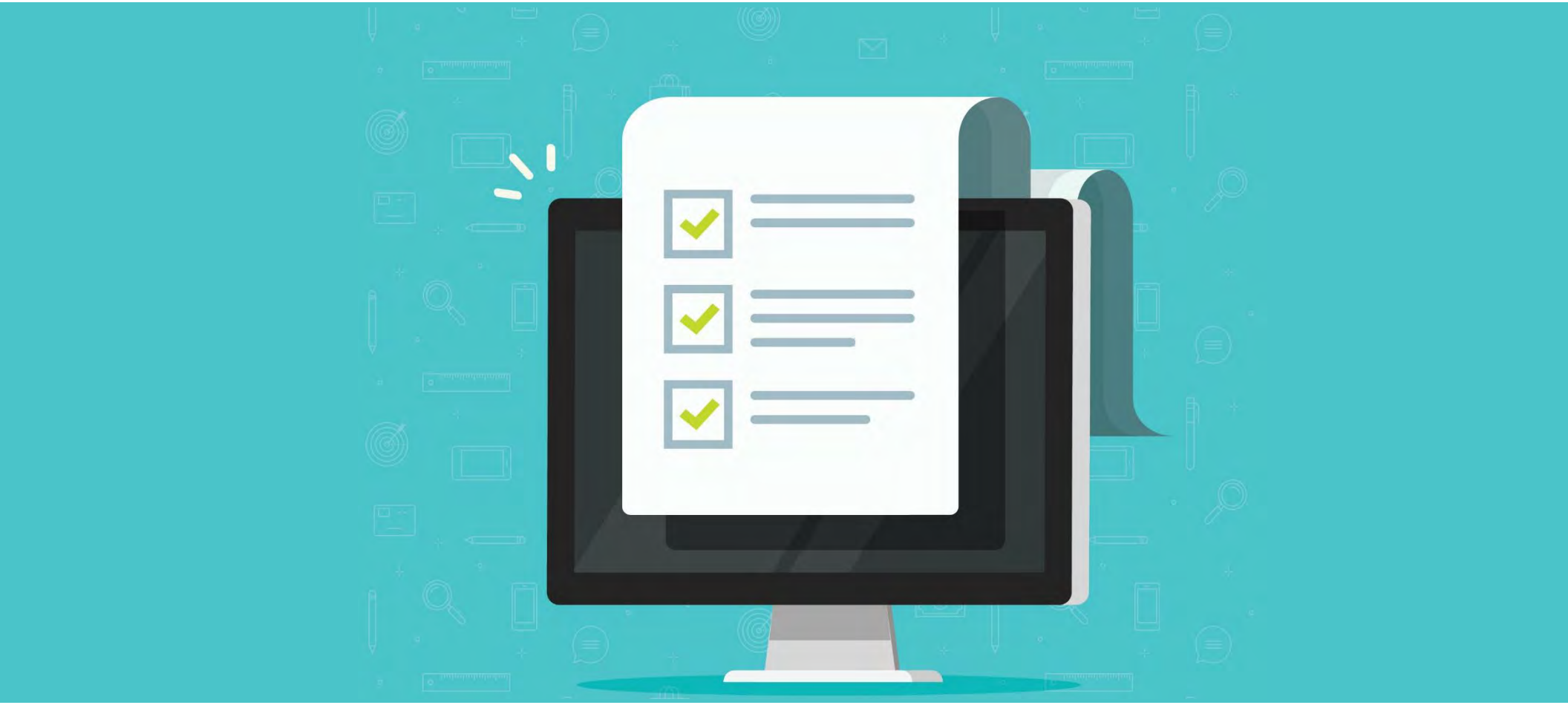
e-SIDES

Ethical and Societal Implications of Data Sciences

# Requirements for the design and use of privacy-preserving big data solutions



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731873

Ethical and Societal Implications of Data Sciences

## 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 organising 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 D4.2 Overview of design requirements

Find more at: https://e-sides.eu/assets/media/e-sides-d4.2-v1.1-1549617745.pdf

## About this white paper

This white paper is based on Deliverable 4.2 of the e-SIDES project, which proposes and discusses four general requirements for the design and use of big data solutions. The requirements are based on previous e-SIDES deliverables (especially, the gap analysis documented in D4.1 and the assessment of classes of privacy-preserving technologies presented in D3.2), a review of related previous work and an analysis of design challenges in the context of privacy-preserving technologies.




Grant Agreement number: 731873

## REQUIREMENT 1 – EMBED SECURITY AND PRIVACY FEATURES

The importance of privacy-preserving technologies grows continuously with information systems becoming increasingly networked within organisations and across organisational boundaries, and datasets becoming larger and more heterogeneous. Security and privacy features based on these technologies need to be embedded in big data solutions rather that provided as extras or optional add-ons. The features need to be activated and, if possible, configured so that they provide a high level of privacy protection by default.

**Key reasons:**

- A tight integration increases the probability that privacy-preserving technologies work effectively and efficiently
- Operators of big data solutions are rather unlikely to purchase and use privacy add-ons
- The pressure exerted by clients, users, data protection authorities and the legal system is still too low
- Some organisational cultures do not seem to be ready for a truly privacy-preserving business conduct

Privacy-preserving technologies need to be combined to be effective. There is no most important technology or most important class of technologies.

If organisations are reasonably transparent about their practices and know to use big data solutions that are inherently privacy preserving, trust in the data economy will most likely increase in the long run.

## REQUIREMENT 2 – TAKE PREVENTIVE MEASURES

Privacy breaches need to be prevented before they happen. Recent data breaches clearly show the limitations of reactive approaches that typically prescribe measures to be taken when privacy is violated or when certain rules are broken. Technologies are considered proactive in the sense that they prevent incidents or rule violations in the first place.

**Key reasons:**

- They may have serious negative implications on the affected individuals
- Significant fines are more and more often the consequence of data breaches
- Considerable negative effects on the reputation of organisations involved not unlikely
- They generally reduce trust in the data economy

Privacy by design, which is closely related to the concept of privacy-preserving technologies as it requires privacy safeguards to be integrated in technological solutions, plays a key role with respect to the prevention of breaches before they happen.

Just as making sure that security and privacy features are embedded in big data solutions, preventing breaches before they happen has the potential to significantly contribute to achieving a higher level of trust in the data economy.

## REQUIREMENT 3 – CONNECT PEOPLE, PROCESSES AND TECHNOLOGY

A combination of technical and non-technical measures is essential. To ensure privacy, knowledgeable people as well as proper processes are important complements to privacy-preserving technologies. For instance, a particular need for increased awareness, improved usability and education targeting was observed.

**Key reasons:**

- Currently, most big data solutions, no matter if they have particular security and privacy features, can only be used correctly by experts
- In general, technologies have limitations and alone are not sufficient

At some level, non-technical measures will always be necessary to make sure a given technology functions as expected. Moreover, data protection officials are needed that are aware and able to assess the impact on privacy or risks regarding personal data that is collected and used by organisations.

The combination of technical and non-technical measures is essential to detect, investigate and prevent data breaches and misuse. Awareness of the fact that people and processes are there to make sure that a given technology functions as expected will help increasing trust in the data economy.

## REQUIREMENT 4 – COMPLY WITH LAWS AND CORPORATE POLICIES

Big data solutions need to comply with laws and corporate policies. At the same time, they must be flexible enough to be adapted to changing demands, not only with respect to the business side but also with respect to the regulatory side. The heavy discussions that recently accompanied the entry into force of the GDPR show that achieving legal compliance is not always trivial, even if the legal text has been available for a long time.

**Key reasons:**

- For big data solutions, it is essential to be not only legally compliant but also in line with corporate policies
- Organisations must be able to demonstrate that they have full control over their data-related activities

Compliance is a prerequisite for trust in the data economy.

## RELATED PREVIOUS WORK

**OECD principles**

In its revised Recommendation Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data ("Privacy Guidelines"), the OECD refers to the following principles: Collection limitation, Data quality, Use limitation, Security safeguards, Openness, Individual participation, Accountability.

**EU GDPR principles**

The following principles are described in the GDPR: Lawfulness, fairness and transparency, Purpose limitation, Data minimisation, Accuracy, Storage limitation, Integrity and confidentiality.

**Protection goals**

Hansen et al. describe the following protection goals for privacy engineering in a 2015 IEEE Security and Privacy Workshop article: Confidentiality, Integrity, Availability, Unlinkability, Transparency and Intervenability.

**Privacy by design**

The recommendations of D'Acquisto et al. published in a 2015 ENISA report are related to: Privacy by design applied, Decentralised versus centralised data subjects, Support and automation of policy enforcement, Transparency and control, and User awareness and promotion of privacy-preserving technologies.

**IEO/IEC 29100:2011 principles**

ISO/IEC describes numerous privacy principles that form the basis of its privacy framework. Among them are: Consent and choice, Purpose legitimacy and specification, Collection limitation, Data minimisation, Use, retention and disclosure limitation, Accuracy and quality, and Openness, transparency and access.

**US FTC fair information practices**

The FTC identified the following core principles of privacy protection: Notice/awareness, Choice/consent, Access/participation, Integrity/security and Enforcement/redress.

**System design**

The system design mistakes described by Hansen in the proceedings of the 2011 IFIP International Summer School are particularly relevant and include: Storage as default, Linkability as default, Real name as default and Function creep as feature.

e-SIDES

**To know more about e-SIDES:**

www.e-sides.eu

**To contact us:**

info@e-sides.eu

@eSIDES_eu

e-SIDES

IDC ANALYZE THE FUTURE

eLAW LEIDEN

Fraunhofer

Grant Agreement number: 731873