How effective are privacy-preserving technologies in addressing ethical and societal issues?
About the e-SIDES project

Data-driven innovation is deeply transforming society and the economy. Although there are potentially enormous economic and social benefits, this innovation also brings new challenges for individual and collective privacy, security, as well as democracy and participation. The main objective of the CSA e-SIDES is to complement the research on privacy-preserving big data technologies, by analysing, mapping and clearly identifying the main societal and ethical challenges emerging from the adoption of big data technologies, conforming to the principles of responsible research and innovation; setting up and organizing a sustainable dialogue between industry, research and social actors, as well as networking with the main Research and Innovation Actions and Large Scale Pilots and other framework program projects interested in these issues. It will investigate stakeholders’ concerns, and collect their input, framing these results in a clear conceptual framework showing the potential trade-offs between conflicting needs and providing a basis to validate privacy-preserving technologies. It will prepare and widely disseminate community shared conclusions and recommendations highlighting the best way to ultimately build confidence of citizens and businesses towards big data and the data economy.

Deliverable D3.2 Assessment of Existing Technologies

Find more at: https://e-sides.eu/resources/deliverable-d32-assessment-of-existing-technologies

About this white paper

This white paper is based on Deliverable D3.2 of the e-SIDES project. Within the scope of the deliverable, privacy-preserving technologies are assessed taking ethical and societal issues into account. The paper also presents an overview of existing privacy-preserving technologies (based on Deliverable 3.1) and of the ethical and societal issues taken into account (based on Deliverable 2.2) in order to describe the results of the assessment of the technologies.

Find more: Deliverable D3.1 Overview of Existing Technologies https://e-sides.eu/resources/deliverable-31-overview-of-existing-technologies
<table>
<thead>
<tr>
<th><strong>Overview of Existing Privacy-Preserving Technologies</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Anonymisation</strong></td>
</tr>
<tr>
<td><strong>Encryption or removal of personally identifiable information</strong></td>
</tr>
<tr>
<td>In the context of big data applications, traditional anonymisation techniques fail because there are hundreds of data points for a single individual. Privacy models that may be used when anonymizing data include k-anonymity, l-diversity, t-closeness and differential privacy.</td>
</tr>
<tr>
<td><strong>Encryption</strong></td>
</tr>
<tr>
<td><strong>Encoding of information so that only authorised parties can access it</strong></td>
</tr>
<tr>
<td>In the context of big data applications, combining the advantages of public key encryption in scalability and key management with the speed and space advantages of symmetric encryption is not enough. Fine-grain sharing policies are necessary that go beyond the <em>encrypt all or nothing</em> model. Relevant cryptographic primitives include attribute-based encryption, proxy re-encryption and functional encryption.</td>
</tr>
<tr>
<td><strong>Sanitisation</strong></td>
</tr>
<tr>
<td><strong>Encryption or removal of sensitive information</strong></td>
</tr>
<tr>
<td>Sanitisation techniques other than encryption and removal of columns include masking data, substitution, shuffling and number variance. A key problem is that it can be difficult to find substitution data in large quantities.</td>
</tr>
<tr>
<td><strong>Access control</strong></td>
</tr>
<tr>
<td><strong>Selective restriction of access to places or resources</strong></td>
</tr>
<tr>
<td>Big data applications typically require fine-grain access control. Approaches such as role-based access control increasingly lack manageability. Attribute-based access control and similar approaches support fine-grain access control policies in big data based on attributes that are evaluated at run-time.</td>
</tr>
<tr>
<td><strong>Multi-party computation (MPC)</strong></td>
</tr>
<tr>
<td><strong>Distribution of data and processing tasks over multiple parties</strong></td>
</tr>
<tr>
<td>MPC is a field of cryptography with the aim to allow securely computing the result of any function without revealing the input data. Although MPC was proven to be theoretically plausible, there are only very few practical applications. Key challenges in the big data context are utility, performance and ease of use.</td>
</tr>
</tbody>
</table>
### Overview of Existing Privacy-Preserving Technologies

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Policy enforcement</strong></td>
<td><strong>Enforcement of rules for the use and handling of resources</strong>&lt;br&gt;Automated policy enforcement mechanisms are particularly important in the big data era as policies get easily lost or neglected in the course of data being transferred between different systems. Data expiration policies, for instance, are already enforced by some big data solutions.</td>
</tr>
<tr>
<td><strong>Accountability</strong></td>
<td><strong>Evaluation of compliance with policies and provision of evidence</strong>&lt;br&gt;A cornerstone of accountability in the context of big data applications is the provision of automated and scalable control and auditing processes that can evaluate the level of compliance with policies. Provable data processing and proofs of retrievability are among the main approaches to cloud data-integrity verification without retrieval.</td>
</tr>
<tr>
<td><strong>Transparency</strong></td>
<td><strong>Explication of information collection and processing</strong>&lt;br&gt;In the context of big data applications, transparency may be achieved by purely textual information, multichannel and layered approaches and standardized icons and pictograms. Transparency is considered critical to allow data subjects informed choices.</td>
</tr>
<tr>
<td><strong>Access and portability</strong></td>
<td><strong>Facilitating the use and handling of data in different contexts</strong>&lt;br&gt;Having access to data means that data subjects can look through and check the data stored. Portability gives data subjects the possibility to change service providers without losing their data.</td>
</tr>
<tr>
<td><strong>User control</strong></td>
<td><strong>Specification and enforcement of rules for data use and handling</strong>&lt;br&gt;Means that allows reaching user control include consent mechanisms, privacy preferences, sticky policies and personal data stores.</td>
</tr>
</tbody>
</table>
# Ethical and Societal Issues of Big Data Technologies

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Privacy</strong></td>
<td>A claim, an entitlement or a right of an individual to determine what information about himself or herself can be communicated to others. Privacy protects the dignity and human identity. It includes self-respect and respect towards all humans by others. The intrusion into people’s privacy and organisations’ business practices is perceived as problematic. Big data has integrated itself into nearly every part of people’s online life and to some extent also in their offline experience.</td>
</tr>
<tr>
<td><strong>Self-determination</strong></td>
<td>The free choice of one’s own acts without external compulsion. Self-determination is the ability to make choices based on personal beliefs, values and preferences. Big data driven profiling practices can limit free will, free choice and be manipulative in raising awareness about, for instance, news, culture, politics and consumption. Thereby, they impede autonomy. The pressure towards conformity is referred to as normalisation. This restricts the breadth of choices and pushes back pluralism and individuality.</td>
</tr>
<tr>
<td><strong>Welfare</strong></td>
<td>The general state of health or the degree of success of a person. Welfare ensures well-being and quality of life. Big data-based calculations in which commercial interests, rather than non-profit-led interests, are prioritised, are examples of situations in which solidarity is under pressure. Detrimental implications can emerge in the contexts of employment, schooling or travelling by various forms of big data-mediated unfair treatment of citizens and adversely affect human welfare. Big data has indirect effects on the environment.</td>
</tr>
<tr>
<td><strong>Interdependency</strong></td>
<td>The dependence of two or more people or organisations on each other. Interdependency highlights the interconnection and interdependence of individuals and organisations. The dependency of people and organisations on organisations and technology leads to a limitation of flexibility. Organisations are strongly dependent on the data as well as the big data technologies they use. Due to the fact that the application of big data technology is not a single linear process, but consists of different stages, with different actors involved, the harms connected to it can have an incremental character that is not only difficult to articulate but also difficult to attribute to any given stage or actor.</td>
</tr>
<tr>
<td><strong>Trustworthiness</strong></td>
<td>The ability to be relied on as honest or truthful. Trustworthiness refers to the degree to which an individual or organisation can be trusted to act in a manner consistent with their commitments. Citizens often do not know how to refuse or rectify their digital profile, in case there are falsely accused. For example, false negatives during biometric identification, false positives during profiling practices. Their trust is then undermined. Data reuse in the world of big data can also have diverse detrimental effects for citizens. This puts non-maleficence as a value under pressure. The risk of abuse is not limited to unauthorized actors alone but also to an overexpansion of the purposes of data use by unauthorized actors.</td>
</tr>
<tr>
<td><strong>Accountability</strong></td>
<td>The properties that ensure that the actions of a person or organisation can be traced uniquely to the person or organisation. Accountability is the ability to be held responsible for one’s actions. Big data technologies challenge to a large extent the possibility of holding accountable any single actor taking part in the different stages of big data gathering, processing and decision making. There is a lack of transparency with respect to organizational algorithms and business practices. Algorithms, for instance, are not only opaque but also mostly unregulated and thus perceived as incontestable.</td>
</tr>
<tr>
<td><strong>Fairness</strong></td>
<td>Impartial and just treatment or behaviour without favouritism or discrimination. Fairness puts constant pressure on the value of justice. As none of the relevant rights is of absolute character, in cases of conflict with another right or interest, a right can be limited pursuant to the principle of proportionality. Not everybody or every organisation is in the same starting position with respect to big data. Discrimination is understood as the unfair treatment of people and organisations based on certain characteristics.</td>
</tr>
<tr>
<td><strong>Legislation</strong></td>
<td>A law or set of laws suggested by a government and made official by a parliament. The current legal framework for protection of human rights displays a number of vulnerabilities in the context of big data applications. It provides for several provisions which shift the obligation of balancing different fundamental rights and interests on the private actors. With little guidance provided, the question is about the legitimacy of such decisions, especially where obligations of private actors lack transparency.</td>
</tr>
</tbody>
</table>
Assessment of Existing Technologies

METHODOLOGY
The assessment, based on interviews and desk research, consists of 2 parts:

- A technology-specific assessment of selected classes of privacy-preserving technologies;
- A more general assessment of the technologies.

INTERVIEWS
The first part of the interviews focused on the assessment of the relevance and applicability of selected privacy-preserving technologies for addressing ethical and societal issues faced in the context of big data applications. The following questions were discussed:

- What societal and ethical issues can be addressed by the technologies?
- How effective are the technologies in addressing the issues?
- What problems may arise when addressing the issues with the technologies?

In the second part of the interviews, the technologies were assessed more generally by discussing the following questions:

- To what extent are the technologies integrated in today's big data solutions?
- Is there a significant demand for big data solutions that include the technologies?
- What role do regional/cultural differences play (mainly concentrating on North America and Europe)?
- To what extent do the technologies need to be complemented by non-technical measures?
- Who along the data value chain is or should be responsible for addressing the issues?

Approximate duration: 30-45 minutes.
Interviewees: renowned experts from research, industry or data protection authorities based in Europe, North America or the Middle East.

DESK RESEARCH
The desk research focused on the same set of questions as the interviews. The desk research took into account articles published in journals and conference proceedings, as well as magazines and newspapers.
MAIN SOCIETAL AND ETHICAL ISSUES ADDRESSED

Privacy
Anonymisation and sanitisation techniques help removing personally identifiable or other sensitive information from datasets, thus improving privacy protection.

Legality
Anonymisation and sanitisation technologies also play a key role with respect to legality. Data protection law, above all the new EU General Data Protection Regulation (GDPR), clearly states what measures have to be taken by the affected organisations.

EFFECTIVENESS

Technologies can be used to protect, anonymise or aggregate data in ways that are effective and efficient. Effective means, in this context, that re-identifying individual information becomes either impossible or costly enough to be unprofitable. Efficient means that the desired transaction can be completed with no additional costs for the involved parties.¹

Most of the interviewees consider anonymisation and sanitisation technologies as the most relevant and mature technologies in the big data context.

KEY PROBLEMS IDENTIFIED

To determine the optimal balance between improved privacy protection through anonymisation and sanitisation, and the usefulness of data for decision making. Sometimes there is no satisfying trade-off: either some utility and very weak privacy, or some privacy and hardly any utility;

The uniqueness of certain characteristics or behaviours;

The fact that it is usually unknown what other information is available to potential adversaries;

The fact that it is not unlikely that data that was anonymised at a specific point in time can be re-identified in the future, given various technological and social changes.

THE INTERVIEWEE’S OPINION

Senior scientist at a multinational technology company headquartered in Europe

Anonymisation is particularly difficult with respect to medical data. The difficulties are caused by free text in data that includes names, indirect descriptions of things such as diseases or treatments, or dates that can easily be cross-related with other data sources.

MAIN SOCIETAL AND ETHICAL ISSUES ADDRESSED

Privacy
Encryption is fundamental for the protection of data and privacy-preserving analysis. It transforms data in a way that only authorised parties can access it, which is ultimately a strong protection measure for personal data.2

Accountability
Technical measures to ensure accountability are typically based on encryption.

Legality
Some laws require organisations to take reasonable measures to avoid data breaches: encryption is fundamental in this regard.

EFFECTIVENESS

- Encryption is a cornerstone of data security;
- It plays an important role in protecting and advancing the freedom of opinion and expression;
- A solution that is considered to maximise privacy and query expressiveness, at least theoretically, is Fully Homomorphic Encryption (FHE). A big advantage is that it does not lead to a loss in data quality, which is why it is particularly important for medical applications.

KEY PROBLEMS IDENTIFIED

- Encryption is not sufficient in isolation. It is suggested to be tightly integrated with other security controls, including endpoint security, network security, application security and physical security systems, which are increasingly being run over IP-based networks;
- Even when end-to-end encryption is used, the exchange of information can be subject to judicially-ordered surveillance;
- Computation cost makes FHE relatively slow compared to other methods;
- Some of the current encryption methods may be useless in the era of quantum computing.

THE INTERVIEWEE’S OPINION

Associate professor focusing on the design, analysis and application of technologies to protect privacy at a European university

Encryption is generally very strong. However, it is important to keep the trust model in mind. As long as fully trusted parties are exchanging encrypted data and related keys, everything is fine. If the parties do not fully trust each other, encryption does not provide any protection.

MAIN SOCIETAL AND ETHICAL ISSUES ADDRESSED

Accountability  
Big data applications typically require fine-grained access control. Access control consists of two main components: authentication and authorisation. Authentication is a technique used to verify that someone is who he or she claims to be. It must be combined with an additional layer of authorisation, which determines whether a user should be allowed to access the data. Therefore, technologies for access control are certainly essential for accountability.

Legality  
Just as encryption, access control is fundamental when measures are taken to avoid data breaches.

EFFECTIVENESS
The responsibility for the effectiveness of access control mechanisms relies on each organisation.

KEY PROBLEMS IDENTIFIED
- Access control is considered inadequate for addressing privacy: those who obtain access to data, legitimately or not, can use the data without restriction;
- Access control gets more difficult when multiple devices are used;
- There are some key challenges to the enforcement of access control:
  - It gets difficult to enforce persistent policies in hybrid environments with a large range of devices;
  - Lack of an appropriate control model: organisations must determine the appropriate access control model to adopt based on the type and sensitivity of data they’re processing;
  - It may be needed to combine multiple technologies to achieve the desired level of access control;
  - Authorisation is still sometimes neglected by organisations;
  - In the past, access control methodologies were often static. It is important, however, to be able to change access control policies dynamically.

THE INTERVIEWEE’S OPINION

Associate professor focusing on the design, analysis and application of technologies to protect privacy at a European university

Some classes of technologies, such as access control, policy enforcement, accountability and transparency rely very much on a strong trust model and thus have a single point of failure. There is usually one entity that is designing the access control system or specific policies, and responding to accountability and transparency requirements. In such cases, non-technical things such as the legal system become relevant as complement or substitute for technologies.

MAIN SOCIETAL AND ETHICAL ISSUES ADDRESSED

Privacy
MPC is often considered as the counterpart of sending encrypted data to a trustworthy third-party who would return the intended result. To avoid having to trust third-parties, MPC can be a key technology providing access to data while ensuring strong privacy protection.

Trustworthiness
MPC technologies are also relevant with respect to trustworthiness, as the misuse of data through one organisation is largely excluded.

EFFECTIVENESS

- MPC can have a large impact on a number of areas: it could help unlocking the large quantity of health data that is currently inaccessible for study due to privacy concerns; it has already been used to evaluate gender pay disparities, detect tax fraud and prevent satellite collisions;
- Examples for practical implementations in which MPC has proved to be quite efficient are trading sugar beet by Danish farmers and smart metering deployment in the Netherlands;
- MPC provides strong security guarantees, as only the result of the computation is revealed.

KEY PROBLEMS IDENTIFIED

- Computer power and bandwidth use can still be potential obstacles;
- Expert involvement is required;
- Secure MPC is less secure than FHE, especially when many untrusted parties are involved;
- Practical implementation problems stand in the way of wide adoption.

THE INTERVIEWEE’S OPINION

While anonymisation allows using standard analytics tools, the value of the data may be decreased by the anonymisation process. MPC allows working with the data as it is but restricts the efficiency of the analysis and the range of tools that can be used.

MAIN SOCIETAL AND ETHICAL ISSUES ADDRESSED

Privacy  
Policy enforcement technologies are highly relevant for protecting privacy as they allow enforcing rules for the use and handling of resources including personal data.

Trustworthiness  
They also increase trustworthiness as rules are enforced automatically. Particularly when data is transferred between systems, the risk that rules get lost or neglected can be reduced.

Accountability  
Accountability benefits from the use of policy enforcement technologies as the rules are specified formally and violations are documented. This increases the transparency of data use and handling significantly.

EFFECTIVENESS

In order to enforce sensitive data security policies, an effective policy enforcement framework has to be flexible and support different data processing requirements.

KEY PROBLEMS IDENTIFIED

- In the big data era, policies get easily lost or neglected in the course of data being transferred between different systems;
- Big data security and privacy are challenges both for users and service providers¹;
- Policy enforcement gets increasingly difficult as the chain of responsibilities becomes longer and the roles more geographically dispersed.

THE INTERVIEWEE’S OPINION

Technology advisor for a national data protection authority in Europe

Alongside anonymisation technologies, technologies for policy enforcement play a key role in the context of big data. It is very hard for data protection authorities to exercise their enforcement power, particularly if actors outside the EU are involved. Technical instruments, which assure proper functioning of things and in such a way offer the possibility of self-enforcement, are thus considered to be highly relevant. Technologies for policy enforcement are not yet mature.

MAIN SOCIETAL AND ETHICAL ISSUES ADDRESSED

**Trustworthiness**  Both accountability and transparency are often considered key prerequisites for trust.

**Accountability**  Accountability requires the evaluation of compliance with policies and the provision of evidence. This is a key task supported by accountability technologies, which, for instance, provide automated control and auditing processes.

**Legality**  Legality, the state of being in accordance with the law, can be more easily evaluated if technical measures are in place that support accountability and transparency.

EFFECTIVENESS

The use of accountability and transparency technologies may not only increase an entity’s trustworthiness but also has the potential to lead to efficiency improvements as deficiencies in processes may become obvious.

KEY PROBLEMS IDENTIFIED

Transparency is very challenging to achieve: in the big data context, explaining what algorithms do with data or previewing what would be the outcome of providing data is extremely difficult.

THE INTERVIEWEE’S OPINION

*Research associate focusing on transparent computer systems at a European university*

Transparency is currently a big issue in the IoT context, and the concept should also include machine learning interpretability. Many machine learning modules that drive IoT devices are black boxes. Such devices may perform in an undesired way and it’s almost impossible to find out why. The performance is not interpretable by non-experts. A key problem is that most IoT devices do typically not have a screen. This makes it difficult to let them explain what they do in a visual manner. Moreover, records of the interactions between devices need to be stored somewhere, ideally at an independent and trusted party.
MAIN SOCIETAL AND ETHICAL ISSUES ADDRESSED

**Trustworthiness**
Being able to attest the origin and authenticity of the data used positively affects trust in the data-driven products and services of an entity.

**Accountability**
Accountability refers to the properties that ensure that the actions of a person or organisation can be traced uniquely to the person or organisation. To be able to do this consequently in a data-driven context, data provenance is essential.

EFFECTIVENESS
The information that data provenance provides is useful for debugging data and transformations, auditing, evaluating the quality of and trust in data, modelling authenticity, and implementing access control for derived data.

KEY PROBLEMS IDENTIFIED
The challenges that are introduced by the volume, variety and velocity of big data, also pose related challenges for provenance and quality of big data, defined as veracity\(^8\):

- The provenance data from big data workflows is too large;
- Provenance collection overhead during workflow execution is too much;
- It is hard to integrate distributed provenance;
- It is hard to reproduce an execution from provenance for big data applications.

THE INTERVIEWEE’S OPINION

Privacy and civil liberties engineer at a software and services company specialised on big data analysis headquartered in North America

The measures taken by the company with focus on data provenance allow users to comprehensively investigate certain values that seem wrong. They cannot only check if, how and when errors were introduced but also implement a fix and, based on the provenance tree, rebuild the dataset with the corrected values.

---

MAIN SOCIETAL AND ETHICAL ISSUES ADDRESSED

Privacy Empowering users, informing them and giving them access to their data is fundamental for privacy protection.

Legality Empowering users to access their data is an obligation for data controllers. The scope of users' access right is specified in Article 15 of the GDPR.

These technologies are also highly relevant for the issues interdependency, self-determination and trustworthiness. Apart from access, portability is a prerequisite for interdependency. Technologies for user control have the potential to increase self-determination and trust.

EFFECTIVENESS
Empowering users, informing them and giving them access to their data is not only for the users' benefit but also for the benefit of the organisation using the data. Nevertheless, users do not practice this opportunity very often. This may be because they are not informed about the opportunity, because the process is too complicated or because they do not care.

KEY PROBLEMS IDENTIFIED
- Portability leads to the duplication of data. Data may be brought from one domain where there are safeguards to another domain that is riskier.
- The informational asymmetry between data subjects and data collectors is a central problem. The data subject cannot understand what his or her data can be used for and how certain database transactions will end up being inefficient or unfair.
- Many people are reluctant to take time and learn about the complexities and the language used to work with even a simple interface which can enhance their control over data.
- Once data is released, asking for the consent of the data subject may be difficult.
- It is very hard to erase data and disregard the impact it had on computations or analyses.

THE INTERVIEWEE’S OPINION

Associate professor focusing on the design, analysis and application of technologies to protect privacy at a European university

It is certainly great to empower people but to some extent it also means that responsibility is pushed to them. I am in favour of designing systems in ways that make it difficult for users to endanger themselves rather than just giving them controls and expecting them to know how to use them. Even for experts it is sometimes difficult to understand what the outcome of certain decisions will be, and thus non-expert users should not be expected to understand them and make meaningful decisions.
TODAY’S SOLUTIONS
- There is wide agreement that technologies are integrated only to a limited extent in today’s big data solutions;
- There are strong solutions in research but there is a big gap when it comes to deployment;
- Companies increasingly try to brand themselves as privacy protectors;
- Traditional legal instruments may not be implemented in big data settings because they are in conflict with business models: privacy by design could be effective as it accompanies the processing with safeguards and does not prevent development.

CUSTOMERS AND USERS
- Users are blinded by the benefits of careless handling of their personal data;
- There is low demand from the customer side for technologies to protect privacy: a change in culture is considered necessary;
- Features protecting users must be embedded in the products rather than provided as add-ons;
- Policy makers and regulators could play an important role with respect to the demand.

REGIONAL DIFFERENCES
- Prime the utility of data;
- Rely on case-based legal decisions;
- Do not seem to trust the government much.
- Prime concerns about privacy;
- Live in a historically more rule-driven environment;
- See governments as important privacy protectors.
ORGANISATIONAL MEASURES

- There is broad consensus that the combination of technical and organisational measures is essential;
- Scandals show the limitations of reactive approaches that prescribe norms or actions;
- Technical solutions that are proactive and prevent breaches or rule violations are needed;
- Awareness and education are key aspects;
- Technologies must be accompanied by processes, laws, agreements and policies.

RESPONSIBILITY

- Consumers need to protect themselves because nobody else will do it for them;
- Organisations using data are responsible for data management and anonymisation;
- The strongest party should be responsible for privacy preservation;
- Data protection must not be considered as somebody else’s problem;
- Supervisory authorities and governments must shape the framework conditions.

Technology advisor for a national data protection authority in Europe

*Both are important for privacy preservation technology and commitment. The technologies are not the key challenge. In order to make them effective, it is not sufficient if just a single person in the organisation has the required expertise, the entire environment must be aware of the technologies and the related opportunities and threats.*

Associate professor at a European university

*The responsibility placed on the user should be as small as Possible.*

Professor at a North American university

*Tools for the individual data owners must be provided to control what happens with their data. The research community must develop these tools and they must be available to data providers.*
To know more about e-SIDES:

www.e-sides.eu

To contact us:

✉️ info@e-sides.eu

🐦 @eSIDES_eu