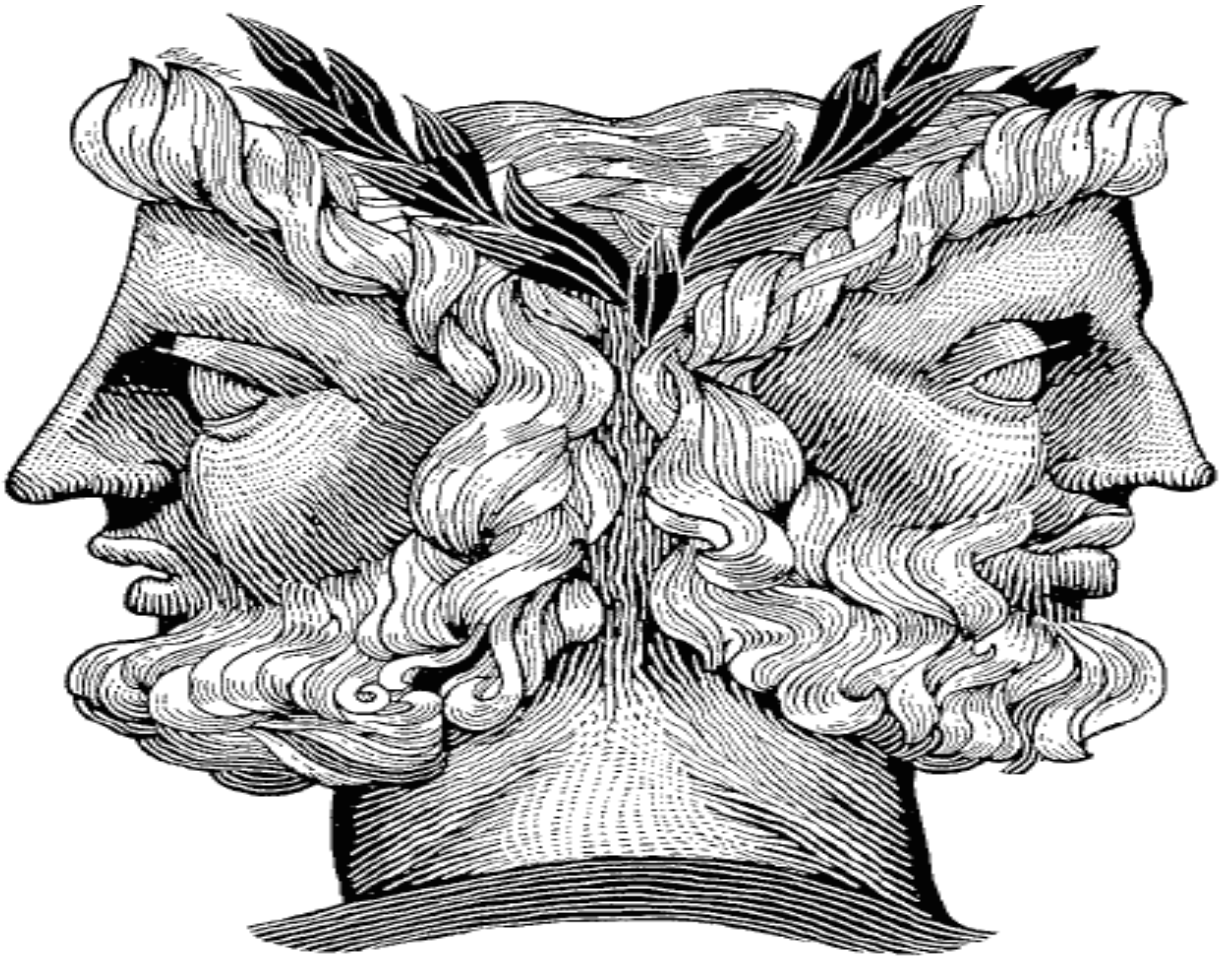
ANONYMIZING YOUR DATASETS: MYTHS & REALITIES

Antonios Roussos
Head of Global E&C Data Privacy & Group DPO
Astellas Pharma Inc.
November 2023

THE CURRENT PRIVACY LANDSCAPE



THE CURRENT PRIVACY LANDSCAPE

Looking Backwards

Looking Forwards

Looking Forwards

Data Governance, Ethics & Privacy

astellas

EU DIGITAL & DATA INITIATIVES: A POLICY MAPPING SNAPSHOT

European Health Data Space
De facto compulsory Electronic Health Records; infrastructure frameworks for maximizing interoperability across member states and data sharing standards

NIS 2 Cybersecurity Directive
The "GDPR" for data security; strengthening security requirements and addressing the security of supply chains. Requiring individual companies to address cybersecurity risks

Artificial Intelligence Act
Develop a European regulatory framework on Artificial Intelligence (AI). It uses a risk-based approach to classify AI systems under four categories ranging from minimal risk to unacceptable risk.

ePrivacy Regulation
A proposal for the regulation of various privacy-related topics, mostly in relation to electronic communications within the European Union, reinforcing trust and security in the digital world

Data Act
A horizontal legislative proposal, affecting all industries similarly. Fostering data-sharing with governments. Increasing the sharing of data generated by devices and machines, including those used in healthcare and research settings

Data Governance Act
It encourages wider reuse of data including personal data held by public sector bodies; using secure processing environments and anonymization techniques

Digital Markets Act
New rules for platforms that act as 'gatekeepers'. The proposal addresses three main problems: high barriers to entry, anti-competitive practices by gatekeepers and fragmented regulation and oversight

Digital Services Act
Regulating the operation of commercial platforms and services they provide, including services offered on the internet and aims to ensure transparency, accountability and regulatory oversight of the EU online space

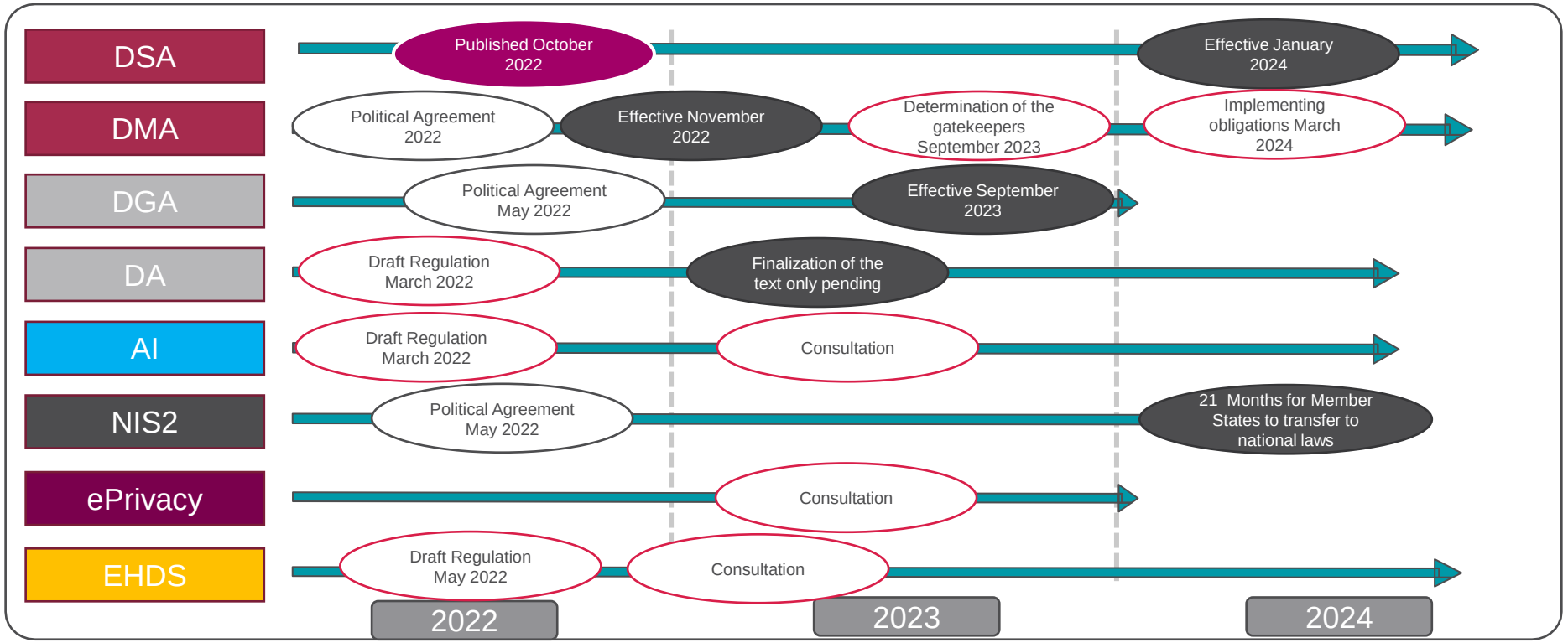
Anticipated major impact in the life science sector

Anticipated medium impact in the life science sector

Anticipated lower impact in the life science sector

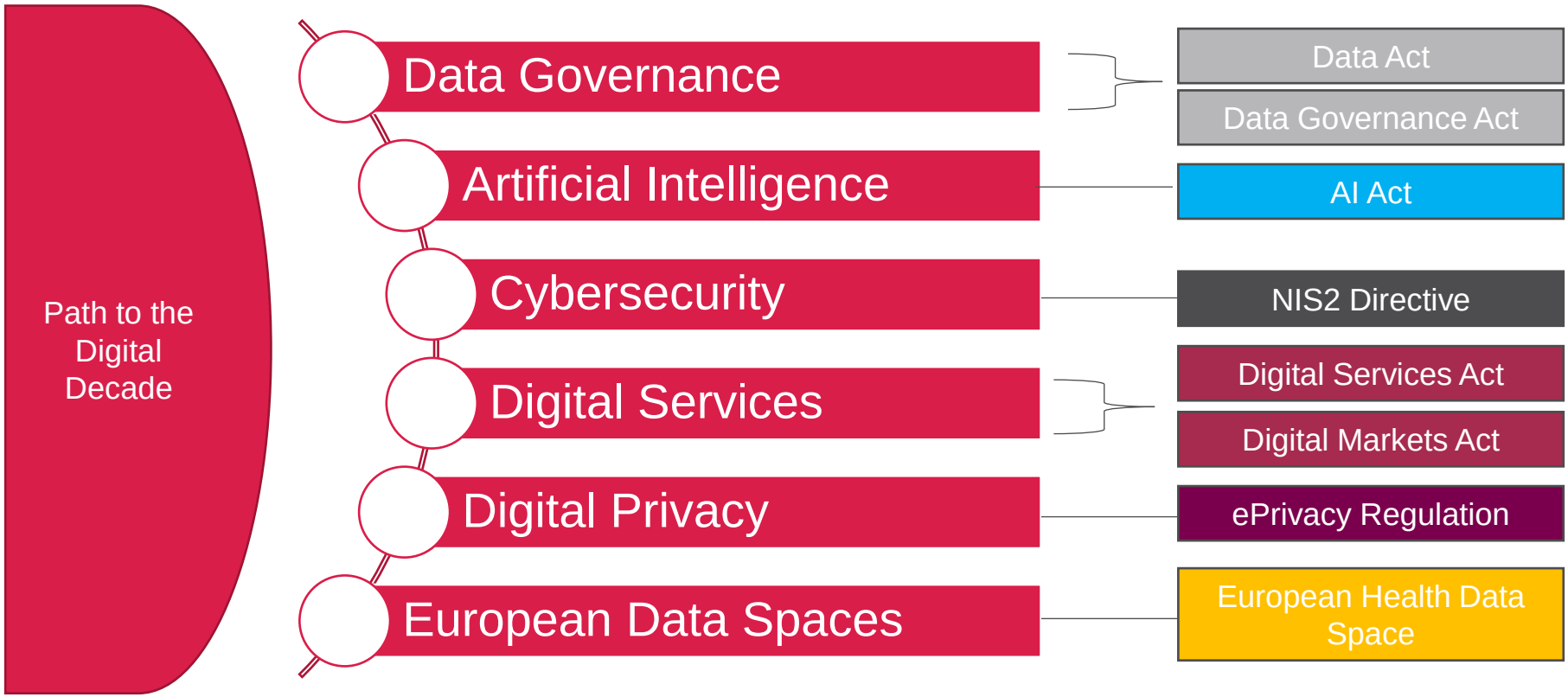
astellas

AN IMPLEMENTATION TIMELINE TO REMEMBER!



NAVIGATING THROUGH THE LABYRINTH!

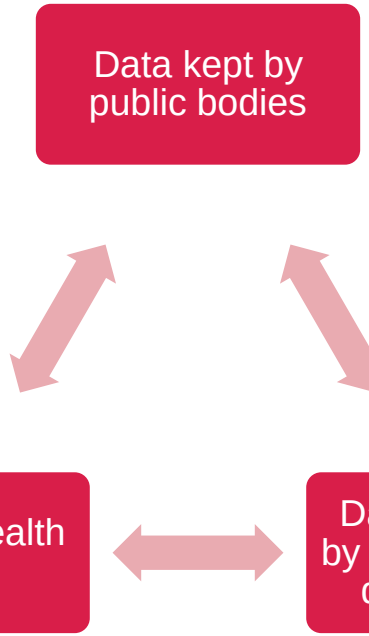
***A classification model to remember the basic categories of those** data & digital initiatives*



WHAT IS THE ANTICIPATED IMPACT ON DATA SHARING IN THE EU?

New opportunities to identify and leverage new data sources in at least three important areas:

Data Governance Act
Effective September 2023



European Health Data Space
Still in consultation

Data Act
Effective within 2025

Looking Backwards

Back to basics: What is personal data and what is anonymous in the era of Artificial Intelligence?

astellas

PERSONAL OR ANONYMOUS? CASE 1

A press release from the EU Anti Fraud Agency (OLAF):

- Uploaded in OLAF's website
- ANONYMOUS
- Leading a research project
- Funded by ERCEA with a specific amount
- In a Greek University
- Where her father was also employed as Professor

Mike Flanagan via CartoonStock - https://www.cartoonstock.com

* CaseT-384/2020, *OC vs. EU Commission* *(only in French)*

PERSONAL OR ANONYMOUS? CASE 2

A doctor uploads a video in YouTube recording a patient surgery*

- Without reference to the

of

- Place and Hospital
- Type of surgery
- Age of patient
- Name of surgeon and his associates

PERSONAL OR ANONYMOUS? CASE 3

An online Form sent by a European Agency to the shareholders and creditors of a Spanish financial institution going through an insolvency procedure*

- 

ıny
- nd
- s to
d did
- PS
- The EDPS found that the replies/comments contained at least “pseudonymized” and not anonymized data
- Appeal before the General Court

Benjamin Schwartz, The New Yorker

* CaseT-557/2020, *SRB vs. EDPS*

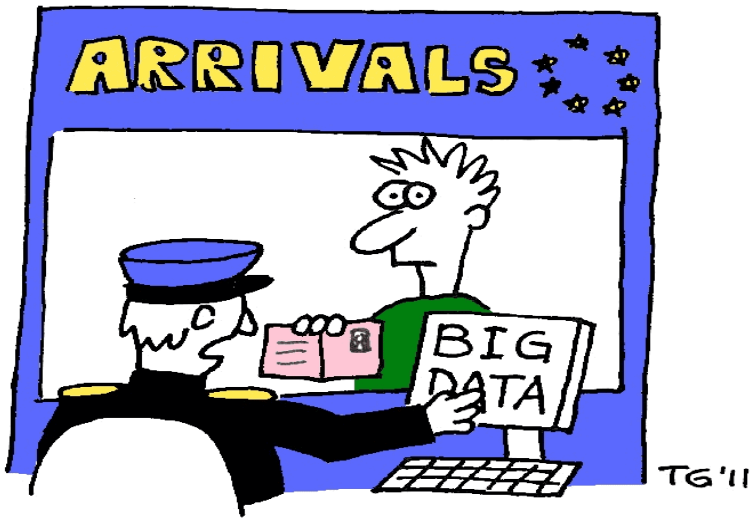
PERSONAL OR ANONYMOUS? CASE 3

Latanaya Sweeney's research*

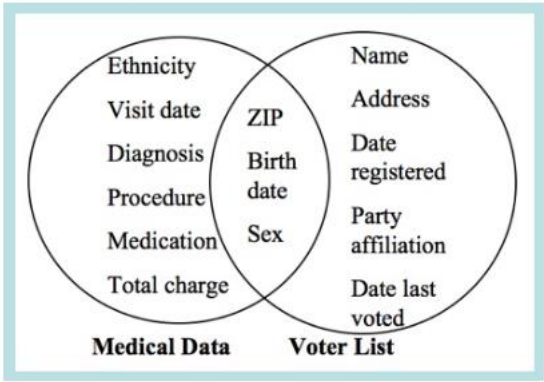
- Latanaya used as the basis
- sex,
- re-

registration list)

- **Reidentification success rate was 87%!**



"Your recent Amazon purchases, Tweet score and location history makes you 23.5% welcome here."



* L. Sweeney. *k*-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10 (5), 2002; 557-570, available here

astellas

WHAT IS ANONYMIZATION IN EUROPE?

Stockphoto

THE ROOT CAUSE OF THE PROBLEM 1/2

GDPR, Recital 26: The "reasonably likely" criterion is not really very helpful!

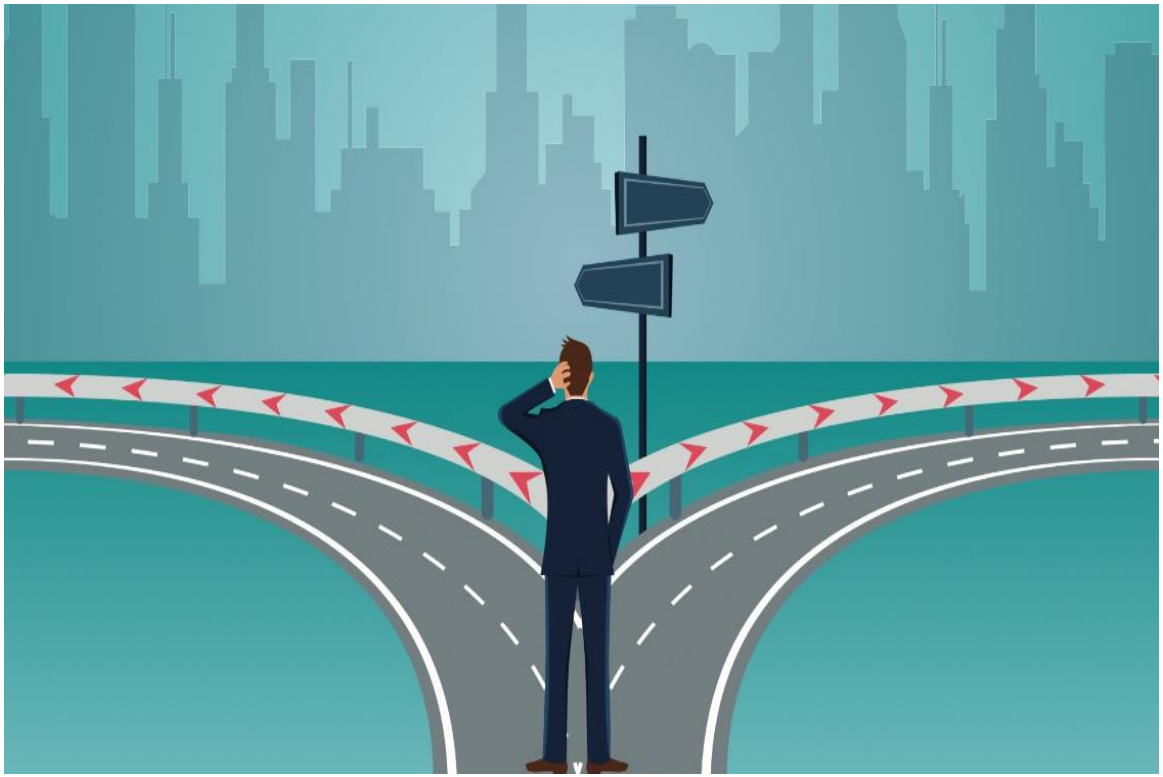
(26) The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

THE ROOT CAUSE OF THE PROBLEM 2/2

Two different methods to determine if identifying an individual is *"reasonably likely"*:

Absolute Anonymization

Re-identification of individuals is impossible!



Relevant Anonymization

Possibility to re-identify an individual cannot be excluded

THE FIRST OPINION OF THE WP29 IN 2007

Opinion 4/2007 on the concept of personal data

Relevant Anonymization

Anonymous data

"Anonymous data" in the sense of the Directive can be defined as any information relating to a natural person where the person cannot be identified, whether by the data controller or by any other person, *taking account of all the means likely reasonably to be used either by the controller or by any other person* to identify that individual. "Anonymised data" would therefore be anonymous data that previously referred to an identifiable person, but where that identification is no longer possible. Recital 26 also refers to this concept when it reads that "*the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable*". Again, the assessment of whether the data allow identification of an individual, and whether the information can be considered as anonymous or not depends on the circumstances, and a case-by-case analysis should be carried out with particular reference to the extent that the means are likely reasonably to be used for identification as described in Recital 26. This is particularly relevant in the case of statistical information, where despite the fact that the information may be presented as aggregated data, the original sample is not sufficiently large and other pieces of information may enable the identification of individuals.

In other areas of research or of the same project, re-identification of the data subject may have been excluded in the design of protocols and procedure, for instance because there is no therapeutical aspects involved. For technical or other reasons, there may still be a way to find out to what persons correspond what clinical data, but the identification is not supposed or expected to take place under any circumstance, and appropriate technical measures (e.g. cryptographic, irreversible hashing) have been put in place to prevent that from happening. In this case, even if identification of certain data subjects may take place despite all those protocols and measures (due to unforeseeable circumstances such as accidental matching of qualities of the data subject that reveal his/her identity), the information processed by the original controller may not be considered to relate to identified or identifiable individuals taking account of *all the means likely reasonably to be used by the controller or by any other person*. Its processing may thus not be subject to the provisions of the Directive. A different matter is that for the new controller who has effectively gained access to the identifiable information, it will undoubtedly be considered to be "personal data".

astellas

Opinion 5/2014 on Anonymization Techniques

Absolute Anonymization

Secondly, "the means likely reasonably to be used to determine whether a person is identifiable" are those to be used "by the controller or by any other person". Thus, it is critical to understand that when a data controller does not delete the original (identifiable) data at event-level, and the data controller hands over part of this dataset (for example after removal or masking of identifiable data), the resulting dataset is still personal data. Only if the data controller would aggregate the data to a level where the individual events are no longer identifiable, the resulting dataset can be qualified as anonymous. For example: if an organisation collects data on individual travel movements, the individual travel patterns at event level would still qualify as personal data for any party, as long as the data controller (or any other party) still has access to the original raw data, even if direct identifiers have been removed from the set provided to third parties. But if the data controller would delete the raw data, and only provide aggregate statistics to third parties on a high level, such as 'on Mondays on trajectory X there are 160% more passengers than on Tuesdays', that would qualify as anonymous data.

CASE LAW 1/2

In favor of a "relative anonymization" standard?

Breyer

- Is an IP address considered personal or anonymous?

43 In so far as that recital refers to the means likely reasonably to be used by both the controller and by 'any other person', its wording suggests that, for information to be treated as 'personal data' within the meaning of Article 2(a) of that directive, it is not required that all the information enabling the identification of the data subject must be in the hands of one person.

44 The fact that the additional data necessary to identify the user of a website are held not by the online media services provider, but by that user's internet service provider does not appear to be such as to exclude that dynamic IP addresses registered by the online media services provider constitute personal data within the meaning of Article 2(a) of Directive 95/46.

45 However, it must be determined whether the possibility to combine a dynamic IP address with the additional data held by the internet service provider constitutes a means likely reasonably to be used to identify the data subject.

46 Thus, as the Advocate General stated essentially in point 68 of his Opinion, that would not be the case if the identification of the data subject was prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant.

* Case C-582/14, *Breyer vs Germany*

astellas

CASE LAW 2/2

In favor of a "relative anonymization" standard?

SRB vs. EDPS

- Different treatment of the same data depending on who is processing them
- SRB had transferred to its consulting firm only coded data
- EDPS considered such data as being automatically "personal" since at least SRB could re-identify the individual
- The General Court rejected this automatic inference
- Under appeal before the Court of Justice

104 It is apparent from paragraph 45 of the judgment of 19 October 2016, *Breyer* (C-582/14, EU:C:2016:779), cited in paragraph 92 above, that it was for the EDPS to determine whether the possibility of combining the information that had been transmitted to Deloitte with the additional information held by the SRB constituted a means likely reasonably to be used by Deloitte to identify the authors of the comments.

105 Therefore, since the EDPS did not investigate whether Deloitte had legal means available to it which could in practice enable it to access the additional information necessary to re-identify the authors of the comments, the EDPS could not conclude that the information transmitted to Deloitte constituted information relating to an 'identifiable natural person' within the meaning of Article 3(1) of Regulation 2018/1725.

* CaseT-557/2020, *SRB vs. EDPS*

astellas

… WHAT ABOUT DATA PROTECTION AUTHORITIES;

Uncertainty, variations and confusion!

France (CNIL)

- Closer to an absolute anonymization standard
- *"… via une évaluation approfondie des risques d'identification, que le risque de ré-identification avec des moyens raisonnables est **nul** …"*

Ireland (Data Protection Commission)

- Closer to the relevant anonymization standard

United Kingdom (ICO)

- New draft guidelines
- Clear preference of the relevant anonymization standard

Spain (AEPD)

- Closer to an absolute anonymization standard
- *"Finally, it must be stated that, in order to anonymise a file, the corresponding data "should be such as not to allow the data subject to be identified via "all" "likely" and "reasonable" means" by the data controller or by any third party. Therefore, anonymisation procedures must ensure that **not even the data controller is capable of re-identifying the data holders in an anonymised file**. "*

IN SEARCH OF REGULATORY CLARITY!

EDPB

- Trying to publish a new Opinion
- Divergence between the DPAs of member states

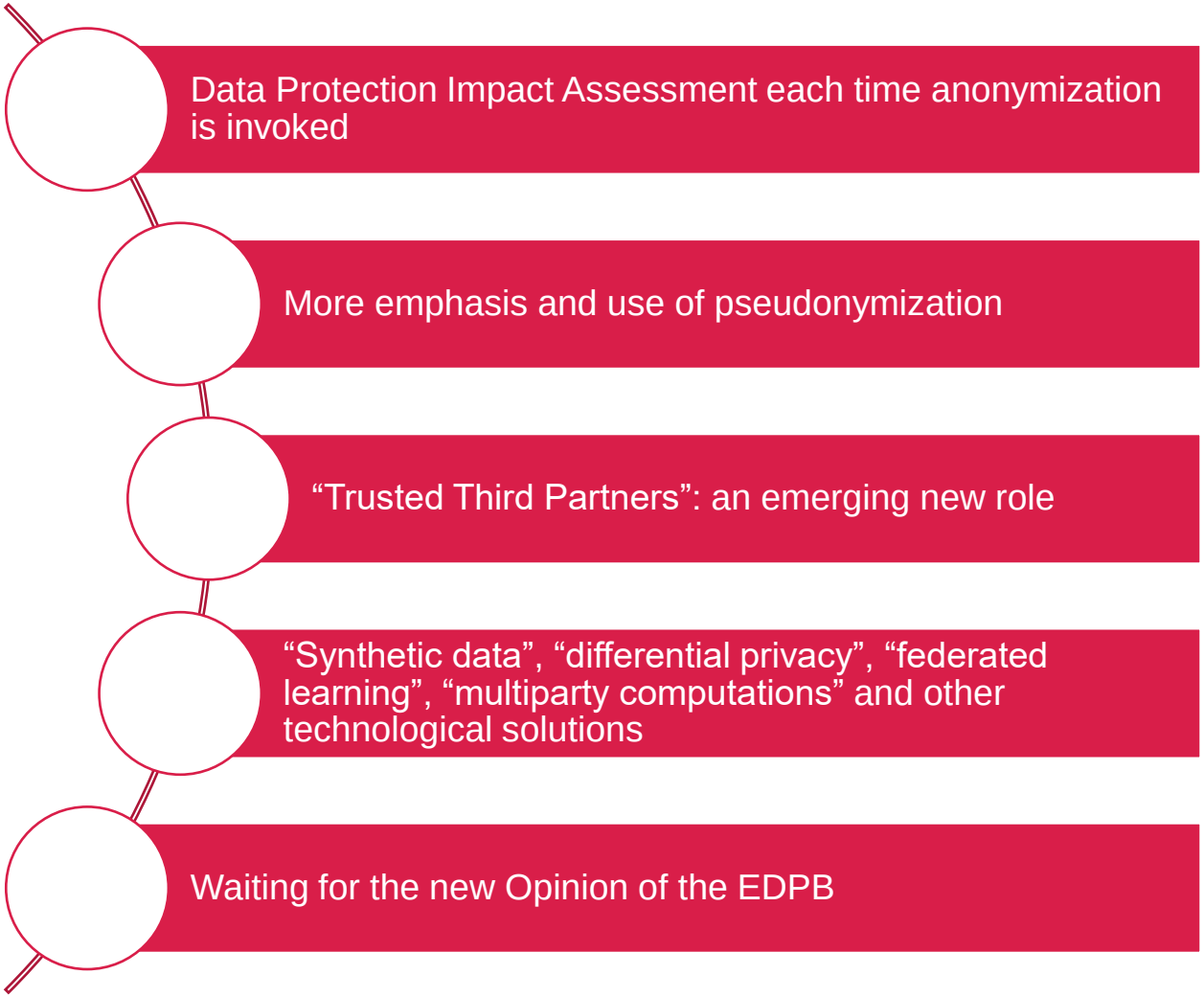
European Health Data Space

- Making the sharing of anonymized data mandatory (under some circumstances)

Data Act

- Sharing of data from interconnected devices (IoT)

WHAT DO WE DO UNTIL MORE CLARITY IS ACHIEVED?



QUESTIONS ?