Clarus

A framework for user centred privacy and security in the cloud

CLARUS White Paper

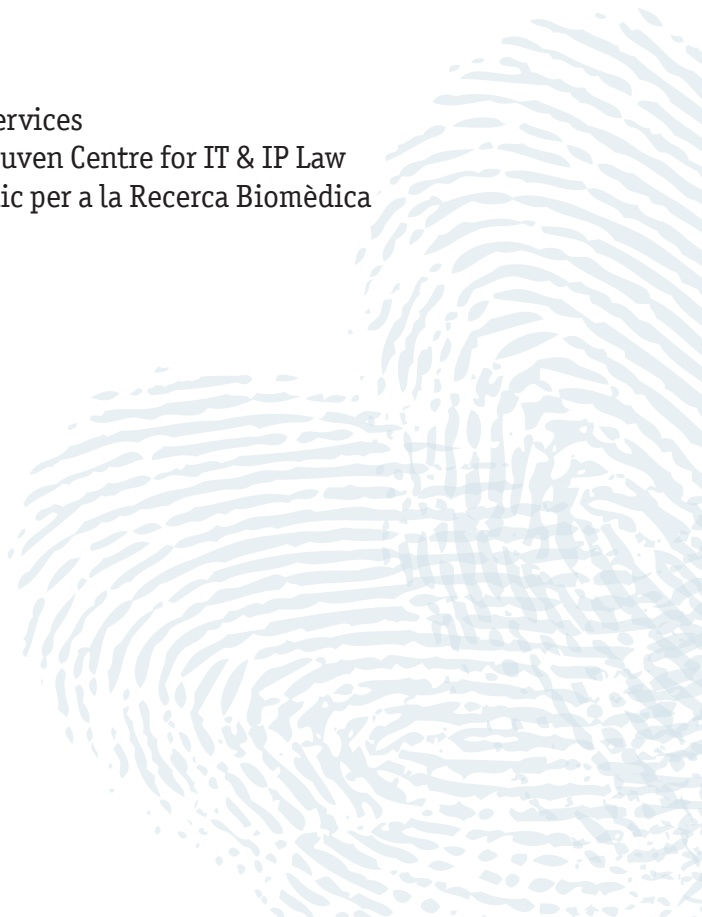
Privacy-preserving techniques for no-compromise security in the cloud

www.clarussecure.eu

@CLARUSecure

Authors:

Stephanie Parker, Trust-IT Services
Plixavra Vogiatzoglou, KU Leuven Centre for IT & IP Law
Rafael Mulero, Fundació Clínic per a la Recerca Biomèdica
Thierry Chevallier, AKKA



Innovative research in security-enabling techniques, attack-tolerant systems and in new architectures for secure delivery in the cloud.

The objective of CLARUS, www.clarussecure.eu, is to enhance trust in cloud computing services by developing a secure framework for storing and processing data outsourced to the cloud. This model change will give control back to data owners and increase transparency about data management, privacy and security. It thus improves levels of acceptance of cloud technology and creates new business opportunities.

The CLARUS service proposition is a proxy in the trusted domain of the end-user to provide a transparent solution to preserve the confidentiality of personal data and guarantee data protection before data is outsourced to the cloud for storage and processing. The proxy relies on the assumption that the Cloud Service Provider is "honest but curious", performing honestly the operations on the data as requested by the user, but it might also attempt to learn from the data. To address the need for privacy while leveraging the computational and storage capabilities of public Cloud Service Providers (CSPs), CLARUS proposes a set of privacy-preserving techniques.

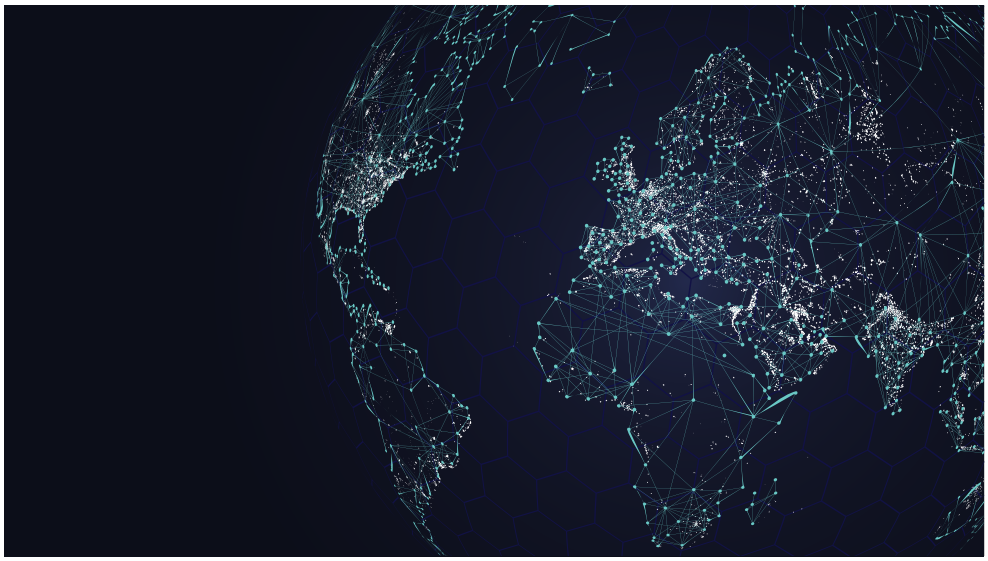
The concept of security as a service is implemented by the CLARUS proxy, which holds the keys and manages the knowledge to restore outsourced and secured data. The security-enabling techniques are a set of cryptographic primitives useful in the cloud context: searchable encryption, access control, homomorphic encryption, and secure multiparty computation. In the context of privacy-preserving techniques, a set of non-cryptographic techniques for the cloud has been defined: statistical disclosure control, data coarsening, data splitting.

Cutting-edge privacy-preserving mechanisms: searchable encryption, k-anonimity, data splitting

Significantly outperforming standard cryptographic techniques in terms of efficiency, flexibility of operations and data access

Essential cloud characteristics: On-demand self-service | Broad network access | Resource pooling | rapid elasticity | Measured Service

Case Study 1 – Geospatial Data Demonstration



The CLARUS solution is demonstrated on sets of geospatial data, which refer to environmental and geographical information. Datasets in the environmental domain possess interesting characteristics like the enormous size of the available data, the different degrees of access rights and the availability of metadata, which must be considered while applying the CLARUS solution. Environmental information is also highly relevant for the Free Flow of Data within the Digital Single Market, as highlighted in the EC Staff Working Document on Building a European Data Economy through free flow of data and cloud computing services[1]..

In general, the nature of the data included in geospatial information varies, in that data may be non-personal or personal, confidential or public, thus requiring diverse levels of access. While the term "personal data" refers to information relating to an identified or an identifiable person, non-personal data is the exact opposite, in the sense that no person may be identified in relation to the data. An example of personal data in the context of geospatial data may refer to people who have accessed, read or downloaded geospatial data or who have used a particular service in relation to environmental information. This way, actors operating in the geospatial scenario might own and manage information, which either may be available in the public domain or may be confidential and therefore must be protected. Security tools are needed for protecting data used in commercial settings by private companies, analysts or other institutions in the public sector.

The CLARUS geospatial demonstration

1 https://ec.europa.eu/digital-single-market/en/news/staff-working-document-free-flow-data-and-emerging-issues-european-data-economy.

case focuses on specific scenarios in the field of environmental data management, where cloud technologies are used and where CLARUS could bring solutions to important security expectations, such as storage of geo-referenced data; geo-publication of groundwater borehole data; geo-processing of mineral concentration data and geo-collaboration on gas supply network data.

» A key advantage of this approach is being able to extend these scenarios to more generic use cases in the field of Geographic Information Systems (GIS), namely storing geospatial data; searching and retrieving geospatial data and performing computations on geospatial data and updated geospatial data.

» Another important aspect of the geospatial use case is the requirement of making services interoperable, and thus for making them compliant with standards. Standards in the geospatial domain are defined by the Open Geospatial Consortium (OGC). Among these standards, the OGC web services standards are of utmost importance for implementing the scenarios described above, namely the Web Map Service (WMS) for serving maps on the web from several geo-referenced data sources, the Web Feature Service (WFS) for exchanging geographical features across the web, and the Web Processing Service (WPS) for invoking transformation services on the Internet.

» While cloud architectures provide actors in the geospatial domain with a high-quality, robust and cost-effective service, some geospatial data is confidential and usage in the cloud raises security issues. Thus, some European public institutions and Data Providers are still reluctant to "move to the cloud", due to the perceived threats on data security, user control of data, and data location.

Securing the publication and the processing of their data is a key challenge for geospatial data providers, who often want to limit access to some of their spatial datasets and data services, due to public security concerns or to commercial concerns. This is notably the case for European geo-survey organisations whose mission includes the management of confidential environmental data, besides the legal obligations to share public data to a large audience.

The geospatial use case is of great interest for potential CLARUS adopters, as it shows how the solution can adapt to a highly-standardised landscape (cf. OGC standards), via its plug-in mechanism for protocol support.

In addition, the geospatial use case applies to data held by public authorities and thus it shows how CLARUS end-users can monitor, audit and retain control of their data while benefitting from the functionality and cost-saving benefits of cloud services.

One of the key features of CLARUS is to support multi-usage scenarios for outsourcing data to the cloud by applying different security techniques. The geospatial use case demonstrates this feature through a variety of scenarios, showing the broad range of technical solutions available, for example:

» Hiding precise location of objects to non-authorised parties thanks to anonymisation/coarsening techniques.

» Protecting geographical features thanks to distributed data splitting among different CSPs.

» Protecting the result of a geo-statistical computation thanks to encryption.

The CLARUS security framework for outsourcing data to the cloud is in line with the security expectations of actors in the geospatial domain. Adding CLARUS to a spatial data cloud infrastructure will mitigate the security threats and strengthen the trust from cloud users, i.e. data providers and data consumers. CLARUS helps geospatial data providers gain confidence in the cloud, providing them with control of their data in the context of honest but curious cloud service providers (CSP).

Among the numerous use cases for datasets and services in the geospatial domain, geo-publication and geo-processing in the cloud are probably the most common scenarios where CLARUS will provide a solution to important confidentiality requirements. The CLARUS solution will address the concern of security in geospatial data sharing, particularly in the event of a regional or national disaster, one of the major reasons cited by organisations for failing to share data e.g. in the case of emergency response.

In addition, as location data may provide for the identification of individuals, including their habits and routines, CLARUS could in the near-future be an answer to the problem of privacy in the use of location based services (i.e. location privacy issues). Other possible applications of CLARUS in the geospatial domain could be satellite imagery (protecting sensitive data in very high-resolution products) and health geo-statistics (privacy-preserving health statistics related to environmental factors).

Case Study 2 - The eHealth Demonstration Case

eHealth is a key vertical for the European Digital Single Market. However, concerns about data privacy and security abound, also in the light of high numbers of data breaches at healthcare facilities in both the U.S. and Europe. It is also important to note that healthcare is a highly regulated vertical, making compliance a key driver for securing sensitive data.

In CLARUS, the eHealth use case concerns a distributed e-health scenario that requires immediate access to medical data outsourced to cloud providers. The main actor in this use case is the hospital responsible for treating the Electronic Medical Records (EMRs) of the patients, which contain information that is highly identifying or confidential. A series of functionalities are needed, like creating, managing and updating medical histories, including results of clinical visits, searching for specific patients/histories, as well as shared and cooperative access to the data based on the defined access policies.

In this scenario, the CLARUS solution will use searchable encryption methods used to ensure robust protection and data retrieval capabilities on outsourced health records, but also anonymisation techniques will be employed to securely outsource medical datasets that are still useful for research (e.g., data analysts outside of CLARUS). The

use case focuses on passive Electronic Health Records (EHR), which refer to information from patients that have no contact with the hospital, for whatever reason, for a period of 5 years or more.

This information, which is stored on premise at the hospital, represents a large amount of data that the hospital needs but does not frequently use, making it a storage consumption problem. In addition, this information can be processed for research purposes, requiring several computation resources that can impact negatively on the performance of the hospital information system.

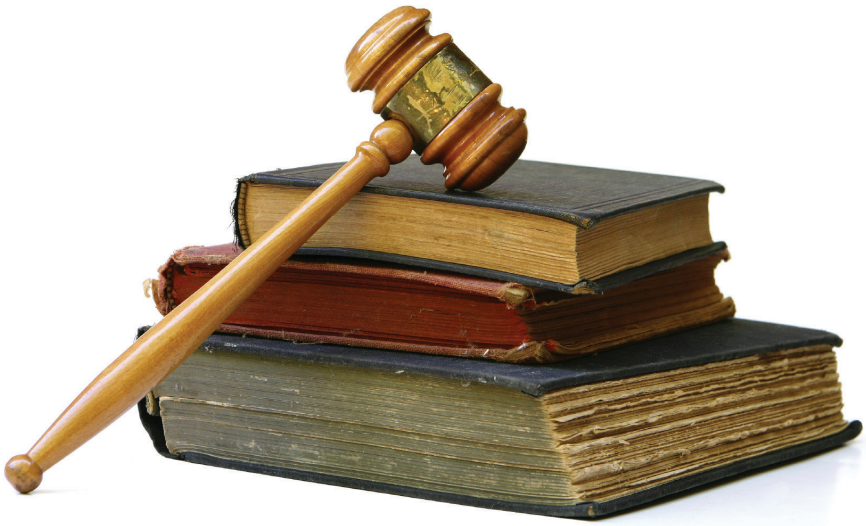
Outsourcing this information to the cloud is an opportunity to solve these problems. The space and computation resources used for this information will be available to handle active EHR data (information about patients who are in contact with the hospital for healthcare purposes).

According to several data privacy laws, clinical data is characterised as sensitive, confidential and private. While there are many cloud solutions available on the market, they are perceived as "honest but curious", and therefore not considered an option for handling this kind of data. The key challenge for the e-Health case is making it possible to store, retrieve and compute sensitive information in a secure way, applying different security mechanisms to avoid that data privacy can be compromised if the cloud is accessed by unwanted users.

The case can be used as an example of how to outsource highly sensitive and confidential data to cloud applying security techniques like searchable encryption and anonymisation (k-anonymity and t-closeness). Searchable encryption allows encrypting data before moving to cloud and then performing the search process directly over the encrypted data without prior decryption, which guarantees that data is shown in clear only to allowed users. Anonymisation techniques (k-anonymity and t-closeness) allows to mask data in a way that data can be processed and computed while privacy is not compromised if the cloud is accessed by unwanted users.

Adopting the CLARUS solution could be an opportunity for the e-Health sector to start using cloud platforms, improving, among others, data sharing between different healthcare entities and the quality of research studies related to different healthcare areas.

Legal analysis of the geo-publication use case



In applying the CLARUS solution to geospatial data there are various legal aspects that need taking into account. As mentioned above, geospatial information is mainly non-personal data but may include personal data, as for instance the personal log-in details to a geo-data related service. Furthermore, different sets of non-personal data fall under different legal obligations of publication or protection. This way, data held by the public sector that are critical for public safety or security or data with a strong business potential and personal data, will need to remain confidential while other environmental information may need to be made available according to national or EU laws.

Access to information held by the public is dictated by national freedom of information (FOI) laws, as at an EU level this is dominated by the principle of subsidiarity. According to this principle, there are areas which do not fall within the EU's exclusive competence but rather remain within the competences of the Member State due to their national character, as it is agreed in the Treaties signed for the birth and function of the European Union[2]. In the said areas,

2 Treaty on European Union, article 5(3), http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012M%2FTXT , Treaty on the Functioning of the European Union, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012E%2FTXT

the Union acts only if and in so far as the objectives of the proposed action cannot be sufficiently achieved by the Member States, but can rather be better achieved at Union level. Even though, on a national level FOI laws may stipulate different conditions for providing access to information, there are three European Directives regarding access related to environmental and spatial data which have significance in relation to the geospatial data being used during the CLARUS project. The ACCESS Directive regulates public access to environmental information[3], the INSPIRE Directive establishes a legal basis for the creation of the Infrastructure for Spatial Information in the European Community[4] and the PSI Re-use Directive refers to the re-use of public sector information[5].

According to the ACCESS Directive, public authorities are required to make environmental information available to the public either through express request or proactively of their own initiative. As such, it ensures that citizens are able to access environmental data in order to participate and assess the governmental decision-making process. This Directive defines environmental information broadly, as information on the state of the elements of the environment, on factors such as energy, on measures such as policies affecting or likely to affect the above, on reports on the implementation of environmental legislation, on economic analyses within this context and on the state of human safety and health. The framework includes the way relevant information should be disseminated, for example through policies, plans and programmes relating to the environment, data or summaries of data derived from the monitoring of activities affecting, or likely to affect, the environment or environmental impact studies and risk assessments concerning the environmental elements. In addition, it provides for grounds not to make this information available, in situations where there is a legal obligation to maintain the confidentiality of the data, as, for instance, under the data protection regime. More specifically, these content related exceptions can only be invoked if the disclosure of the information would "adversely affect" the interests that are protected and they must be interpreted in a restrictive way in a balancing of the respective interests, *in casu* the right to the protection of personal data.

The INSPIRE Directive focuses on the exchange of spatial data between public authorities regarding the performance of public tasks related to the environment and the facilitation of public access to this information to the point necessary. 'Spatial data', as defined in this Directive, is a narrower term relating to data with a direct or indirect reference to a specific location or geographical area, while 'spatial data set' means an identifiable collection of spatial data. As such, there is a small overlap with the above-mentioned ACCESS Directive.

3 Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC, L 41/26.

4 Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), L 108/1.

5 Directive 2003/98/EC of the European Parliament and of the Council on the re-use of public sector information, L 345/90, 17 November 2003 as amended by Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 on the re-use of public sector information, L 175/1.

The latter prevails over the INSPIRE Directive in case of conflict though. However, the INSPIRE Directive goes further in creating detailed rules on the availability of high quality metadata for all data sets and services. In fact, 'metadata' within the framework of this Directive, refers to information on the conformity of spatial data sets with the implementing rules, to the conditions applying access to and use of spatial data sets and services, to the quality and validity of spatial data sets, to the public authorities responsible for the establishment, management, maintenance and distribution of spatial data sets and services and to limitations on public access. Limitations are defined depending on the service information is used for. In this way, public access to data sets provided for discovery may be limited only for severe reasons while public access to data sets provided for other services can be limited for additional reasons that are the same as the ones provided for by the ACCESS Directive.

Finally, the PSI Re-use Directive provides the minimum rules for public authorities to make their data available for non-commercial reuse of existing and public-sector information that is generally available. The rationale behind this Directive is that the public sector collects, produces, reproduces and disseminates a wide range of information in many areas of activity, such as social, economic, geographical, weather, tourist, business, patent and educational information. Making public all generally available documents held by the public sector — concerning not only the political process but also the legal and administrative process — is a fundamental instrument for extending the right to knowledge. However, safeguards must be implemented to protect confidential information, as it is the case with the aforementioned legal instruments. Under this legal framework, the dissemination of these sets of data must not interfere with national security and third parties' intellectual property and data protection rights.

These directives aim at promoting the accessibility of publicly held information to the public and thus stimulating the EU information services market, taking into account the data protection safeguards when this information includes personal data. To that end, the European Commission adopted the European 'Free Flow of Data' initiative regarding non-personal data, as one of the actions within the Digital Single Market strategy[6]. Non-personal data is data that do not relate to an identified or identifiable natural person, such as anonymized data. At the moment, there is no comprehensive legal framework regulating non-personal data amongst Member States, while on the contrary there is a plethora of national laws imposing technical and legal barriers to their free movement across the EU. In particular, the main problem identified is data localisation restrictions, i.e. rules or practices that specify a particular, often geographically defined, area where specific data needs to be collected, processed or stored, while issues like data ownership, data portability and access to and transfer of data are similarly troubling.

As it is pointed out in the EC Communication and Staff Working Document on Building a European

6 Free Flow of Data Inception Impact Assessment (IIA), November 2016, available at http://ec.europa.eu/smart-regulation/roadmaps/docs/2016_cnect_001_free_flow_data_en.pdf

data economy, data localisation restrictions facilitate scrutiny and access by competent authorities as well as security of the data but they also become financially and practically cumbersome for businesses[7]. In the context of cloud computing, data localisation restrictions hamper the very nature of cloud computing, while ensuring data portability guarantees an enhanced use of cloud computing services. At the same time, as vast amounts of data are generated by machines or processes based on emerging technologies, such as the Internet of Things, access to those data and possibility of transferring them should be provided for in order to extract maximum value out of them. Limitations to protect confidentiality, personal data, intellectual property and so on should also be imposed as a counterbalance however.

In order to tackle these issues, the European Commission is taking actions towards the abolishment of unnecessary national data localisation restrictions and is engaging in dialogues with the stakeholders to explore manifold solution. This initiative is also complemented by the European Cloud Initiative in enhancing the digital economy and the free movement of data[8]. The CLARUS solution is set to benefit from these initiatives as the barriers on cloud computing will be mitigated, as well as promote them, as its technology can contribute to the different degrees of access to data, the secure transfer of data and data portability.

Legal analysis of the eHealth use case.

The eHealth use cas includes medical data and in this sense, personal data and more specifically special categories of personal data, also known as sensitive data. This term refers to data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation. The data protection regime, as it was regulated by Directive 95/46/EC, soon to be replaced by the General Data Protection Regulation 20016/679/EU, has introduced a wide range of rules that will be applicable in this use case[9].

7 EC Communication, "Building a European Data Economy", COM(2017) 9, 10.01.2017, available at https://ec.europa.eu/digital-single-market/en/news/communication-building-european-data-economy and EC Staff Working Document on the free flow of data and emerging issues of the European data economy Accompanying the document Communication Building a European data economy, 10.1.2017 SWD(2017) 2 final, available at https://ec.europa.eu/digital-single-market/en/news/staff-working-document-free-flow-data-and-emerging-issues-european-data-economy

8 More information on the site of the European Commission available at https://ec.europa.eu/digital-single-market/en/cloud

9 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (Data Protection Directive), O.J. L 281, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML , Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing

It is important to emphasise, however, that although the aim of the GDPR is to harmonise the legal framework, the laws of the Member States are allowed to diverge from the Regulation, when explicitly foreseen. For example, regarding the processing of sensitive data the Regulation provides a margin of manoeuvre for Member States' to restrict or specify its rules. Thus Member States are allowed to specify or introduce further conditions for the processing depending, inter alia, on the nature of the data concerned.

The different messaging and format standards used by different medical institutions makes it difficult to exchange information in a common way between hospitals, the GDPR also addresses this issue in Article 20, which establishes the new right to data portability, under certain conditions. In particular, where controllers process personal data through automated means, data subjects have the right to receive the personal data concerning them from the controllers in a structured, commonly used, machine-readable and interoperable format, whenever data subjects provide the personal data and the processing of this personal data is based on their consent, the processing is carried out by automatic means or the processing is necessary for the performance of a contract.

As far as the processing of sensitive data for research purposes is concerned, the GDPR aims to promote innovation and encourage research. Thus, it defines the term "research" broadly (recital 159) by stipulating that research "include(s) for example technological development and demonstration, fundamental research, applied research and privately funded research(..)".

More specifically, regarding the primary use of research data relating to health, meaning when personal data is originally collected for research purposes, the legal grounds for processing the data will be with the consent of the patient. Nevertheless, consent is not always a prerequisite for processing health data for research purposes. For instance, the Belgian Data Protection Act determines that health data may also be processed if necessary for substantial reasons of public interest or when necessary for population screening.

Regarding the secondary use of research data, meaning the further processing of data for historical, statistical or scientific purposes, the GDPR addresses the issue of compatible use more extensively compared to the current legal framework set by the Directive 95/46. It explicitly mentions that "further processing for scientific, historical and research purposes shall not be considered incompatible with the initial purposes and foresees specific conditions regarding compatibility." Therefore, at the European level, the mechanism for further processing of data for research purposes can be summarised as follows. When the purposes of the research can be fulfilled by further processing data which do not permit or do not any longer permit the identification of data subjects, the research should be fulfilled in this manner: Pseudonymisation can be included as a technical measure, as long as it allows the purpose of the

Directive 95/46/EC (General Data Protection Regulation), http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1495790670928&uri=CELEX:32016R0679

research to be met. But if the latter is not met, then other appropriate safeguards (incumbent to the Member States to define) should be put in place to protect the rights and freedoms of the data subjects.

As mentioned above, under the national regime for research, a Member State law may also foresee derogations to the right of the data subjects to access data processed on them, to request rectification, to restrict processing, and to object, unless the research is of significant public interest. Also, reflecting the difficulties of pure anonymisation, the GDPR encourages the pseudonymisation technique, to which it refers in numerous provisions.

Finally, regarding the access of public data by law enforcement agencies and the respective authorisation procedures, they can vary significantly across jurisdictions with differing oversight mechanisms since this is an area largely regulated by national legislation. Thus, it should be reiterated that such agencies will generally be afforded express powers by Statute to operate and gain access under certain circumstances, and particular controls. For many countries, this involves the exercising of some form of warrant dependent on, inter alia, the type of information to be accessed and the urgency of the matter (i.e., a matter of national security). Furthermore, article 48 of the GDPR includes a provision concerning the recognition and enforcement of 'any judgment of a court or tribunal and any decision of an administrative authority of a third country requiring a controller or processor to transfer or disclose personal data'. Therefore, such judgments or decisions may only be recognised or enforceable in any manner, if based on an international agreement, such as a mutual legal assistance treaty, in force between the requesting third country and the Union or a Member State.

As a general conclusion, the GDPR aims to strengthen data subject's rights, but also to promote innovation by encouraging research initiatives. To this end, it provides a broad definition for research, as well as numerous exceptions from the purpose limitation principle and other data subjects' rights in this context. However, in order to benefit from the above-mentioned exceptions, both researchers and member states must have in place adequate safeguards for the effective protection of personal data. With particular reference to research, the key changes introduced by the GDPR can be summarised as follows.

(a) Increased responsibilities for research organisations: Under the GDPR the principles of privacy by design and by default, in respect of the main principle of data minimisation will be the standard approach for data collection and use. Controllers and processors will now have more enhanced accountability obligations to maintain extensive records on data processing activities. In addition, organisations will have to undertake privacy impact assessments, to notify risky data breaches to the DPAs and to affected data subjects, in cases of high risks and damage caused by a breach, as well as to appoint data protection officers, when the organisation is involved in regular and systematic monitoring or processing of sensitive personal data on a large scale.

(b) Profiling: The GDPR explicitly prohibits the use of an individual's sensitive personal data for profiling purposes, unless (a) that individual

has given his/her explicit consent (except where a law provides that such prohibition cannot be lifted by the individual's consent); or(b) such profiling is necessary for reasons of public interest.

(c) Consent: Consent must be specific and evidenced by clear affirmative action and explicit consent is required from individuals to process special categories of data (i.e. health related data). All information notices including privacy policies and research consent forms must be written in plain and intelligible language, while consent must be as easy to withdraw as it is to give. It is noteworthy to mention at this point that data for research purposes can also be processed by relying on the "legitimate interests of the data controller", thus without the need to obtain consent from the data subject, under the condition that this does not override the rights of individuals. At this point, it should be noted that consent is a matter usually addressed also on a national level by ethics committees, which provide for additional standards that need taking into account.

(d) National regimes for scientific, statistical and historical research: A considerable margin of manoeuvre is provided to the Member States to derogate from the obligations of the GDPR regarding research purposes, under the condition that they provide adequate safeguards. This possibility provides a dispensation from data subject rights to access, rectification of inaccurate data, restriction of processing and to object, including processing for research purposes. However, it should be highlighted that in that case, research must be done in line with recognised ethical research standards and by implementing appropriate technical and organisational safeguards, such as data minimisation and pseudonymisation.

(f) Penalties and fines: The significant penalties for non-compliance with fines of up to 4% of worldwide turnover or €20 million, point out the significance of the obligation to comply with the rules as set out in the GDPR.

European Commission

CLARUS has received funding from the European Union's Horizon 2020 programme – DG CONNECT Software, Services & Cloud. Contract No. 644024