about as the gold standard in data protection in a number of ways. It is also helpful as a tool to learn good practice, even if the legislation does not directly apply in your particular area of the world.

Learning objectives – by the end of this section, your participants will:

- *Have an understanding of the current legislation governing data protection across the European Union.*
- *Have an understanding of the individual rights of data subjects and the responsibilities of data controllers and processors.*
- *Think about how to apply these rights and responsibilities to their own data in their own contexts.*
- *Understand how to complete a data mapping exercise to identify what data is held and where.*
- *Understand what a Subject Access Request (SAR) is and how to comply with one, including redactions.*

The European Legislative Framework: An introduction to the General Data Protection Regulation (GDPR)

The EU ***General Data Protection Regulation*** (GDPR) is new legislation that provides a single data privacy law for the European Union. It builds on and strengthens existing Data Protection legislation in the following key areas:

Definitions:

- A "Data Subject" is any person whose personal data is processed by your organisation.
- A "Data Controller" is the entity that determines the purposes, conditions and means of the processing of personal data.
- A "Data Processor" is the entity that processes data on behalf of the Data Controller.

➔ ***Individual rights*** – under the GDPR, the rights of the data subject are strengthened or enhanced in a number of areas. These include:

- **Right to be informed** – data subjects have the right to know who is doing what with their data.
- **Right of access** – data subjects have the right to access the personal data you hold on them – this includes giving them a copy of their data and supplying this free of charge and within a reasonable time frame (*also see below on Subject Access Requests*).
- **Right to rectification** – data subjects can demand changes to the data you hold on them where they deem it to be false, out of date, or incomplete.
- **Right to erasure** – data subjects now have the right to demand that their information is erased – this is also known as the "right to be forgotten".
- **Right to restrict processing** – data subjects can request the blocking or suppression of processing of their personal data. In these circumstances, it may be required that data processors continue to store data relating to the data subject in order to support this. (An example of this might be where a donor or supporter has requested an organisation no longer contact them asking for monetary support. The data then held would be the minimum value to support the request to no longer be contacted).
- **Right to data portability** – allowing data subjects to obtain, move, reuse data across services or for their own purposes.
- **Right to object** – data subjects have the right to object to processing of their data including for marketing or profiling.
- **Rights related to automated decision making, including profiling** – there are specific requirements in order to be compatible with the legislation regarding automated decision making.

➔ The GDPR also strengthens or enhances the ***accountability of data controllers***, where they are expected to put into place comprehensive governance measures, and to promote accountability and transparency.

➔ This also includes wider reaching obligations on ensuring ***compliance of processors***, including contractors.

➔ It also includes ***obligations regarding reporting data breaches***, and for having a responsible person (Data Protection Officer) in large organisations, public bodies and those who do large-scale processing of personal data.

➔ Data protection by design/default – thinking about implementing measures to protect data in the design of any new system before data collection takes place.

➔ Privacy impact assessments – due consideration being given to any possible impacts on individuals privacy of all processing activities.

A word about Subject Access Requests….

Under current and proposed data protection legislation any *data subject* has the right to request any personal data about them held by any organisation. This means anyone can request a copy or to view the information a data processor or controller holds on them, and compliance with the request cannot be refused. In complying with any such request, it is important to uphold the principles (individual rights) of data subjects – which may be redacting some information where that information is gathered from 3rd party sources but forms a part of the record you hold.

Ask your participants to think about how they would comply with a subject access request from someone who uses their services. What might they need to consider in complying with such a request? Have them think of examples where a record may contain information that requires redacting prior to complying with a request? Discuss these as a group.

Data mapping

In order to properly protect data, and comply with the points above, you first of all need to know what data you hold, and where. It is a good idea to complete a data mapping exercise, so that you and those in your organisation are clear on what data is held, where it is stored, how long it is kept, and when and how it is destroyed.

Things to consider….

WHAT personal data do you have? In order to answer this question, you will need to do a full review and audit of ALL the personal data you collect, for e.g.:

- Do you have information written on scraps of paper in a drawer in your desk?
- Do you have personal information written in a paper diary?
- Are you able to comply with a Subject Access Request (SAR)?

WHERE is it stored?

- In a database?
- On shared/network drives? Who has access?
- In paper diaries? On scraps of paper in your desk?
- At home? In emails?

WHERE is it sent?

- Can you confirm you only send client information within your own country?
- Do you have explicit, written/documented permission to send the data? Can you evidence this?

HOW is it processed?

- On computers?
- Pieces of paper?
- How are you protecting any data in transit?

WHAT do you tell people/the data subjects about processing?

- Do you have freely available information for data subjects?
- Identify 3rd party processors?
- Does the data cross any national/international boarders?

Clinical Information Assurance: Is it the right information? Accurate? Do you audit and update regularly? Is there evidence?

The following exercise will help your participants think holistically about what data they collect and hold, and how and where it is stored. This will help them to both know what data may need protecting, but also what sort of safeguards may need to be put into place regarding different types of data. It will also highlight where there might be gaps in appropriate data management. By the end of the exercise your participants should:

-have an understanding of the various locations and forms their data takes.

-think about where and how it is stored, and who has access.

-think about how the above points are critical in the right to erasure or in responding to a subject access request.

-start to think about retention schedules.

Exercise 3: Data mapping

Choose a dataset relevant to your centre (e.g. client information) and complete the following data mapping exercise by answering the following questions:

- *What data is collected?*
- *Is consent gained at the point of collection?*
- *Where is the information stored? In a database, paper file, computer file?*

- *What is the data used for?*
- *Who can access it?*
- *Is it ever shared externally?*
- *How/when is the data reviewed/added to/updated?*
- *How long is it kept for?*
- *When and how is it deleted?*

Learning from others: Data breaches in the news…

National Alzheimer's charity found to have serious failings in the way it handled sensitive personal data, including discovering volunteers were using personal email addresses to receive and share information about people who use the charity, storing unencrypted data on their home computers and failing to keep paper records locked away. (ICO, 2016-01-07).

Health clinic fined £180k,000 (204,000 EUR, 253,000 USD) for data breach when it accidentally sent a newsletter with recipients' email addresses in the "To" field, instead of the "BCC" field, and thereby effectively revealing the HIV status of recipients. (ICO, 2016-05-09).

USB stick found in West London contained airport security data (Register, 2017-10-30).

Personal details of adopted children, parents and social workers accidently emailed to party invitees (Chronical Live, 2017-12-26).
Filing cabinet full of confidential government papers ends up at second hand store. (Guardian, 2018-02-02).

Confidentiality Dos and Don'ts

Dos

- **Do** safeguard the confidentiality of all person-identifiable or confidential information that you come into contact with.
- **Do** be aware that any recorded information about an individual should be protected – this includes notes and diaries.
- **Do** clear your desk at the end of each day, keeping all portable records containing person-identifiable or confidential information in recognised filing and storage places that are locked at times when access is not directly controlled or supervised.

- **Do** switch off computers with access to person-identifiable or business confidential information, or put them into a password-protected mode, if you leave your desk for any length of time.
- **Do** ensure that you cannot be overheard when discussing confidential matters.
- **Do** challenge and verify where necessary the identity of any person who is making a request for person-identifiable or confidential information and ensure they have a need to know.
- **Do** share only the minimum information necessary.
- **Do** take care when sending fax or email correspondence and where applicable retain a delivery/read receipt.
- **Do** transfer person-identifiable or confidential information securely - e.g. by using email encryption.
- **Do** seek advice if you need to share person-identifiable information without the identifiable person's consent and record the decision and any action taken.
- **Do** report any actual or suspected breaches of confidentiality.
- **Do** participate in induction, training and awareness raising sessions on confidentiality issues.

Don'ts

- **Don't** share passwords or leave them lying around for others to see.
- **Don't** share information without the consent of the person to which the information relates, unless there are statutory grounds to do so.
- **Don't** use person-identifiable information unless absolutely necessary; anonymise the information where possible.
- **Don't** collect, hold or process more information than you need, and do not keep it for longer than necessary.
- **Don't** think that comments or notes you make will only be for your eyes only; individuals have the right to access information kept about them by making a Subject Access Request (SAR).
- **Don't** leave information unattended on your desk.
- **NEVER** leave files or information in the car, on the bus or when working from home, ensure that information is not accessible to anyone other than YOU.

The following exercise will help your participants think critically about current practices in their own organisation by using some common data protection dilemmas. It should get them thinking about all the ways in which they and their colleagues handle, store and transfer sensitive information. They should consider different working practices and the risk to data that may be associated with those. They should consider possible safeguards in order to minimise risk to data. You may wish to ask participants to come up with some scenarios

themselves, directly relating to their own organisational or practice. By the end of the exercise your participants should:

-be able to identify potential risks associated with their particular organizational practices.

-be able to identify safeguards to minimise the risks they identified above.

Exercise 4: Things to think about

Think about what you might do in the event of the following scenarios:

You are in a rush to leave the centre. You want your notes and documents to get started on the MLR, which you plan to do on your laptop on the long train journey home.

Consider: What are the possible risks to data protection?

How can you minimise them?

1. *You are working at home one evening on an MLR draft which you want reviewed the next day. How can you minimise the risks to data protection?*
2. *You find some paper copies of client sensitive materials at home. How are you going to deal with this?*
3. *You realise you have accidentally emailed some client information to a colleague's home email address. What are you going to do? Who will you inform of the error?*

The following exercise will help your participants understand how to comply with a Subject Access Request under the Data Protection legislative framework outlined above. Participants will need to relate back to the earlier data mapping exercise (Exercise 3) in order to locate all of the information they may have on an individual, whether in hard-copy or electronic format. They should also consider if any of the information they hold contains 3rd party information, and how they might go about redacting this in their response to a SAR. They should also think through the possible consequences for the individual, and ensure they are as clear as possible about this (e.g. sending sensitive health data via unsecured email, etc.). By the end of the exercise your participants should know:
-what steps to carry out in order to deal with a SAR.

Exercise 5: SAR compliance

You receive an email from someone claiming they are the solicitor of a client of the centre, and they are requesting a copy of the records you hold on them as a subject access request. How do you comply with the request? What do you need to consider in order to comply appropriately and in good time?

The following exercise is to get your participants to think about what constitutes a data protection breach or breach of confidentiality, and to try and see what value they may place on the severity of such breaches. In making their value judgement, they should provide you with an explanation as to how they decided one incident was more important or severe than another. You should also get them think about their own context and add their own incidents to the list. When breaches happen, how will they ensure that such incidents provide learning (and aren't repeated? By the end of the exercise your participants should be able to:

-identify what constitutes a data breach.

-identify different ranges of severity of different data breaches.

Exercise 6: Rank the following scenarios in order of severity, from greatest to least

- *You find copies of documents left on a photocopier in the office – these include letters and a list of a person's mental and physical health concerns.*
- *A colleague tells you they emailed a client's information including personal sensitive health data to the wrong email address.*
- *A colleague tells you they were travelling with client data and they accidentally left it on the bus/train/somewhere they can't remember.*
- *A very concerned client comes to you and tells you that their information, and the information of a number of other clients, has been found online. This information includes sensitive health data.*
- *You discover your office was broken into overnight. One of the filing cabinets that contains client information has been broken open and files are missing.*
- *You discover a colleague has emailed a client's full medical report to a number of external people to be used as a training exercise, without the written consent of the client.*
- *An administrator accidentally sent client information (names, demographics) to a solicitor's firm where papers accidentally were posted out along with a completed medico-legal report.*
- *A medico-legal report containing highly sensitive personal data has been distributed to training delegates without proper anonymization.*

- *An administrator gave out information about a client over the phone without the necessary checks in place as to whether the individual had consent for the information.*

4. Responsible data management: The bigger picture beyond just data protection

So far, you have discussed data protection, and the types of data that are relevant under certain legislative frameworks. In this section you will introduce Responsible Data Management to your participants – which covers a wider set of thoughts, behaviours, considerations that apply to the whole data life cycle, and not just in the specific context of personally-identifiable datasets.

Learning objectives – by the end of this section, your participants will:

- *Be introduced to the concept of 'responsible data'.*
- *Develop an understanding of the wider implications of responsible and ethical data in overall data management.*
- *Explore the possible consequences of not practicing responsible data management.*

What is responsible data?

The collective duty to account for unintended consequences of working with data by:

1) Prioritising people's rights to consent, privacy, security and ownership when using data in social change and advocacy efforts,

2) Implementing values and practices of transparency and openness.

What is responsible data management?

Responsible data management is about treating the data that we collect with respect and upholding the rights of people whose data we collect.

It is being responsible and mindful of impacts on people in all aspects of data management including **collection, handling, storage** and **use**.

Being responsible with other peoples' data from the point of collection to report publication.

Responsible Data considerations

Key elements of practising responsible data include:

Power dynamics: The least powerful actors in any situation are often the first to see unintended consequences of data collected about them. Processes like co-designing or ensuring that people from diverse backgrounds are involved in data collection or analysis processes can mitigate against this.

For example, in humanitarian crises, the people from whom data is being collected hold far less power than those who are asking for their data. How could that power asymmetry affect their willingness to give their data over?

Diversity and bias: Considering questions like, "who makes the decisions? What perspectives are missing? How can we include a diversity of thought and approach?" can highlight blind spots, and areas where adding additional voices would be valuable.

We believe diversity of all kinds strengthens our projects and our approach. We've seen projects, products and organisations suffer from homogenous staff or communities – and often, the negative effects of data are first seen and experienced by marginalised communities. We need to include those voices and provide ways to improve as a result.

Unknown unknowns: We can't see into the future, but we can build in checks and balances to alert us if something unexpected is happening.

Often, "But we didn't know" is the first thing heard when there are unintended negative consequences of a data-related project. It's our responsibility to think about how we can build in proxies for particularly important or impactful unintended consequences.

Precautionary principle: Just because we can use data in a certain way, doesn't necessarily mean we should. If we can't sufficiently evaluate the risk and understand the harms when handling data, then perhaps we should pause for a minute and re-evaluate what we're doing and why.

Technology offers us all sorts of possibilities. Not all of these are smart, and not all of these will have good effects on the world. If we're working in social change, our priority is to respect and protect people's rights – and that requires us to be thoughtful about our own actions.

Thoughtful innovation: For new ideas to have the best possible chance of succeeding – and for everyone to benefit from those new ideas and projects – innovation needs to be approached with care and thought, not just speed.

Innovation is about finding better and more effective solutions to better meet needs. To do that, we must first take the time to think about what those needs are – perhaps through research, perhaps in other ways. Then we must think about what possible solutions could meet those needs and, crucially, would have positive impacts (without unintended negative side-effects) on the people we're trying to support in the long term.

Holding ourselves to higher standards: In many cases, legal and regulatory frameworks have not yet caught up to the real-world effects of data and technology. How can we push ourselves to have higher standards and to lead by example?

Working in social change and advocacy means we hold ourselves to a certain set of ideals. Profit isn't our goal – positive social change is. In many areas of the world, regulatory frameworks have loopholes that allow projects that, if we think about them again, might be considered exploitative. Different countries have very different standards of legal protections for privacy – like the strongly rights-protecting upcoming General Data Protection Regulation across the European Union.

Building better behaviours: There is no one-size-fits-all for responsible data. Existing culture, context and behaviours change the implications and ways in which data is used.

Responsible data is not a prescriptive practice – sadly there aren't any checklists to go through and then be considered 'responsible'. Much of this is about building better-informed approaches to working with data – which might include regularly reviewing the decisions that were made, given new information. Practising responsible data isn't just a task for those who directly handle data – it's an operational issue that everyone, from leadership to staff, needs to be thinking about.

If it was you, and your data, how would you like it to be treated/respected/kept safe?

Confidentiality – Access to data shall be confined to those with appropriate authority.

Integrity – Information shall be complete and accurate. All systems, and assets shall operate as expected.

Availability – Information shall be available and delivered to the right person, at the right time, when it is needed.

Ask your participants to consider the following scenario, which serves to highlight issues around both data integrity *and of* confidentiality*. By the end of the exercise your participants should be able to:* *-identify these issues of data integrity and confidentiality.*

-anticipate what possible consequences there might be for the client.

Exercise 7: Data integrity and confidentiality

Scenario

JP is a client of the centre and is attending regular appointments for psychological therapy in relation to torture he received when detained by state police.

Because of a data entry error, an administrator mistakenly calls his work number rather than his personal number and as JP is in a meeting, once of his colleagues picks up the phone. Thinking JP has answered, the administrator goes on to ask him if he can re-arrange his appointment to a later time.
As the administrator discloses where the call is coming from, it is immediately apparent to JP's colleague that he is receiving therapy from the centre, and he goes on to further disclose this information to his other colleagues.
Questions:

- *What lessons can be learnt from the above scenario?*
- *What are the possible consequences for JP?*

Hints:

- *The work number is where the personal number should be.*
- *Confidentiality is breached when the administrator tells someone other than JP his personal information.*

5. The Data Life Cycle

In this section, you will introduce the data lifecycle to your participants and get them to think about how responsible data considerations fit into each step. You will ask your participants to go through some practical exercises in order to practice responsible data management techniques. Participants should think about the different types of risks and threats to be considered in different stages of the data management lifecycle.

Learning objectives – by the end of this section, users will:

- *Understand the data lifecycle, and how responsible data considerations fit into each step.*
- *Understand how to assess risk when it comes to data management and complete a risk assessment looking at possible safeguards.*
- *Understand how to give consideration to people's privacy when considering data management and what a privacy impact assessment is.*

Understanding your data life cycle

The data lifecycle is the sequence of stages that a particular unit of data goes through from its initial generation or capture to its eventual archival and or deletion. There are often 6 or more stages identified in the typical data lifecycle, covering the **collection**, **handling**, **storage** and **use** of data.

I. Collection

1. Planning – try to think about what outcomes you are trying to achieve and what types of information you will need to collect in order to get there.
2. Assessing Risk – are you asking the question because it will enhance your data set, or is there another reason?
3. Training staff and other data collectors – ensuring staff understand and apply responsible data handling techniques.
4. Consent – it is considered a gold standard of good practice to ensure proper informed consent when obtaining and using personal data.

II. Handling & III. Storage

5. Management – How will you collect, store, use, share? How will you ensure your data is accurate? Do you need to put into place measures to regularly review, cleanse, purge data?

IV. Use

6. Use – how will you use what has been collected?
7. Feedback – a key to ethical data management is to ensure those who have trusted to give you their personal data are given feedback wherever possible on the good things that have been achieved using their data.
8. Retention & destruction – how long do I need to keep my data for? Do I need to keep all data, or only retain some? Can I anonymise it?

I. Collection

Planning: Risk assessments, privacy impact assessments, informed consent, training.

Step 1: Make a **plan**. Clearly define the purpose of collecting the data. The benefits you expect from collecting the data should be proportional to the risks. You should be guided by the interests and wellbeing of the individuals about whom you are collecting the data. DO not collect more data than necessary. Plan anonymisation by design wherever possible. Plan how you will get informed consent before you start. Check for any bias in your collection methods. How will you check the accuracy and cleanliness of your data?

Step 2: Do a **risk assessment.** Collecting data can put people at risk. Assess the risks and take action to avoid negative consequences, e.g. by assuring data security and confidentiality.

Privacy risk is the risk of harm arising through an intrusion into privacy. Some of the ways this risk can arise is through personal information being:

- Inaccurate, insufficient or out of date;
- Excessive or irrelevant;
- Kept for too long;
- Disclosed to those who the person it is about does not want to have it;
- Used in ways that are unacceptable to or unexpected by the person it is about; or
- Not kept securely.

Harm can present itself in different ways. Sometimes it will be tangible and quantifiable, for example financial loss or losing a job. At other times it will be less defined, for example damage to personal relationships and social standing arising from disclosure of confidential or sensitive information.

A privacy impact assessment attempts to assess and report on the possible impacts on individual's privacy when organisations process personal data.

Things to consider when conducting privacy impact assessments:

- What is the reason for collecting person-identifiable information?
- Do the individuals whom I collect the data about exercise choice and consent to the processing?

- Do I have policies that govern the use and processing of personally identifiable information? Are these policies robust and fit for purpose?
- Do staff receive support and regular training?
- Are their clear audit trails of where your data is stored and transmitted?
- Are there clear information sharing protocols in place?

The following exercise will help your participants think about all aspects of data management before starting a new project or service. Your participants should clearly choose a new project or process and consider all aspects of responsible data management (RDM) in their plan. This exercise should highlight the importance of thinking about RDM before starting any new project, process or service, so that RDM principles can be incorporated and communicated from the beginning.

Your participants should start to think about the end of the data cycle from the beginning – and consider what sort of questions they are trying to answer from the beginning – for example, free text versus categorized data, and the pros & cons of each. By the end of the exercise your participants should be able to:

-identify RDM practices across the lifecycle of the data from the very beginning to the very end.

Exercise 8: Make a Plan

Ask your participants to get into pairs or small groups, and to devise a plan for a new project or service their organisation will be delivering. Ensure they address the following in their plans:

- *What is the project or service trying to achieve?*
- *What is the purpose of collecting the data, and what will you do with it?*
- *What methods will you use to collect the data?*
- *How will you get informed consent?*
- *How will you train your team?*
- *What are the risks and how will you manage them?*
- *In what form will the data take? Categorised, uncategorised, free-text?*
- *What form does the data need to be in, in order to comply with the sort of analysis you wish to do, if any? Does need to comply with other datasets? (e.g. agreement of categories or groupings)*

- *How will you deal with responses to categorised data where those responses don't exist (e.g. someone answers "neither" to the question "what is your gender?" and the responses in your system are "male, female")*
- *What safeguards do you need to put into place regarding access, transfer, storage or sharing of data?*
- *How long will you retain/archive/dispose of data? Do you have a way to anonymise data prior to archive?*

The following exercise will help your participants think practically about all the risks associated with various types of data handling and put into place safeguards to address those risks. By the end of the exercise your participants should be able to:

- to weigh up and balance the risk of collecting data with the benefits of using that data.

Exercise 9: Data handling risk assessment

Have your participants complete the table below; assessing risks and assigning a risk score for each. Ask them to complete what existing controls they might already have in place and try and identify what further controls they might need to be put in place. Some possible suggestions have been inserted into the table for them. Add additional risks to the assessment as they identify them.

Identification of risk & potential impact(s)	Risk type	Pre-control risk score: Likelihood x impact (min 0, max 25	Existing controls /assurances	Post control risk score L x I	Further planned actions
Unauthorised viewing/access loss of hard-copy documents while in transit					
Staff/ volunteers holding documents at home/off-site					
Emails being sent from personal email addresses					
Documents emailed to the wrong recipient					
Loss or theft of laptops or USB drives					
Inaccurate or incomplete data					
Duplicate data					

Step 3: **Train** your staff. Ensure your staff are aware of and understand data protection and responsible data considerations. Ensure they do risk assessments. Ensure they know how to obtain informed consent. Ensure they are training in data security best practice.

Step 4: Get **informed consent.** Tell respondents how you will use their data and why you need it.

- ➔ **Explain**. Clearly explain to people how you will use their personal information and point them to additional information about this – for example, on your organisation's website, in a leaflet, or on a poster.
- ➔ **Give choice**. Give people a choice about how their information is used and tell them whether that choice will affect the services offered to them.
- ➔ **Meet expectations**. Only use personal information in ways that people would reasonably expect.

In this exercise, your participants will think about how they might gather informed consent in their own context. They should think about all the ways in which they use client data and ensure this is included on any consent form. Your participants should give consideration to taking consent from vulnerable people who may need things explained in a variety of different ways, who may not always fully understand what consent is, and who may later withdraw consent previously given. By the end of the exercise your participants should be able to:

-ensure that their consent forms reflect the ways in which their organization uses (and plans to use) client information.

-take into account consent issues relating to vulnerable people.

Exercise 10: Informed consent

a) With your own services and context in mind, develop a consent form that covers the different uses of data in your organisation. Ensure your form is clear, concise and accurately reflects where there is choice to be granted.

You may wish to also ensure your form covers common data sharing arrangements where your service may regularly share data with others (e.g. other health services, etc.)

b) With a partner/in pairs, practice going through the consent form and gain informed consent.

II. Handling & Storage

Step 5: **Manage your data**

➔ Ensuring data quality: Clean, protect, enhance.

Consider what you will need to put in place to ensure your records and data have integrity: they are clean, duplication-free, and error-free.

Things to think about:

What are the possible consequences of unclean data? Think specifically about both day-to-day management of a service (see above example re: JP) as well as any sort of analysis or decision-making that may come from the participant's data. What sort of information or conclusions are they, or others drawing about this data? What policies or processes might they need to put into place to ensure good quality data?

In the following exercise, your participants will think about the value of clean and standardised data, and the possible risks with poor quality data. They will learn to standardise a small set and how that contributes to clean data and may reduce errors in analysis. By the end of the exercise your participants should be able to:

-clean data.

-understand how cleaning/standardizing data contributes to data integrity.

Exercise 11: Cleaning/standardising data

Ask your participants to look at the following dataset and suggest ways to clean it or standardise it.

Jane Doe	New York	Clinical therapy	1978-01-15
Ahmed Assan	N.Y	Therapy	15th Jan 1978
Georgeau Constantine	ny	clinical	01/15/1978

➔ Transfer – take care when using portable devices or when moving hard-copy records. Encrypt digital records. Restrict access. Ensure devices are encrypted. Take extra care when moving paper records which are at risk of loss or theft.

➔ Access – Ensure access is restricted on a 'need to know' basis. Restrict systems and records access. Ensure hard-copy records are locked away at all times and access is controlled.

➔ Store – ensure you know and understand where your information is stored, and

➔ Share – will you need to share all, or part of your data? For example, individual records in support of a client, or, larger data sets with a research partner?

Your participants will learn how to put together a basic data sharing agreement, taking into consideration data protection issues and the possible value of sharing data. Participants should consider whether anonymization techniques may be appropriate in the context of their chosen example. By the end of the exercise your participants should be able to:

-understand data protection issues when sharing data.

-create a simple data-sharing agreement.

Exercise 12: Data sharing agreement

Have your participants think about what partners they work with and where they may need to share data. Develop a mock data-sharing agreement for this.

The following scenarios are aimed to get your participants thinking about how they might balance the needs of the service with data protection principles. By the end of the exercise your participants should be able to:

-balance any conflicting interests when considering responsible data management.

Exercise 13: things to think about

How would you respond or advise your colleagues to respond to the following scenarios:

- *You're taking intake information from a survivor and they ask what is going to happen to the information they are providing you.*
- *You are taking history information from a survivor through a translator, and although the survivor gave a very long and emotional answer to a question, the translator clearly gave a summary of what was said.*
- *A partner organisation wants to do a collaborative project which would mean data sharing. What might you need to think about/put in place when considering this?*
- *A new colleague joins your organisation and asks what your policy is around clients giving or refusing consent to use their data for certain activities. What do you tell them?*

III. Use

Step 6: **Use** your data! This might be lobbying, advocacy, or learning applied to your own services. Think about who is being represented in the data. Have you considered any bias? Gender balance? Are you confident in the quality of the data? In the quality of the analysis?

Step 7: Provide **feedback**. It is good practice to involve those from whom you collect data in the use of it. For example, if you have used your data in writing a report for advocacy purposes, it is good practice to share that report with respondents wherever possible.

Step 8: **Retention & destruction**. Ensure you have appropriate retention & destruction policies in place. Do you need to keep data in its current form? Do you have to keep personal data? Is there a way to keep only aggregate data? Or to anonymise it? If not, are you able to retrieve individual records to destroy (see 'right to erasure' above under GDPR). When choosing to delete, ensuring deletion is permanent and all copies/versions are deleted as well.

Effective and responsible data management extends over the entire lifecycle of the data.

Conclusion: where to begin? Things you can start doing now

- Raise awareness amongst your colleagues, contractors and partners regarding safe data handling – particularly with the most senior people in your organisation.
- Train and communicate with your staff on responsible data management and data security measures, including policies to ensure access is blocked when staff depart from your organisation.
- Ensure you have robust policies and procedures regarding the safe handling and protection of personal data. Policies: Data Protection, Information security, confidentiality, data sharing, incident (breach) reporting, recovery, disclosure.
- Ensure clear lines of accountability when it comes to handling data.
- Ensure transparency in data handling and ensure those whose personal data you handle are clear on how it will be used.
- Implement a system of risk assessment and privacy impact assessments whenever looking at collecting new forms of data or collecting data in a different

way. A plan will go a long way to helping you think through the risks and possible consequences in the event of a data breach.

- Use a risk-based approach and think about what could go wrong in order to try and put preventative measures in place to reduce the risk.
- Duty of candour: be prepared to disclose when there are breaches or unintended loss, or corruption of data occurs.
- Ensure you have robust contracts in place to cover issues around data sharing, or where you sub-contract any data processing.
- Review IT security, and ensure appropriate measures are in place for encryption, back-ups, updates, BYOD (bring your own device), etc.
- Ensure you have an incident response plan in the event of data breach.

6. Glossary

Aggregate – form of grouping into a class or cluster for the purposes of higher-level analysis or anonymisation.

Data breach – a security incident in which sensitive, protected or confidential **data** is copied, transmitted, viewed, stolen or used by an individual unauthorized to do so.

Data compliance – the completeness of a data set.

Data cleansing/cleaning – the process of detecting and correcting (or removing) corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

Data controller – a person/organisation who determines the purposes for which and the manner in which any personal data are, or are to be, processed.

Data governance – defined process(es) of an organisation to ensure that high quality data exists throughout its lifecycle.

Data hygiene – the collective processes conducted to ensure the cleanliness of data. Data is considered clean if it is relatively error-free. Dirty data can be caused by a number of factors including duplicate records, incomplete or outdated data, and the improper parsing of record fields from disparate systems.

Data lifecycle – the flow of information through a system, from creation and storage to deletion.

Data processor – any person/organisation who processes personal data on behalf of the data controller.

Data protection officer – the individuals(s) responsible for ensuring that an organisation(s) comply with data protection legislation, and often their own internal data protection policies.

Data set – a collection of related data.

Data sharing agreement – an agreement or framework for the sharing of data which sets out how data will be transmitted, stored and used.

Data standardization – the critical process of bringing data into a common format that allows for collaborative research, large-scale analytics, and sharing of sophisticated tools and methodologies.

Data subject – an individual who is the subject of personal data.

Encryption – the process of encoding messages or information in such a way that only authorised parties can read it. Encryption does not of itself prevent interception but denies the message content to the interceptor.

GDPR – General Data Protection Regulation – the new legislation covering data protection of data subjects within the European Union.

Information commissioner – the authority responsible for oversight and enforcement of data protection legislation in the United Kingdom.

Information ethics –"the branch of ethics that focuses on the relationship between the creation, organization, dissemination, and use of information, and the ethical standards and moral codes governing human conduct in society".

Information governance – the management of information at an organisation. Information governance balances the use and security of information.

Informed consent – a voluntarily and freely given agreement, based upon a clear appreciation and understanding of the fact, implications and possible future consequences of an action.

Personal data – data that relates to a living individual who can be identified from those data, or from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller.

Phishing – the attempt to obtain sensitive information such as usernames, passwords, or credit card details (and sometimes indirectly, money) often for malicious reasons, by masquerading as a trustworthy entity in an electronic communication.

Processing – collecting, amending, handling, storing, or disclosing personal information.

Privacy impact assessment – an assessment undertaken to ascertain the impact of any new processing on the privacy of data subjects. A tool to identify and reduce privacy risks.

Pseudonym – a fictitious name or an alias.

Records management – the field of management responsible for the efficient and systematic control for the creation, receipt, maintenance, use and disposition of records. This includes identifying, classifying, storing, securing, retrieving, tracking and destroying or permanently preserving records.

Redaction – the censoring or obscuring part of a text or information for legal, security, privacy or anonymization purposes.

Sensitive personal data – refers to data about:

- Racial or ethnic origin.
- Political affiliations.
- Religion or similar beliefs.
- Trade union membership.
- Physical or mental health.
- Sexuality.
- Criminal record or proceedings.

Subject Access Request (SAR) – any request (in writing) by a data subject for the personal information on them held by an organisation.

7. References / Further Resources

- Care Quality Commission. *Safe Data, Safe Care*. https://www.cqc.org.uk/sites/default/files/20160701%20Data%20security%20review%20FINAL%20for%20web.pdf
- Data Protection Commissioner Ireland. *GDPR and you* http://gdprandyou.ie/
- DLA Piper, *Data Protection Laws of the World.* https://www.dlapiperdataprotection.com/index.html?t=world-map&c=DK
- General Data Protection Regulation https://www.eugdpr.org/
- Geraghty, R. (2016). *Anonymisation and social research*. https://www.slideshare.net/ISSDA/anonymisation-and-social-research?qid=fa5a5338-8766-4b0b-9bf6-105f852d5932&v=&b=&from_search=1
- Information Commissioner's Office https://ico.org.uk/
- Information Commissioner's Office (2017). *Preparing for the General Data Protection Regulation (GDPR): 12 Steps to take now*. https://ico.org.uk/media/1624219/preparing-for-the-gdpr-12-steps.pdf
- Information Commissioner's Office (2017). *Subject Access Code of Practice: Dealing with requests from individuals for personal information*. https://ico.org.uk/media/for-organisations/documents/2014223/subject-access-code-of-practice.pdf
- Information Commissioner's Office. *Bring Your Own Device (BYOD)*. https://ico.org.uk/media/for-organisations/documents/1563/ico_bring_your_own_device_byod_guidance.pdf
- ICRC. *Professional Standards for Protection Work*. https://www.icrc.org/eng/assets/files/other/icrc-002-0999.pdf
- National Data Guardian for Health and Care. *Review of Data Security, Consent and Opt-outs*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF
- NHS Digital. *Data Security Awareness Level 1*. https://www.igt.hscic.gov.uk/
- NHS England. *Information Governance Handbook*.
- Responsible Data Forum: https://responsibledata.io/
- UK Data Service. *Anonymisation*. https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation

Appendix A

Training Agenda Template

Training objectives: develop skills and knowledge regarding responsible data management (collection, handling, storage, use), and, develop skills to train others.

Day 1

0900-10:30 (1.5h)

Introduction to responsible data management

- Why does it need protecting?
- Think about how you would like your own data handled

What data?

- Personal data, sensitive data
  → Exercise: Identifying and categorising personal/sensitive/confidential data
- Anonymisation & pseudo-anonymization
- Aggregation
  → Exercise: list commonly processed datasets in your own context; categorisation of own datasets; start to think about information sharing

10:30: BREAK (15)

10:45-12:30 (1.75h)

Introduction to EU GDPR / legislative framework

- GDPR: The EU legal framework. The wider context and legal framework: European General Data Protection Regulation (GDPR)
- The rights of data subjects; obligations of data controllers
  →Exercise: data mapping
  →Exercise: data protections scenarios & getting people to think how they would respond. Getting people to think of possible scenarios in their own context

12:30: LUNCH (60)

13:30-15:00 (1.5h)

GDPR continued

- Subject Access Requests – what they are and how to comply
  → Exercise: Compliance with Subject access requests, 3rd party information, redaction
  → Exercise: Data protection scenarios

15:00: BREAK (15m)

15:15-16:30 (1.25h)

Continue with exercises, getting participants to try and come up with real-world data sets that relate to their everyday work, and real-world scenarios that relate to their everyday work
Think about risk-based decision making

16:30: Recap of day & touch on next day's topics

Day 2

09:00-10:30 (1.5h)
Recap previous day's topics
Introduction to responsible data management

- What is RDM?
- Why do we want to practice ethical/RDM?
- Considerations in RDM
- Confidentiality, integrity, availability
  → Exercise: scenario & discussion

10:30: BREAK (15)

10:45-12:15 (1.5h)
The Data Lifecycle: Overview of the data lifecycle

- A common life cycle – steps in the data lifecycle and risks at different stages
- Collection, storage, handling, use – the 8 steps within this framework
- I. Collection: Responsible data collection
  - Planning overview, things to think about when planning data or a new project or service, privacy impact assessments, informed consent, ensuring appropriate skills in data collectors, standardised data sets
  - Assessing risk in RDM – being prepared for worst case scenarios / learning from others

  → Exercise: Make a plan: think about a core service activity or new project, and how you might go about conducting a privacy impact assessment regarding the data you are collecting. Remember to think about the differences between Pd & non-Pd
  → Exercise: Do a risk assessment
  - Training your staff
  - Consent: informed, appropriate, flexible. Informed consent: what is it, what does it look like, what is the purpose of collection, planning for withdrawal of consent or if people change their minds

  → Exercise: devise a consent form which will be presented to people who access your services at the point of access. You will need to think about how to convey what you will be using the information for, and why you need it.
  → Exercise: Practice getting informed consent

12:15 – recap & touch on Day 3 topics

Day 3

09:30-11:00 (1.5h)
Recap of topics covered so far / Q&A's

The Data lifecycle continued

- II. Handling and III. Storage
    - Data management. Ensuring data integrity, data quality, standardisation, validity, comparability; how this related to ethical data. Setting up appropriate infrastructure; controlled access, safe storage (physical/electronic), safe sharing/transfer – anonymization, pseudo-anonymization. Encryption. (Refer back to SAR day 1).

    → Exercise: data cleaning
    - Data management continued

    → Exercise: Data sharing agreements

    → Exercise: scenarios/things to think about

11:00: BREAK

11:15-12:30 (1.25h)

- IV. Responsible data use
    - Use: Data→ information→ action: lobbying, advocacy, programme evaluation. Quality improvement. Change. Consideration given to how data can be *mis*-used What do you use data for? How can you ensure it is not mis-used?
    - Feedback: Providing feedback where possible closing the loop and feeding back results and analysis to those who own the data – showing them what happened/what was achieved.

      → Exercise: How might you provide feedback to people who use your services on what you have done with their data? (Think report sharing, or sharing outcomes of a lobbying effort – possible media: groups/website)

LUNCH (60m)

13:30-14:30 (1h)

- Retention & disposal. – retention policies, & appropriate processes in place. The data shadow - understanding the location of your data: locally/networks/cloud. Relate retention back to SAR

  → Exercise: Draw up a retention schedule for the various forms of data you keep
- Embedding RDM into your organisation

  → Exercise: develop an action plan for how you will take the skills you've learned over the past week & apply them

BREAK (15)

14:45-15:45 (1h)

- Continue action planning
- Feedback to the group on your action plan going forward

15:45-16:00 recap, last Q&A's