COMMENTS TO UK ICO ON
Draft Anonymisation, Pseudonymisation and Privacy Enhancing Technologies Guidance

By email to: anonymisation@ico.org.uk

On behalf of the International Pharmaceutical and Medical Device Privacy Consortium (IPMPC), MedTech Europe, and the Association of British HealthTech Industries (ABHI), we are pleased to submit these comments on the Information Commissioner's Office *Draft Anonymisation, Pseudonymisation and Privacy Enhancing Technologies Guidance*.

The IPMPC is comprised of chief privacy officers and other privacy, data protection and security professionals from a number of research-based, global pharmaceutical companies and medical device manufacturers. The IPMPC strives to be a leading voice in the global pharmaceutical and medical device industries to advance innovative privacy solutions to protect patients, enhance healthcare, and support business enablement.[1]

MedTech Europe is the European trade association for the medical technology industry including diagnostics, medical devices and digital health. MedTech Europe's mission is to make innovative medical technology available to more people, while helping healthcare systems move towards a sustainable path.[2]

ABHI is the UK's leading industry association for health technology supporting the HealthTech community to save and enhance lives. HealthTech plays a key role in supporting delivery of healthcare and is a significant contributor to the UK's economic growth.[3]

The medical technology ('medtech') and pharmaceutical industries share a long history of and commitment to advancing medical science. By turning scientific research into solutions for patients, healthcare professionals (HCPs), and health systems, the medtech and pharmaceutical industries have contributed to better outcomes for patients and greater efficiency in healthcare.

We welcome the efforts of the ICO to update its guidance on anonymisation, pseudonymisation, and privacy enhancing technologies. The updated guidance will bring further clarity on data protection requirements applicable to many of our research and development activities. In doing so, the guidance will help us achieve our ultimate goals of improving patient

[1] More information about IPMPC is available at www.ipmpc.org. This filing reflects the position of the IPMPC as an organization and should not be construed to reflect the positions of any individual member.

[2] More information about MedTech Europe is available at medtecheurope.org.

[3] More information about ABHI is available at https://www.abhi.org.uk/.

16 September 2022

outcomes, improving access of patients and health care providers to novel therapies, and increasing efficiency and long-term sustainability of the health care system.

Our comments are broken down by chapter. In addition, we provide a few general comments below:

- We support the inclusion of the various decision-tree flow charts throughout the chapters. These are very helpful tools to aid in summarizing the information presented and decisional processes described.
- We encourage the incorporation of additional applied examples throughout the chapters. We believe additional examples would aid in understanding how to apply the guidance in practice.
- It is important to recognise that both data controllers and processors may have legitimate interests in utilising anonymisation and pseudonymisation techniques to enable scientific research. The Guidance should provide examples of situations in which processors may need to use anonymised or pseudonymised data in order to improve their products and services and explain how such processing can take place in compliance with data protection requirements (*i.e.*, as a new, independent controller and in accordance with the permissions granted by the original controller). Such additional guidance would be especially useful for medtech companies, who frequently sell devices and services to hospitals and health systems but rely on the ability to conduct secondary data analyses in order to make product/service improvements.

We appreciate that the draft Guidance has been released by the ICO prior to the UK Government presenting to Parliament, on 18 July 2022, the Data Protection and Digital Information Bill, aiming to reform the UK GDPR and other privacy legislation. We share the UK Government's overarching objective to simplify the rules around research and to create a new pro-growth data protection framework that reduces burdens on businesses and boosts the economy. We believe that a clearer regulatory environment for personal data use will allow for responsibly harnessing the power of data to advance research and innovation in medical technologies and services, thereby benefiting British citizens and the health system. Clear rules on anonymisation, pseudonymisation, and privacy enhancing technologies will be an important part of this new framework.

For this reason, we encourage the ICO to modify the Guidance to reflect the clarification of the concept of 'personal data' in line with the Bill. This clarification is consistent with the existing law but important to make clear that organisations only need to consider identifiability (i) at the time of processing rather than to speculate as to future risks, and (ii) within a reasonable degree of parties involved or that may reasonably access the data as a result of the processing. We see this as a welcome development that has potential to remove part of the legal uncertainty currently surrounding the use and sharing of data for research purposes.

Chapter 1: Introduction to anonymisation

- **Page 9:** We support the ICO's view that 'the same information can be personal data to one organisation, but anonymous information in the hands of another organisation. Its

status depends greatly on its circumstances, both from your perspective and in the context of its disclosure'. This is a critical concept but not one which is well understood among many organisations. We encourage the ICO to include additional examples of scenarios in which information may be personal data to one organization but anonymous information in the hands of another. Useful data sharing scenarios to include and address whether it might be possible to structure the activity so that the recipient is viewed as receiving anonymous information include:

- A scenario in which information is personal data as to the controller but a service provider who needs access to the data has no reasonable means to identify data subjects.
- A clinical trial scenario in which a sponsor of the trial has no ability to re-identify participants in the normal course aside from very limited circumstances (i.e., on-site monitoring of trial sites with prohibitions on taking identifiable data off-site, and compensating participants in the event of subject injuries).
- One or more scenarios addressing sharing of pseudonymized health data for research purposes and in which the recipient has no access to the re-identification key. For example, it would be helpful to include a scenario addressing the use of longitudinal patient data for further research and innovation - presenting the need to have identifiers in the datasets that safely allow a third party to link data to individual patients over time and possibly also over different therapies and providers. This may be essential to determine outcomes achieved or relationships between treatment and pre- or post-treatment factors, as well as to identify potential bias in treatment. This type of research is increasingly important for improving and 'personalizing' therapies using more refined understanding of disease and treatment pathways harnessing large data sets.

- **Page 12**: We understand that the ICO views applying anonymisation techniques to turn personal data into anonymous information as a data processing activity that itself is subject to compliance with data protection requirements. However, it should also be emphasised that anonymisation is a privacy-preserving technique, and such processing should typically be viewed as 'compatible' with the initial purpose of collection (and, to the extent a separate legal basis for processing is necessary, an organisation's legitimate interests in anonymising data to enable secondary uses will typically override other interests).[4]

[4] This is consistent with the historical views of the Article 29 Working Party when examining these issues under Directive 95/46/EC. The Working Party states in its Opinion on Anonymisation: 'Anonymisation constitutes a further processing of personal data; as such, it must satisfy the requirement of compatibility by having regard to the legal grounds and circumstances of the further processing.' And later, '[T]he anonymisation process, meaning the processing of such personal data to achieve their anonymisation, is an instance of "further processing". As such, this processing must comply with the test of compatibility in accordance with the guidelines provided by the Working Party in its Opinion 03/2013 on purpose limitation. This means that, in principle, the legal basis for anonymisation can be found in any of the grounds mentioned in Article 7 [of the Data Protection Directive] (including the data controller's legitimate interest) provided the data quality requirements of Article 6 of the Directive are also met and

- This is especially true in the context of scientific research, where anonymisation of personal data should always be considered compatible with the initial purpose of processing and consistent with the public interest in advancing scientific understanding.

Chapter 2: How do we ensure anonymisation is effective?

- **Pages 5-6:** We are concerned that the explanation of 'singling out' and 'linkability' in the discussion concerning 'key indicators of identifiability' may lead to overly broad interpretations of what is 'personal data'. The concepts of 'singling out' and 'linkability' are useful in determining 'identifiability', but they must be appropriately contextualized. Whereas the draft Guidance states that one needs to consider information other than identifiers and how such information may be used to provide context that can single out an individual, it is also true that data fields that are possible identifiers in one context may not be in another. 'Identification' of a data subject (*i.e.,* discovery of the 'identity' of a data subject) ultimately involves more than simply being able to differentiate records relating to one otherwise anonymous individual in a dataset from records relating to another otherwise anonymous individual. Just like identifiability lives on a spectrum, so does the analysis of whether 'singling out' can render an individual 'identified' or 'identifiable'. For example, information should not necessarily be considered 'identified' or 'identifiable' by virtue of the fact that unique, random numbers have been assigned to *distinguish* data about one person from another. A medical record that has been otherwise stripped of identifiers does not become identifiable merely because a unique, random code has subsequently been assigned to it. A DNA sequence, although potentially unique to an individual, should not be deemed to identify that individual in the absence of a database or similar available record source that links this sequence to the 'identity' of the individual (*i.e.*, some socially consequential distinguishing characteristics, such as a name or contact information) and to which the researcher has access.

  To be clear, we are not suggesting that in order for data to be considered 'identifiable', it must ultimately be possible to link or re-link the data to a named individual. However, we suggest to allow for a contextual, risk-based assessment of considerations such as the impact to or ability to impact a specific individual (e.g., to make contact with or to make decisions about) in order for 'identification' of an individual to have any useful meaning.[5]

---

with due regard to the specific circumstances and all the factors mentioned in the Working Party's opinion on purpose limitation. . . . [T]he Working Party considers that anonymisation as an instance of further processing of personal data can be considered to be compatible with the original purposes of the processing but only on condition the anonymisation process is such as to reliably produce anonymised information in the sense described in this paper.'

[5] For example, the *Guide to Basic Anonymisation*, as recently released by the Singapore PDPC, in its Annex D, 'Assessing the risk of re-identification', allows for such an approach: "*The main factors affecting the risk threshold should include the harm that could be caused to the data subject, as well as the harm to the organisation, if re-identification takes place; but, it also takes into consideration what*

Proposed Language:
What is 'singling out'?
You need to consider whether singling out is reasonably possible, both by you and by another party. Understanding that identifiability runs on a spectrum, consideration should be given to whether such singling actually identifies an individual, and if so, whether it may negatively impact the individual. This should be part of your assessment of the effectiveness of your anonymisation processes.

What is 'linkability'?
Common techniques to mitigate linkability include masking and tokenisation of seemingly identifying key variables, for example, sex, age, occupation, place of residence, country of birth. Where these mitigating processes can support linkability of effectively anonymized data sets (that is, data sets where the individuals cannot be identified), use of linked data sets may be permissible subject to additional risk assessments, as may be necessary.

- **Pages 13-14:** We agree that an assessment of identifiability risk involves a contextual analysis, and the level of identifiability risk that is acceptable may depend on whether there are restrictions on recipients' uses and sharing, the sensitivity of the data, and similar factors. It would be helpful to include – or even just to cite – some examples of levels of risk that have been considered acceptable in past scenarios. We would find it helpful to understand, for example, acceptable re-identification risk thresholds for public data releases versus non-public data releases of health information. There are established risk thresholds adopted in other jurisdictions which can be used as a guide. For example, in the U.S., the Centers for Medicare & Medicaid Services has used an acceptable reidentification risk at 0.04[6]. Health Canada uses a reidentification risk of 0.09 in one guidance[7], and the European Medicines Agency sets a 0.09 reidentification threshold[8].[9]

other controls have been put in place to mitigate any residual risks. The higher the potential harm, the higher the risk threshold should be."

[6] See U.S. Centers for Medicare & Medicaid Services, 2008 Basic Stand Alone Medicare Claims Public Use Files.

[7] See https://www.canada.ca/en/health-canada/programs/consultation-public-release-clinical-information-drug-submissions-medical-device-applications/draft-guidance.html

[8] See https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data_en-1.pdf

[9] For additional guidance, we also recommend referencing *Anonymizing Health Data: Case Studies and Methods to Get You Started* by Khaled El Emam & Luk Arbuckle and *De-identification Methods for Open Health Data: The Case of the Heritage Health Prize Dataset* by Khaled El Emam. El Emam's books provide case studies for conducting thorough risk assessments in the health care space and address, inter alia, longitudinal studies, EMR records, and secure linkability of medical records.

- **Page 18:** Regarding the example included concerning the sharing of health data, it would be useful to further expand on the context and protections required. We suggest including a scenario in which a hospital shares patient data with a third-party that intends to use the data for research and development purposes (processing within that third party's environment, not to be shared further), and the additional information required to be able to identify data subjects is not allowed to be shared with that third party and/or not part of the data agreed to be shared with that third party.

- **Page 19:** We agree with and support the inclusion of the statement that '[i]t is reasonable to conclude that certain professionals with prior knowledge, are not likely to be motivated intruders (e.g., doctors). This could apply where it is clear that the profession in question imposes confidentiality rules and requires ethical conduct.' We also agree that 'a relevant factor is whether someone would learn anything new.' These are, again, critical points where we often find that there is confusion. A doctor, for example, might recognise a patient record that has been anonymised for public disclosure as relating to his/her own patient. One should not impute from this that the anonymisation was necessarily flawed/inadequate, especially where the doctor is not learning anything new. As noted by the ICO on page 12, 'effective anonymisation is about finding the right balance', and where the recipient already has access to the data, the overall risk is clearly mitigated by existing knowledge.

- **Page 20:** We encourage some further clarification of the section concerning 'What is the difference between information, established fact and knowledge?'. Because these terms are not defined, they may be subject to different interpretations. For example, we interpret this section as drawing a distinction between publicly available information, proprietary information, and first-hand and second-hand personal knowledge. If that is correct, some further discussion of these concepts and how they impact the 'motivated intruder' test would be helpful.

- **Pages 22-23:** We agree with the discussion concerning how to assess whether disclosure of data between organisations qualifies as a disclosure of personal data. We encourage the inclusion of additional examples, in particular examples addressing sharing among corporate affiliates. This would help to better understand what technical and organisational controls can be put in place to avoid a disclosure of data from one legal entity to an affiliated legal entity being viewed as a disclosure of personal data. For example, the U.S. NIST Framework uses the model of 'trusted data recipient', which it defines as recipients who are bound by additional administrative controls such as data use agreements and/or who are bound by confidentiality or ethical rules within their professions (e.g., doctors or attorneys). It would be reasonable for the UK to consider and adopt a similar concept.

Chapter 3: Pseudonymisation

- **Pages 15-17:** The discussion concerning 'How should we approach pseudonymisation?' identifies a number of steps for controllers to take to establish that personal data has been

adequately pseudonymised. We encourage the ICO to clarify that an approach to pseudonymisation can cover a number of similar processing activities that present similar data protection risks without requiring separate documentation to support the pseudonymisation of each dataset.[10] For example, a sponsor of clinical studies may adopt a process for pseudonymisation of the data collected at the study sites and reported to the sponsor via case report forms. A single pseudonymisation procedure should be able to be followed for all clinical studies that present similar risks.

Chapter 4: Accountability and governance

- As a general comment, we would encourage the ICO to provide, in addition to the theoretical examples, some practical use cases. We believe this could help organisations to better understand this chapter.

- **Page 7 (How should we work with other organisations, where necessary?) and 14 (How should we mitigate re-identification risk due to a security incident?):** The ICO should provide further guidance on what recipients of anonymised data sets are obligated to do to ensure that the data is/remains anonymous. In these situations, the provider of the data set controls the anonymisation procedure. Is it sufficient for the recipient to rely on the provider's representations?

- **Pages 11-12:** The ICO should provide examples of notices that provide acceptable transparency concerning an organisation's anonymisation practices. As the ICO appears to recognise in Chapter 2 of the draft guidance, the approach that an organisation takes to anonymisation of a particular dataset involves a contextual analysis. While an organisation may adopt a standard approach for a particular type of dataset, across the many types of datasets processed by an organisation, there may be many different approaches. Is it acceptable for an organization to simply state that it follows relevant anonymisation guidance, including the UK ICO guidance? Further, it is important that the guidance provide examples of the level of specificity that is needed with respect to the explanation for anonymising individuals' personal data.

- **Pages 12-13:** The draft guidance states that '[i]t is important that your members of staff who are involved in decisions about creating and disclosing anonymous information have a clear understanding of the anonymisation techniques you use; any risks involved; and how to mitigate these risks.' We encourage the ICO to provide further information on what background qualifications are needed by individuals performing anonymisation functions, including those responsible for assessing reidentification risks.

Chapter 5: Privacy-enhancing technologies (PETs)

- As general comments, we endorse the premise that PETs have the ability to mitigate security and privacy concerns, promote data protection principles, and, in some cases, achieve compliance with data protection law. We also recognize that NHS is using PETs

[10] Ideally, this should also be applicable to anonymisation processes as well.

for linking patient data across different organizational domains. The inclusion of more such example use cases of governmental organizations successfully using PETs would encourage industry to further embrace and study these PETs. Finally, we encourage the ICO to standardize these PETs so that they can be easily applied in the business environment.

- **Pages 35-36:** With regard to synthetic data, we encourage the ICO to recognise that such data can be used in a broad range of scenarios, including – but not limited to:
    - Product improvement, development, and testing: Synthetic data can be used in situations where data is needed for testing a product to be released but such data does not yet exist or is not available to the testers.
    - Machine learning.
    - Clinical trials: Synthetic data can be used as a baseline for future studies and testing when no real data yet exists. The US Food and Drug Administration (FDA) has already approved a number of studies where synthetic data has been used as a control arm.
    - Marketing: Synthetic data can be used by marketing units to run detailed, individual-level simulations to improve their marketing spend.

  Given its broad range of possible uses and benefits, we encourage the ICO to provide further clarification on the legal implications of the creation and use of synthetic data. In particular, further guidance would be helpful on what data protection requirements apply to the process of creating synthetic data sets from original data sets.

We appreciate the opportunity to provide comments on the draft guidance. Please do not hesitate to reach out to us should any of our comments require further clarification.

About IPMPC

The IPMPC is a trade association representing multinational pharmaceutical, biotech, and medical device companies. The IPMPC strives to be a leading voice in the global pharmaceutical and medical device industries to advance innovative privacy solutions to protect patients, enhance healthcare, and support business enablement.

www.ipmpc.org

For more information, please contact Peter Blenkinsop at the IPMPC Secretariat (peter.blenkinsop@faegredrinker.com).

About MedTech Europe

MedTech Europe is the European trade association for the medical technology industry including diagnostics, medical devices and digital health. Our members are national, European and multinational companies as well as a network of national medical technology associations who research, develop, manufacture, distribute and supply health-related technologies, services and solutions.

www.medtecheurope.org.

For more information, please contact Aline Lautenberg, General Counsel (a.lautenberg@medtecheurope.org).

About ABHI

ABHI is the UK's leading industry association for health technology supporting the HealthTech community to save and enhance lives. HealthTech plays a key role in supporting delivery of healthcare and is a significant contributor to the UK's economic growth.

www.abhi.org.uk

For more information, please contact Andrew Davies, Digital Health Lead (Andrew.Davies@abhi.org.uk).