הרשות להגנת הפרטיות
THE PRIVACY PROTECTION AUTHORITY
سلطة حماية الخصوصية

משרד המשפטים
MINISTRY OF JUSTICE | وزارة العدل

Guide to Privacy-Enhancing Technologies

April 2025

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 ת.ד. קרית הממשלה,

Introduction

Privacy-Enhancing Technologies (PETs) refer to a set of approaches and digital solutions designed to safeguard personal data. The use of PETs allows for the extraction of the required value from data while preserving privacy and protecting personal data by (1) obfuscating the personal data required for use and reducing their level of detail, (2) reducing the risk of personal data exposure during processing, and (3) enhancing control over personal data usage.

In recent years, the advancement and deployment of PETs have been a shared focus among data protection authorities worldwide[1]. Integrating PETs into the design of digital systems supports compliance with privacy and data security laws, fulfills the core obligation to protect personal data, and serves as a foundation for building trust with users.

Objectives of this document

1. To provide an overview of both established and emerging Privacy-Enhancing Technologies (PETs).
2. To present examples and considerations for integrating PETs as part of addressing privacy protection in processes, systems, and projects.

Scope of this document

This guide focuses on a selection of representative and prominent technologies from the variety of PETs currently available. It outlines the foundational operating principles of each technology and presents key considerations in making the decision to use a particular technology as part of the overall measures for protecting personal information.

The document does not specify general technologies in the areas of information security. Furthermore, the document does not deal extensively with the legal norms

[1] Further details appear in the 'Key sources of information' section later in this chapter and in the references in the body of this document.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

(laws and regulations) or organizational processes (such as privacy impact assessments) that affect the integration of PETs.

PETs and the Privacy by Design principle

The Privacy by Design principle refers to the integration of privacy protection and information security considerations into the design of systems and processes, starting from the requirements and architecture concept phase and continuing throughout the entire life cycle of those systems and processes. This principle has been adopted in numerous decisions over the years[2] and has been included as part of the requirements of the European Union's General Data Protection Regulation (GDPR)[3]. It has since been incorporated into laws and regulatory frameworks in various jurisdictions worldwide.

The use of PETs is part of the means for optimal protection of personal information as part of the implementation of the Privacy by Design principle in the processes of data collection, storage, and access control. PETs also have the potential to significantly expand the possibilities of using personal information, including highly sensitive personal information in particularly large-scale datasets. The decision to adopt PETs may arise from a Privacy Impact Assessment (PIA)[4] conducted prior to initiating data processing, or as part of continuous risk management practices throughout a system's lifecycle. Privacy-enhancing technologies complement legal and organizational processes and measures available to developers of systems and technological tools, assist in the optimal implementation of privacy protection law provisions and promote the fair use of personal information.

---

[2] For example, A Resolution on Privacy by Design adopted at the 32nd International Conference of Data Protection and Privacy Commissioners held in Jerusalem in 2010.

[3] Art. 25 GDPR: Data protection by design and by default.

[4] For additional details refer to the Methodological Reference Guide for conducting Privacy Impact Assessment (Hebrew) published by the Israeli Privacy Protection Authority in 2022.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

How the technologies are presented

The descriptions of technologies in this document include explanations of the core principles behind each privacy-enhancing technology and how they function, accompanied by examples or illustrations where relevant. Each section presents key considerations for implementation, as well as applications across various sectors and domains. The document also highlights the challenges and limitations associated with the use of these technologies, along with risks and issues that require special attention. The examples provided demonstrate the principle and mode of operation of privacy-enhancing technologies in a simplified and accessible manner for the purpose of illustration and without specifying all the details and technical processes required for practical implementation. The examples present common or prominent uses from diverse fields and do not prescribe a particular course of action or restrict the potential applications of the technologies. Practical implementation of privacy-enhancing technologies requires thorough analysis of a wide variety of topics and factors, while taking into account the unique characteristics of each case and project.

Throughout the document, readers will find links to additional resources, including publications by researchers and industry experts from Israel and abroad, as well as official materials from government and regulatory bodies worldwide. These references are intended to support deeper exploration, provide further examples, and encourage continued learning on the topics discussed.

Intended audience

This guide is intended for professionals responsible for assessing privacy risks and implementing appropriate safeguards within development projects of digital systems and services. No less so, the document may support development teams working in digital domains in the implementation of privacy-enhancing technologies, from the initial planning stages through the entire lifecycle of a project. Specifically, the document is relevant for the following roles:

1. Data Protection Officers (DPOs) and legal advisors involved in privacy protection.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

2. Product managers and project managers overseeing the development, implementation, and operation of information systems, services and digital products.

The use of the document does not require a technical background or technological expertise. As a result, the level of detail with respect to each technology is only sufficient to convey its core purpose and to support an assessment of its relevance and applicability to specific use cases. Readers are encouraged to consult additional resources for implementation details and response to challenges and risks, and for this purpose, the links provided at the end of each chapter throughout the document can be used. It should be noted that this document is not a substitute for consultation with the relevant domain experts when evaluating a specific technological product.

Key sources of information

This guide is based in part on materials on PETs that were published by the OECD (See policy paper on Emerging privacy-enhancing technologies), the UN (See PET Guide for Official Statistics), and other data protection authorities worldwide. Some of the references in the document were adopted from a comprehensive report published by Great Britain's Information Commissioner's Office (ICO) (See Privacy-enhancing technologies (PETs) guide). Other materials were adopted from the repository of publications from official sources on privacy-enhancing technologies on the Future of Privacy Forum (FPF) website. Further materials that form the basis for specific sections of this document are referenced both within the main text and in the footnotes.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Contents

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Introduction to Privacy-Enhancing Technologies

Privacy-Enhancing Technologies (PETs) are part of a set of measures designed to help reduce risks to privacy and the security of personal data. Integrating privacy-enhancing technologies is part of the toolbox of the Data Protection Officer (DPO) in an organization, alongside other tools such as Privacy Impact Assessment[5]. PETs are a complementary element to existing legal and organizational tools as part of processes for controlling and monitoring privacy and information security risks.

Definition of PETs

Over time, various definitions for PETs have emerged[6]. For the purposes of this document, the following definition is applied:

*"**Privacy-Enhancing Technologies (PETs)**" – set of methods, processes and digital tools designed to support the protection of personal data. PETs allow to obfuscate[7] personal data and reduce its level of details, reduce the risk of exposure of personal data during processing, and enable greater control over how personal data are used.*

The relevance and value of PETs can be better understood within the framework of the legal definition of "personal data", which is the cornerstone definition of Israeli Privacy Protection Law, 5781-1981. This definition was entirely amended within Amendment No. 13 to the Privacy Law Protection, enacted in 2024. According to the new definition, personal data are defined as:

***"Personal data"** – data relating to an identified or identifiable person; for the purposes of this definition, an "identifiable person" is one who can be identified with*

[5] For additional details refer to the Methodological Reference Guide for conducting Privacy Impact Assessment (Hebrew) published by the Israeli Privacy Protection Authority in 2022.

[6] See the OECD's guide on Emerging privacy-enhancing technologies for a survey on PETs definitions.

[7] 'Data obfuscation' term follows the definition in OECD's guide on Emerging privacy-enhancing technologies.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

reasonable effort, directly or indirectly, including through an identifying detail such as name, ID number, biometric identifier, location data, online identifier, or one or more details relating to their physical, health, economic, social or cultural status."[8]

Accordingly, when PETs are used in a way that data can no longer be attributed with reasonable effort to an identifiable person, whether directly or indirectly, including through re-identification, the provisions of the Privacy Protection Law and its regulations shall not apply on such data, and its use would not constitute a privacy infringement.

However, it is important to clarify that the processing of personal data by an organization, in itself, including processing by the use of PETs, is subject to the provisions of the Privacy Protection Law. Moreover, while certain PETs (such as anonymisation techniques) can significantly enhance data protection, they are nevertheless subject to various risks that would, in general, preclude the data from being excluded from the scope of personal data as defined under the Privacy Protection Law.

Categories of PETs

PETs are a diverse family of methods, processes, and digital tools that are appropriate for different stages in the information life cycle, and in particular for three broad areas:

1. Collecting data and preparing it for use.
2. Using the data.
3. Controlling the use of data.

[8] It should be noted that the Privacy Protection Law (Amendment No. 13), 5774-2024, which constitutes a comprehensive reform of the law, is scheduled to enter into force on August 14, 2025.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

PETs differ in their mode of operation and form of implementation, but they can be divided into three broad categories according to their operating principle:

1. Obfuscating the personal data required for use and reducing their level of detail.
2. Reducing the exposure of personal data during use.
3. Monitoring access to personal data.

The division according to stages in the data life cycle and operating principle produces three main categories of PETs, according to Figure 1:

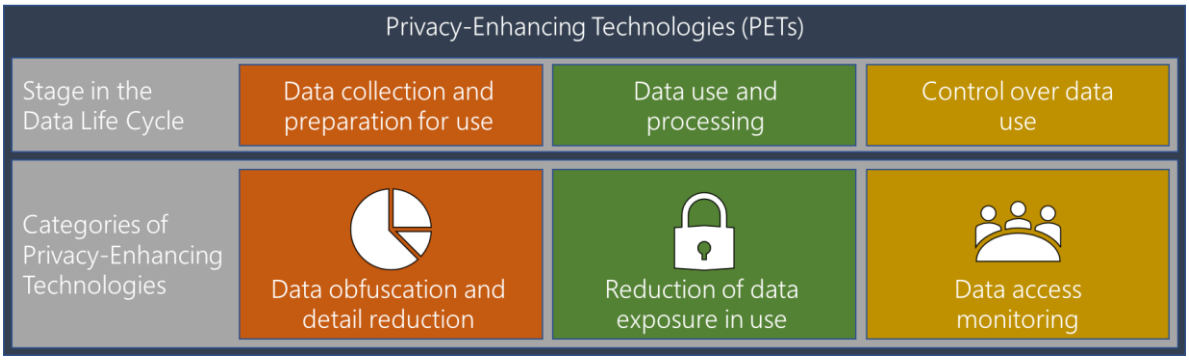

Figure 1: Primary categories of PETs

The following is a breakdown of the main categories of PETs:

1. **Data collection and preparation for use**: Obfuscation of personal data required for use or reduction of their level of detail, that may include removing identifying details, changing the data, blurring the precise values, or adding "noise"[9]. Examples of technologies in this area include data anonymisation and synthetic data.

[9] Adding "noise" to data is similar to adding multiple voices to a recording to prevent the information from being identified in a conversation. Adding "noise" is the addition of random data or random modification of some of the data to make it more difficult to identify the original data. When added "noise" is done appropriately, the usefulness of the data can be maintained while better protecting the information.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

2. **Data use and processing**: Reducing the exposure of personal data during processing and even using data without viewing them during processing. Examples of technologies in this area include homomorphic encryption and multi-party computing.
3. **Control over data use**: Definition of rules and permissions for access to personal data and display of data relating to the identity of the person accessing the data, the type of data, and the time of access. Examples in this area include personal[10] data stores and documentation and transparency tools.

Mapping of key PETs to categories

The table below lists the privacy-enhancing technologies according to the categories defined in the previous section. For each technology in the table, the following are briefly described: the operating principle; examples of applications; and challenges and limitations. Further details can be found in the appropriate chapter for each technology later in this document:

[10] 'Personal' in this regard refers to the "Data Subject" – the person that the data relates to (the definition is set out in Regulation 1 of the Privacy Protection (Information Security) Regulations, 2017, Hebrew).

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Technologies for Data Obfuscation and Detail Reduction

Technology	Principle of operation	Example use cases	Challenges and limitations
Anonymization	Removal and blurring of identifying characteristics	Omitting or generalizing values, masking data	○ Data loss and downscaling ○ Fidelity to the original data, biases and usability ○ Technological expertise in operating the tools
Synthetic Data	Generating new datasets with a statistical relationship to the source data	Producing data for training artificial intelligence models	
Differential Privacy	Adding random noise to the data	Making datasets accessible for statistical research	

Technologies for Reduction of Data Exposure in Use

Technology	Principle of operation	Example use cases	Challenges and limitations
Homomorphic Encryption	Performing calculations on data while it is encrypted	Performing calculations on sensitive data	○ Computational load ○ Technological expertise in implementation

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Technology	Principle of operation	Example use cases	Challenges and limitations
Zero Knowledge Proof	Proving the correctness of data without revealing the data itself	Verifying attributes (such as age or experience)	○ Limited and dedicated applications ○ Network traffic ○ Computational load ○ Technological expertise in implementation
Multi-Party Computation	Joint data processing without any party sharing its data with the other parties	Performing calculations that require collecting sensitive data from multiple participants	
Private Set Intersection	Finding common data between datasets without revealing non-shared data	○ Finding shared contacts ○ Monitoring contact between people	
Federated Learning	Distributed machine learning while minimizing the data shared between the parties	Training machine learning models	○ Accuracy of results ○ Technological expertise in implementation
Trusted Execution Environment	Processing data in an isolated and secure part of the computer system	○ Secure payments ○ Biometric identity verification	○ Computational load ○ Expertise in hardware and software ○ Trust in the manufacturer / developer of the environment

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Technologies for Data Access Monitoring

Technology	Principle of operation	Example use cases	Challenges and limitations
Personal Data Stores	Managing access permissions to personal data and documenting actual access events	Storing personal data and managing access permissions	Responsibility of the dataset owner for information management, access permissions and data security
Documentation and Transparency Tools	Accurate and complete documentation of access to personal data while allowing the data subject to track its use	Displaying details of access events to personal data in a dataset	Implementation in organizations

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה

Technologies for Data Obfuscation and Detail Reduction

As noted, data that is no longer identifiable, meaning it cannot be linked to an identified individual, even through re-identification using reasonable effort – does not constitute "personal data" under the updated definition of the Israeli Privacy Protection Law under Amendment No. 13 (2024).

Moreover, the principle of data minimisation[11] is a foundational element of privacy law. According to this principle, only the minimum necessary amount of personal data should be collected, stored, and used, considering the volume, type, and retention duration of the data, in relation to the specific purpose of collection or the objective of the dataset.

PETs can support this principle by obfuscating personal data or reducing its level of detail. In some cases, they may even render data non-identifiable while still fulfilling the underlying purpose of the data collection or data storage. Beyond their legal relevance, reducing the level of detail and scope of personal data plays an integral role in Privacy by Design strategies and contributes significantly to mitigating information security risks.

Anonymisation

Anonymisation refers to the removal of attributes or the alteration of data values to prevent or significantly reduce the likelihood of identifying a data subject. Common approaches to anonymisation include:

- **Removal of direct identifiers** (such as names, ID numbers, or other unique identifiers listed in the amended definition of "personal data");

[11] See Israel Privacy Protection Authority policy paper draft (Hebrew) on Data Minimisation (2021).

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

- **Removal of quasi-identifiers**, which, when combined with other data, could lead to re-identification (e.g., occupation or workplace);
- **Data generalisation**, such as reducing the precision of individual data points or grouping them into broader categories;
- **Adding random noise** to obscure personal data values.

While reducing the level of detail and scope of personal data strengthens privacy protections, it may also affect the value and usability of the data for their intended applications.

Common anonymisation techniques and examples

1. **Attribute Suppression**
   Removing data values (columns) that are unnecessary for the intended processing.
   Example: When a dataset contains student's name, teacher's name and a grade, for analysing student grades by teacher, the student's name column can be removed, retaining only the teacher's name and the grades.
2. **Record Suppression**
   Removing unique or outlier records (rows) that may be easily re-identified.
   Example: When a dataset contains income data by age, and there is only one resident over the age of 100, their record may be easily re-identified and therefore could be removed to protect privacy, assuming this would not significantly affect the sample.
3. **Character Masking**
   Obscuring parts of textual data by replacing some of the characters with fixed symbols.
   Example: Changing a postal code from 96554 to 96XXX reduces the location specificity so a specific region could not be identified, but in some postal code systems may allow for the identification of a city or a district.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

4. **Pseudonymisation**

   Replacing direct identifiers with artificial substitutes or tokens.

   Example: Substituting an ID number with a randomly generated string. Pseudonyms may be either reversible (e.g., using lookup tables[12] or encryption methods) or irreversible. This technique allows linkage of personal data across datasets without revealing person's identity (e.g., analysing a patient's hospitalization history within numerous institutions).

5. **Generalisation**

   Reducing the precision of a datum or a set of data[13].

   Example: Replacing an exact income (15,980 ILS) with a range (10,000 - 20,000 ILS), or reporting "length of hospital stay in days" instead of exact admission and discharge dates.

6. **Shuffling**

   Randomly reordering column values across records in a dataset, assuming intra-record dependencies are not relevant for the purpose of the processing.

   Example: Randomly shuffling body weight values between individuals in a dataset.

7. **Noise Addition**

   Modifying data values that might lead to re-identification, especially by linking to additional or external sources, via introducing controlled random noise or rounding values.

   Example: Rounding values (e.g., changing 18.1 ILS and 44.9 ILS to 18 ILS and 45 ILS, respectively).

[12] Along with reducing personal data through pseudonymization, keeping lookup tables increases the number of tables in the database. Using a lookup table can increase the risk of retrieving personal data and linking personal data to pseudonymized information in the event of an information security incident.

[13] Generalization is effective provided that there is a sufficiently large group of individuals or alternatively that the generalization does not allow for the identification of a group of individuals within a large population.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

8. **Data Aggregation**

   Replacing individual-level data with aggregate metrics such as sums, averages, or standard deviations.

   Example: Instead of disclosing individual donations with exact amounts and dates, reporting the number and total value of donations per month.

Examples of applying these anonymisation techniques are provided in Figure 2**.**

Method	Description	Before anonymisation	After anonymisation
Attribute suppression	Deleting a value (column from a table)	Moshe | Male Dana | Female	Moshe Dana
Record suppression	Deleting a record (row from a table)	Moshe | Male Dana | Female	Dana | Female
Data masking	Replacing part of the description with a constant character	Moshe | moshe@a.com Dana | dana@a.com	Moshe | *****@a.com Dana | ****@a.com
Pseudonomization	Replacing identifying information with "fictitious" information	Moshe | Male Dana | Female	A12c19V0 | Male LVV098C | Female
Generalisation	Reducing the level of precision of a data item or data set	Moshe | 15,000NIS Dana | 22,000NIS	Moshe | 10-20K NIS Dana | 20-30K NIS
Data perturbation	Swapping column values between records	Moshe | 68kg Dana | 59kg	Moshe | 59kg Dana | 68kg
Noise addition	Distorting values by adding noise or rounding the information	Moshe | 68kg Dana | 59kg	Moshe | 70kg Dana | 55kg
Aggregative data	Replacing information with an aggregative measure (such as a sum or average over time)	Name | Date | Sum Moshe | 12.10.23 | 50NIS Dana | 11.6.24 | 13NIS	Name | Year | Total Moshe | 2023 | 170NIS Dana | 2024 | 180NIS

Figure 2: Examples of the application of anonymisation techniques

Key Privacy Risks in the Use of Anonymised Datasets include identity disclosure (separating a specific individual within the dataset from the other data subjects) and attribute disclosure (inferring information about an individual's attribute):

1. **Identity disclosure** – Full identification of a data subject's record within the dataset. This is often enabled when records contain unique characteristics. Assessing the likelihood of identity disclosure depends on the data features in

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

the dataset and the number of categories that may be used in separating a specific individual.

2. **Attribute disclosure** – The attribution of a characteristic from the dataset to a specific individual. For example, if an anonymised attendance log shows that all employees in a particular department arrived after 10 a.m. on a given day, and it is known that a specific individual works in that department, it may be inferred that this person was late on that day, even if their name does not explicitly appear in the dataset.

It is important to note that commonly used anonymisation techniques, such as record suppression, data masking, or generalization, are not a guarantee from partial or full re-identification. Over the years, many cases have been published in which it was possible to identify an attribute or draw conclusions about a person from anonymized information, for example by cross-referencing with another database.

The challenge of distinguishing personal from anonymous data is substantial and has been examined in depth, including in the report of the inter-ministerial committee on Artificial Intelligence in the Financial Sector[14]. According to this report, anonymization has significant advantages in protecting the privacy of data subjects, but it does not guarantee complete protection of personal information and the privacy of data subjects, especially in the era of artificial intelligence and its capabilities, which facilitate the re-identification of data subjects from anonymized data.

[14] See Artificial Intelligence in the Financial Sector – interim report for public consultation (Hebrew), 2024.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Case Study: The Netflix prize and re-identification of anonymised data

In 2006, Netflix released an anonymised dataset containing movie ratings as part of a competition to improve its recommendation algorithms, without any additional data about the users except for a unique pseudonymised user ID that indicated numerous ratings of the same user. The dataset included:

1. User ID;
2. Movie details;
3. Rating dates;
4. Movie ratings (on a scale of 1–5).

In 2007, researchers at the University of Texas demonstrated that they could re-identify certain users by correlating Netflix's anonymised data with publicly available ratings from IMDb (Internet Movie Database). Matching the details was done by similarity of the rating dates for the same movies across both platforms. The level of matching reported in some cases was extremely high.

The severity of the privacy violation in this case stemmed from the fact that Netflix ratings were given anonymously, and some users referred to films from which conclusions could be drawn about their political opinions, religious beliefs, and other personal details. Linking the ratings on IMDb made it possible to obtain identifying details of the raters and, accordingly, to infer sensitive data about them.

Case Study: Re-identification using a linkage attack – Governor of Massachusetts case

In the 1990s, anonymised hospital admission data for public employees in the state of Massachusetts was released for research purposes. All direct identifiers (e.g., name, address, Social Security Number) were removed from the published data.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Researcher Latanya Sweeney obtained voter registration records from the city of Cambridge, where the Governor of Massachusetts resided. By cross-referencing the two datasets, she discovered that only six people in the city shared the Governor's birth date; half of them were men, and only one lived in the Governor's ZIP code area. This way, by using open information, she was able to link sensitive medical information to the details of a specific person. The overlapping data in both datasets is shown in Figure 3.

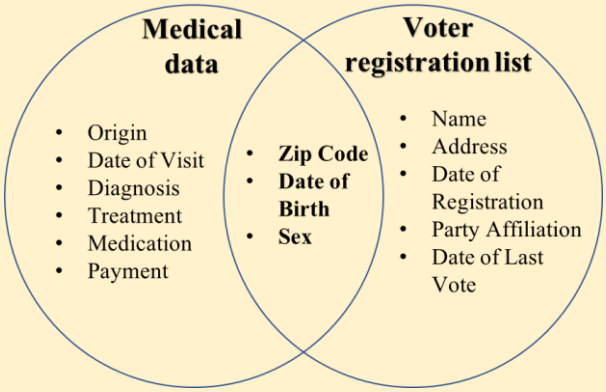

Figure 3: Overlapping data in both voter registration records and hospital admission data

In a follow-up study[15], Sweeney demonstrated that approximately 87% of the U.S. population could be uniquely identified using just three quasi-identifiers: date of birth, gender, and ZIP code.

The risks of re-identifying or drawing conclusions from anonymised data have increased dramatically in the era of Big Data, given the accessibility of many databases and the technological development in the field. Accordingly, adequate anonymisation of data and assessment of the risk of re-identification constitute a significant challenge.

[15] L. Sweeney, Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

One approach to ensuring data anonymisation appears in the U.S. Health Insurance Portability and Accountability Act (HIPAA)[16], in section 164.514 of the Act, and is based on two possible channels:

1. **Expert determination** for data anonymization: A determination by a person with appropriate knowledge and experience applying accepted statistical and scientific principles and methods that the risk that the information could be used, alone or in combination with other reasonably available information, to identify the data subject is very small, while documenting the methods and results of the analysis that justify this.
2. **Safe Harbor approach:** removing 18 families of direct and indirect identifiers (such as addresses, license plates, dates, web browsing identifiers, etc.)

The ability to extract personal data from anonymised datasets depends on several factors. Various models and metrics have been developed to evaluate the risk of re-identification and to compare the levels of details across different datasets.

The k-Anonymity model is one of the models that may be used to assess the level of re-identification risk of data in a database. The reference database includes direct identifiers (fields that allow identification of a person – such as name or email address), indirect identifiers (fields that, in combination with additional information, allow identification of a person – such as date of birth or neighborhood of residence) and a sensitive attribute (data that the identity of the person to whom it belongs needs to be protected). With respect to this database, and assuming that all direct identifiers have

[16]Section 164-514: https://www.govinfo.gov/content/pkg/CFR-2017-title45-vol1/pdf/CFR-2017-title45-vol1-sec164-514.pdf .
Further info: Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, US Department of Health and Human Services, Office for Civil Rights, 2010

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

been removed from it, the k-Anonymity method examines the smallest number of records with identical indirect identifiers.

In a database that meets the definitions of k-Anonymity, each record has at least k records with same indirect identifiers (assuming that the database does not include direct identifiers). This allows for better protection against attribute linkage attacks, since it is not possible to link the sensitive attribute to a particular record out of the k records without additional information.

Example: The k-Anonymity model

The table in Figure 4 presents a portion of a dataset containing three quasi-identifiers (age, gender, and marital status), alongside a sensitive attribute (number of children). In this dataset, a group of three individuals: all male, married, and aged between 40 and 50 – meets the requirement for k-anonymity with k=3. This group is highlighted with a thick border.

Quasi-identifiers			Sensitive attribute
Age	Marital status	Gender	Number of children
[40-50]	Married	M	2
[40-50]	Married	M	3
[40-50]	Married	M	4
[30-40]	Married	W	3
[30-40]	Married	W	2
[30-40]	Married	W	3

Figure 4: Example for a dataset that satisfies the requirement for 3-Anonimity

If, as in Figure 4, every unique combination of quasi-identifiers (age, gender, and marital status) in the dataset appears in at least three records, the dataset as a whole can be said to satisfy 3-anonymity with respect to the sensitive attribute (number of children).

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

This means that no single record may be uniquely identifiable based on the combination of quasi-identifiers; rather, each such combination appears in three or more records, thereby reducing the risk of re-identification through linkage attacks.

As the k value in the *k-anonymity* model increases, the likelihood of re-identification decreases, but the level of generalization rises, which may reduce the usefulness of the data. In addition, the k-Anonymity model assumes that each individual has only one record in the database. If this assumption is not met (for example, for a hospitalization data database that may include multiple records for the same individual), it is necessary to increase the k value significantly, so that the chance that all or most of the values in the k group will belong to the same individual is reduced.

The range of values for k-Anonymity is determined according to the characteristics of the information on a case-by-case basis, since the likelihood of re-identification varies from case to case and depends on many parameters. It is not possible to determine a specific general k parameter, and its definition must be in accordance with the privacy risk assessment for each case, process and system.

The definition of the mechanism and parameters for anonymisation should be done as part of an overall concept of Privacy by Design and as part of complete organizational processes that include ongoing evaluation and Privacy Impact Assessment. These mechanisms and parameters should be re-examined periodically and whenever the relevant conditions in the organization and its environment change.

Sources for further reading on anonymization
1. A guide by the Personal Data Protection Commission of Singapore (PDPC): Guide to Basic Anonymization. 2. A framework for evaluating the risks of re-identification by the Office of the Australian Information Commissioner (OAIC): De-identification Decision-Making Framework.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Sources for further reading on anonymization
3. Guidelines on removing identifying data from structured datasets by the Information and Privacy Commissioner of Ontario (Canada): De-identification Guidelines for Structured Data. 4. A public consultation draft by the European Data Protection Board (EDPB) on pseudonymization: Guidelines 01/2025 on Pseudonymisation.

Synthetic Data (SD)

Synthetic data refers to data that is artificially generated to replicate the statistical patterns and characteristics of real-world datasets that may contain personal data. The primary objective of synthetic data is to enable similar analytical outcomes to those that would be derived from actual personal data, while avoiding the use of real personal data.

Beyond privacy protection, synthetic data can serve a range of additional purposes, including the generation of large-scale datasets for training artificial intelligence (AI) models and the creation of data that reflects a wide variety of edge cases and rare events. Synthetic data may be used in combination with real (non-synthetic) data or as its complete replacement.

Common methods for generating synthetic data:

1. **Parametric generation**: Developing a model that captures the statistical properties of the source data (e.g., mean, standard deviation, correlation) and producing synthetic records based on that model.
2. **Machine learning and artificial intelligence**: Generating synthetic data from learned patterns and features extracted during the training process on real datasets.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Example: Synthetic dataset by the World Bank[17]

In 2023, the World Bank released a synthetic dataset for simulation and training purposes. This dataset includes data on 8,000 households, representing a sample of the population of a fictional, middle-income country. The data includes variables commonly collected in population censuses (e.g., education, occupation, housing characteristics, and family characteristics) and household surveys (e.g., expenditures or asset ownership).

The dataset was generated using artificial intelligence, based on data resampled and re-encoded in a manner that prevents any linkage between the synthetic records and the original data. The dataset is intended for educational and simulation purposes and does not represent any specific country.

Records in a synthetic dataset should not correspond to real individuals. While incidental matches may occur by chance (e.g., due to random sampling), these should be rare. However, under certain circumstances, it may still be possible to infer identifiable personal information from synthetic data.

Key threats to synthetic data:

1. **Database reconstruction attack** – Reconstructing personal data that was used in the generation of the synthetic dataset.
2. **Attribute inference attack** – Inferring a specific personal attribute in the data used to generate the synthetic dataset.
3. **Membership inference attack** – Determining whether the data of an individual (or a group) was included in the original dataset used for synthesis.

The use of synthetic data inherently reduces the privacy risks and allows for further uses of the data. However, under certain conditions, there may be situations where it

[17] World Bank. (2023). Synthetic Data for an Imaginary Country, Sample, 2023 [Data set]. World Bank, Development Data Group.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

will be possible to re-identify personal information (for example, in case of many unique records). Therefore, even when using synthetic data, privacy risk analysis and additional measures to protect the information (such as k-Anonymity or differential privacy) are required.

Sources for further reading on synthetic data
1. A guide for generating synthetic data by the Singapore Personal Data Protection Commission (PDPC): Proposed Guide on Synthetic Data Generation. 2. A guide for generating synthetic data by the Alan Turing Institute and the Royal Society (UK): Synthetic Data - What, Why and How? 3. An article discussing synthetic data by the International Association of Privacy Professionals (IAPP): Synthetic Data: What Operational Privacy Professionals Need to Know.

Differential Privacy (DP)

Differential privacy is an approach designed for the analysis of datasets that include personal data but are intended for statistical processing rather than for disclosure of individual-level data. The core principle of differential privacy is that the inclusion, removal, or modification of a single record in a dataset should have a negligible and ideally unnoticeable effect on the output of the statistical analysis performed on that dataset.

Rather than removing or masking identifiers, differential privacy is achieved by injecting random noise into the data. This added noise ensures that individual-level data cannot be inferred from the aggregate result. Despite the added noise, differential privacy techniques are designed to preserve the overall utility of the data, enabling accurate insights at a statistic level.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Example: Differential privacy by randomized response mechanism

A classic method to implement differential privacy is the randomized response mechanism, which enables the collection of sensitive data while protecting the privacy of individual responses.

For example, consider a survey question: *"Did you vote for Party X or Party Y?"* Each respondent flips a coin in private and answers according to the result:

1. If the coin lands heads: they provide a random answer (Party X or Party Y, with 50% probability for each).
2. If the coin lands tails: they provide their truthful response.

Only the respondent knows whether their answer is truthful or random. As a result, the data collector cannot determine any individual's true response. However, with a sufficiently large sample size, the overall distribution of answers approximates the true distribution in the population.

To demonstrate how the result is calculated without knowing the respondents' true answers, suppose that the true distribution is 60% for Party X and 40% for Party Y. Half of the respondents will answer truthfully (yielding a 60:40 split), and the other half will respond randomly (resulting in a 50:50 split). We will not be able to tell which of the respondents answered randomly and which gave a true answer, but only the overall distribution (which is the average between the two distributions, i.e. approximately 45%:55%).

This is sufficient to calculate the true distribution, which can be done by multiplying the gap (10%) by a factor of two. This gives us a true gap of 20%, meaning the true distribution will be approximately 60% versus 40%, as we set in the example.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Differential privacy is based on the idea of a "small difference" between two datasets that differ in one record, so that personal data about a particular person cannot be inferred by comparing the results. This concept describes extracting data (querying) the datasets in two cases:

1. When a data item of a particular person is removed from the data set.
2. When a data item of that person is kept as it is in the data set.

Differential privacy ensures that the result of the query will be very similar in the two cases above. That is, the impact of adding a person's details will be negligible, and therefore the chance of personal data leaking from comparing the results of the processing with or without the data about that person will be very small.

Example: An illustration of a case where differential privacy is not preserved[18]

A university publishes aggregate statistics on students' monthly incomes. According to its publications, in April, 304 students were enrolled, 30 of whom earned over 30,000 ILS. In May, 303 students were enrolled, and 29 earned over that amount.

Although these are aggregate data, it can be concluded that in May a student whose income was over 30,000 ILS per month left the university. As a result, the classmates of that student may be aware of their income level.

This example describes a situation in which differential privacy is not preserved, meaning that information can be inferred from a comparison of information for groups that differ from each other in the details of only one person.

An important component of differential privacy is the epsilon (ε) value, or "privacy budget," which determines the level of noise that will be added to the dataset. A

[18] Following: A. Wood, M. Altman, A. Bembenek, M. Bun, M. Gaboardi, J. Honaker, K. Nissim, D. R. OBrien, T. Steinke, and S. Vadhan. "Differential privacy: A primer for a non-technical audience." Vanderbilt Journal of Entertainment & Technology Law 21, no. 1 (2018): 209-275.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

formal definition of differential privacy is that the effect of adding or removing a single item will not change the query value beyond the parameter ε. The smaller ε, the higher the privacy of users, but the use of the tool will require stronger noise and therefore the data may be less useful.

Example: The impact of ε on insurance premiums[19]

Consider a woman named Ayelet, who holds a life insurance policy worth 100,000 ILS. Based on actuarial data, her annual premium is 1,000 ILS, reflecting a 1% estimated risk of death.

Ayelet participates in a medical study that shows a 50% probability of death due to stroke in the coming year. If these data were disclosed to the insurer, her premium could rise dramatically to 50,000 ILS and more.

However, if the study's data are released aggregately under differential privacy constraints with ε = 0.01, the insurer's revised estimate of her risk could increase by no more than 1% × (1 + 0.01) = 1.01%. Hence, her premium could rise by a maximum of 10 ILS.

Considerations and limitations

Differential privacy is a tool that depends on the data, the type of processing, and the mechanism for publishing the data. For data with high variability, it will be necessary to add stronger noise to achieve a similar level of data protection.

Differential privacy allows for anonymity as long as an adequate level of noise is added, and as long as the answer to the query is anonymous data (e.g., aggregated data). The outcomes of differential privacy are effective for statistical analysis of general trends, but are less suitable for identifying anomalies or patterns within the data itself due to the additional noise. It should be emphasized that differential privacy products will not necessarily yield anonymous information, and therefore each case must be examined

[19] See footnote 18.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

on its own merits. If the tool is not implemented properly, there is a risk of information leakage through repeated queries by an attacker.

Sources for further reading on differential privacy
1. An introductory resource from Harvard University: Differential Privacy: A Primer for a Non-Technical Audience. 2. An open-source project by Harvard University for developing differential privacy tools in various contexts: OpenDP. 3. An introduction and resources on differential privacy by Dr. Damien Desfontaines: A Friendly, Non-Technical Introduction to Differential Privacy.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Technologies for Reduction of Data Exposure in Use

While the protection of data at rest and in transit is a fundamental element of modern digital infrastructures, PETs extend the scope of protection to data in use. These technologies aim to minimise the exposure of personal data during processing and thereby reduce associated risks through diverse methods, processes, and digital tools.

Homomorphic Encryption (HE)

Homomorphic is an algebraic term that means 'of same (equal) shape', referring to the (partial) preservation of plaintext structure in ciphertext. This special algebraic structure enables performing specific algebraic operations on ciphertext (without opening it), and under certain conditions obtaining a result that corresponds to same operation on the plaintext.

Therefore, homomorphic encryption allows the use of data while it is encrypted without the need to open the encryption, and accordingly reduces the risk of the information being exposed to an unauthorized party during processing.

The process of working with homomorphic encryption begins with encrypting the information and creating an *evaluation key* that allows working on the encrypted data. Then, a calculation is performed on the data in its encrypted form (without revealing it). Finally, the result can be opened using the *secret key*. In this process, the information is protected (encrypted) in use and this method is particularly useful in scenarios where data must be outsourced to untrusted environments, such as remote or edge devices.

Example: Illustrative example: principle of homomorphic encryption

Homomorphic encryption involves the use of unique encryption protocols. To illustrate the principle of using homomorphic encryption, we will present a naive

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

scheme (which does not follow modern encryption mechanisms). In the example, we will refer to sending two numbers for the purpose of performing a multiplication operation on an external server, when we do not want the server to be exposed to the numbers themselves or to the result of the multiplication. The encryption that we will use in this scheme will be raising the numbers to the power of a secret number that will change for each calculation operation.

For the purpose of the demonstration, we will choose the numbers 2 and 3. We will apply the 'encryption' by raising them to the power of a secret number (for the sake of the example – power of 2), so that the information that will be sent to the server will be the numbers 4 and 9. The server will perform the multiplication operation and return the result – 36. On this number it will be possible to perform the square root operation and thus obtain the correct result – 6.

In this example, the data are encrypted before it is sent from the client for processing on the server and the server operates on the data in their encrypted form. The result is sent from the server to the client, who decrypts the data to receive the result, so that at no stage is the data exposed to any party other than the client.

As mentioned, this implementation scheme serves to illustrate the principle of operation only, and the implementation of homomorphic encryption requires the implementation of all relevant algorithms and data security mechanisms, the details of which are beyond the scope of this document.

Implementation protocols for homomorphic encryption support diverse degrees of computational flexibility:

1. **Partially Homomorphic Encryption (PHE)**, **Somewhat Homomorphic Encryption (SHE)**, and **Leveled Fully Homomorphic Encryption (LFHE)** allow only specific types of operations or a limited number of operations.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

2. **Fully Homomorphic Encryption (FHE)** enables unlimited and arbitrary computations on encrypted data, representing the most advanced form of this technology.

Homomorphic encryption produces a significant computational load required by working with the information in its encrypted form compared to processing unencrypted data. Therefore, this method is less efficient when it is necessary to process a large amount of data, and accordingly it is recommended to focus the use of homomorphic encryption on data whose encryption will help to the greatest extent to reduce the sensitivity and risk to personal data. Technological developments that will contribute to higher computing speeds and the development of more efficient algorithms are expected to contribute to reducing the computational limitations in implementing homomorphic encryption and accordingly to expanding the use of this technology.

The use of homomorphic encryption requires the precise use of appropriate tools and infrastructure. It is important to ensure the use of tested algorithms and software, appropriate encryption levels (key size) and the protection of private keys, as well as defining processes for generating an additional private key and re-encrypting the information that has already been encrypted in the event that the original key is exposed.

Sources for further reading on homomorphic encryption
1. OpenMined Blog on Homomorphic Encryption and Its Implementation**:** What is Homomorphic Encryption?. 2. An introduction to fully homomorphic encryption by prof. Boaz Barak (Harvard University): Fully Homomorphic Encryption: Introduction and Bootstrapping.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Zero Knowledge Proof (ZKP)

A Zero Knowledge Proof is a technique that, in certain cases and in response to specific queries, allows the proof of a property or the validity of data without revealing the data itself. ZKPs can be used to demonstrate attributes such as age (without disclosing the birth date), financial status (without revealing financial data), ownership of property (without disclosing transaction details), and, in principle, can support identification methods such as facial recognition, fingerprints, and voice confirmation.

ZKPs are implemented through a series of actions whose results depend on specific data. If, the outcome presented by the prover is correct consistently over multiple queries, one can assert with a high degree of confidence that the prover possesses the required data. In this manner, the knowledge of certain data can be demonstrated without the data itself being exchanged between the parties.

Example: ZKP – possible implementation

A potential implementation of a ZKP is a request for a function value that corresponds to a specific input. If the user consistently returns the correct value of the function over multiple requests, it strongly indicates that the user possesses knowledge of the function, as the probability of guessing the correct value decreases with each additional round of requests.

In the example illustrated in Figure 5, the client (left figure) requests the server to provide the values of a secret function known to the client at certain points (the blue circles in the diagram). The server (center figure) performs the computation and sends the values back to the client for verification. The client (right figure) matches the values received from the server to those of the secret function in its possession. If the received values are sufficiently and repeatedly close to those of the function, the client can conclude that the server knows the secret function.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

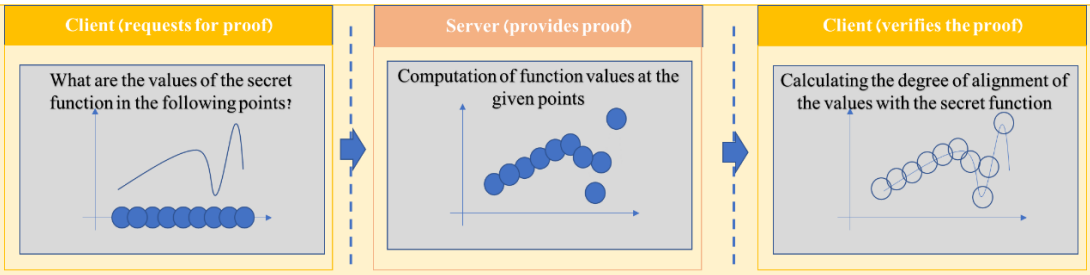

Figure 5: Zero Knowledge Proof – Example of Implementation

This process enables the verification that the server possesses knowledge or a property (in this case, knowledge of the function), without the client needing to present or transmit this information to a third party.

ZKPs reduces the exposure of personal data, as this mechanism allows the desired objective to be achieved without sharing the full data with another party. However, careful protocol design is required for each specific case, and, as such, this method is suitable for focused queries and the protection of specific sensitive and valuable information. Since ZKPs provide a certain degree of probability rather than absolute certainty about the accuracy of the data, it is crucial to tailor the required level of certainty (and, accordingly, the number of computations or difficulty of the evaluation task) to the specific application.

Sources for further reading on zero knowledge proofs
1. Lecture by Professor Alon Rosen (Reichman University) on Zero Knowledge Proofs: Introduction to Zero Knowledge. 2. Technical lecture on Zero Knowledge Proofs by Professor Shafi Goldwasser (MIT, Weizmann Institute): Introduction and History of ZKP.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Multi-Party Computation (MPC)

Multi-party computation is a cryptographic protocol that enables multiple participants to compute or perform operations on personal data while ensuring that the data of each participant remains private. Each participant can access the result of the shared computation, depending on the specific protocol and processing method implemented. A classic example of MPC is the "Millionaire's Problem", where a group of individuals wishes to know who has the highest wealth, without revealing any data, including their own wealth, to others.

A possible implementation of MPC is secret sharing. This technique divides the data that needs protection into several parts, distributing the parts among the participants in the computation. No party can access another's data unless all pieces distributed to the parties are combined. This method reduces the likelihood that the data will be exposed, as all parties must collaborate to access the complete data.

Example of MPC implementation

Suppose three employees (Ayelet, Ben, and Jessica) wish to calculate the average of their salaries without revealing their own salaries to each other. To do so, each employee divides their salary into three random parts (so that the sum of these parts equals their salary). For example, if Ayelet's salary is 15 thousands ILS, she could divide it into three parts: 9, 2, and 4 thousands ILS. A table of possible values for the three employees is shown below:

Employee	Salary	Part A	Part B	Part C
Ayelet	15	9	2	4
Ben	18	20	10	-12
Jessica	12	8	1	3

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Each employee keeps one part for themselves and sends the other two parts to the other participants (e.g., Ayelet sends Part B to Ben and Part C to Jessica). Now, the parts with the participants are:

Employee	Salary	Part A (original)	Part B (received)	Part C (received)
Ayelet	(secret)	9	10 (from Ben)	1 (from Jessica)
Ben	(secret)	20	2 (from Ayelet)	3 (from Jessica)
Jessica	(secret)	8	4 (from Ayelet)	-12 (from Ben)

Finally, each employee computes the sum of the parts they got (Ayelet – 20, Ben – 25, and Jessica – 0 in our example), and the average is calculated by dividing the sum by the number of participants (3). In this case, the correct result (15) is obtained, without any participant being exposed to the salary data of the others or being able to infer it from the data they received.

The implementation of a MPC protocol can be done in several configurations, depending on the number of participants in the computation and their level of trust. Some configurations allow overcoming a number of untrusted participants and even those who deliberately launch attacks on the algorithm. This tool is implemented in a large number of practical applications and allows a high level of information protection.

MPC may require significant computational and communication effort as part of the process of distributing secrets among the participants. The tool can include homomorphic encryption or other mathematical mechanisms for decentralizing data. Potential uses of MPC include complex and significant tasks such as electronic voting or data mining.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Sources for further reading on multi-party computation
1. Description of MPC on Institute of Electrical and Electronics Engineers (IEEE): What Is Multiparty Computation? 2. Lecture by Prof. Tal Rabin (University of Pennsylvania) on Secure Multiparty Computation. 3. Article on Medium about MPC: A Crash Course on MPC.

Private Set Intersection (PSI)

Private Set Intersection is a special case of multi-party computation that enables two parties to identify common elements between their datasets, without revealing data about non-matching elements. This technique can be used to calculate the size of the intersection (the number of matching data points between the two parties) and perform statistical analysis on the common set.

PSI can be performed in two ways:

1. **Traditional PSI**: The parties communicate directly, each holding a copy of their dataset.
2. **Delegated PSI**: The computational or storage operations of the datasets are outsourced to a third party.

Example of PSI implementation
Suppose two parties wish to compare their lists of names: • Party 1: {Eden, Dan, Noya} • Party 2: {Adam, Rona, Dan}

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

A naive algorithm (without applying appropriate modern cryptographic mechanisms) for PSI would use a hashing function to compare the hashed values and identify matching items. For example, the hashing function values for shared name "Dan" are highlighted:

- Party 1: {12AS33, **BCD343**, 14MHGH}
- Party 2: {67887Q, 4ERT67, **BCD343**}

This example illustrates the basic principle of the operation; however, implementing PSI requires using proper cryptographic algorithms and security mechanisms that are beyond the scope of this document.

PSI is commonly used in various applications, such as identifying contacts in messaging applications or social networks, i.e., finding mutual friends without revealing non-mutual friends. More generally, PSI applications include matching (e.g., human genome matching), database merging (e.g., contact lists), or measuring data conversion rates in online advertising.

PSI implementation relies on following precise cryptographic protocols. When implementing PSI, it is essential to consider the size of the data sets, the level of trust between the parties sharing the data, and the integration of additional privacy-enhancing technologies.

Sources for further reading on private set intersection
1. Lecture on PSI, available on the website of the U.S. National Institute of Standards and Technology (NIST): A Brief Overview of Private Set Intersection. 2. A technical analysis of PSI Applications, available on the website of University of California: What are we PSIgning up for? Analyzing Applications of Two-Party Private Set Intersection.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Federated Learning (FL)

Federated learning enables multiple parties to collaboratively train artificial intelligence (AI) models on their local data, and then aggregate the patterns identified in the local models into a global, accurate model, eliminating the need to share the raw data each party used locally for training.

Federated learning can be implemented in the following approaches:

1. **Centralized Federated Learning**: A central server generates an algorithm or a model, and sends it to distributed data sources. The model is updated according to the local data sources and sent back to the central server, which creates a unified weighted model.
2. **Decentralized Federated Learning**: In this model, there is no central server involved. The parties communicate directly with each other and update the model based on their local data.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | 6133401 תל-אביב ,33503 .ת.ד ,קרית הממשלה

Example of centralized federated learning implementation scheme

Many companies are using federated learning to train AI systems while preserving users' personal information and privacy. A common centralized distributed learning process is shown in Figure 6:

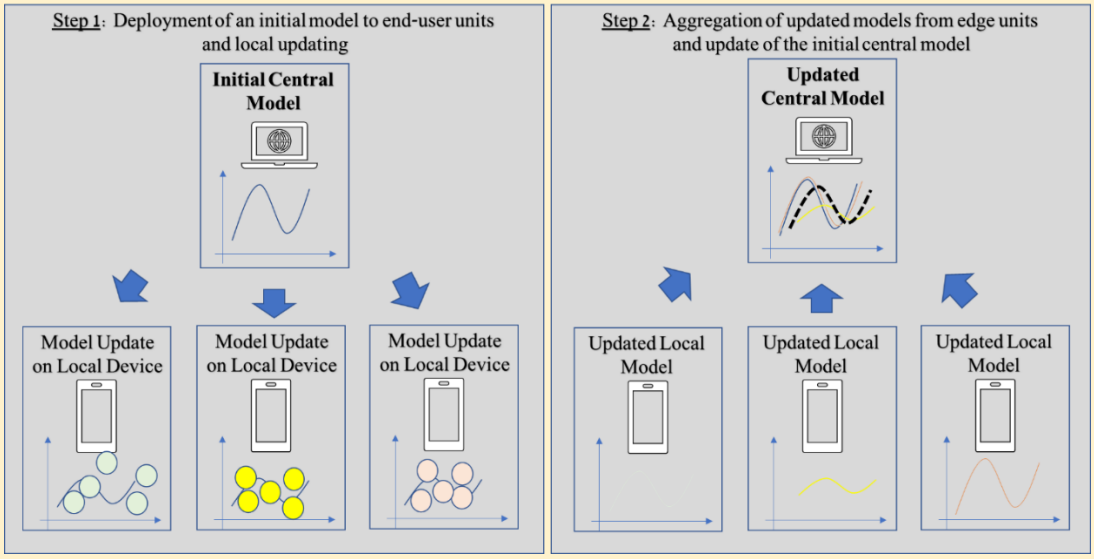

Figure 6: Illustration of a Centralized Federated Learning scheme

Step 1 (Figure 6 left) is building the initial model and distribute it to edge devices. The edge devices then train and complete the model based on the data they hold.

Step 2 (Figure 6 right) is sending the updated model from the edge devices back to the central unit. The central unit aggregates the models and produces a new, revised model.

This process may repeat multiple times to refine the models, for example as new data are received or following user input and instructions.

Federated learning may require significant computational loads, especially when processing large-scale data. To address these challenges, practical implementation of federated learning should account for data set characteristics, the risks of re-

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

identification from models, and the integration of additional privacy-enhancing technologies to enhance personal data protection.

Sources for further reading on federated learning
1. Dr. Aviv Keren's lecture at the Yuval Ne'eman Science, Technology, and Security Workshop at Tel Aviv University: Federated Learning in the Real World. 2. OpenMined Blog on Federated Learning: Design a Federated Learning System in Seven Steps.

Trusted Execution Environment (TEE)

A Trusted Execution Environment is a secure area in a processor, server, or mobile device used to process sensitive data. It allows code to run and access data in physical or logical isolation from other parts of the system, ensuring that unauthorized software or users outside of the TEE cannot access the processed data. Data stored and processed within a TEE is encrypted, ensuring that even if an attacker gains access to the device, they cannot easily extract or modify sensitive data.

Example – TEE: infrastructure and uses
Many industrial systems utilize TEEs. The infrastructure for TEEs is embedded in processors by companies such as Intel, AMD, NVIDIA, and others. The use of TEEs is also common in mobile devices and operating systems, such as iOS and Android, which support this technology. The development of TEEs initially relied on hardware-based mechanisms, incorporating dedicated encryption capabilities and keys. In recent years, software-based systems and mechanisms for identification, authentication, and encryption have been developed, enabling the operation of software-based TEEs.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Common TEE applications include:

1. **Biometric identification**: Matching biometric data against identity engines in biometric authentication processes (face recognition, fingerprints, and voice recognition). TEE provides protection for biometric information and creates a barrier against insecure applications running on the system.
2. **Secure computation**: Performing computations without providing access to the data or code used for those computations. A secure environment for executing sensitive code, such as encryption algorithms or information-sensitive software ensures that sensitive code is not exposed or modified by other software or attackers.
3. **Financial services**: Providing secure access to digital wallets and credit card details for payments, especially in mobile devices. Digital wallets, such as Google Pay or Apple Pay, use a trusted execution environment to store and protect the encryption keys required to make payments securely.
4. **Data analysis and machine learning**: Ensuring privacy during computations such as Multi-Party Computation and Federated Learning.

Sources for further reading on trusted execution environments
1. The Confidential Computing Consortium on TEE: Confidential Computing: Hardware-Based Trusted Execution for Applications and Data. 2. Eurostat's project on using TEE for statistical processing: Project ESTAT.2019.0232.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Technologies for Data Access Monitoring

The protection of personal data can be enhanced through mechanisms that enable the data subject to have better control over access permissions to their personal data, as well as to monitor access events related to their data.

Personal Data Stores (PDS)

Personal Data Stores are tools that incorporate mechanisms to track access to personal information. These are technological solutions that allow users to manage, store and use their personal data in an organized and secure manner. PDS are designed to hold information for various applications and services, and allow users to access and monitor the use of their information.

PDS are designed to store data for various applications and services, and may include data from diverse sources, such as contact details, medical history, account data, or documents. Many of these stores operate on cloud platforms, where data are encrypted and protected to ensure access is restricted to authorized users. Alternatively, local storage on personal devices (such as computers or smartphones) with local data security measures may be employed.

PDS users can control and manage access permissions to their information. This allows users to share data with others securely if they choose to do so. In some cases, these repositories allow information to be shared with other services or applications in a controlled manner, with the user's explicit consent. Users can delete or update their information as needed and to maintain their privacy.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Example: Potential use case for personal data stores

Traditionally, financial information is stored by the relevant organization that generates or utilizes it, such as a bank, an employer, or a credit card company. In this method, the transfer of data between different institutions is incomplete and must be carried out each time the other institution requires the data.

Utilizing PDS for financial information would enable individuals to keep their data in one place while granting access to authorized bodies as needed. For instance, a user could grant viewing access to salary and bank transaction data to mortgage lenders for assessing loan eligibility. This approach enables the user to be better informed about the use of their data, providing the ability to inquire about access events lacking clear context, or following such unresolved events – revoke access permissions to prevent potential misuse in the future.

While PDS offer improved convenience and better control, they also require users to take significant responsibility in managing access permissions and monitoring subsequent activity. Insufficient attention to these details may lead to unauthorized access or misuse of personal data.

Sources for further reading on personal data stores

1. U.S. Department of Health & Human Services: Personal Data Stores (PDS): A Review.
2. BBC Report on a Personal Data Store system that manages a unified media profile for the viewer: Personal data stores: building and trialling trusted data services.
3. Article about Personal Data Stores on Medium: What IS a Personal Data Store?

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

Documentation and Transparency Tools (DTT)

The protection of personal data can be strengthened by implementing mechanisms that ensure accurate and immutable documentation of data access, along with effective tools that enable data subjects to monitor how their personal data are used. The integration of such mechanisms contributes to building trust, enhancing transparency and promoting a culture of privacy protection by ensuring that data processing activities are recorded and these records are made accessible to the individuals themselves.

DTTs are evolving privacy-enhancing solutions designed to provide data subjects with new capabilities to monitor, oversee, and control the processing of their personal data.

Example: Data Tracker Portal

The government of Estonia operates a Data Tracker Portal[20], which allows individuals to track the use of their personal data within government-managed repositories. Through the portal, users can view a detailed access history, showing when their personal data was accessed, from which dataset, by which unit or organization, and for what purpose or type of operation.

Publicized cases of improper data use in Estonia, brought to light through the country's digital infrastructure, which enables citizens to monitor access to their personal data, include incidents such as a police officer accessing personal data about his future spouse and a medic reviewing why an ambulance had been sent for a specific address at the request of a curious neighbor[21].

Another example of a documentation and transparency mechanism is Australia's national health records system. This system enables individuals to receive real-time notifications whenever their health records are accessed or when specific types of

[20] Described in further details at e-Estonia site: Data tracker – tool that builds trust in institutions.

[21] Reported at e-Estonia site - I spy with my little eye…privacy!

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401

changes are made. Because the alert is triggered at the time of access, the entity accessing the data cannot subsequently deny having done so[22].

Summary

The use of privacy-enhancing technologies (PETs) supports privacy protection through a range of methods and tools. Integrating these technologies into products and systems should be guided by a systemic view of the data in use and its lifecycle, associated privacy risks, processing activities, data sensitivity, and the relevant organizational and legal frameworks. Broadening the adoption of PETs can strengthen privacy safeguards as part of a comprehensive strategy to protect data in the digital era, complementing existing legal measures for enforcing privacy laws.

[22] Further information is available in the Notifications and Message Settings section of the Australian Digital Health Agency website.

ppa@justice.gov.il | 073-3928529 | *3103
www.go.gov.il/ppa | קרית הממשלה, ת.ד. 33503, תל-אביב 6133401