

CIRCL

Computer Incident
Response Center
Luxembourg

CSIRT activities and GDPR compliance Data
Data Science & Engineering GDPR meeting
C3 workshop

May 15th 2018

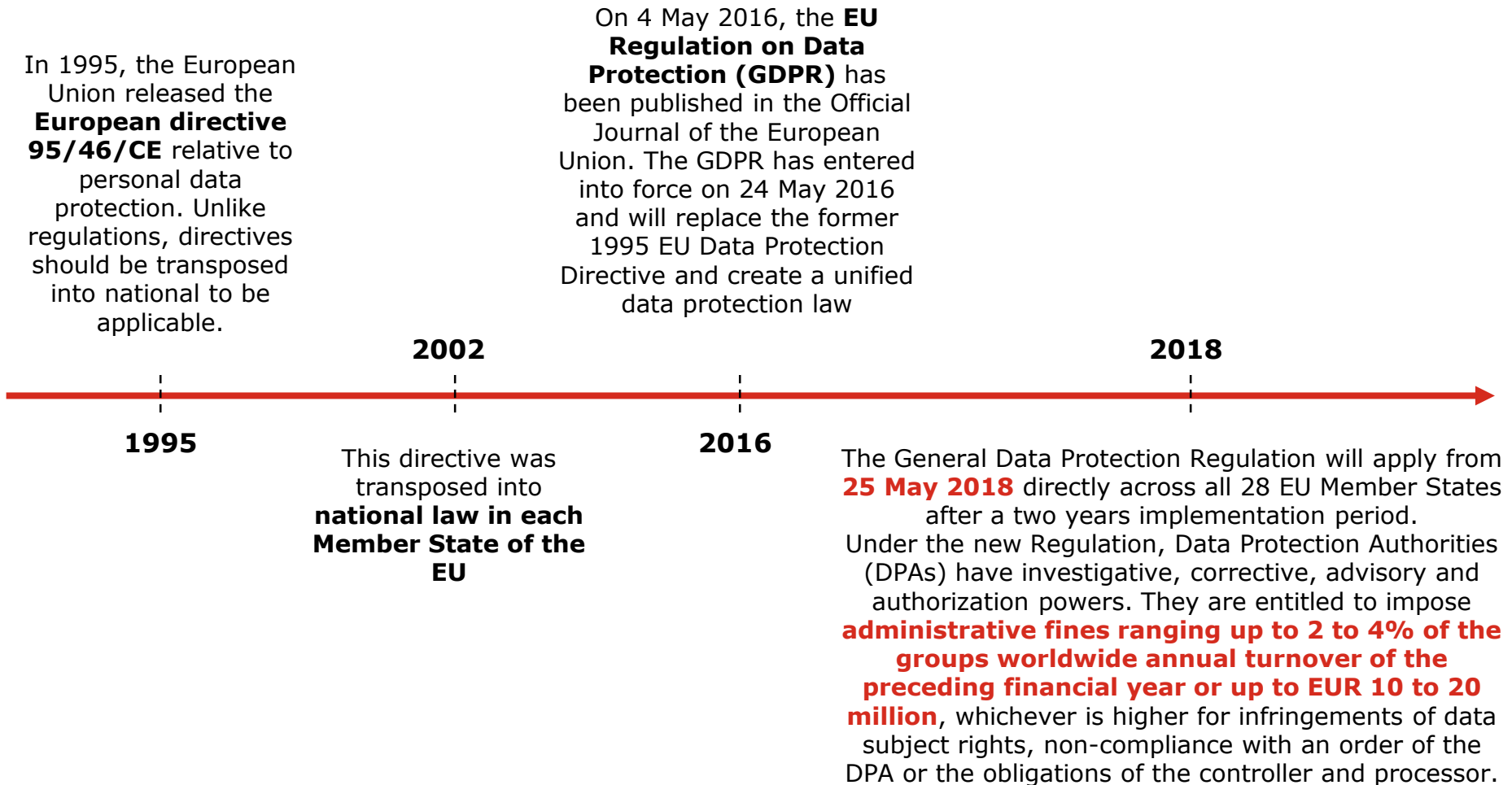


Co-financed by the European Union
Connecting Europe Facility

GDPR refresh

GDPR refresh

Timeline



GDPR refresh

Personal data & lifecycle

Any information relating to an identified or identifiable natural person (the 'data subject') [...]

Relating	Identification	Data subject	Reference
<ul style="list-style-type: none">• Content• Purpose• Result	<ul style="list-style-type: none">• Direct• Indirect	<ul style="list-style-type: none">• Not dead• Not unborn• Not legal person	<ul style="list-style-type: none">• Name• ID number• Location data• Online identifier• ...

'processing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction




GDPR refresh

Know What is personal data (e.g. IP addresses)


IP addresses are personal data (To be on safe side, always consider IP addresses as personal data)




ECJ (Scarlet Extended case): IP addresses processed by ISPs are personal data as they allow users to be precisely identified



WP29, opinion 1/2008: unless controller is position to distinguish with certainty that data correspond to users that cannot be identified, will have to treat all IP info as personal data



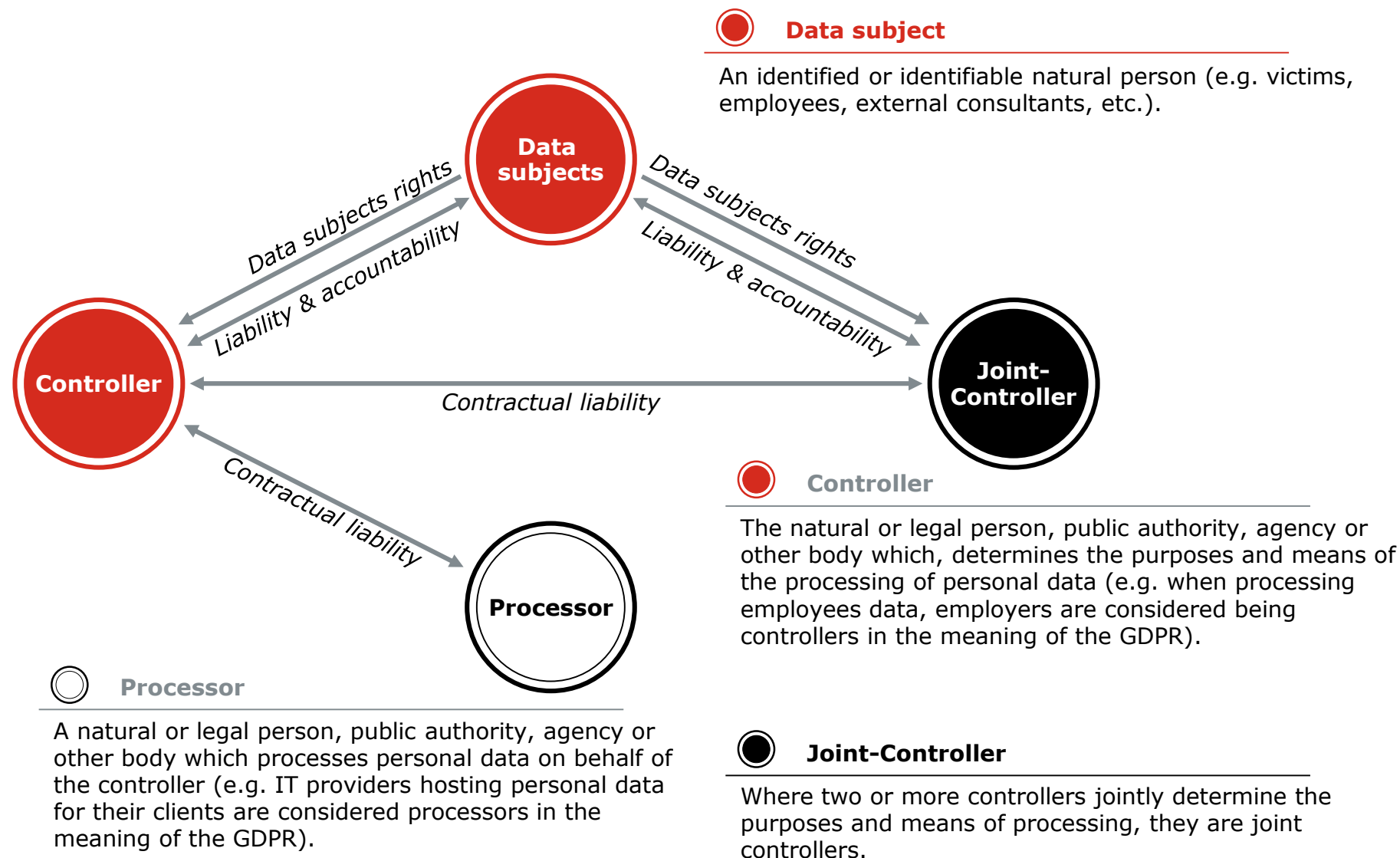
ECJ (Patrick Beyer case): dynamic IP addresses, also if only ISP has additional data to identify data subject, are personal data – unless identification is prohibited by law or would require a disproportionate effort



Same applies to user activity data, server logs and traffic data.

GDPR refresh

Roles introduced by the GDPR



GDPR implementation

Examples for CSIRTs

GDPR Implementation

Examples of gaps & resolution approach for CSIRTs (1/3)

	Data controller & processor	Data retention	Legal ground*
Some ways to tackle the gap	<ul style="list-style-type: none">Assess whether CSIRT is controller or processor	<ul style="list-style-type: none">Consider data retention period and draft data retention policy	<ul style="list-style-type: none">Be aware of legal ground for processing of personal data
Starting points	<ul style="list-style-type: none">CSIRT is controller when it processes data based on mandate (not on behalf of another body)When CSIRT acts on behalf of law enforcement agency, other CSIRT, etc., it is processor as it does not decide purposes/means.	<ul style="list-style-type: none">Ideally keep personal data for a short periodIf you want to store personal data for a longer period, pseudonymise and/or anonymise itStore personal data only for as long as it is needed by law enforcement and/or judiciary.	<ul style="list-style-type: none">CSIRTs are more likely to use legal obligation (Art. 6(c)) and public interest (Art. 6(e)) is they work under a mandate defined by national law or NISD;legitimate interest (Art. 6(f)) as mentionned in Recital 49, e.g. in case of a private CSIRT.

* For more information, please refer to "Information sharing and cooperation enabled by GDPR" (http://misp-project.org/compliance/gdpr/information_sharing_and_cooperation_gdpr.html)

GDPR Implementation

Examples of gaps & resolution approach for CSIRTs (2/3)

	Policies & procedures	Roles & Responsibilities	Record of Processing Activities	Transparency & Information
Some ways to tackle the gap	<ul style="list-style-type: none">▪ Review of existing policies & procedures▪ Update or create when needed	<ul style="list-style-type: none">▪ Define roles & responsibilities▪ Designate (if applicable) a DPO or coordinator for GDPR	<ul style="list-style-type: none">▪ Identify and document your processing activities of personal data▪ Classify those processes in a risk-based approach	<ul style="list-style-type: none">▪ Review and update privacy notices and other channels of communication
Starting points		<ul style="list-style-type: none">▪ Public CSIRTs need a DPO (Art. 37 (1)(a)). The DPO can be shared with other public entities (Art. 37 (3))▪ Private CSIRTs do not explicitly require a DPO but designation of responsible person	<ul style="list-style-type: none">▪ An example of templates for the records specifically for CSIRTs is provided by CIRCL: https://github.com/CIRCL/compliance▪ Risk Methodology (e.g. based on WP29 and/or ENISA Guidelines*)	<ul style="list-style-type: none">▪ Privacy Notice templates. Examples of such templates can be found on other CSIRT websites, e.g. CIRCL's privacy notice under creative commons https://www.circl.lu/privacy▪ CSIRT can do a privacy notice for services with specific data retention requirements e.g. PGP servers (https://pgp.circl.lu/)

* ENISA, "Guidelines for SMEs on the security of personal data processing", December 2016

ENISA, "Recommendations for a methodology of the assessment of severity of personal data breaches", December 2013

Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679

GDPR Implementation

Examples of gaps & resolution approach for CSIRTs (3/3)

	Training & Awareness	Data Subject Rights	Data Breach Incident Management	International Data Transfers
Some ways to tackle the gap	<ul style="list-style-type: none">Organize dedicated training sessions / implement an awareness programRegularly communicate	<ul style="list-style-type: none">Define, document and implement the processes to manage data subject rights	<ul style="list-style-type: none">Define / update and implement incident management procedures, covering the personal data aspects	<ul style="list-style-type: none">Identify these transfers based on the processing activitiesEstablish the expected level of complianceImplement required safeguards
Starting points	<ul style="list-style-type: none">Tailored awareness sessions & trainingsFor CSIRTs in the CSIRT network, training materials are available through the different Connecting Europe Facility projects	<ul style="list-style-type: none">Procedure to handle data subject rightsMany CSIRTs already have the tools and expertise to handle constituency requests. For example, CSIRTs can create custom email templates in their incident management tool (e.g. RTIR*)	<ul style="list-style-type: none">Many CSIRTs already have a formal incident management process. In those cases, an update to those processes is sufficient.	<ul style="list-style-type: none">CSIRTs are usually sharing information outside the EU, usually for public interest. International data transfers should be documented.

* RTIR: Request Tracker for Incident Response

Open data and data mining in the context of GDPR

What is open data?

Definitions

- “**Open data**” is data that **anyone can access, use and share**. It must be provided in a **common, machine-readable format** and it must **be licensed**. Its license must permit people to **use the data in any way they want**, including **transforming, combining and sharing it with others**, even commercially. For data to be open, **it should have no limitations that prevent it from being used in any particular way** (*European Data Portal*).
- Open data is **mostly made available by public sector bodies** (e.g. through open data portal on national or EU level such as the European Data Portal).
- Open data can also be **published by private entities**.

Example: CIRCL **publishes open data** on common vulnerabilities and operational statistics related to incidents response. The license used is under Creative Commons license (“international CC BY 4.0”) which is also a recommended license by the Luxembourg Government Open Data portal.

- **Open data “re-use”** from the public sector means *“the use by persons or legal entities of documents held by public sector bodies, for commercial or non-commercial purposes other than the initial purpose within the public task for which the documents were produced”* (Art 2(4)) *The Public Sector Information (PSI) Directive (2013/37/EU)* .
- However, private entities can consult open data sources published by other private actors.

Example: An information security researcher may access open data on malware statistics from CIRCL and use it further for academic research.

What is the legal context?

Does it apply to CSIRTs?

Public Sector Information (PSI) Directive:

- The PSI Directive 2013/37/EU and the new Proposal for a revised PSI Directive **allow for the “re-use of public sector information for private or commercial purposes with minimal or no legal, technical or financial constraints** (Recital 2, PSI Directive).
- The PSI directive currently does **not** cover scientific and research data, however **the new Proposal for a revised PSI Directive** (COM(2018) 234 final) **will extend the scope to include research data** such as statistics, results of experiments, measurements, observations resulting from fieldwork, survey results etc.

In cases when public CSIRTs want to publish open data, the PSI Directive and the GDPR apply.

The General Data Protection Regulation (GDPR):

- The GDPR applies to all entities processing personal data. Furthermore, the data protection framework is compatible with the PSI Directive and the Proposal for a revised PSI Directive, hence it applies both to **public sector bodies** and the **re-users of open data for datasets** which **contain personal data**,
- *Note: other sector-specific limitations may apply for publishing open data (e.g. in the financial sector).*

In practice that means that the **GDPR applies to public or private CSIRTs** who publish open data containing personal data as well as to **re-users of open data**.

How can CSIRTs publish open data in line with the GDPR?

Guidelines on open data and the GDPR

In its “*Opinion 6/2013 on open data and public sector information ('PSI') reuse*”, the **Article 29 Working Party (WP)** **recommends to public sector bodies** the following **measures prior to making PSI publicly available if it contains personal data**:

- Apply the “**data protection by design and by default**” principles “at the earliest occasion when considering making PSI publicly available”;
- Conduct **Data Protection Impact Assessments (DPIA)** prior to making datasets containing personal data publicly available. DPIAs should equally be conducted in cases of anonymized datasets derived from personal data;
- Carry out **re-identification testing on anonymised datasets** to assess the risk of re-identification and identify appropriate safeguards;
- **Apply the safeguards identified through the re-identification assessment** including technical, legal and organisational measures (e.g. license terms, technical measures to prevent bulk download of data, anonymization techniques etc.). The assessment could also lead to a decision not to make the data publicly available;
- Include a **data protection clause in the terms of license to re-use the PSI** when it contains personal data, including for anonymized datasets derived from personal data;
- Ensure that the **personal data is adequately anonymized** and the **license conditions** specifically **prohibit the re-identification of individuals** and re-use of personal data. Hence, **refrain from making PSI publicly available when the DPIA assessment concludes that the open license is not sufficient** to address the data protection risks involved.

Private CSIRTs could adopt a similar risk-based approach.

How can CSIRTs publish open data in line with the GDPR?

Guidelines on open data and the GDPR - continued

Determining the controller and the processor in the context of open data:

The GDPR defines a **controller** as "*the natural or legal person, public authority, agency or other body which, alone or jointly with others, **determines the purposes and means of the processing of personal data***" (Article 4(7)).

Example: **CSIRT A** publishes open data in compliance with the GDPR. The data are consequently used and combined with data from other sources by **CSIRT B** for academic research. The latter is the data controller as they determine the new purpose of the re-use.

How can CSIRTs use open data and data mining for research purposes in line with the GDPR

Challenges

- **Re-identification of individuals:** In the context of big data and deep learning, large amounts of PSI and open data can be re-used to uncover correlations between seemingly unrelated events/parameters and potentially lead to the re-identification of individuals.
- **Data subject's rights:** The GDPR expands the set of data subject's rights compared to the EU Data Protection Directive and introduces further obligations for entities processing personal data. These provisions may pose a challenge for data science.

Example: The data subject has right "to **object**, on grounds relating to his or her particular situation, at any time **to processing of personal data** concerning him or her which is based on point (e) or (f) of Article 6(1), **including profiling** based on those provisions. The controller shall no longer process the personal data **unless the controller demonstrates compelling legitimate grounds** for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims."

- However, **processing of personal data for research purposes** is subject to special conditions and exemptions under the GDPR.

How can CSIRTs use open data and data mining for research purposes in line with the GDPR

What are the research activities under the GDPR?

The GDPR has specific exemptions for the following **categories of research activities**:

- **Scientific research** refers to “*technological development and demonstration, fundamental research, applied research, and privately funded research.*” (Recital 159)

Example of scientific research as a CSIRT: Training algorithms with data in order to improve a CSIRT’s security services.

- **Historical research** includes “*historical research and research for genealogical purposes, bearing in mind that this Regulation should not apply to deceased persons*” (Recital 160)

Example of a historical research or archiving purposes as a CSIRT: Information on incidents is kept in MISP for archiving purposes in the public interest longer after the last occurrence of specific attacks, in order for example to discover attack patterns and produce statistics.

- **Statistical research** means “*any operation of collection and the processing of personal data necessary for statistical surveys or for the production of statistical results [...] The statistical purpose implies that the result of processing for statistical purposes is not personal data, but aggregate data, and that this result or the personal data are not used in support of measures or decisions regarding any particular natural person*” (Recital 162)

Example of statistical research as a CSIRT: CSIRTs can publish statistical research from honeypots about malware infections.

How can CSIRTs use open data and data mining for research purposes in line with the GDPR

Legal grounds for processing personal data for the research purposes

Article 29 WP Opinion specifies that the PSI does not automatically give lawful grounds of processing and re-use of PSI. Hence, **any re-user of open data** must have **an appropriate legal basis under the GDPR**.

The processing of personal data for **archiving, scientific, historical** and **statistical research purposes** is compatible with the GDPR (Art 5) and could be considered as lawful processing of personal data provided that:

- It is performed in the **public interest**

Example of a public interest for CSIRTs: If a honeypot is operated by national, governmental or sectorial CSIRT to increase the level of security of their constituency.

- Represents the **legitimate interest** of the controller

Example of a legitimate interest for CSIRTs: A private organisation operates a honeypot on its network to detect potential attack

- **The appropriate safeguards** have been put in place **to prevent the re-identification of the data subject** (e.g. pseudo anonymization) (Art 89). The precise nature of safeguards for the exemptions is up to the Member States to decide and might vary per country.

How can CSIRTs use open data and data mining for research purposes in line with the GDPR

Exemptions under the GDPR for research activities

The GDPR recognizes that the processing of personal data for scientific or historical research purposes or statistical purposes, could be considered as “**the legitimate expectations of society for an increase of knowledge.**”(Recital 113).

Research activities are subject to the following **exceptions and exemptions**:

- Exemption from **the obligation to provide information where personal data have not been obtained from the data subject** (Art 14) if “*the provision of such information would be impossible or would involve a **disproportionate effort***” (e.g. if it is obtained from a publically available source) with the appropriate safeguards applied. Furthermore, Recital 33 allows for consent to be obtained only to certain areas or parts of the scientific research.

Example of a disproportionate effort for CSIRTs: Performing a re-identification test on datasets obtained from a public source might constitute a disproportionate effort.

- **Transfer data to third countries could be allowed** if there is a **legitimate interest** or the **processing of the personal data for scientific and research purposes is performed “for important reasons of public interest”** (Art 49(1(d))).

How can CSIRTs use open data and data mining for research purposes in line with the GDPR

Exemptions under the GDPR for research activities - continued

Automated individual decision-making, including profiling (Art 22) is allowed under the following conditions:

- It is necessary **for entering into, or performance of, a contract between the data subject and a data controller**;
- **It is authorised by Union or Member State law** to which the controller is subject and which also lays down suitable **measures to safeguard the data subject's rights and freedoms** and **legitimate interests**; or
- It is based on the data subject's **explicit consent**.

Example of a legitimate interest for CSIRTs in the context of profiling: Training machine-learning algorithms for research on improving incident response and detecting threat actors could constitute a legitimate interest of a CSIRT.

- **Processing of sensitive data** for research or scientific purposes is possible if:
 - The data subject has given explicit consent or **if this data was explicitly made public by the data subject** (Article 9(2)(a), Article 9(2)(e))

or:

- It is performed in accordance with Article 89, based on a Union or Member State law allowing it, **it respects the proportionality principle** and has **ensured that the necessary safeguards are adopted**.

How can CSIRTs use open data and data mining for research purposes in line with the GDPR

Exemptions under the GDPR for research activities - continued

The following exemptions further apply:

- Exemption from **the purpose limitation** for research purposes (Art 5(1)(b))
- Exemption from **the right to erasure** ("Right to be forgotten") (Art 17(3) (d))
- Exceptions for the following data subject rights when **processing of personal data for archiving, scientific or research purposes** (Art 89):
 - Right of access by the data subjects (Art 15);
 - Right to rectification (Art 16);
 - Right to restriction of processing (Art 18);
 - Notification obligation regarding rectification or erasure of personal data or restriction of processing (Art 19) when the data has been processed for **archiving purposes**;
 - Right to data portability (Art 20) when the data has been processed for **archiving purposes**;
 - Right to object to processing (Art 21(6) if the research activities are conducted for reasons of public interest;

It is to be noted however, that **other principles such as data minimisation** remain unchanged.

It is also recommended that **the necessary safeguards such as re-identification tests and DPIAs are performed on anonymized datasets** in order to be sure that the results of a conducted research do not lead to directly or indirectly identifying individuals.

Conclusions

The GDPR as an enabler for data science and research

Researchers **may use open data** containing personal data and apply data mining techniques **under certain conditions of the GDPR**.

The GDPR is **not an obstacle to research**, but seeks to **find a balance between privacy** and **social welfare and economic growth** through scientific advancement.

Q&A