



e-SIDES

Ethical and Societal Implications of Data Sciences

Which data technologies play a key role to preserve privacy and security?





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 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 D3.1 Overview of Existing Technologies

Find more at: <http://www.e-sides.eu/resources/deliverable-31-overview-of-existing-technologies>



Overview of Existing Technologies



Anonymisation

Encryption or removal of personally identifiable information



Sanitisation

Encryption or removal of sensitive information



Multi-party computation

Distribution of data and processing tasks over multiple parties



Accountability

Evaluation of compliance with policies and provision of evidence



Data provenance

Attesting of the origin and authenticity of the information



User control

Specification and enforcement of rules for data use and handling



Encryption

Encoding of information so that only authorised parties can access it



Access control

Selective restriction of access to places or resources



Policy enforcement

Enforcement of rules for the use and handling of resources



Transparency

Explication of information collection and processing



Access and portability

Facilitating the use and handling of data in different contexts



Anonymisation

Encryption or removal of personally identifiable information

MAIN TECHNIQUES

Randomisation: removing strong links between attributes and the identity of the data subject. This can be achieved through:

- **Adding of noise:** masking the true attribute values of a dataset.
- **Permutation:** shuffling the values of attributes in a table so that some of them are artificially linked to different data objects.
- **Differential privacy:** only the results of a query are anonymised and not the dataset itself. This is usually achieved by randomly distorting the data after each query by adding random noise to the true result of the query.

Generalisation: generalizing or diluting the attributes of the data subjects by changing their scale or magnitude. This is achieved through:

- **k-anonymity:** preventing two datasets from being joined through generalization of quasi-identifiers (attributes which, in combination, can be linked with external information to re-identify subjects) or suppression of sensitive attributes in order to avoid re-identification. Typical methods to achieve this are perturbation (addition of dummy datasets that contain no useful information) or the deletion of attributes or datasets that are not necessary for the specific analysis.
- **l-diversity:** an extension of k-anonymity with reduced granularity of a data representation, i.e. sensitive attributes must be “diverse” within each quasi-identifier equivalence class.
- **t-closeness:** a refinement of l-diversity. The idea behind t-closeness is that the distribution of sensitive attributes within each group of quasi-identifiers should be “close” to their distribution in the entire original database.

EXAMPLE

Anonymised table with suppressed unique identifiers (patient names) and generalised quasi-identifiers (date of birth, gender, Zip Code).

<i>Unique identifier</i>	<i>Quasi-identifiers</i>			<i>Sensitive attribute</i>
Name	DOB	Gender	Zip Code	Disease
*	'65	Male	7909*	Heart Disease
*	'65	Male	7909*	Hepatitis
*	'65	Male	7909*	Bronchitis
*	'81-'83	Male	7933*	Broken Arm
*	'81-'83	Male	7933*	Flu
*	'81-'83	Female	7933*	Diabetes
*	'81-'83	Female	7933*	Gastritis



Encryption

Encoding of information so that only authorised parties can access it

■ MAIN TECHNIQUES

Database encryption:

- **Attribute-based encryption (ABE):** the secret key used for the encryption and the ciphertext depend upon certain attributes (e.g., the individual's country, job or habit). The decryption can be performed only if the presented set of attributes matches the attributes of the ciphertext.
- **Identity-based encryption (IBE) / hierarchical (HIBE):** using any string (e.g. a name or an e-mail address) as the public key, the corresponding decryption keys are issued by a trusted party. HIBE is a variant of IBE allowing multiple private key generators arranged in a hierarchical form to easily handle the task of private key generation.
- **Proxy re-encryption (PRE):** allowing a semi-trusted proxy to convert the ciphertext encrypted under the public key of one user into a ciphertext that can be decrypted through another user's private key.
- **Functional encryption:** where a user with certain attributes has a key that enables him/her to access a particular function of the encrypted data.
- **Predicate encryption / hierarchical predicate encryption (HPE):** the secret keys correspond to the predicates and these secret keys are used to decrypt the ciphertext associated with the attributes corresponding to the predicate. HPE is a cryptographic primitive that facilitates the delegation of the search capabilities.
- **Storage-path encryption:** instead of encrypting the data, only the storage path is encrypted.

Encrypted search:

- **Property-preserving encryption (PPE):** encrypting data in a way that some property is preserved.
- **Structured encryption:** the basis of Boolean keyword search on encrypted data.

Encrypted computation:

- **Fully homomorphic encryption (FHE):** a particular type of encryption that permits computations on ciphertexts and also results are obtained in an encrypted form.
- **Oblivious RAM (ORAM):** algorithms allow a client to store large amounts of data (e.g., in the cloud) while hiding the identities of the entities being accessed.

■ EXAMPLE

One of the most important needs for electronic forensic analysis is an “audit log” containing a detailed account of all activity on the system or network to be protected. Audit log entries could be annotated with attributes such as the name of the user, the date and time of the user action, and the type of data modified or accessed by the user action. A forensic analyst charged with some investigation would be issued a secret key allowing for a particular kind of encrypted search; such a key would only open audit log records whose attributes satisfied a certain condition.



Sanitisation

Encryption or removal of sensitive information

MAIN TECHNIQUES

- Encryption
- Removal of columns
- Data masking
- Substitution
- Shuffling
- Number variance

In the big data era, for instance, it can be difficult to find substitution data in large quantities.

EXAMPLE

The tables below represent the original dataset and its sanitised version. The columns named “Email” and “Card” have been sanitised through the technique of data masking: most of the characters have been replaced by a Mask character (in this case, by X or *).

Order	First Name	Last Name	Address	State	Zip	Email	Card
1289	Reggy	Mackie	838 River Ave., Tiffin	WA	93706	mackie@dsa.com	4024-0071-8423-5933
1234	Justin	Doe	22 The Grove, Newport	WA	91768	doedoe@chann.com	5157-5834-6560-1768
1223	Kathy	Abrams	2 High Street, Wheeling	MO	32068	kittykatty@fugie.com	4024-0071-8423-6700
1233	John	Bradley	509 Acacia Avenue, Brompton	FL	18974	m2a2@mii.com	5131-1060-4724-1415

Order	First Name	Last Name	Address	State	Zip	Email	Card
1223	Kathy	Abrams	2 High Street, Wheeling	MO	32068	k*****@*****.***m	XXXX-XXXX-XXXX-6700
1233	John	Bradley	509 Acacia Avenue, Brompton	FL	18974	m***@***.***m	XXXX-XXXX-XXXX-1415
1234	Justin	Doe	22 The Grove, Newport	WA	91768	d*****@*****.***m	XXXX-XXXX-XXXX-1768
1289	Reggy	Mackie	838 River Ave., Tiffin	WA	93706	m*****@***.***m	XXXX-XXXX-XXXX-5933

Source: <https://www.datasunrise.com/data-masking-made-simple/>



Access control

Selective restriction of access to places or resources

▪ 2 STEPS

- **Identification:** usually based on identity information.
- **Authentication:** typically relying on passwords, access tokens, biometrics or combinations of the three methods. Methods involving two independent ways for verifying identity are referred to as two-factor authentication methods.

▪ MAIN APPROACHES

- **Role-based access control (RBAC):** determining access to resources based on the user's role within the organization.
- **User-based access control:** often implemented through access control lists (ACLs), tables that list each user and determine its access to a resource.
- **Attribute-based access control (ABAC):** determining access to resources based on policies taking attributes of the user, the resource and the environment state into account. ABAC is sometimes referred to as policy-based access control (PBAC).

▪ EXAMPLE



MongoDB (<https://www.mongodb.com/>), one of the most popular NoSQL data stores, employs Role-Based Access Control (RBAC) to govern access to a MongoDB system. A user is granted one or more roles that determine the user's access to database resources and operations. Outside of role assignments, the user has no access to the system. A role grants privileges to perform the specified actions on resource. Each privilege is either specified explicitly in the role or inherited from another role or both. A privilege consists of a specified resource and the actions permitted on the resource.

Source: <https://docs.mongodb.com/manual/core/authorization/>



Multi-party computation

Distribution of data and processing tasks over multiple parties

■ MAIN FEATURES

- Multi-party computation (MPC) relies on the distribution of data and processing tasks over multiple parties.
- MPC is a field of **cryptology** 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 still no practical solutions. Key challenges in the big data context are utility, performance and ease of use.
- MPC is relevant in the context of **Privacy-Preserving Data Mining (PPDM)**. The common aim of all PPDM approaches is to protect sensitive data used in data mining algorithms. In **traditional PPDM**, there is one data owner, and another party performs the data mining. In **distributed PPDM**, there are multiple data owners, whose combined data is mined in a collaborative fashion, either by the owners themselves or by one or more third parties. Privacy-preserving querying takes place if the data owner does the mining itself.
- Distributed PPDM addresses the problem to perform data mining on sensitive datasets from multiple parties when there is no single party that is trusted to hold all of the data. Based on who performs the mining, **three deployment models** are described:
 - A single miner collecting data from many clients, where each holds an individual record about itself.
 - A relatively small number of parties each holding a sub-database, where joint data mining is performed and each party contributes computing power.
 - A larger number of parties holding a sub-database but the joint data mining is outsourced to a smaller number of computing parties that do not necessarily provide input themselves.
- In theory, every algorithm can be performed with MPC to fully protect the privacy of sensitive inputs. However, because of the large performance penalty this induces, in practice, often a hybrid approach is used in which the data providers combine MPC with local computation. Such approaches range from ones where the local computation is complex and the MPC aggregation easy, to ones where the local computation is easy and for the bulk of the work MPC is used.

■ EXAMPLE

In their study, Kukkal et al. (2015) aimed to provide a graph construction protocol proven to be correct, private and robust, in accordance to the requirements of a standard MPC protocol. According to their protocol, each individual participating in the protocol, henceforth referred to as a party, reports her adjacency list (all her outgoing edges) as her private input. The algorithm guarantees that the parties do not learn any additional information apart from the data that can be gathered from just their input and output.



Policy enforcement

Enforcement of rules for the use and handling of resources

■ MAIN FEATURES

- **Automated policy enforcement mechanisms** are considered 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.
- **Access control** is the most common approach to ensure policy enforcement: by only permitting access to users who are policy compliant, policy can be enforced by default.
- One limitation of access control as a strategy for policy enforcement is its “once-and-for-all”-character and application specificity. This can become a problem in particular in the Internet of Things (IoT) context.

As Pasquier et al. note, decisions over whether users get access to some data are, by definition, taken before they gain access, but thereafter access-control systems foresee no further action (monitoring, enforcement, etc.). Moreover, traditionally, access-control policies tend to be application-specific. This is particularly problematic in large-scale distributed systems like IoT, where many heterogeneous users and applications interact. Under these circumstances, application-specific (rather than system-wide) access policies are liable to lead to policy conflicts. As an enhancement of access control for policy enforcement (and audit) in the IoT, they therefore propose information flow control, to enable continuous, data-centric, cross-application and end-to-end control of data flows. Concretely, two-component tags (a form of metadata) are attached to each data item, and each application. Data are then only allowed flow to apps with matching labels.

Thomas F.J.M. Pasquier et al., “Managing Big Data with Information Flow Control,” in Proceedings of the IEEE 8th International Conference on Cloud Computing, 524–31 (IEEE, 2015), <https://www.cl.cam.ac.uk/research/srg/opera/publications/papers/2015ieeeCloud.pdf> (accessed December 14, 2017)

■ EXAMPLE

One important challenge in access control for big data is implementing sufficiently fine-grained access controls, so as to enable differential access depending on the sensitivity of the data and the authorisation level of the users. According to Ulusoy et al. (2015) especially access-control approaches for MapReduce - one of the most widely used big data programming models for processing and storing large datasets - mostly have an all-or-nothing quality; either permitting access to the entire dataset, or no access at all. To overcome this problem, they propose a modular framework named GuardMR that enforces fine-grained security policies at the key-value level (instead of the higher-level file level) of MapReduce. Based on the organisational role (security clearance) of the users, GuardMR permits different levels of access to the data.

Huseyin Ulusoy et al., “GuardMR: Fine-grained Security Policy Enforcement for MapReduce Systems,” in Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security - ASIA CCS '15, ed. Feng Bao et al., 285–96 (New York, New York, USA: ACM Press, 2015)



Accountability

Evaluation of compliance with policies and provision of evidence

MAIN FEATURES

- Accountability is a key concept in data protection law whose importance will be further reinforced when the EU GDPR comes into force in May 2018. A cornerstone of accountable information systems is the provision of **automated and scalable control and auditing processes** that can evaluate the level of compliance with privacy policies against predefined machine-readable rules.

MAIN CHALLENGES TO AUDITING COMPLIANCE WITH PRIVACY POLICIES AND SUGGESTED SOLUTIONS

Challenges

IoT: Given the IoT's nature as a vast scale network of interconnected "things", traditional application (thing)-centric approaches to auditing are liable to constitute an obstacle to understanding system-wide behaviour.

Cloud computing: The spread of cloud computing has created further requirements for audit, namely verifying the integrity of data stored in the cloud. 2 closely linked challenges: 1) designing mechanisms to audit cloud-stored data efficiently; 2) designing audit processes so as to protect the privacy of the cloud users.

Solutions

An information-centric audit mechanism built around provenance data represented as a directed acyclic graph (DAG)

- 1) Retrieve the entire stored data from the cloud and then verify it.
- 2) Outsource the audit process to a third party and sought to develop ways of verifying the data's integrity without retrieving all, or even any of it.

MAIN APPROACHES

PDP and POR are the two main approaches to **cloud data-integrity verification without retrieval**. The basic idea of both is to split the data file stored in the cloud into blocks, and then attach some metadata to the data, in form of either a homomorphic linear authenticator or a homomorphic verifiable tag. This metadata is then used to verify the underlying data blocks. Of the two schemes, PDP is said to be the safer and more efficient.

EXAMPLE

Sen et al. (2017) have developed a workflow for privacy compliance in MapReduce-like big data systems. The workflow consists of two components; a formal language called LEGALESE, and GROK, a data-inventory mapper for MapReduce-like systems. The prototype system has been applied to the backend data-analytics pipeline of the search engine Bing.

Sen et al., "Bootstrapping Privacy Compliance in Big Data Systems," <https://www.andrew.cmu.edu/user/danupam/sen-guha-datta-oakland14.pdf> (accessed December 14, 2017)



Transparency

Explication of information collection and processing

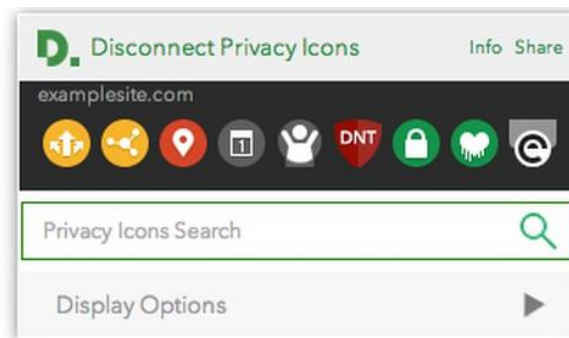
■ MAIN APPROACHES

Transparency in big data may be achieved by:

- **Purely textual information:** this approach does not seem to cope with the evolution of services and to comprehensively inform users on the processing of data occurring in the complex big data value chain.
- **Multichannel and layered approaches:** this approach has been suggested in order to improve the effectiveness of information.
- **Standardized icons and pictograms:** e.g., Disconnect Privacy Icons, Mozilla privacy icons. This is a promising emerging approach for transparency in big data, however, measures need to ensure that people are able to understand the graphic scheme as well as the pictographic parts of icons.

■ EXAMPLE

In June 2014, TRUSTe (a well-known privacy/trust seal) released together with Disconnect (a privacy-advocacy and open source software company) a set of Privacy Icons designed with the aim to help people quickly understand how websites handle their personal data. More specifically, Disconnect Privacy Icons is a browser extension providing users with a set of 9 icons informing them about the most important data practices of a website. These icons and badges are presented in search results (as an overlay to native search engine results) and while browsing (inside the extension). The icons indicate information related to use of data, data collection, user location tracking, data retention periods, children's privacy, Do Not Track compliance, SSL support, Heartbleed vulnerability, TrustE certified.



<https://disconnect.me/help#privacy-icons>

<https://www.pcworld.com/article/2366840/new-software-targets-hardtounderstand-privacy-policies.html>

Edwards, Abel, "The Use of Privacy Icons and Standard Contract Terms for Generating Consumer Trust and Confidence in Digital Services", CREATE Working Paper 2014/15 (31 October 2014), available on:

<http://www.create.ac.uk/publications/>



Data provenance

Attesting of the origin and authenticity of the information

■ MAIN FEATURES

- Provenance provides a record of the sources and transformations (**processing history**) of a piece of data.
- It aims to answer the questions: Where do data come from? Who manipulated the data? What transformations were applied?
- It is commonly represented in the form of a **DAG** showing the interactions between the data items, processes and agents in question.
- It can enable the re-use and reproduction of experiments, debugging, process optimisation and performance prediction.
- It facilitates judgements about the trustworthiness of data and thus helps improve decision-making.
- Scholars have also proposed using provenance for audit and accountability, and as a tool for access control.

■ MAIN CHALLENGES AND SUGGESTED SOLUTIONS

Challenges

Privacy and security:

there is an inherent trade-off between the utility provided by provenance information, and the level of privacy/security guarantees maintained. More utility tends to equal lower privacy/security guarantees, and vice-versa.

Scalability and the size of the provenance data collected:

obtaining a fine-grained provenance tracking of a big data workflow can easily lead to provenance data that is several times larger than the original dataset.

Solutions

Abstract the provenance graph

Access controls

Extending cryptographic techniques

Risk-management approaches

Stream processing

Storing and processing at rest

Reducing the amount of provenance collected

Graph compression

Development of new decentralised provenance systems

■ EXAMPLE

Korolev and Joshi (2014) have proposed a tool, named PROB, to track provenance in big data experiments, with the aim to achieve reproducibility of experiments that involve big data workflows and ultimately increase the collaboration between computational and biological scientific communities. PROB utilises Git-Annex to store only the hash values of datasets, not the datasets themselves, thereby enabling sharing of workflows without having to share (possibly restricted) datasets.

Vlad Korolev and Anupam Joshi, "PROB: A tool for tracking provenance and reproducibility of big data experiments," in *Proceedings of the 20th IEEE International Symposium on High Performance Computer Architecture* (Orlando, Florida, USA: IEEE, 2014), http://ebiquity.umbc.edu/file_directory/papers/693.pdf (accessed December 14, 2017)



Access and portability

Facilitating the use and handling of data in different contexts

■ MAIN FEATURES

- **Access:** providing access to users on their data is an important privacy condition as well as an obligation of data controllers.
- **Portability:** an additional and important tool for users, giving the possibility to change service providers without losing their data.

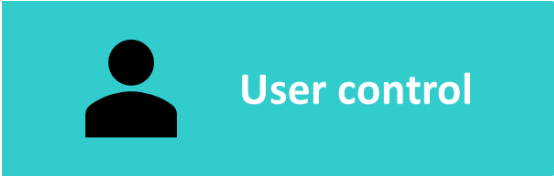
■ EXAMPLE

Among the relevant projects dealing with data and portability are the Midata UK initiative, providing access in transactions and consumption for the energy, finance, telecommunications and retail sectors, and the French MesInfos platform for access to financial, communication, health, insurance and energy data.

- The Midata initiative (announced for the first time in 2011) was launched by the UK government as a voluntary programme involving businesses, consumer bodies and regulators, with the aim to give consumers increasing access to their personal data in a portable, electronic format. The overarching goal of the initiative was to make individuals use this data to gain insights into their own behaviour, make more informed choices about products and services, and manage their lives more efficiently. Moreover, the initiative was aimed to encourage sustainable economic growth by boosting competition between companies and driving innovation; create opportunities for businesses through improved dialogue with consumers and increased trust, and the opportunity to provide innovative new personal information services and tools.



- The MesInfos project was launched in 2012 with the aim to explore and implement the “self-data” concept in France. Self-data is defined by the project as “the collection, use and sharing of personal data by and for individuals, under their complete control and designed to fulfil their own needs and aspirations”. The project aims to engage organisations in giving the customer data they have back to their customers, so that to restore trust and the balance of powers. This constitutes a “paradigm shift in a digital economy which is mostly based on business models such as tracking people and using (or even selling) their data for the sole gain of the company”.



Specification and enforcement of rules for data use and handling

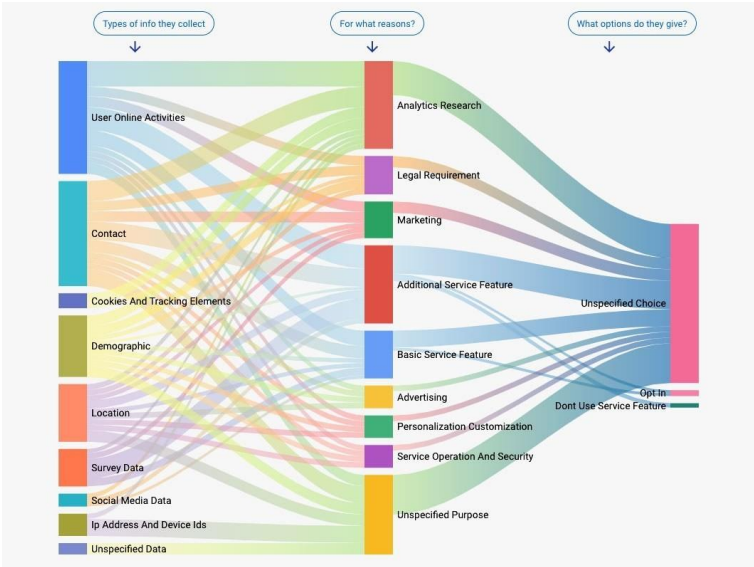
MAIN APPROACHES






































- **Consent mechanisms:** new usable and practical mechanisms to collect consent are needed that do not constitute a barrier for the usability of services. Engineered banner solutions, for instance, provide a basis for user-friendly consent mechanisms. They are increasingly used to give users of websites and apps information about the use of cookies and other forms of local storage. Software agents providing consent on behalf of the user based on the properties of certain applications could be a topic to explore in the future. Moreover, positive user actions such as specific gestures or motions could be used to constitute consent.
- **Privacy preferences and sticky policies:** sticky policies can provide a mechanism for attaching privacy preferences to specific datasets and accordingly drive data processing decisions. Privacy policies and requirements cannot only be expressed by data subjects but also by other parties.
- **Personal data stores:** technical solutions which give data subjects’ increased control over their data through transition from distributed data models to user-centric models. Such solutions (alongside with personal data vaults, personal data lockers and personal clouds) enable individuals to gather, store, update, correct, analyse and/or share personal data.










EXAMPLE

According to data provided by the World Economic Forum, despite internet users claiming to have read a privacy policy, only about 20% actually do. This is primarily linked to the fact that policies are usually long and difficult to comprehend: to address this issue, researchers from the Federal Institute of Technology at Lausanne, Switzerland (EPFL), the University of Wisconsin, and the University of Michigan developed Polisis (short for “privacy policy analysis”), a machine-learning AI able to read the privacy policy of any website inside a minute. Furthermore, Polisis can provide users with a readable summary of the policy, a flow chart of what exactly is going to be shared and where that information will be used, as well as information regarding opting out of information sharing. The developers also created Pribot, a chatbot able to provide answers on any privacy policy for free.

https://pribot.org/files/Polisis_Technical_Report.pdf
<https://www.analyticsvidhya.com/blog/2018/02/polisis-uses-ai-to-protect-your-online-privacy/>
https://www.weforum.org/agenda/2018/02/we-know-you-arent-reading-your-privacy-policy-but-this-ai-will-do-it-for-you?utm_content=buffer146ba&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer




Project	Objectives	Key technology classes			
 <p>SPECIAL – ICT-18-2016 https://www.specialprivacy.eu/</p>	Address the contradiction between big data innovation and data protection by proposing a technical solution that makes both of these goals more realistic.	Encryption		Accountability	
		MPC		Data provenance	
		Access control		Transparency	
		Policy enforcement		User control	
 <p>SODA – ICT-18-2016 https://www.soda-project.eu/</p>	Enable practical privacy-preserving analytics of information from multiple data assets using MPC techniques. The second aim is to combine MPC with a multidisciplinary approach towards privacy.	MPC		User control	
		Anonymisation		Access control	
		Sanitisation		Policy enforcement	
		Encryption		Transparency	
 <p>MHMD – ICT-18-2016 http://www.myhealthmydata.eu/</p>	Build and test new models of privacy and data protection that meet the requirements of the biomedical sector, in which issues of data subjects' privacy and data security represent a crucial challenge.	Anonymisation		Accountability	
		Encryption		User control	
 <p>SLIPO – ICT-14-2016 http://www.slipo.eu/</p>	Deliver technologies to address the data integration challenges of POI data in terms of coverage, timeliness, accuracy and richness.	Data provenance		Access and portability	
 <p>DAIAD – ICT-15-2016 http://daiad.eu/</p>	Develop innovative low cost, inclusive technologies for real-time, high-granularity water monitoring and knowledge extraction.	Anonymisation		User control	
 <p>AEGIS – ICT-14-2016 https://www.aegis-bigdata.eu/</p>	Create a curated, semantically enhanced, interlinked and multilingual repository for safety-related data that allowing organisations in the public safety and personal security sector to provide better services to their customers.	Anonymisation		Data provenance	
		Sanitisation		Access and portability	
		Access control			
 <p>TT – ICT-15-2016 https://transformingtransport.eu/</p>	Demonstrate the transformations that big data will bring to the mobility and logistics market in terms of, for instance, increased operational efficiency, improved customer experience and new business models.	Anonymisation			

Project	Objectives
 <p>EW-Shopp – ICT-14-2016 http://www.ew-shopp.eu/</p>	Deploy and host a platform to ease data integration tasks in the e-commerce, retail and marketing domain. The project integrates contextual variables (e.g., weather conditions, calendar events, holidays) into the analysis of consumer behaviour.
 <p>Data Pitch – ICT-14-2016 https://datapitch.eu/</p>	An open innovation programme that aims at bringing together corporate and public-sector organisations that have data with start-ups and SMEs that work with data.
 <p>QROWD – ICT-14-2016 http://qrowd-project.eu/</p>	Offer innovative solutions to improve mobility, reduce traffic congestion and make navigation safer and more efficient. To achieve that, QROWD integrates geographic, transport, meteorological, cross-domain and news data with the goal of maximizing the value of big data in planning and managing urban traffic and mobility.
 <p>EuBusinessGraph – ICT-14-2016 http://eubusinessgraph.eu/</p>	Create a business graph, which is understood as a highly interconnected graph of Europe-wide company-related information both from authoritative and non-authoritative sources (including data from both the public and the private sector). The project strives to provide a data marketplace that enables the creation of a set of data-driven products and services via a set of six business cases.
 <p>FashionBrain – ICT-14-2016 https://fashionbrain-project.eu/</p>	Combine data from different sources to support different fashion industry players (i.e., the retailers and the customers) by predicting upcoming fashion trends from social media as well as by providing personalised recommendations and advanced fashion item search to customers.
 <p>BigDataOcean – ICT-14-2016 http://www.bigdataocean.eu/</p>	Enable maritime big data scenarios through a multi-segment platform that combines data of different velocity, variety and volume under an interlinked, trusted and multilingual engine.
 <p>DataBio – ICT-15-2016 https://www.databio.eu/</p>	Demonstrate the benefits of big data technologies in terms of sustainable improvement and productivity of bioeconomy industry raw materials. For this purpose, a data platform is built based on existing software components that is suitable for different industries and user profiles to ensure effective utilisation of existing datasets, effective participation of the ICT industry and easy setup of new multivendor applications. The project deploys pilots in the fields of agriculture, forestry and fishery.
 <p>BDVe - Big Data Value ecosystem ICT-17-2016 http://www.big-data-value.eu/</p>	Support the Big Data Value Public-Private Partnership (PPP) in realising a vibrant data-driven EU economy. It strives to be accurately informed about the most important facts in big data as a solid basis to support the decision-making process in the PPP and develop a lively community. Furthermore, it supports the implementation of the PPP from an operational point of view, as well as the development of a European network of infrastructures and centres of excellence around big data.
 <p>K-PLEX - Knowledge Complexity ICT-35-2016 https://kplex-project.com/</p>	Investigate the strategies humanities and cultural researchers have developed to deal with their typically unstructured data. It aims to use humanities knowledge to explore bias in big data approaches to knowledge creation.

To know more about e-SIDES:

www.e-sides.eu

To contact us:

 info@e-sides.eu

 [@eSIDES_eu](https://twitter.com/eSIDES_eu)



e-SIDES