Co-funded by
the Health Programme
of the European Union

JASEHN

Joint Action to Support the eHealth Network

REPORT

on

How to handle health data for purposes other than patient care

Document Information:

Document status:	For discussion to the members of the eHealth Network at their 11th meeting on 09 May 2017
Approved by JAseHN sPSC	Yes
Document Version:	V4.0
Document Number:	D7.2.2
Document produced by:	Joint Action to support the eHealth Network • WP7: Exchange of Knowledge • Task7.2: Secondary use of Health Data
Author(s):	Jeremy Thorp, HSCIC (United Kingdom) Greg Fletcher, HSCIC (United Kingdom) Jurgen Wehnert, GEMATIK (Germany) Christof Gessner, GEMATIK (Germany)
Member State Contributor(s):	Belgium, Finland, France, Germany, Greece, Hungary, Sweden, Italy, Luxembourg, Norway, Portugal, UK
Stakeholder Contributor(s):	HL7, ENISA, COCIR

TABLE OF CHANGE HISTORY

VERSION	DATE	SUBJECT	MODIFIED BY
1.0	2016-09-20	Draft submitted "for review" by JaseHN WP3	JEREMY THORP
2.0	2016-10-17	Draft 2.0 submitted for SPSC review	JEREMY THORP
2.2	2017-02-02	Draft 2.1 following OECD Recommendation	JEREMY THORP
2.3	2017-02-17	Minor changes following KT member review	JEREMY THORP
2.3	2017-02-21	WP3 quality check	MARA TIMOFE
3.0	2017-03-15	Updates following quality review	JEREMY THORP
3.1	2017-03-24	Updates following additional review comments	JEREMY THORP
3.3	2017-03-29	Minor updates following final minor comments	JEREMY THORP
4.0	2017-04-18	Final version for submission to the eHN	JEREMY THORP

LIST OF TABLES

Table 1. Examples of approaches to data management (Source OECD [14]) 13
Table 2. Controls / standards to be applied 18
Table 3. Edited extract from code for sharing of public health data 20

LIST OF FIGURES

Figure 1. Examples of Existing Secondary Uses of Data 9
Figure 2. Outline of process steps 14
Figure 3. Deciding when and how to release anonymised data 23

TABLE OF CONTENTS

1. Introduction 4

2. Background 5

2.1 Deliverable 7.2.1 5

2.2 Update on Digital Single Market, Cloud Computing, Open Data and Big Data 5

2.3. OECD Council Recommendations on Health Data Governance 6

2.4 Scope of secondary uses of health data (breadth of potential uses) 6

2.5 The aim of the task and of the second deliverable 7

3. Opportunities for re-use of health data 8

3.1 Types of purpose and analysis 8

4. Implications and Responsibilities 11

4.1 Legal and Regulatory context (e.g. GDPR, eIDAS) 11

4.2 Policy aspects (governance) 11

4.3 Information (data standards for consistency, aggregation) 12

4.4 IT infrastructure (cyber security, anonymisation) 12

4.5 Applications (processing and data management) 12

5. How to Handle Secondary Use of Health Data 14

5.1 Adopt Principles 15

5.2 Apply Process 16

5.3 Assess Privacy Risks 17

5.4 Agree controls 18

5.5 Assure Compliance 19

6. De-identification 22

Annex A: Source Materials 24

Annex B: Glossary (including source) 27

Annex C: Overview 30

1. Introduction

This report forms part of Task 7.2 of the Joint Action on eHealth. The Description of Work introduces the work by saying "*This task will address the following subjects:*

- *The pros and cons of the use of cloud computing in health,*
- *Publication of a code of conduct on how to handle secondary use of health data*
- *Recommendation on de-identification of data for secondary use*".

The deliverables from this task are as follows:

D7.2.1 Report on the use of cloud computing in health (month of delivery: M7, November 2015)

D7.2.2 Code of conduct on how to handle health data for purposes other than patient care (month of delivery: M13, May 2016).

The latter has been held back pending ratification of the OECD recommendations on Health Data Governance (January 2017) [16] to ensure consistency and alignment of key messages. Given the many relevant supporting publications, section 6 provides a list of supporting materials, and section 7 provides an extensive glossary.

These outputs are aimed at Member States, National Competence Centres and Data Protection Authorities with responsibility for commissioning or providing reporting facilities relating to purposes other than direct care for a patient.

Article 40 of the General Data Protection Regulation encourages Member States, the supervisory authorities and the Commission to draw up codes of conduct intended to contribute to the proper application of the Regulation, *„taking account of the specific features of the various processing sectors and the specific needs of micro, small and medium-sized enterprises"*.

This deliverable provides a description of good practices rather than a formal code of conduct. It is therefore recommended that each MS makes their own assessment of the suggestions provided here.

2. Background

2.1 Deliverable 7.2.1

In the deliverable 7.2.1, it was noted that cloud computing offers many opportunities for providing high-quality, high-bandwidth, resilient access to computing resource that can be rapidly provided or released as required. The report introduced different models of cloud computing, in terms of the levels of service and the implementation approach and went on to consider the pros and cons of cloud computing for eHealth, with a focus on the uses of data other than for the direct care of an individual patient.

The advantages include the resilience to the speed of technological change, the need for continuous re-investment, the huge growth in storage requirements and the pressures on recruitment and retention of skilled staff. It is necessary, however, to address the sensitivity of medical data, the legal and ethical obligations on security and data protection and the budgetary constraints within which healthcare organisations typically operate.

Any approach to the use of cloud needs to comply and be seen to comply with these requirements and hence gain the confidence of the public. The report made a series of proposals, for use in developing contractual agreements, and set out a number of steps from a business, legal and technical perspective.

This follow-on deliverable develops a code of conduct on how to handle secondary use of health data. However, the code is not dependent on a cloud-based deployment and applies independent of the form of provision.

2.2 Update on Digital Single Market, Cloud Computing, Open Data and Big Data

The topic of cloud computing continues to be very popular and relevant, as does the issue of secondary uses of patient data. Taken together, these form a top-priority agenda.

In April 2016, the European Commission issued its communication on cloud computing [2]. During the drafting of this deliverable, the Commission commissioned a study into big data in health [11] and ENISA published „Technical Guidelines for the implementation of minimum security measures for Digital Service Providers" [10]. This is based on the requirements of the Directive on security of network and information systems (NIS Directive) to Cloud Service Providers, Also ENISA has been working on a white paper on Cloud security for eHealth services due to be released mid 2017.

The Cloud Select Industry Group (CSIG), a group composed of representatives from European and multinational industry, public administrations and other, has been helping the Commission to gather feedback about major cloud-related issues from a variety of perspectives. In February 2017 the CISG handed over the latest draft of the code of conduct on data protection for cloud service providers to the Code's newly established General Assembly.

This JAseHN task 7.2.2 has aimed to ensure consistency with these parallel activities whilst building on previous work and adding its own value to the topic in hand, in order to provide eHealth Network members with practical advice for moving forward.

2.3. OECD Council Recommendations on Health Data Governance

In January 2017, the OECD´s governing body, the Council, adopted a recommendation on health data governance [16]. The Council, has the power to adopt Decisions and Recommendations**,** usually referred to as "the OECD Acts". Recommendations are not legally binding, but practice accords them great moral force as representing the political will of Member countries and there is an expectation that Member countries will do their best to implement a Recommendation. 22 EU countries are members of the OECD. The Recommendation is open to non-Members adherence also. The Recommendation on health data governance has been jointly developed by the Committee of Digital Economy Policy including its Working Party on Security and Privacy in the Digital Economy and the Health Committee advised by the Health Care Quality Indicators Expert Croup. The Recommendation has been built on the best practices identified including the EU regulations [1]. The OECD Committees will follow up the implementation of the Recommendation and report it to the OECD Council.

This Deliverable 7.2.2. has been validated against the OECD Council Recommendation to ensure there is no conflicting content.

2.4 Scope of secondary uses of health data (breadth of potential uses)

The following section describes the concept of secondary uses, referring broadly to data uses other than the direct care of an individual patient. Recent developments have greatly increased the scope and hence potential of such uses. For instance:

- The era of „Big data“ indicates the new sources and modes of data available. The report of the EC study [11] has alluded to the characteristics of volume, velocity and variety of big data and the importance of securing the quality and safety of new sources of data

- The opportunities for linking data sources which previously could not be combined through matching techniques and hence affording new opportunities to examine the cause of disease and the effectiveness of treatments

- The potential, across Europe, for analysing data on a much wider scale, particularly for areas such as rare diseases where the size of the population across the European Union has rendered such studies viable

- The increased availability of cost effective data storage presents the opportunity to retain richer data sets spanning greater periods

- At the same time, the push for transparency of data has further reinforced the opportunities and responsibilities of sharing the value of such anlaysis with a wider public.

2.5 The aim of the task and of the second deliverable

This deliverable D7.2.2 begins by considering both the opportunities and the responsibilities of the handling of health data. It then introduces a proposed approach, consisting of a set of principles, processes, standards, activities and compliance mechanisms. This is followed by consideration and recommendations on de-identification.

3. Opportunities for re-use of health data

3.1 Types of purpose and analysis

A considerable amount of information is collected during the provision of care and treatment, some of it specific to the patient being treated, some of it not. The „primary" purpose of this information is to support and improve individual patient care and much of it is held under professional and legal obligations of confidentiality. This information, often in conjunction with other records is of value for many other purposes to support healthcare. In practice, such „secondary" uses covers a very wide spectrum including:

- Improving the quality of local clinical care, for example through the audit of clinical practice
- Protecting the health of the public through surveillance and immediate response to infectious disease and other environmental threats to health, monitoring adverse effects of therapeutic interventions, and informing and evaluating screening
- Improving the management of the health system, for example by supporting the more efficient commissioning of services and payment by results
- Identifying patients who interact with multiple parts of the health system in order to monitor equity of access and provision
- Ensuring that health policy is evidence-based through carrying out empirical research
- Providing better information to the general public about healthy lifestyles
- Improving the quality and safety of care or reducing the impact of new risks to population health
- (through big data analytics) processing large amounts of data from multiple sources to predict best possible care protocols to individual patients in decision support applications
- Research by others using data collected by the care team but involving no contact with the team's patients
- Research which requires further contact with patients or former patients.

Figure 1 (overleaf) gives examples of such "secondary" activity.

There are many benefits from the secondary use of information. The health and well-being of the population may be improved by activities such as disease surveillance, screening and needs assessment and preventive activities. Secondary use creates possibilities for cost effective research and gives new possibilities for scientific innovations from new sources of large amounts of data such as bio-banks and data from genomic technologies. Secondary use of data that has been created during health care processes is needed for improving quality and safety of care. It is also important for clinical and medicine safety when assessing long term effects of drugs and other care methods.

Checking quality of care
• Testing the safety and effectiveness of new treatments and comparing the cost-effectiveness and quality of treatments in use
• Clinical audit (at either local or national level)
• Supporting audit studies for Cancer, Heart Disease, Diabetes, etc.
• Comparative performance analysis across clinical networks
• Ensuring the needs of patients within special groups are being met e.g. Children at risk, chronically sick, frail and elderly
Protecting the health of the general public
• Drug surveillance (pharmacovigilance) and other research-based evidence to support the regulators
• Surveillance of disease and exposures to environmental hazards or infections and immediate response to detected threats or events
• Vaccine safety reviews
• Safety monitoring of devices used in healthcare
• Linking with existing National Registries for diseases / conditions
• Analysis of outcomes following certain health interventions
• Monitoring the incidence of ill health and identifying associated behavioral or lifestyle risk factors;
• Identifying groups of patients most at risk of a condition that could benefit from targeted treatment or other intervention.
Managing healthcare spending
• Data for reimbursement
• Data for payment based around outcome indicators
• Detection of fraud and deliberate misuse of resources
Managing the health service
• Capacity and demand planning
• Commissioning
• Data for Standards and Performance Monitoring
• Clinical indicators
• Information to support the work of national regulators
• Evidence to support the work of developing national guidelines
• Measuring and monitoring waiting times
• Data to support Productivity Initiatives
• Benchmarking
• Data to support testing, validation and verification of new medical device features or services
Investigating concerns or complaints about healthcare
Teaching healthcare workers
Supporting research
• Assessing the feasibility of specific clinical trials designed to test the safety and/or effectiveness and/or cost-effectiveness of healthcare interventions;
• Assessing the impact of consent on recruitment and selection bias on clinical trial consenters and non-consenters
• Identification of potential participants in specific clinical trials, to seek their consent;
• Providing data from routine care for analysis according to epidemiological principles, to identify trends and unusual patterns indicative of more detailed research
• Providing specific datasets for defined approved research projects
• Biosurveillance
• Investigating and developing effective next generation processes for security, including anonymisation and de-identification"

Figure 1. Examples of Existing Secondary Uses of Data

It is critical to ensure that any access to data has a defensible legal basis, which can include legitimate use and compatibility test with primary use collection, appropriate technical, organizational and physical security controls, and appropriate contracts that apply data minimization principles (encryption, masking, deidentification) to ensure the data subject privacy and confidentiality is maintained and compliant with global regulatory or legal obligations for secondary uses.

4. Implications and Responsibilities

4.1 Legal and Regulatory context (e.g. GDPR, eIDAS)

The previous section identified some of the opportunities of analysing health data. It is imperative, however, that any such activity takes place only where there is a sound legal basis.

The Data Protection Directive (95/46/EC) on the protection of individuals with regard to the processing of personal data went a long way to harmonising the data protection frameworks across European Member States. It did however, leave some aspects open to local determination, and hence there were some inconsistencies in interpretation.

The General Data Protection Regulation [1] has brought much greater clarity and brought some new provisions (such as the right to be forgotten); it is important to plan future action in the light of the GDPR, even though full implementation is not due until 2018. Aspects such as a person's rights over who sees their personal data and the ability to check who has accessed their data will be difficult to meet unless facilities have been designed from the start. The ENISA study [8] provides a description of how privacy-enhancing technologies might be applied.

Similarly, the eIDAS regulation [4] brings a set of obligation as regarding the authentication and validation of both citizens and health professionals. This creates a basis of trust for the sharing of data, but this assumes a system of authorisation in which a user has legitimate reason to see data to which they have access.

The Directive on security of network and information systems (NIS Directive) [5] is the first piece of EU-wide legislation on cybersecurity; some of the implications are considered in the infrastructure section below.

4.2 Policy aspects (governance)

There is good understanding of the obligations within any organisation with regard to the handling of personal data and the mechanisms to be put in place to safeguard and assure the data within their care. Audits and, in some cases, regulatory activity may be applied to provide independent assurance.

As indicated in the previous section, there are increasing opportunities of combining data from different sources. As European projects such as EHR4CR and eTRIKS found, such collaborative efforts bring additional concerns and challenges over governance of data. Whilst the enabling secondary use of medical data by healthcare professionals and researchers is important to improve the quality of health care and research effectiveness, it is necessary at the same time to protect patient privacy and to ensure that no harm is done to a patient through the use of the data. The projects therefore worked to produce a Code of Practice [12] which aimed to provide a set of harmonised rules applicable to secondary use of medical data. It is intended to be useful to research projects involving multiple legal entities established in one or more EU member countries.

Without security there can be no privacy. Without assuring privacy through accountability and providing transparency, it will be difficult to establish an appropriate level of public trust.

In addition to ensuring data is secure in transit and at rest, good security practice requires maintaining appropriate levels of confidentiality, integrity and availability [35].

Privacy is ensuring the collection and use of information is appropriate, being transparent with individuals, and respecting their rights and choices made by individuals. Privacy can only exist if security is in place.

Ensuring responsible persons are accountable for assuring compliance with privacy and security principles, creates an appropriate environment for establishing and maintaining trust.

The process of de-identification, by which identifiers are removed from the health information, mitigates privacy risks to individuals, and thereby supports the secondary use of data for comparative effectiveness studies, policy assessment, life sciences research, and other purposes.

4.3 Information (data standards for consistency, aggregation)

The role of semantic standards to underpin cross-border exchange of data for individual patients is well-understood. Equally, the informed analysis of data in a comparable and meaningful manner is dependent on the application of standard terminologies in which the context and meaning are consistently understood.

4.4 IT infrastructure (cyber security, anonymisation)

It is increasingly understood that the risks of inappropriate disclosure require sophisticated security techniques and appropriate levels of de-identification of personal data to mitigate the threats.

For clarification, anonymization is distinct from de-identification. Anonymization has mainly technical implications. See

- the Article 29 working party opinion on anonymization [7]
- ENISA's documents on Privacy Enhancing Technologies [9]
- the ENISA report on Big Data privacy by design [8]

The NIS Directive will require other than security measures to be implemented, such as incident reporting schemes.

4.5 Applications (processing and data management)

A recent study by the OECD [14] identified a number of pertinent examples in both the opportunities and the precautionary measures needed to ensure a well-governed approach to data management.

- Ten countries have 70% or more of the national datasets necessary for understanding health care pathways and outcomes.
- Eight countries have independent research ethics review boards advise on decisions to process personal health data.
- Finland and Iceland publish approval decisions for individual data linkage projects on a website.
- Australia and United Kingdom (Scotland) have accreditation for health data processors that ensures high data protection standards are met.
- Nine countries provide a website where the process to follow to become approved to access to de-identified linked data is explained.
- The United States and the United Kingdom (England) consider the data security environment and the data use when deciding the degree of data de-identification required.
- Switzerland, the United States and the United Kingdom provide examples of engaging external experts to test data security.
- Secure, real-time, remote data access systems are available in the United Kingdom (Scotland and Wales), Netherlands and the United States and are being developed in Denmark and Korea.
- Secure research data centres are in use in Canada, Japan, Singapore, the Netherlands and the United States.
- Fourteen countries require a signed obligation to legally bind data recipients to the rules to be followed to protect the data
- A fine or criminal conviction can be imposed for deliberate misuse of data in Korea, Norway and the United Kingdom, and among statistical authorities in Canada and the United States**.**

Table 1. Examples of approaches to data management (Source OECD [14])

5. How to Handle Secondary Use of Health Data

The governance of the secondary use of health data can be achieved by using a set of agreed measures:

1) Adopt principles: the principles come from the EU General Data Protection Regulation [1]
2) Apply process: a structured process that reviews current and planned arrangements for handling of personal data. The practical handling of health data creates specific needs for solutions
3) Assess privacy risks: the application of the process and agreed controls will require specific considerations of individual systems and data flows
4) Agree controls: the application of the process needs to be undertaken in a consistent way including agreed controls (standards)
5) Assure compliance: and finally it is necessary to provide mechanisms for assuring that all plans have been completed and actions undertaken satisfactorily.

Figure 2 below provides an overview of the approach.

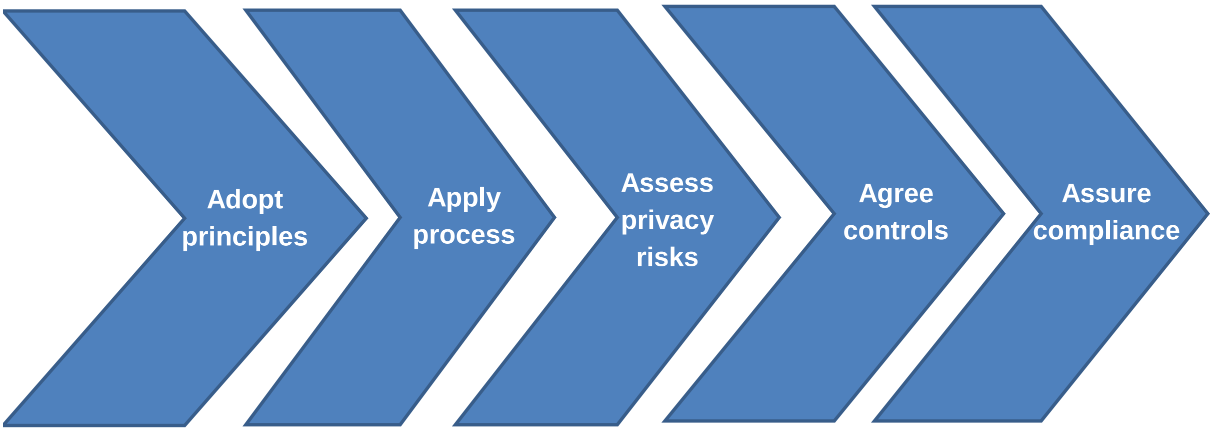


Figure 2. Outline of process steps

The following sub-sections develop each aspect of the process.

5.1 Adopt Principles

The starting point is the founding principles of handling personal data. These have been restated in the General Data Protection Regulation [1] as follows:

1.Personal data shall be:

(a) processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');

(b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');

(c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');

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

(e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');

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

2.The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ('accountability')

These have been included here in full as they set out the key concepts to underpin handling of personal data. It is important to note that the principles apply regardless of whether:

- The focus is the „primary" or „secondary" use of data
- The technical platform is private or public, cloud-based or not.

Increasingly the role of the patient is coming centre-stage, to ensure transparency and communications with patients. The ICO in the UK has produced guidance on Privacy notices, transparency and control [22].

5.2 Apply Process

The principles are well known and understood, but they are a starting point. The application of the principles requires a structured process that reviews current and planned arrangements for the handling of personal data. The approach below is taken from the UK's Health and Social Care Information Centre (HSCIC) code of practice [24] and consists of seven steps that may be applied to data and flows.

1. Establish the purpose of arrangements to handle confidential information
2. Establish and use standards for handling health data
3. Recognising objections to the handling of confidential information
4. Implement systems for handling confidential information
5. Adopt sound analysis of confidential data
6. Share information
7. Dispose of information once it is no longer required

The steps refer to "confidential" information. In general, there is a duty of confidentiality when one gives information to another person in circumstances where it is reasonable to expect that the information will be kept confidential. Special types of information are always considered confidential and this is the case when handling personal health data and data directly related to the health data.

It is important to consider the purpose for which information is collected and hence the purposes for which it may be used. The Article 29 working Party published an opinion on purpose limitation [6].

When applying the process it is important to know the different pieces of information that can make it possible to identify persons. Below is a list of data that can make it possible to identify a person:

- Name
- Address
- full post code
- date of birth
- health identifer, plus:
- Any other contact information that may allow them to be identified, for example, a phone number or email address, a photograph, video or audio tape or other image
- Anything else that may be used to identify them directly or indirectly, for example, rare diseases, drug treatments or statistical analyses within a small population.

The application of a process such as this from the start can enable privacy by design.

5.3 Assess Privacy Risks

The application of the process and the standards will typically require specific consideration of individual systems and data check flows. A Privacy Impact Assessment (PIA) [17] is a process which assists organisations in identifying and minimising the privacy risks of new projects or policies. A PIA will help to ensure that potential problems are identified at an early stage, when addressing them will often be simpler and less costly. The PIA process is a flexible one that can be integrated with an organisation's existing approach to managing projects. The time and resources dedicated to a PIA should be scaled to fit the nature of the project. A PIA should begin early in the life of a project, and run alongside the development process. Typically it would involve the following steps:

- Describe the information flows of the project. Explain what information is used, what it is used for, who it is obtained from and disclosed to, who will have access, and any other necessary information.
- Identify the privacy and related risks: Some will be risks to individuals, e.g. damage caused by inaccurate data or a security breach, or upset caused by an unnecessary intrusion on privacy. Some risks will be to the organisation, e.g. damage to reputation or the financial costs. Legal compliance risks include Data Protection and the Human Rights Act.
- Identify and evaluate the privacy solutions: Explain how to address each risk. Some might be eliminated altogether. Other risks might be reduced. Most projects will require you to accept some level of risk, and will have some impact on privacy. Evaluate the likely costs and benefits of each approach. Think about the available resources, and the need to deliver a project which is still effective.
- Sign off and record the PIA outcomes: Make sure that the privacy risks have been signed-off at an appropriate level. This can be done as part of the wider project approval. A PIA report should summarise the process, and the steps taken to reduce the risks to privacy. It should also record the decisions taken to eliminate, mitigate, or accept the identified risks. Publishing a PIA report will improve transparency and accountability, and lets individuals learn more about how your project affects them. However, no information on the level of detail that would help in planning a security breach should be published.
- Integrate the outcomes with delivery: The PIA findings and actions should be integrated with the project plan. It might be necessary to return to the PIA at various stages of the project's development and implementation. Large projects are more likely to benefit from a more formal review process.
- Consult with internal and external stakeholders as needed throughout the process, including appropriate consideration of
  - Communicating privacy information
  - Informing Individuals of their rights
  - Managing Consent
  - Managing subject access requests
  - Handling data breaches.

5.4 Agree controls

The application of the process needs to be undertaken in a consistent way and it would be helpful to frame this in the context of agreed controls (often using standards). The following have been proposed [26], acknowledging that the way that they apply will vary according to the type and size of organisation.

Leadership Obligation 1: People: Ensure staff are equipped to handle information respectfully and safely, according to the Principles. Data Security Standard 1. All staff ensure that personal confidential data is handled, stored and transmitted securely, whether in electronic or paper form. Personal confidential data is only shared for lawful and appropriate purposes Data Security Standard 2. All staff understand their responsibilities, including their obligation to handle information responsibly and their personal accountability for deliberate or avoidable breaches. Data Security Standard 3. All staff complete appropriate annual data security training and pass a mandatory test, provided through the revised Information Governance Toolkit [21]
Leadership Obligation 2: Process: Ensure the organisation proactively prevents data security breaches and responds appropriately to incidents or near misses. Data Security Standard 4. Personal confidential data is only accessible to staff who need it for their current role and access is removed as soon as it is no longer required. All access to personal confidential data on IT systems can be attributed to individuals. Data Security Standard 5. Processes are reviewed at least annually to identify and improve processes which have caused breaches or near misses, or which force staff to use workarounds which compromise data security. Data Security Standard 6. Cyber-attacks against services are identified and resisted and security advice is responded to. Action is taken immediately following a data breach or a near miss, with a report made to senior management within 12 hours of detection. Data Security Standard 7. A continuity plan is in place to respond to threats to data security, including significant data breaches or near misses, and it is tested once a year as a minimum, with a report to senior management.
Leadership Obligation 3: Technology: Ensure technology is secure and up-to-date. Data Security Standard 8. No unsupported operating systems, software or internet browsers are used within the IT estate. Data Security Standard 9. A strategy is in place for protecting IT systems from cyber threats which is based on a proven cyber security framework. This is reviewed at least annually. Data Security Standard 10. IT suppliers are held accountable via contracts for protecting the personal confidential data they process and meeting the National Data Guardian's Data Security Standards.

Table 2. Controls / standards to be applied

5.5 Assure Compliance

The actions described in the previous sections will lead to the development of plans and actions. It is necessary to provide mechanisms for assuring that such plans have been completed and actions undertaken satisfactorily. The application of a code of conduct with accompanying audit report might be one way of achieving this.

The World Health Organisation and several funders of public health research, led by the Wellcome Trust supported the development of a code of conduct [30] to encourage greater sharing of public health data. This code is relevant to all data of potential public health significance, focusing on research and routine data collection systems.

The eTRIKS project [12], funded by the EC produced a code of practice. Table 3 below is an extract from this, edited to reflect the requirements in scope for this paper.

	Rule
1	The Data Controller … shall verify that the data collection complied with the applicable legal and ethical requirements, and the secondary use / transfer of the data meet current legal and ethical requirements
2	The entity which provides personal data to recipient(s) of another legal entity for the purposes of a scientific project shall inform the project of any restriction of use or obligation applicable to these data
3	The secondary use of health data in scientific research projects shall be made with anonymised or pseudonymised data always when possible
4	... where a controller takes action in order to make personal data available to a third party located outside the European Union shall be considered a transfer in the meaning of Directive 95/46/EC Article 25 ... [to be updated to GDPR Article 44]
5	Pseudonymised data can be shared outside the EU if handled in line with 10.
6	Transfer of personal data to third parties shall comply with 1
7	… state of the art deidentification measures have to be taken to remove and/or to conceal sufficient direct and/or indirect identifiers
8	Initial pseudonymisation and anonymisation of directly identifiable medical data shall only be performed by someone bound to medical or clinical secrecy (e.g., a health care professional) or another person with the legal position and professionalism for such de-identification
9	Further de-identification of data shall only be done by persons authorised to have access ...
10	Personal Data which have been pseudonymised shall not be considered as anonymous data and thus rules on handling personal data still apply
11	Anonymisation of check personal data is considered further processing which is compatible with the purpose for which the data were originally processed and the identifying key destroyed.
12	Aggregated data are considered anonymous data provided safeguards are taken to avoid the risk of re-identification of data subjects (e.g., in case of low cell counts, rare disease).
13	Anonymised medical data can be used for research without consent or legal basis. They should however remain protected, considering future risks of identification
14	Data controllers prospectively collecting personal data shall inform the study participants comprehensively and appropriately about the …[arrangements]:
15	The data controller shall ensure that study participants provide their informed consent,

	preferably in writing, on the basis of the information defined in 14.
16	In order to enable new research on the basis of secondary use of medical data, consent forms used in projects collecting data should cover secondary use [see note 1 below]
17	Results or outcomes from a research project should be made available to study participants in a manner allowing non-specialists to understand the study results
18	The research project shall, where applicable, define rules to deal with incidental findings (e.g., to communicate the finding to the initial controller).
19	Personal medical data collected for health care purposes shall only be re-used for research projects if the secondary use is covered by the patient's consent, or is permitted by an applicable law [see note 1 below]
20	Personal medical data already lawfully collected for research purposes ... can be re-used in another project if the initial consent covers the possibility of re-use ... [see note 1 below]
21	Check Genetic data which is rich enough to single-out a person (e.g., a whole genome sequence) shall ... be subject to the safeguards applicable to personal data).
22	Genetic data which is rich enough to single-out a person shall be re-used for purposes not covered by the initial check consent only if the new purpose is of substantial public interest, and, data have been de-identified …
23	Study participants have the right to withdraw their consent at any time without justification.
24	If study participants withdraw their consent for use of their data, data and derived results shall be erased from all research project databases except where law requires (e.g. adverse event reporting).
29	Appropriate technical and organizational measures shall be implemented to protect personal data against accidental destruction or loss, alteration and unauthorized disclosure or access
30	The carrying out of processing by way of a processor shall be governed by a contract or legal act binding the processor to the controller
31	The controller shall ensure that the processor provides sufficient guarantees regarding technical and organizational security measures
32	Any secondary use or sharing of data with third parties shall be documented
33	Research data related to scientific publications should be retained long enough to ensure reproducibility and verifiability of the findings …
34	Retention periods should be defined before processing personal medical data according to its purpose and applicable law
35	Pseudonymised research data may be stored on the computer systems of member organisations or of their authorised partners as long as it may be required …
36	Anonymous research data may be stored as long as necessary.
37	Pseudonymised data shall not be publicly disclosed.
38	Anonymised data disclosure shall follow state of the art de-identification methods …
39	Accountability for maintaining disclosed data anonymous (i.e., not re-identifiable) shall be clearly assumed by the relevant data controller(s)
40	All partners of a collaborative scientific research project shall sign a binding agreement setting out different roles and responsibilities
41	Each entity processing personal medical data should … transpose personal data protection rules into binding requirements

Table 3. Edited extract from code for sharing of public health data

Note 1: care should be taken regarding the scope of consent. According to Art. 4 Nr. 11 GDPR a consent has to be for specific purpose. Recital 33 foresees that data subject can give its consent for a specific area of scientific research, if this is happening under the respective ethical standards for scientific research.

6. De-identification

The third strand of Joint Action task 7.2 is to provide recommendations on de-identification of data for secondary use. Many de-identification methods exists, including the ISO standard ISO/FDIS 25237 (Health informatics – Pseudonymization), regulatory agency anonymisation guidelines [21], and others published in scientific reviews [13] or on professional organisation‘s websites. The Article 29 Working Party have published an opinion paper on Anonymisation Techniques [7].

Anonymisation, De-identification and Pseudonymisation are well recognized terms documented within global laws and regulations as mitigating controls that enable privacy and confidentiality of individuals when applied appropriately. These terms are often used interchangeably but they are different. This report uses the definition provided in the OECD paper [16] in which de-identification means „a process by which a set of personal health data is altered, so that the resulting information cannot be readily associated with particular individuals“.

The OECD paper highlights the opportunities for deriving value from health and other data, as the volume and variety of data increase, and different data sets can be linked and merged across the many organisations that collect them.

Whilst the aim is to make the information no longer "about an identifiable individual", it is recognised that de-identified information can reveal a detailed profile and is not free from re-identification risks, depending on the motive, effort and skill set of the recipient of that data

In practice, de-identification is about the management of the risk of subsequent re-identification. Rather than a black or white approach, it is more of a spectrum of approaches, in which the core principle is that, having established the legal basis for data access, the minimum amount of identifiable data be provided to enable the business purpose to be supported. The risk of re-identification is dynamic and may change as technology improves and costs decrease. This requires balancing the level of de-identification, the motives and capacity to re-identify, the risk of invasion of privacy (potential impact on person’s privacy) and the level of controls/ security measures in place.

The following criteria may help to define which de-identification measures of health data shall be used.

- The legal basis for handling the data should be checked. If there is no legal justification, managing aggregated data such as data from public statistics is the only option.
- Consider the needs of the secondary data user-case. If the task can be performed with aggregated data that is the first option.
- If data on individual level (“row data”) is needed the first option is fully anonymised data.
- If there is a need to keep the possibility for re-identification for conducting future analysis or data linkage or informing the data subjects of specific conditions or research outcomes, the method of de-identification with pseudonymisation should be considered. The key can be kept by a third party.

- If there is a need for data linkage the first option is to do the linkage by the original data controllers and give the linked data to the secondary data user as pseudonymised data or fully anonymised data if no future de-identification is needed.
- The secondary use of health data without de-identification can be done only if no other option is possibly and all the possible safety measures are in use.

The application of the criteria may be assisted by considering factors such as:

- How many people have access to the data?
- Are these people bound to confidentiality?
- How high is the interest in re-identification?
- What is the level of the organizational and technical protection measures?
- How long will the data be stored?
- Will the data set be enriched over time?

Figure 4 below illustrates a decision tree to assist consideration of the release of anonymised data (derived from ICO guidance on Data Anonymization [19].

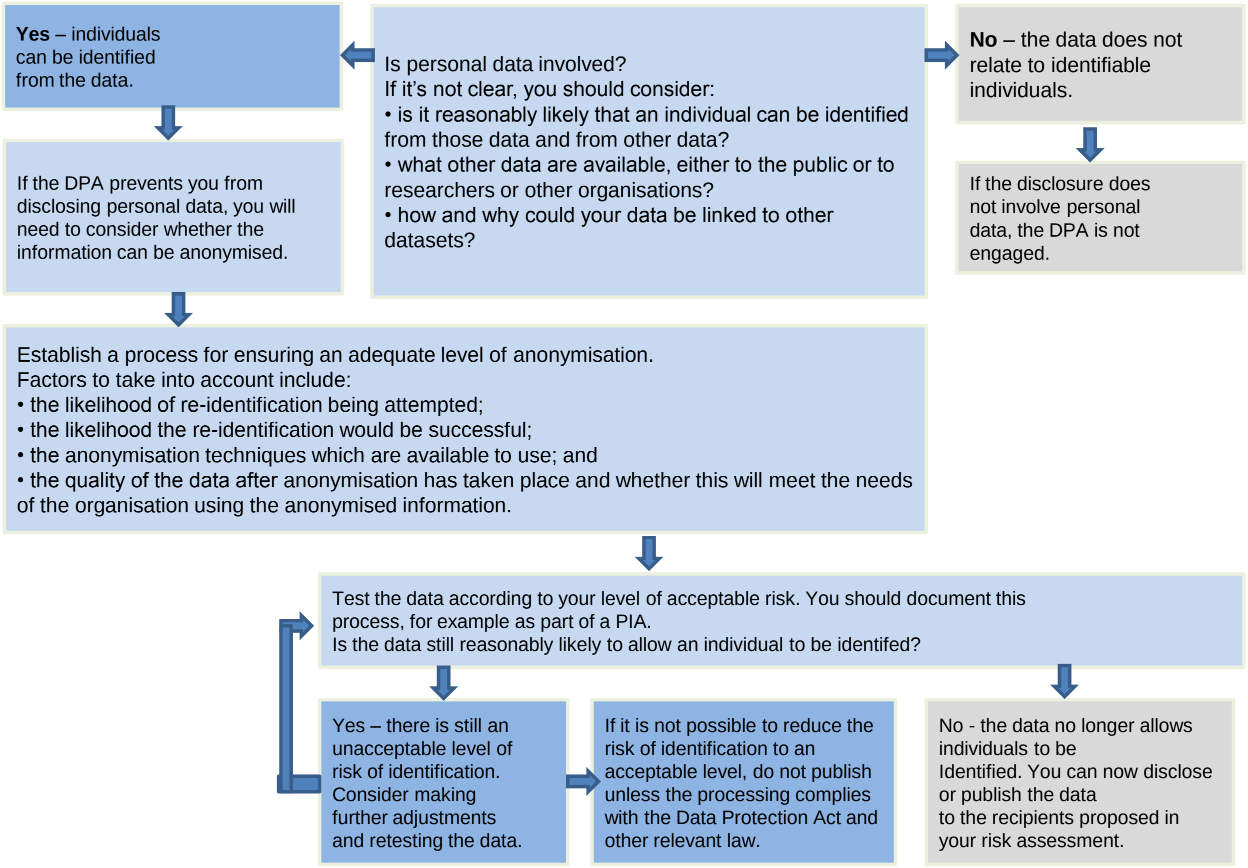

Yes – individuals can be identified from the data.
Is personal data involved? If it's not clear, you should consider: • is it reasonably likely that an individual can be identified from those data and from other data? • what other data are available, either to the public or to researchers or other organisations? • how and why could your data be linked to other datasets?
No – the data does not relate to identifiable individuals.
If the DPA prevents you from disclosing personal data, you will need to consider whether the information can be anonymised.
If the disclosure does not involve personal data, the DPA is not engaged.
Establish a process for ensuring an adequate level of anonymisation. Factors to take into account include: • the likelihood of re-identification being attempted; • the likelihood the re-identification would be successful; • the anonymisation techniques which are available to use; and • the quality of the data after anonymisation has taken place and whether this will meet the needs of the organisation using the anonymised information.
Test the data according to your level of acceptable risk. You should document this process, for example as part of a PIA. Is the data still reasonably likely to allow an individual to be identifed?
Yes – there is still an unacceptable level of risk of identification. Consider making further adjustments and retesting the data.
If it is not possible to reduce the risk of identification to an acceptable level, do not publish unless the processing complies with the Data Protection Act and other relevant law.
No - the data no longer allows individuals to be Identified. You can now disclose or publish the data to the recipients proposed in your risk assessment.

Figure 3. Deciding when and how to release anonymised data

Annex A: Source Materials

European Commission

[1] REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

[2] Brussels, 19.4.2016 COM(2016) 178 final COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS European Cloud Initiative - Building a competitive data and knowledge economy in Europe

[3] eHN paper : draft code of conduct on privacy for mobile health applications, May 2016

[4] The Regulation (EU) N°910/2014 on electronic identification and trust services for electronic transactions in the internal market (eIDAS Regulation) provides a predictable regulatory environment to enable secure and seamless electronic interactions between businesses, citizens and public authoritiesProfessional Qualification DIR

[5] DIRECTIVE (EU) 2016/1148 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union

European Article 29 Working Party

[6] Opinion 03/2013 on purpose limitation, ARTICLE 29 DATA PROTECTION WORKING PARTY 00569/13/EN, April 2013

[7] Opinion 05/2014 on Anonymisation Techniques, ARTICLE 29 DATA PROTECTION WORKING PARTY 0829/14/EN, April 2014

ENISA

[8] Privacy by design in big data, An overview of privacy enhancing technologies in the era of big data analytics, ENISA, December 2015

[9] Readiness Analysis for the Adoption and Evolution of Privacy Enhancing Technologies: Methodology, Pilot Assessment, and Continuity Plan, ENISA, December 2015

[10] Technical Guidelines for the implementation of minimum security measures for Digital Service Providers, ENISA, February 2017

European Projects

[11] Study on Big Data in Public Health, Telemedicine and Healthcare, *Final Report,* Gesundheit Österreich Forschungs- und Planungs GmbH, November 2016

[12] Code of Practice on Secondary Use of Medical Data in Scientific Reserach Projects, eTRIKS, 2014

[13] Code of practice on secondary use of medical data in European scientific research projects, Anne Bahr and Irene Schlünder

OECD

[14] OECD Health Policy Studies: Health Data Governance - PRIVACY, MONITORING AND RESEARCH, OECD, 2015

[15] Health Data Governance: Privacy, Monitoring and Research - Policy Brief October 2015

[16] Recommendation of the OECD Council on Health Data Governance, January 2017

UK Information Commissioner's Office, UK

[17] Conducting privacy impact assessments code of practice, ICO, November 2014

[18] How to disclose information safely: Removing personal data from information requests and datasets, ICO, October 2015

[19] Preparing for the GDPR: 12 STEPS TO TAKE NOW, ICO, 2016

[20] The ICO present a set of data protection principles: https://ico.org.uk/for-organisations/guide-to-data-protection/data-protection-principles/.

[21] They also offer guidance on Data Anonymization: https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf.

[22] Privacy notices, transparency and control: A code of practice on communicating privacy information to individuals, ICO, October 2016

Department of Health / Health and Social Care Information Centre, UK

[23] IG Toolkit: https://www.igt.hscic.gov.uk/

[24] Code of practice on confidential information, HSCIC, December 2014

[25] Information Management Strategy: Ensuring that all our staff are committed to the safe and efficient handling of information, HSCIC, May 2016

[26] Review of Data Security, Consent and Opt-outs, National Data Guardian for Health and Care, 2016

[27] Code of Practice on confidential information, HSCIC, 2014

Other UK

[28] Protecting Patient Confidentiality: NHS Scotland Code of Practice, 2012

[29] A guidance document published by the Privacy and Consumer Advisory Group presents nine principles for Identity Assurance https://www.gov.uk/government/groups/privacy-and-consumer-advisory-group

[30] Statement from WHO and Wellcome Trust in support of a code of conductregarding the use of data for research, https://wellcome.ac.uk/what-we-do/our-work/sharing-research-data-improve-public-health-full-joint-statement-funders-health

United States of America

[31] Guide to Privacy and Security of Electronic Health Information, US ONC, 2015

[32] Big Data and Privacy: A technological perspective, US President's Council of Advisors on Science and technology, 2014

[33] International Data Privacy Law (2015) 5 (4): 279-291 first published online September 19, 2015 doi:10.1093/idpl/ipv018

[34] The Open Group have documented a Cloud Ecosystem Reference Model (http://www.opengroup.org/cloud/cloud/cloud_ecosystem_rm/index.htm)

[35] Framework for Improving Critical Infrastructure Cybersecurity, Draft version 1.1, National Institute of Standards and Technology (NIST), January 2017, https://www.nist.gov/sites/default/files/documents/2017/01/30/draft-cybersecurity-framework-v1.1.pdf

Republic of Ireland

[36] International review of secondary use of personal health information, HIQA, Republic of Ireland, 2012

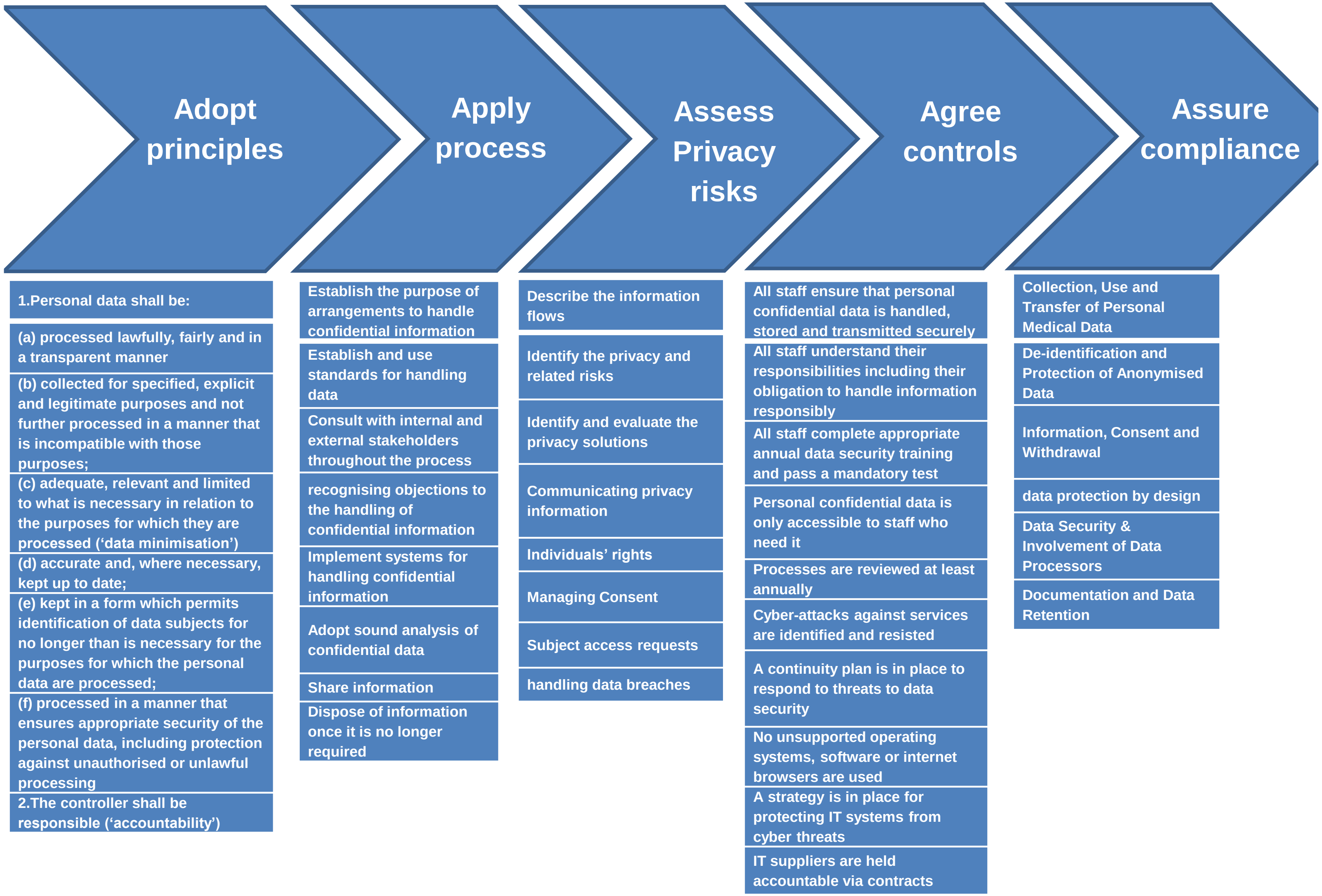
Annex B: Glossary (including source)

Definition	Source
AGGREGATED DATA: Data of several individuals that have been combined to show general trends or values.	WHO
Anonymisation: The process of rendering data into a form which does not identify individuals and where identification is not likely to take place.	ICO
Anonymised data: Data in a form that does not identify individuals and where identification through its combination with other data is not likely to take place.	ICO
Cloud services: Any resource that is provided over the internet.	NDG
Consent: The informed agreement for something to happen after consideration by the individual. For consent to be legally valid, the individual must be informed, must have the capacity to make the decision in question and must give consent voluntarily.	NDG
Cyber threat: The possibility of a malicious attempt to damage or disrupt a computer network or system.	NDG
Data breach: Any failure to meet the requirements of the Data Protection Act, including but not limited to an unlawful disclosure or misuse of personal data.	NDG
DATA CONTROLLER: The natural or legal person, or any other body, which alone or jointly with others determines the purposes and means of the processing of personal data	WHO
Data linkage: A technique that involves bringing together and analysing data from a variety of sources, typically data that relates to the same individual.	ICO
DATA PROCESSOR (or Processor): The natural or legal person, or any other body, which processes personal data on behalf of the controller.	WHO
Data protection: Technical and social regimen for negotiating, managing and ensuring informational privacy, confidentiality and security	NDG
Data quality: The correctness, timeliness, accuracy, completeness, relevance and accessibility that make data appropriate for their use.	NDG
Data security: Protecting data, such as a database, from destructive forces and from the unwanted actions of unauthorised users	NDG
Data sharing: The disclosure of data from one or more organisations to a third party, or the sharing of data between different parts of an organisation	NDG
Data subject: An individual who is the subject of personal data.	ICO
De-identification means a process by which a set of personal health data is altered, so that the resulting information cannot be readily associated with particular individuals.	OECD
De-identified: This refers to personal confidential data, which has been through anonymisation in a manner conforming to the ICO Anonymisation code of practice	NDG
Direct care: A clinical, social or public health activity concerned with the prevention, investigation and treatment of illness and the alleviation of suffering of individuals.	NDG
Disclosure: The act of making data available to one or more third parties.	ICO
Encryption: The process of transforming information (referred to as 'plain text' or 'in the clear') using an algorithm (called a 'cipher') to make it unreadable to anyone except those possessing special knowledge, usually referred to as a 'key'.	NDG
General Data Protection Regulation (GDPR): The General Data Protection Regulation (GDPR) is the EU Regulation 2016/679 adopted by the European Parliament and Council, to strengthen and unify data protection for individuals within the European Union.	NDG
Genome: The total genetic complement of an individual.	NDG
HEALTH DATA: Any data concerning patient‘s or study participant‘s health,	WHO

collected within the context of health care or clinical trials (e.g. name, address, living conditions, health data, life style habits, social security number, image data…)	(AMD)
Incident management: A term describing the activities of an organisation to identify, analyse and correct hazards to prevent a future re-occurrence.	NDG
Incident reporting: A method or means of documenting any unusual problem, occurrence, or other situation that is likely to lead to undesirable effects or that is not in accordance with established policies, procedures or practices.	NDG
Information Governance (IG): The set of multi-disciplinary structures, policies, procedures, processes and controls implemented to manage information at an enterprise level, supporting an organisation's immediate and future regulatory, legal, risk, environmental and operational requirements.	NDG
ISO/IEC27000 series: Information security standards published jointly by the International Organisation for Standardization (ISO) and the International Electrotechnical Commission (IEC) ISO 27799:2016(en) Health informatics — Information security management in health using ISO/IEC 27002	NDG
Linked data: The result of merging data from two or more sources with the object of consolidating facts concerning an individual or an event that are not available in any separate record.	NDG
Longitudinal study: A study that involves linking data about the same individual over a period of time, eg to study an individual's health episodes.	ICO
NISD: DIRECTIVE (EU) 2016/1148 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union	EC
Open Data: Data that meets the following criteria: • accessible (ideally via the internet) at no more than the cost of reproduction, without limitations based on user identity or intent; • in a digital, machine readable format for interoperation with other data; and • free of restriction on use or redistribution in its licensing conditions.	ICO
Opt-out: The option for an individual to choose not to allow their data to be used for the purposes described.	NDG
Personal Confidential Data (PCD): Personal information about identified or identifiable individuals, which should be kept private or secret. 'Personal' includes the DPA definition of personal data, but it is adapted to include dead as well as living people and 'confidential' includes both information 'given in confidence' and 'that which is owed a duty of confidence' and is adapted to include 'sensitive' as defined in the DPA	NDG
Personal data: Data which relate to a living individual who can be identified (a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.	ICO
Personal health data means any information relating to an identified or identifiable individual that concerns their health, and includes any other associated personal data.	OECD
Perturbation: The alteration of values within a data set to guard against data-linkage.	ICO
Pseudonymisation: The process of distinguishing individuals in a dataset by using a unique identifier which does not reveal their 'real world' identity.	ICO
Records Management: The practice of maintaining the records of an organisation from the time they are created up to their eventual disposal.	NDG

Re-identification means a process by which information is attributed to de-identified data in order to identify the individual to whom the de-identified data relate.	OECD
SECONDARY USE OF health DATA (or Data Re-Use): Processing of already existing medical data for a purpose different from the purpose for which they have been initially collected	WHO
THIRD PARTY: Any natural or legal person other than the data subject, the controller, the processor and the persons who, under the direct authority of the controller or the processor, are authorised to process the data.	WHO

Annex C: Overview



Adopt principles

- 1.Personal data shall be:
- (a) processed lawfully, fairly and in a transparent manner
- (b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes;
- (c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (‘data minimisation’)
- (d) accurate and, where necessary, kept up to date;
- (e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed;
- (f) processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing
- 2.The controller shall be responsible (‘accountability’)

Apply process

- Establish the purpose of arrangements to handle confidential information
- Establish and use standards for handling data
- Consult with internal and external stakeholders throughout the process
- recognising objections to the handling of confidential information
- Implement systems for handling confidential information
- Adopt sound analysis of confidential data
- Share information
- Dispose of information once it is no longer required

Assess Privacy risks

- Describe the information flows
- Identify the privacy and related risks
- Identify and evaluate the privacy solutions
- Communicating privacy information
- Individuals’ rights
- Managing Consent
- Subject access requests
- handling data breaches

Agree controls

- All staff ensure that personal confidential data is handled, stored and transmitted securely
- All staff understand their responsibilities including their obligation to handle information responsibly
- All staff complete appropriate annual data security training and pass a mandatory test
- Personal confidential data is only accessible to staff who need it
- Processes are reviewed at least annually
- Cyber-attacks against services are identified and resisted
- A continuity plan is in place to respond to threats to data security
- No unsupported operating systems, software or internet browsers are used
- A strategy is in place for protecting IT systems from cyber threats
- IT suppliers are held accountable via contracts

Assure compliance

- Collection, Use and Transfer of Personal Medical Data
- De-identification and Protection of Anonymised Data
- Information, Consent and Withdrawal
- data protection by design
- Data Security & Involvement of Data Processors
- Documentation and Data Retention