Online Training Module:
Considerations in sharing data

Slide 1
Welcome to the training module 'Considerations in sharing data'. This video is approximately 30 minutes. We have found that organisations often have ideas about how sharing data could improve insights into their work, and would benefit from an overview of the key issues and complexities before getting started.

Slide 2
My name is Ben Ritchie and I've been supporting data management and analysis for CORC members and partner organisations since 2015. I'm delighted to share our knowledge and experience in this module of the CORC Online Training Programme.

The module has two parts and we hope it will provide two learning outcomes. First, an understanding of key concepts and terms for managing and sharing data. Second, an ability to identify important areas to take into account when establishing data sharing arrangements, or resolving issues that may arise.

Slide 3
Although data sharing can provide great benefits for individuals and populations, it is an area that is often clouded by uncertainty, with ethical, legal and technical complexities. We've therefore concentrated this section of the module on key concepts and terms. Gaining familiarity with these is crucial for being able to clearly think through and communicate the options and steps for data sharing. We hope you will also gain an appreciation of where uncertainties exist, and the challenges that arise when translating the definitions into practice.

Slide 4
We will first provide an overview of purposes for sharing data, privacy and confidentiality. Next, we will review different ways of describing data, which may seem a little uninspiring on their own, but are vital for understanding subsequent topics. Anonymisation will be discussed in further detail, given it is a powerful technique to reduce data protection risk and support data sharing. In the final part of this section, we will briefly explain the meaning of the terms data controller and data processor, which carry significance in regards to data protection compliance.

Slide 5
There are many scenarios where sharing of data might be needed. These are some examples from members of CORC.

In the first example, a counselling service and its partner schools want to understand the academic outcomes of pupils who access mental health counselling support. Depending on where this question is to be looked into, education or counselling data would need to be shared between organisations.

The second is a continuing question of interest for children and young people's mental health services. To compare outcomes between services, they would need to share data with each other, or with a trusted central organisation that can collate it in a common format for comparison.

Slide 6

Later in this module we will discuss the importance of being clear on the purpose or purposes for sharing data.

In the context of working towards improvements in health or wellbeing, purposes for sharing and using data can be thought of as falling under two broad categories. When data are used for the benefit of people as individuals, the data will inform decisions or actions with respect to particular individuals. This type of purpose is sometimes called 'direct care'. When data are used for the benefit of people in general, the data will be used to support improvements in support or services for groups of people, such as the users of a service, or the population of a geographical area. This type of purpose may be referred to as 'indirect care', 'secondary uses of data' or 'using data to improve health, care and services through research and planning'.

It's useful to be aware that these broad purposes have different requirements in relation to the data items needed, compliance, and permissions to be sought. However, a recent survey by the National Data Guardian found that the distinction between the two is not always well understood or easy to make in practice, and may be blurred by new technologies and ways of working.

Slide 7

Privacy and confidentiality are important concepts in relation to data sharing. Although the terms are sometimes used interchangeably, it can be useful to disaggregate their separate meanings.

Privacy refers to a person's freedom from interference or intrusion in their activities or information. Confidentiality relates to the rules or practices that prevent unwanted disclosure of information. In this sense, confidentiality can be thought of as one of the principles that helps to achieve privacy.

When planning a project that involves sharing of data, it is important to respect privacy and confidentiality. We can see from the quote from the Nuffield Council on Bioethics that it can be challenging to meet the dual demands of extending access to data and protecting privacy. This module provides an introduction to areas to consider to meet this challenge.

Slide 8

An aspect of confidentiality often encountered in relation to data sharing is the duty of confidentiality. This applies to any information provided with the expectation that it would not be disclosed by the recipient. For example, if a young person shared information about their mental health with their clinician, there would be an expectation that this wouldn’t be disclosed to their school.

The general position is that the information that a duty of confidence applies to should not be shared further without the permission of the person who originally provided it. However, what constitutes ‘permission’, and how this should be sought, may vary in different contexts.

The point about the information being identifiable is worth highlighting, because in cases where data relates to a person, “once information is effectively anonymised it is no longer confidential” (HSCIC (2013) Guide to confidentiality in health and social care: references). We will go on to cover the concepts of identifiability and anonymisation in more detail.

Slide 9

We now turn our attention to different ways of describing data. We start by considering the concepts of person-level and aggregate data. With person-level data, information relating to one person can be distinguished from another person. This form of data is usually structured as a table, with at least one row per person. Person-level data is also called ‘individual-level data’ or ‘record-level data’. In this module we will refer to the individuals to whom the data relate as the ‘data subjects’.

Person-level data can be combined together in different ways, such as by counting or taking an average. This results in aggregate data, which provide information about many people, such as the users of a service or the pupils in a school year group. Aggregate data are usually presented in the form of summary tables, charts or statistics.

The distinction between person-level and aggregate data is important, because certain purposes can’t be met with aggregate data. For instance, person-level data would be required to explore the effects of different personal characteristics on treatment outcomes.

Slide 10

The next few slides cover what it means for data to be considered identifiable or anonymised. Before we look at the technical term ‘personal data’ that appears in data protection regulation, we will consider the terms ‘identifiable’ and ‘anonymised’ as they are commonly understood, in a non-technical sense.

Here we have an example screenshot of some made-up data from a school wellbeing survey. The data here are considered identifiable because the people they relate to can be identified, in this case by their name, school and class.

Slide 11

When we talk about anonymised data, we usually mean data about people that by itself does not identify who they are. There may still be a possibility of ascertaining the identity of a person in the data, but not from the anonymised data alone.

Turning identifiable data into anonymised data can be achieved in different ways. In the example here, names have been removed and the pupil_id number has been replaced by a different type of id number, called a pseudo_id.

Slide 12

Another way of turning identifiable data into anonymised data is to aggregate it. In the example here the responses to survey questions about life satisfaction have been categorised into two ranges and the number of pupils counted in each range.

These examples demonstrate that anonymised data may take the form of person-level data or aggregate data.

Slide 13

We now turn to the definition of 'personal data', which is a fundamental concept in data protection regulation. The GDPR and UK Data Protection Act use 'personal data' to describe any information relating to an identified or identifiable natural person, where an identifiable person is one who can be identified, directly or indirectly. 'Natural person' just means a human being, as opposed to a company, which may sometimes be referred to as a 'legal person'.

One way of understanding the definition of personal data is to break it down into two steps. First, can people be directly identified in the data, for example because their names or contact details are included? If the answer is yes, the data would be classed as personal data. If the answer is no, the second step is to consider whether people are indirectly identifiable. In these cases there is often uncertainty, and further discussion and deliberation may be required to decide whether the data can be linked to a living person.

Slide 14

The table on this slide is our previous example of the identifiable data from a school wellbeing survey. Pupils can be directly identified from the table, and so the data are clearly personal data.

Slide 15

This table is one of our previous examples of data that are commonly understood to be anonymised, since it doesn't contain any fields that directly identify people. There is often uncertainty about whether data like this would fall under the definition of personal data. To determine whether these data are personal data, we would need to assess whether people can be indirectly identified.

This is not always obvious, and may not be apparent from looking at the data in isolation. It's likely to require further investigation, with attention paid to the context in which the data are held. For example, if this table of data is held within the school, it would be relevant to consider whether the school has retained a 'look-up' file that links pseudo_ids to the original pupil_ids. The existence of such a look-up file would make it relatively straightforward to re-identify a pupil.

If this table of data has been shared with another organisation that does not have access to a look-up file, other considerations would arise. For example, could knowing the school, year group and ethnicity of any pupils be used to indirectly identify them? If this is theoretically possible, what would be the time, effort and additional information requirements to do so? Recital 26 of the GDPR provides further guidance on this, and talks about the need to take account of whether there are any "means reasonably likely to be used" to indirectly identify someone.

Slide 16

From the previous slides we have seen that data commonly understood to be identifiable are always personal data. We have also seen that data commonly understood to be anonymised may or may not be personal data, depending on the likelihood of a person being indirectly identified, which depends on the nature of the data and the context in which it is held.

It therefore follows that one way of approaching anonymisation is to not only think about making changes to the data, but also to check or introduce controls in the environment in which the data are to be shared. These controls reduce the risk of a person being indirectly identified. For example, restricting access to a small number of users who will maintain data anonymity as part of their role, or as part of agreements they have signed up to. Taking this approach to anonymisation can be a useful way of supporting the purposes for data sharing, while protecting the rights of data subjects, including privacy, and preserving confidentiality where it applies.

In a scenario of sharing person-level data between organisations, it may be that controls to reduce the risk of indirect identification in the recipient organisation's environment are sufficient to ensure that the data are non-personal data. Alternatively, in a scenario of publishing information from a project, the public environment will not have any controls to reduce the risk of indirect identification. Therefore, more substantial changes to the data would be required, including aggregation and presentation of results in a way that ensures individuals cannot be re-identified.

Slide 17

This slide briefly covers the meaning of other data protection terms commonly encountered in data sharing.

Data processing simply means any operation performed on personal data, including its collection, storage, use, transfer, analysis or deletion.

When personal data are processed by an organisation, from a data protection perspective the organisation can have one of two roles. If the organisation determines the purposes and means for processing the data, it is the data controller. If the organisation processes the data on behalf of, or on the instructions of, a data controller, it will be a data processor. The Information Commissioner's Office acknowledge that these roles can sometimes be difficult to distinguish in practice and have guidance on their website to help organisations decide.

Both data controllers and processors have data protection compliance responsibilities. Data controllers have more obligations, and are responsible for the compliance of their data processor or processors.

When personal data are shared, it can be from a data controller to a data processor, or from a data controller to another organisation that takes on data controller responsibilities with respect to the data. Therefore, data sharing may or may not involve a data processor.

Slide 18

Equipped with an understanding of key concepts and terms, we now move to the second part of this module to discuss important areas for consideration and decision making in sharing data.

Slide 19

In our experience, the main areas to take into account are (i) the purpose for data sharing; (ii) the people that the data relate to, referred to as the 'data subjects'; (iii) the particular data to be shared; (iv) the legal basis for sharing the data; (v) roles and data handling processes; (vi) information for the data subjects and (vii) building relationships with relevant stakeholders and gaining permissions.

These areas do not necessarily need to be considered in the order presented, however we have often found it beneficial to begin with clarifying the purpose. It's likely that working through the areas will be an iterative process, whereby developing an understanding in one area will inform others. For example, thinking through the expectations of data subjects with regard to the data to be shared may help to refine the purpose, or the appropriate processes of data handling.

We will now talk through the 7 areas in more detail.

Slide 20

The first area to clarify is the specific purpose or purposes for sharing the data, and to ensure this is understood among all stakeholders. In our experience, it may take some time and effort to obtain a clear statement of the purpose and get stakeholders on the same page. An understanding of the key concepts covered in the first part of this module will support these conversations.

It's likely that those responsible for the data will have privacy or other concerns related to data sharing. It may be helpful to reassure them about both the intended purpose, and purposes that are out-of-scope. For example, if the purpose of sharing the data is to improve support or services for the benefit of people in general, they are likely to be reassured to know that any actions taken in relation to people as individuals is out-of-scope. When establishing a new data sharing initiative, it may also be sensible to take small steps, perhaps starting with a feasibility exercise, or an initial purpose that those responsible for the data are more comfortable with.

The purpose should be captured in writing, for inclusion in documentation such as privacy notices and data sharing agreements. It's hard to go back and change these at a later date, so it's important to get the purpose statement right at the earliest stage possible.

Slide 21

The second area involves identifying exactly which people the data to be shared relate to. For privacy reasons it's important to be precise here and to exclude any unnecessary groups.

It's useful to consider what the expectations of the data subjects are for the use of data that relate to them, including any expectations about confidentiality. To establish what these are, it may be valuable to revisit any information or privacy notices provided to the data subjects about how the data would be used. If the data subjects include children, parent or carer expectations should be taken into account in addition to those of the child.

The GDPR talks about the need for a new purpose for data processing to be "not incompatible" with the original purpose when the data were collected. Scientific research purposes and statistical purposes are considered compatible, however for other data sharing purposes the expectations of data subjects need to be taken into account as part of deciding whether the data sharing purpose is compatible.

Slide 22

The third area involves defining the data necessary for the purpose. It's useful here to apply the data protection principle of minimisation. This means finding the least identifiable and fewest data items that would still allow the purpose to be met. For example, if the purpose can be achieved using anonymised aggregate data, there will be no need to share personal data.

It may be the case that achieving the purpose entails an initial step of linking data, for example between datasets or between time points, and that linking data may require identifiers. In these cases it can be helpful to distinguish between the data items needed for linkage and the main data items required for the purpose, and whether the linkage items can be deleted once data are linked.

For some of our projects we have found it helpful to draw up a data specification, which is a description of each data item. Data specifications may form a useful part of data sharing agreements.

Slide 23

We now turn to the legal aspects of data sharing. It's not within the scope of this webinar to go into the detail of legislation or provide legal advice. Given the complexities of the law around data protection and confidentiality we would suggest consulting with data protection colleagues or the Information Commissioner's Office.

The purpose of this slide is to highlight general areas for consideration. The main examples from our experience are the General Data Protection Regulation (GDPR) and UK Data Protection Act, and the common law duty of confidentiality.

To find out what requirements apply, it is necessary to determine how the data needed for the purpose would be classified in GDPR terms. For example, personal data concerning health fall under the definition of 'special category data' and require an additional condition to be met for lawful processing.

One of the reasons why effective anonymisation can be a useful tool is that requirements differ for personal data and anonymised data. The ICO Anonymisation Code of Practice is a useful reference document that goes into further detail.

Slide 24

The fifth area involves defining roles and processes. Roles should be set out at an organisation level, and a personnel level. When personal data are being shared, the role of the recipient organisation should be defined in terms of data controller or data processor. This will clarify each organisation's responsibilities to data subjects and supervisory authorities. At a personnel level, it involves deciding who will be involved in the different processes for preparing and sharing the data.

By processes, in this context we're referring to how the data will be collected (if not already collected), anonymised (where applicable), transferred, stored and retained. It's helpful to develop a data handling plan or protocol and to keep this up to date. Writing down and thinking through each step of data handling can be a useful exercise, as it will help to both avoid data breaches and improve the efficiency of the processes, which should save time if they are to be carried out more than once.

You should aim to clarify exactly what will happen to the data, in terms of how and when it is transferred, what the recipient's data handling processes are, and when the data will be deleted. Other aspects to be determined are the appropriate data security controls, such as the use of password-protected files or servers.

We discussed the benefits of anonymisation in the first part of this module, and it's worth highlighting here that it's good practice to anonymise data at the earliest opportunity in the process.

Slide 25

The sixth area for consideration is the information that will be provided to data subjects about the sharing of data relating to them. If privacy notices or webpages already exist, it may be possible to update these rather than creating new sources of information.

A useful principle to follow is that there should be 'no surprises' for data subjects about how the data relating to them are shared and used. People generally want to know who has responsibility for the purposes for which the data are being used, and, where person-level data are shared, what the controls are on data access.

Where personal data are being shared, it's the responsibility of the data controller organisation to provide data subjects with specific details, including areas we have already covered in this module such as the purpose, the type of data involved, the legal basis and the roles and processes. Please see the Information Commissioner's Office website for guidance.

Slide 26

The final area is the relationships and permissions required to support and approve the data sharing. In our experience, it can be beneficial to engage with those responsible for protecting data and sharing data appropriately at an early stage, prior to requesting their sign-off. To make effective use of their time, it's advisable to have put together some initial thinking on the other areas of consideration to serve as a starting point for the conversation.

Depending on the scale of the data sharing being planned, it may also be worth establishing other professional relationships to facilitate the process, such as with organisations that represent the interests of the data subjects, or those supporting the data subjects.

Once everyone is on board, the usual process to gain permission is to put all of the relevant information into the appropriate form of agreement for comments and sign-off. The good news is that you should have the necessary information from working through the previous areas of consideration.

Slide 27

In summary, we have covered 7 areas of consideration for data sharing. We do not claim that these areas will be fully comprehensive for every possible scenario of sharing data. However, we think they offer a helpful general framework.

Getting these areas clearly worked out and agreed will place your data sharing plans in a strong position. Similarly, if progress has faltered with a data sharing initiative, it may be worth assessing the situation in relation to each area, to pinpoint where issues lie.

We hope this module has also provided an appreciation of where uncertainty exists around data governance and data sharing. In light of this we would highlight the importance of consultation with different stakeholders throughout the process. For example, consulting with users of the data on the precise purpose, consulting with relevant colleagues on the legal and technical aspects, and with data subjects on what they would like to know.

Thank you for listening to this module. If you would like to further your understanding of any of the topics covered, please see the webpage for links to additional resources.

Slide 28 – no audio

END TRANSCRIPT