Co-financed by the Connecting Europe Facility of the European Union

CONNECTING EUROPE FACILITY

EU-WIDE Legal Text Mining Using Big Data Infrastructures

[Deliverable 1.3: Report on Legal and Ethical Implications]

Deliverable Form	
Project Reference No.	INEA/CEF/ICT/A2017/1567047
Deliverable No.	D1.3
Relevant Activity:	Activity 1: Conceptual Framework Definition and User Requirements
Nature:	R (R=Report, P=Prototype, D=Demonstrator, O=Other)
Dissemination Level:	CO PU = Public, PP = Restricted to other programme participants (including the Commission Services), RE = Restricted to a group specified by the consortium (including the Commission Services), CO = Confidential, only for members of the consortium (including the Commission Services)
Document version:	Draft V5
Date:	27/09/2019
Authors:	Anna-Sophie Novak, Danube University Krems (Austria) / Shefali Virkar, Danube University Krems (Austria), Michalis-Avgerinos Loutsaris, University of the Aegean (Greece) / Charalampos Alexopoulos, University of the Aegean (Greece) / Yannis Charalabidis, University of the Aegean (Greece) / Euripidis Loukis, University of the Aegean (Greece) / Lilian Mitrou, University of the Aegean (Greece) / Ilias Romas, Intrasoft International S.A. (Luxembourg)/ Stefania Stavropoulou, Intrasoft

	International S.A. (Luxembourg)/ Sofia Tsekeridou, Intrasoft International S.A. (Luxembourg)
Document description:	The purpose of this report is to provide a set of legal specifications to be considered for the ManyLaws project to comply with relevant data protection laws. Ethical concerns have also been critically examined.

Document History

Version	Date	Author (Partner)	Remarks
Draft v0.10	18/04/2019	Anna-Sophie Novak, Shefali Virkar	Initial Structure
Draft v0.20	23/06/2019	Anna-Sophie Novak, Shefali Virkar	
Draft v0.30	15/07/2019	Shefali Virkar, Sofia Tsekeridou	Updated Structure
Draft v0.40	17/07/2019	Anna-Sophie Novak, Shefali Virkar, Michalis-Avgerinos Loutsaris, Charalampos Alexopoulos, Yannis Charalabidis, Euripidis Loukis, Lilian Mitrou, Ilias Romas, Stefania Stavropoulou, Sofia Tsekeridou	
Draft v0.50	27/09/2019	Shefali Virkar	
…	…		
Final v1.0	DD/MM/YYYY		

Executive Summary

New technologies enable the automated computational analysis of information in digital form - such as text, sounds, images or data - generally known as text and data mining. Text and data mining technologies are prevalent across the digital economy; however, there is widespread acknowledgment that text and data mining can, in particular, benefit the research community and, in so doing, support business innovation. In the European Union, research institutions and business organisations are confronted with legal uncertainty as to the extent to which they can perform text and data mining of publicly available content. In certain instances, text and data mining can involve acts protected by copyright, by the *sui generis* database right or by both (in particular, the reproduction of works or other subject matter), the extraction of content from a database or both – all of which occur, for example, when the data are transformed by these advanced informatics techniques. Where no exception or limitation applies, consent to undertake such acts is required from right holders. Depending on the nature of the legal datasets processed during the lifecycle of the project, the sources from whence they are collected, and the manner in which they are aggregated, different laws might be applicable at any given time. In order to ascertain whether the appropriate means are reasonably likely to be used in the large-scale processing of big open legal data, account should also be taken of all objective factors; such as the costs of and the amount of time required for identification of all possible stakeholders, the availability of the technology at the time of processing, and current technological developments and trends. The purpose of this report is to, first and foremost, therefore, assess the legal implications of sophisticated data analysis techniques within this context, and to provide a comprehensive understanding of the legal principles and obligations that might come to influence the ManyLaws project, with a view to recommending adequate remedies and safeguards in the future.

The second part of this report will focus on the projected impact of so-called 'invisible' programming values on the ethical design, deployment and subsequent use of the ManyLaws platform. This section identifies and discusses the numerous ethical considerations that system developers need to be mindful of during the various stages of system development; and their roots in traditional moral philosophy. It may be argued that the speed, complexity, and multi-functionality afforded by newer technologies – and their deployment as decision support systems that not only complement, but have a direct impact on political, economic and societal outcomes – have necessitated the closer examination of hitherto neglected moral questions associated with the actions of their creators, developers and promoters, especially before the technology becomes publicly available. As society becomes more and more reliant on information created, collected, collated, and communicated through the use of technology, it becomes imperative that those individuals responsible for the planning, development and operation of these systems are aware of their moral obligation to guarantee information integrity and universal access. Indeed, the average user has to trust the software professional to develop a system that respects personal privacy, protects individual freedoms, and delivers accurate results. A final ethical consideration is thus concerned with equipping programmers with a moral compass to inform their professional activities; wherein the express purpose of such a cognitive framework is to guide complex decision-making while taking into consideration the relevant contextual factors surrounding the given decision point.

The final section of this report describes a half-day workshop organised by the ManyLaws project team under the auspices of JURIX 2018: The 31st International Conference on Legal Knowledge and Information Systems, held on 12 December 2018 in Groningen, The Netherlands. The aim of the workshop was to critically explore the factors contributing to the effective delivery of the services proposed by the project, and to identify emerging legal and ethical implications associated with the application of advanced computing technologies to the acquisition, storage, and processing of legal information. Particular emphasis was placed on data protection and copyright laws, and attempts made to find possible solutions for those concerns. The need for a moral compass to guide system development was highlighted.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF TERMS AND ABBREVIATIONS

Term/Abbreviation	Definition
AI	Artificial Intelligence
API	Application Programming Interface
BOLD	Big Open Legal Data
CKAN	Comprehensive Knowledge Archive Network
DCAT	Data Catalog Vocabulary
DSI	Digital Service Infrastructure
DUK	Danube University Krems, Austria (ManyLaws Consortium Member)
EC	European Commission
EDP	European Data Portal
EEA	European Economic Area
ELI	European Legislation Identifier
EU	European Union
GDPR	General Data Protection Regulation
HPC	High Performance Computing
IATE	Interactive Terminology for Europe (Database)
RIS	Rechtsinformationssystem des Bundes (Database)
UAEGEAN	University of the Aegean, Greece (ManyLaws Consortium Member)

1. INTRODUCTION

1.1 Purpose and Scope

New technologies enable the automated computational analysis of information in digital form, such as text, sounds, images or data, generally known as text and data mining. Text and data mining makes the processing of large amounts of information with a view to gaining new knowledge and discovering new trends possible. Text and data mining technologies are prevalent across the digital economy; however, there is widespread acknowledgment that text and data mining can, in particular, benefit the research community and, in so doing, support innovation. Such technologies benefit universities and other research organisations. However, in the European Union (EU), such organisations and institutions are confronted with legal uncertainty as to the extent to which they can perform text and data mining of content. In certain instances, text and data mining can involve acts protected by copyright, by the sui generis database right or by both, in particular, the reproduction of works or other subject matter, the extraction of contents from a database or both which occur for example when the data are transformed in the process of text and data mining. Where no exception or limitation applies, consent to undertake such acts is required from right holders. As research is increasingly carried out with the assistance of digital technology, there is a risk that the EU's competitive position as a research area will suffer, unless steps are taken to address the legal uncertainty concerning text and data mining.

The aim of the ManyLaws project is to deliver a novel set of services for citizens, businesses and administrations of the European Union, built upon text mining, advanced processing and semantic analysis of legal information. The Action will attempt to build the proper environment and vision of semantically annotated Big Open Legal Data (BOLD), easily searchable and exploitable with proper visualization techniques. The ultimate objective is to provide the technical foundation and the tools for making legal information available to everyone, in a customizable, structured and easy to handle way. To this end, big legal data will need to be accessed, which is currently produced and published in multiple national or EU public databases (e.g. RIS, EUR-Lex) or privately-owned legal databases (e.g. NOMOS). Those datasets will need to be extracted, linked and transformed into a structured relational, open database to prepare them for the mining process.

1.2 Methodology and Structure of the Deliverable

The ManyLaws project includes the development of services that also involve the deployment of large-scale technical processes facilitated by super computing technologies. It is for this reason that exploring the legal and ethical implications of using legal data and of the technical processes involved is both relevant and timely. The report attempts to take a critical look at some of the following questions: what are the central issues involved in the application of text mining tools and techniques to legal data sources and artefacts? What are some of the risks implicit in creating a big open legal database (BOLD), and how can they be mitigated? What are the associated ethical concerns that arise through the use of high performance computing in the development of a legal information retrieval system? How can these concerns be approached and potentially addressed? What is a Moral Compass, and what are the benefits of equipping software developers with guiding ethical principles? This report aims at giving an overview of the potential legal risks relating to Text and Data Mining. The legal fields that could be affected by the mining process include copyright law, database protection law and data protection law. Since the type of data (legislation, case law, social media posts and journal articles) needed for the project are already known, they will be considered, where relevant. This report does not aim at giving comprehensive legal advice.

The content of the deliverable was determined by the project services. It is the aim of the project to mine legislation, case law, social media posts and journal articles over the course of the project. Therefore, this deliverable aims at giving an overview of the legal implications of mining these data types. Since concrete data sources are missing in some cases (case law, social media posts and journal articles) or are still under discussion for others, this deliverable examines the mining process of the data types and the mining process of concrete sources, where those were known. This report is structured into four main parts: Chapter 2 critically discusses the legal concerns associated with the mining of legal information data sources. Chapter 3 then critically examines some of the ethical considerations associated with the ManyLaws project. A workshop held by ManyLaws researchers to explore both these sets of issues is then described, and its outcomes analysed, in Chapter 4. The final chapter, Chapter 5, presents a summary of major themes, issues, and conclusions.

2. MANYLAWS: LEGAL CONCERNS OF MINING SOURCES OF LEGAL INFORMATION

2.1 Overview of Types of Legal Data and Main Data Sources

The architecture for the mining process of legislation is finite. ManyLaws focus lies on the European, Greek and Austrian legal framework. To be more specific, ManyLaws will use five sources of legal information. The identified Greek legal sources are the Greek National Printing office and the Hellenic Parliament Portal while the Austrian are the RIS and the Austrian National Parliament. The Greek Legal Framework contains many types of legal documents, such as laws, presidential degrees etc. Further, the Rechtsinformationssystem (RIS) and Austrian National Parliament will be used for Austrian legal framework, as Austrian sources. Austrian Laws are separated into federal and state law. Finally, EUR-Lex will be used for European Legislation. European Legislation includes Directives, Regulations, etc. but ManyLaws will use only Directives. A detailed analysis of the mining process of these data sources will follow in a separate report.

2.2 Legal and Technical Guidelines

2.2.1 Technical background: Description of the Technical Process

ManyLaws aims to address the challenge of fragmented information in the legal domain, by delivering a set of key services that facilitate seamless and ubiquitous access to legal data to citizens, businesses and administrations. These services will be built upon acquiring, storing, integrating and processing large amounts of legal information, at an advanced level in various languages, using the power of text mining, information processing and visual analytics. The process of the project is presented in the following figure, Figure 1 below.

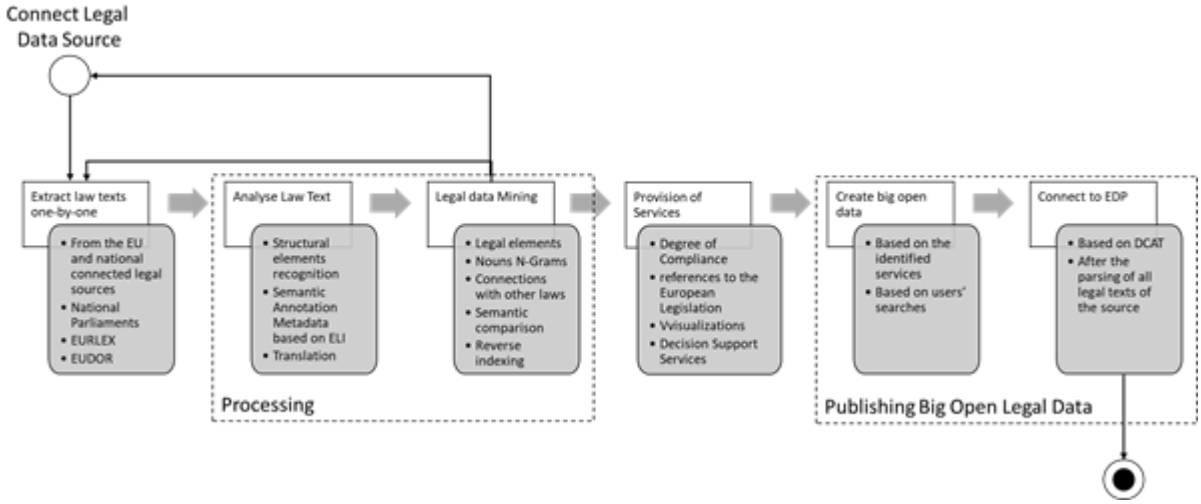

Figure 1: Diagrammatic Representation of the ManyLaws Process

The information processing stage of the infrastructure will make use of massively parallel computing tools, balancing the load between batch and real-time service modes. Two sub-stages are envisaged:

Stage 1: Analysis of Legal Texts

The "Analysis of Legal Texts" stage includes (a) the recognition of the structural elements, (b) the semantic annotation of the metadata, and (c) the translation phase. The ELI, DCAT and the Akoma Ntoso ontology standards will be used in order to facilitate the identification and the annotation of the metadata fields. For the translation of legal documents in the English language to be achieved a comprehensive research was held for the identification of the available online EU official tools capable of translate any legal element automatically. Towards this purpose, IATE and eTranslation DSI APIs will be

used. Generally, the first stage of the information processing stage of the infrastructure is the basis for the second stage "Legal Data Mining".

Stage 2: Legal Data Mining

The second stage includes all the text mining algorithms which will be used for the development of the services. Various algorithms will be applied in different processing tasks, relying on a super-computing infrastructure, in order to produce service – oriented intermediate results. Based on the results of the first stage in combination with the requirements of the target user groups, various tasks which require data mining techniques were identified. More specifically, data mining techniques will be used for:

1. The metadata extraction (based on ELI, DCAT and Akoma Ntoso)
2. The identification of the nouns of each legal document
3. The creation of various n-grams
4. The Law decomposition. The specific task refers only to the Greek legal documents as it is mentioned above (see 2.1.1 Data Sources)
5. The extraction of the correlations among the legal documents.

The data are processed via the Text Mining Tool RapidMiner. Moreover, for the above tasks, tools and algorithms capable of handling large volumes and high velocity of data, are employed on high-performance parallel computers so that data mining can analyse massive databases in a reasonable time. Specifically, an Apache Hadoop Cluster will be used towards the above mentioned reason. Furthermore, the Apache Hadoop YARN (component of the Hadoop Cluster) combines a central resource manager with containers, application coordinators and node-level agents that monitor processing operations in individual cluster nodes. Thus, resources can dynamically allocated to applications as needed, a capability which can improve resource utilization and application performance.

2.2.2 Text Mining Process: Description and Issues Identification

Text Mining is the process by which information and connections are obtained from large amounts of text using algorithms. Text Mining generally begins with defining which are the text serves as input for further processing. For storage purposes the differently formatted texts are firstly converted and secondly split up in sentences and word[1]. Contrary to databases, where the data is already structured, raw input texts are usually not. For the extraction of knowledge from the texts, the characteristic concepts for the application have to be identified, brought into a systematic context and instantiated[2]. Text mining allows for automatic or semi-automatic structuring of large amount of text. Following the storage and transformation of the identified texts, statistical and pattern-based procedures are applied: With statistical approaches, relevant features that are assigned to texts, using linguistic statistical laws[3]. By means of difference-analysis of the used vocabulary, one can identify the relevant discipline-specific terminology[4]. Semantic dependencies between terms are calculated with the help of co-occurrence-analysis and for example cluster-analysis[5]. Pattern-based analysis follows a different approach: Within texts, generally-valid and relevant patterns are found and passages of texts are identified on the basis of previously determined patterns[6]. The result of the analysis is a relational database that can be used for further applications.

[1] Cf. G. Heyer, U. Quasthoff, T. Wittig (2006). Text Mining: Wissensrohstoff Text: Konzepte, Algorithmen, Ergebnisse, W3L AG 2006.

[2] Ibid.

[3] Ibid.

[4] Ibid.

[5] Ibid.

[6] Ibid.

For the proposed application, the standard Text Analysis pipeline performs several levels of analysis: morphological, syntactic, semantic, and discourse[7]. The morphological and syntactic analysis is usually performed with a syntactic parser[8], which recognizes the syntactic word classes such as nouns and verbs, and the syntactic dependency structure of the constituents of a sentence (main verb, subject, object, etc.). In general language processing, recognizing the basic semantic roles of a sentence constituents, i.e., the “who”, “does what”, “where”, “when”, and "how" constituents, is a well-established task for English. Co-reference resolution is identifying when two mentions of an entity or event refer to the same underlying person, place, thing or event in the real world. The proposed ManyLaws ICT architecture supports the integration of the below mentioned pool of identified services while keeping the structure flexible allowing the inclusion of further services and data sources. A layered approach (*see* Figure 2) supporting the data flow, from source data to visualized outputs will handle the large volumes of data. When it comes to the information processing layer, various Text Mining algorithms are applied in different processing tasks, relying on a super-computing infrastructure, in order to produce service-oriented intermediate results such as: creation of reverse indexing, occurrence and frequency tables for millions of words, creation of various n-grams for the identification of important terms or phrases, semantic comparison of different law sets (e.g. EU Directive against national legal framework) performing full word-level and document-to-document comparison for billions of pages. Therefore the power of thousands of processors is needed.

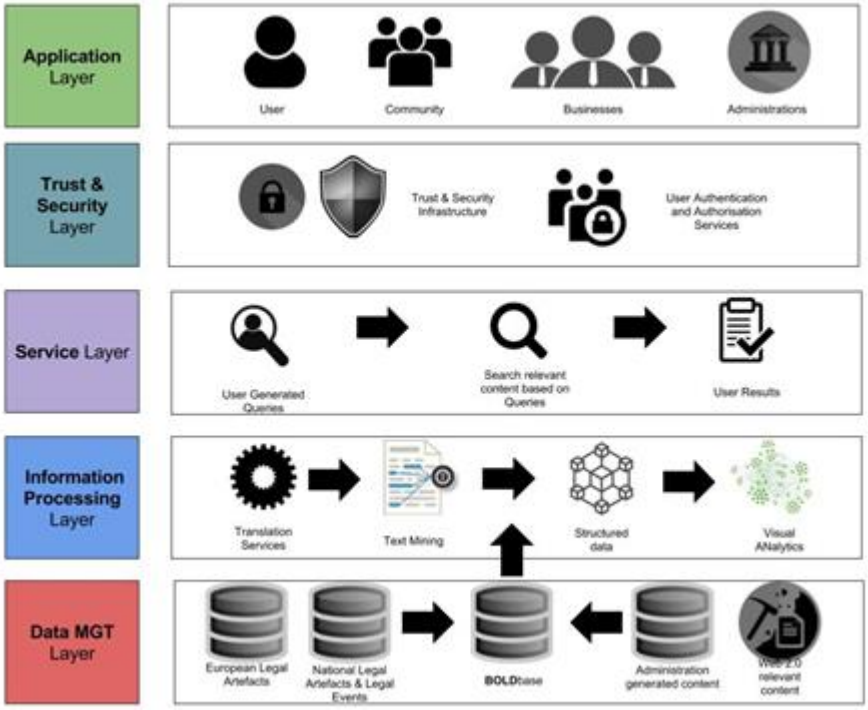

Figure 2: ManyLaws High Level Architecture

Possible applications of Text Mining include efficient research in text bases, identifying similar documents, finding resemblances of terms, locating definitions and references in large amount of texts. With regard to the legal field possible applications include the creation of a legal information retrieval system offering users multiple novel services which, in the case of the ManyLaws project, include functionalities to facilitate research through legal corpora, analysis of the alignment of national legislation with EU legislation, comparisons between national laws which target the same life events, analyses of the references to European legislation by national laws, analyses of related laws within the same Member State, timeline

[7] Cf. M.C. Lacity, M. A. Janson (1994). Understanding qualitative data: A framework of text analysis methods. *Journal of Management Information Systems* vol. 11 iss. 2, pp. 137-155.
[8] N. Lagos, M. Gallé, A. Chernov, A. (2017). U.S. Patent Application No. 14/850,060

analysis for all legal acts, the visualization of progress and current status of a specific national or European piece of legislation, and the sentiment analysis of new legislation.

2.2.2.1 Services Identification

In order to design the appropriate text mining process, it is deemed necessary to first identify the service layer. Services are to be provided towards citizens, businesses and administrations, based on the most common needs of each user type. Through a user interface supporting simplification or advanced usage, these are some of the services to be provided in real time by the ManyLaws project:

• Parallel search in EU member-state legal frameworks (through parallel translation of search terms), using simple keywords. For the current project lifecycle, the focus lies on the European, Greek and Austrian legal framework.

• Assessment of the degree of transposition of an EU Directive in a National Legal Framework.

• Analysis of references to the European Legislation by National Laws.

• Comparative analysis of equivalent or relevant laws from different EU member states or from the same member state.

• Timeline analysis for all legal elements, visualising the progress and current status of a specific national or European legislation (after amendment/extensions) over time including pre-preparatory acts and agreements.

• Visualizations of correlations, dependencies and conflicts between different laws.

• Decision Support Services (e.g. Impact Assessment) within legal procedures.

Based on the above services and sub-products, a variety of add-on services can be developed after capturing new requirements from citizens, businesses and administrations. The system consists of three service components:

(a) **User-generated queries**: the system infrastructure stores search queries made by users for analysis and optimization purposes. The terms and structure that make up the user query are feeding into the semantic search engine to enhance the relevance of the results based on inferred concepts and semantic annotations.

(b) **Search of relevant content based on queries**: the search engine is used for searching through the system's triple store using a scalable Solr-based semantic search engine. Furthermore, the search engine is taking into account semantic relations between search terms and stored entities (e.g. synonyms). Best practices such as faceted search are also used to present the user with more search options, relevant to the search terms by semantic association.

(c) **Search results**: This component is responsible for retrieving and presenting to the user the search results in an efficient and user-friendly manner.

2.2.2.2 Design of the Appropriate Text Mining Process

The *information processing* layer of the ManyLaws architecture, visualised above in Figure 2, deals with the text mining process of the identified legislative databases and could be described in four distinct steps:

a) **Data Preparation and Translation Services:** This is the stage where data is acquired and prepared for the text mining tools to follow. The stages include data reading and initial cleansing, anonymization if needed, semantic annotation (so that to be indexed at the European Open Data Portal), and formulation for processing. Due to the diversified origin of the texts to be acquired, a large amount of effort and computational power will have to be devoted to the translation in English, if any legal document is written in a different member state official language. Translation will be based on eTranslation and IATE which APIs can be used for an automated translation to be achieved. Both eTranslation DSI and IATE will be used for all translation services, for various indexes and n-grams.

b) **Text Mining:** Various algorithms will be applied in different processing tasks, relying on a super-computing infrastructure, in order to produce service-oriented intermediate results. More specifically, the exact processing tasks are:

• The creation of reverse indexing, occurrence and frequency tables for millions of words.

• The creation of various n-grams for the identification of important terms or phrases.

• The extraction of the legal documents' metadata based on ELI, DCAT and Akoma Ntoso.

• The semantic comparison of different law sets (e.g. EU Directive against national legal framework) performing full word-level, document-to-document comparison for billions of pages – needing the power of thousands of processors.

• The identification of the interrelations of all original and translated terms and texts.

• The term extraction analysis of news, social media, blogs and other content.

c) **Structured Data**: This component represents the information collected from the various identified (and already mentioned above) sources and adhering to a common model and format via text mining algorithms. The *structured data* component is of great importance since the results will be used more effectively by the visual analytics service subsequently described. The data will be stored in a database (MongoDB) and will include any harvested and derived information that is necessary to realize the project's use cases.

d) **Visual Analytics Service:** The *visual analytics service* provides the ability to access the entire data transformation pipeline from raw or semantic data to interactive visual representations. The main goal is to enable user-centered and comprehensible solutions for getting insights and knowledge about the entire domain.

2.3 Applicable Law

2.3.1 Copyright Law

One has to differ between the question concerning which court has jurisdiction from the question which law the judge has to apply. This chapter will discuss the applicable law, once a court has jurisdiction to decide the case. In intellectual property law matters with cross-border implications, the function of International Intellectual Property Law is to assign those cases to the relevant legal system for the purposes of their assessment.

Thus, the Rome II Regulation is used to determine which law is applicable to non-contractual obligations[9] in civil and commercial matters, which have a connection to the law of different states. The Rome II Regulation is fully applicable to cross-border cases as a European Regulation both in Austria and in Greece. According to Art. 8 (1) Rome II Regulation, the law applicable to a non-contractual obligation arising from an infringement of an intellectual property right shall be the law of the country for which protection is claimed. That is to say, under the law of the State for whose territory intellectual property rights are claimed, not the country in which the infringement was committed. According to Article 8 (3) Rome II Regulation, the applicable law is beyond the discretion of the parties[10].

Thus, the origin[11], the content, the scope of protection and also the expiration of intellectual property rights are subject to the legal system of the state in which protection is claimed (principle of *lex loci protectionis*).[12] This inevitably leads to the application of different legal systems when referring to acts of exploitation in several states.[13] It is therefore necessary to determine the law applicable to each alleged act of exploitation in different nations.[14] This also applies if the users are confronted with a multitude of legal systems[15].

[9] The term "non-contractual obligations" is to be interpreted in a regulation-autonomous way.
[10] Handig (2009). Das Leck im Schutzlandprinzip, ecolex 2009, 776: "From this perspective, it is a clarification if Art. 8 para. 1 Rome II enshrines the lex loci protectionis. This clarification serves in particular the legal certainty of largely non-codified conflict rules of the 26 member states."
[11] OGH 28.9.1993, 4 Ob 125/93, MR 1994, 26: "This also applies to the assessment of whether it is a free work."
[12] Handig (2009). Das Leck im Schutzlandprinzip, ecolex 2009, 775.
[13] Urheberrechtsverletzung im In- und Ausland wu.ac.at/fileadmin/wu/d/i/privatrecht/Spitzer/Publikationen/hoeller_evbl_2013_145.pdf, ÖJZ (2013) 22, S 1022 ff; RIS-Justiz RS0076849.
[14] Urheberrechtsverletzung im In- und Ausland wu.ac.at/fileadmin/wu/d/i/privatrecht/Spitzer/Publikationen/hoeller_evbl_2013_145.pdf, ÖJZ (2013) 22, S 1022 ff; RIS-Justiz RS0076849.
[15] OGH 28.9.1993, 4 Ob 125/93, MR 1994, 26.

The application of the *lex loci protectionis* is partially justified by the validity of the territoriality principle in copyright law.[16] Accordingly, the rights of the authors are limited to the individual states[17]. Since the national regulations are limited to the borders of the national territory, some argue, that the *lex loci protectionis* results for conflicts of laws (regarding the applicable law)[18].

Conclusion

Since the origin[19], the content, the scope of protection and also the expiration of intellectual property rights are subject to the legal system of the state in which protection is claimed (principle of *lex loci protectionis*), this concludes for this research report, that it is not possible to say which national legal systems will be applicable to a potential case.

2.4 Mining of Legislation and Case Law[20]

2.4.1 Copyright Law

The Austrian Copyright Act protects individual and intellectual creations in the fields of literature, sound art, fine arts and film art. In the context of the Text Mining process, it depends firstly on the Text Mining technique used and secondly on the selected texts whether there is an infringement of copyrights of third parties. Most Text Mining techniques, as does the proposed technique above, rely heavily on copying of the texts for them to be analysed and annotated. The act of copying protected texts is covered by Section 15 of the Austrian Copyright Act. It states that solely the author has the right to copy his or her work. The right to copy is understood broadly and includes even the technically required automated copying.[21]

Conclusion

Therefore, the described mining process would be unlawful, if the used texts are protected and no exception applies. There are exceptions to Copyright Law that could find application on the Text Mining Process. European Union law provides for certain exceptions and limitations covering uses for scientific research purposes which may apply to acts of text and data mining. However, these exceptions and limitations are optional and not fully adapted to the use of technologies in scientific research. Moreover, where researchers have lawful access to content, for example through subscriptions to publications or open access licences, the terms of the licences could exclude text and data mining.[22] These exceptions are discussed as follows:

The research exception[23]

Includes solely non-commercial research. Further, the researcher has to attribute the source of the used data, but only if this is feasible. Organizational structure and finances are not relevant for determining whether the research is commercial or non-commercial.[24] Relevant is the research purpose.[25]

The temporary copies exception

[16] Handig (2009). Das Leck im Schutzlandprinzip, ecolex 2009, 775.
[17] Handig (2009). Das Leck im Schutzlandprinzip, ecolex 2009, 776.
[18] Handig (2009). Das Leck im Schutzlandprinzip, ecolex 2009, 776.
[19] OGH 28.9.1993, 4 Ob 125/93, MR 1994, 26: "This also applies to the assessment of whether it is a free work."
[20]This section was published in Jusletter IT: A.-S. Novak, C. Udokwu, C. Alexopoulos, M.A. Loutsaris, S. Virkar (2019). Mining Legislation: An Analysis of Legal and Technical Implications. In E. Schweighofer, F. Kummer, A. Saarenpää (eds.) Internet of Things – Proceedings of the 22nd International Legal Informatics Symposium IRIS 2019, Bern: Weblaw, pp. 635-638.
[21]W. Dillenz, D. Gutmann (2004). Praxiskommentar zum Urheberrecht. Österreichisches Urheberrechtsgesetz & Verwertungs-gesellschaftsgesetz[2], Wien 2004, § 14 Rz 22.
[22] Recital 10 Directive (EU) 2019/790.
[23] § 42 (2) Urheberrechtsgesetz
[24] A. Zemann (2018). In G. Kuscko, C. Handig, urheber.recht. systematischer kommentar zum urheberrechtsgesetz[2], Wien 2018, § 42 Rz 23.
[25] W. Dillenz, D. Gutmann (2004). UrhG und VerwGesG[2], § 42 Rz 10.

Allows the copying if it is an essential and integral part of a technical process and itself not of economic importance. As the name of the exception suggests, the copies are allowed to be saved temporarily as long as the technical process requires them. The establishment of a permanent corpus with the original elements is therefore not possible under this exception.

Background to the Text Mining Exceptions

The Text Mining exceptions exist in the directive on copyright[26], which is already in force but has not been transposed yet. Especially in the fields of research and innovation, digital technologies permit new types of uses that are not clearly covered by the existing Union rules on exceptions and limitations. In addition, the optional nature of exceptions and limitations provided for in those fields could negatively impact the functioning of the internal market. This is particularly relevant as regards cross-border uses, which are becoming increasingly important in the digital environment. Therefore, the existing exceptions and limitations in Union law that are relevant for scientific research should be reassessed in the light of those new uses. Mandatory exceptions or limitations for uses of text and data mining technologies should be introduced. The existing exceptions and limitations in Union law should continue to apply, including to text and data mining, as long as they do not limit the scope of the mandatory exceptions or limitations provided for in this Directive, which need to be implemented by Member States in their national law. The exceptions and limitations provided for in the directive seek to achieve a fair balance between the rights and interests of authors and other right holders, on the one hand, and of users on the other. The exceptions and limitations can be applied only in certain special cases that do not conflict with the normal exploitation of the works or other subject matter and do not unreasonably prejudice the legitimate interests of the right holders.

The legal uncertainty concerning text and data mining should be addressed by providing for a mandatory exception for universities and other research organisations, to the exclusive right of reproduction and to the right to prevent extraction from a database. In line with the existing European Union research policy, which encourages universities and research institutes to collaborate with the private sector, research organisations should also benefit from such an exception when their research activities are carried out in the framework of public-private partnerships. While research organisations and cultural heritage institutions should continue to be the beneficiaries of that exception, they should also be able to rely on their private partners for carrying out text and data mining, including by using their technological tools.[27]

Art 3 of the directive regulates, the text and data mining process for the purposes of scientific research. Therefore an obligatory exception to the rights of the right holders will be provided for by the member states. The exception includes reproductions and extractions of protected works made by research organisations, to which they have lawful access to, for the purposes of scientific research. Further Art 3 provides for the storage of the copies: Copies of works or other subject matter made in compliance with above can be stored with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results. Further Art 3 regulates the security measures right holders are allowed to apply to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.

Research organisations: Despite different legal forms and structures, research organisations in the Member States generally have in common that they act either on a not-for-profit basis or in the context of a public-interest mission recognised by the State. Such a public-interest mission could, for example, be reflected through public funding or through provisions in national laws or public contracts. Conversely, organisations upon which commercial undertakings have a decisive influence allowing such undertakings to exercise control because of structural situations, such as through their quality of shareholder or member, which could result in preferential access to the results of the research, should not be considered research organisations.[28]

[26] Directive (EU) 2019/790.
[27] Recital 11 Directive (EU) 2019/790.
[28] Recital 12 Directive (EU) 2019/790.

Lawful access: Research organisations including the persons attached thereto, should be covered by the text and data mining exception with regard to content to which they have lawful access. Lawful access should be understood as covering access to content based on an open access policy or through contractual arrangements such as subscriptions, or through other lawful means. For instance, in the case of subscriptions taken by research organisations the persons attached and covered by those subscriptions should be deemed to have lawful access. Lawful access should also cover access to content that is freely available online.[29]

Storage of data

Research organisations could in certain cases, for example for subsequent verification of scientific research results, need to retain copies made under the exception for the purposes of carrying out text and data mining. In such cases, the copies should be stored in a secure environment.[30]

Technical measures

In view of a potentially high number of access requests to, and downloads of, their works or other subject matter, right holders are allowed to apply measures when there is a risk that the security and integrity of their systems or databases could be jeopardised. Such measures could, for example, be used to ensure that only persons having lawful access to their data can access them, including through IP address validation or user authentication. Those measures should remain proportionate to the risks involved, and should not exceed what is necessary to pursue the objective of ensuring the security and integrity of the system and should not undermine the effective application of the exception. Contracts that contradict the exception are unenforceable.

Apart from the research exception, Art 4 of the directive regulates an exception or limitation for text and data mining. Outside of research, text and data mining techniques are widely used both by private and public entities to analyse large amounts of data in different areas of life and for various purposes, including for government services, complex business decisions and the development of new applications or technologies. For these mining activities, right holders should remain able to license the uses of their works. At the same time, users of text and data mining could be faced with legal uncertainty as to whether reproductions and extractions made for the purposes of text and data mining can be carried out on lawfully accessed works or other subject matter. In order to provide for more legal certainty in such cases and to encourage innovation also in the private sector, the directive provided, under certain conditions, for an exception or limitation for reproductions and extractions of works or other subject matter, for the purposes of text and data mining, and allows the copies made to be retained for as long as is necessary for those text and data mining purposes.[31]

Therefore Art 4 of the directive regulates, that member states must provide for an exception or limitation to the rights of the right holders for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining. This exception or limitation must also regulate data storage: Reproductions and extractions made pursuant to above can be retained for as long as is necessary for the purposes of text and data mining.

This exception or limitation to the rights of the right holders applies on condition that the use of works and other subject matter has not been expressly reserved by their right holders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.

Austrian Law

According to section 7 paragraph 1 Austrian Copy Rights Act, laws, regulations, official decrees, notices and court judgements do not fall under the Copyright Act. Since the copying and of legislative data and case law falls under this exception, copyright law is not applicable to the extraction of legislative information or case law. Concerning the explanatory materials from the website of the Austrian Parliament one has to make sure that those are also covered by section 7 paragraph 1 Copyright Act. Explanatory Materials are of considerable importance for interpreting the law.

[29] Recital 14 Directive (EU) 2019/790.
[30] Recital 15 Directive (EU) 2019/790.
[31] Recital 18 Directive (EU) 2019/790.

Furthermore, they are attributable to the parliament, an authority with public authority tasks. As a result, copyright law is not applicable on explanatory materials, which can be subsumed as official notices.

Conclusion

Both legislation and case law do not fall under the Austrian Copyright Act according to section 7 paragraph 1 Austrian Copy Rights Act.

2.4.2 Database Law

Should the texts for the mining process be part of a database, copyright law as well as the sui generis database right have to be considered. Those two rights are independent of each other and can apply on the same database. According to the 17th recital of the directive on the legal protection of databases, the term 'database' should be understood to include literary, artistic, musical or other collections of works or collections of other material such as texts, sound, images, numbers, facts, and data, collections of independent works, data or other materials which are systematically or methodically arranged and can be individually accessed. According to the European Court of Justice, databases 'in any form', whether they are in electronic or non-electronic formats are covered by the directive. Below we will discuss general database protection and then go on to our specific goal of mining governmental legislative databases.

Copyright Protection of Databases

The copyright protection of databases does not extend to their contents. The selection or arrangement of the contents, mark the author's own intellectual creation and are protected as such by copyright. No temporary or permanent reproduction by any means and in any form, in whole or in part can be made without the consent of the author. Art 6 Paragraph 2 of the directive allows member states to make certain exceptions. In Austria, Section 42 Paragraph 2 and 40 h Paragraph 2 of the Copyright Act state that the reproduction for non-commercial research purposes is admissible. It is however doubtful whether the text mining process includes the extraction of the protected selection and arrangement of the contents, as the mining process concentrates on the content.

Sui Generis Protection of Databases

Section 76c of the Austrian Copyright Act protects investments made in databases. According to Article 7 of the directive, the maker of a database who has made qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents has the right to prevent extraction and/or re-utilization of the whole or of a substantial part of the contents of that database. The extraction and re-utilization of an insubstantial part of such a database is therefore allowed, as long as they are not executed repeatedly and systematically. According to § 76d Paragraph 3 Number 2 Copyright Act, the extraction and re-utilization of a substantial part of a publicly available database is lawful for non-commercial research purposes, to a justifiable extent, as long as the source is named.

As we have now discussed the general application of database law on the text mining process, the following section will give an assessment of the applicable laws when mining Austrian legislative databases. As neither the selection nor the arrangement of the contents of the Austrian legislative databases mark an intellectual creation, those are not protected as such by copyright protection of databases. Nonetheless, in the imprints on the website ris.bka.gv.at one will find a copyright notice in favour of the Federal Ministry of Digitalization and Economy and on the website parlament.gv.at one will find a copyright in favour of the Austrian Parliament. Concerning the sui generis protection of databases, one has to check if a substantial investment in either the obtaining, verification or presentation of the contents was made. A qualitative and quantitatively substantial investment was made especially in the presentation of the contents of the Austrian legislative databases. According to the CJEU, the investment going into the creation of the single entries is not taken into account. According to the Austrian Supreme Court, the definition of an investment does not depend on whether the data is given to the maker or if law prescribes the presentation of the data. Rather those expenses that run into the presentation and the updating of the database content are to be considered as investment. Therefore, the maker of the websites ris.bka.gv.at and parlament.gv.at have the right to prevent extraction and/or re-utilization of the whole or of a substantial part of the

contents of that database via the website. An application of section 7 Copyright Act on protected databases by analogy was denied by the Austrian Supreme Court.

Conclusion

Concerning the *sui generis* protection of databases, one has to check if a substantial investment in either the obtaining, verification or presentation of the contents was made. A qualitative and quantitatively substantial investment was made especially in the presentation of the contents of the Austrian legislative databases. Therefore, the maker of the websites ris.bka.gv.at and parlament.gv.at have the right to prevent extraction and/or re-utilization of the whole or of a substantial part of the contents of that database via the website. Since both websites offer some of their data on the national open data portal (data.gv.at) this right does not constitute an insurmountable barrier for the mining process.

2.5 Mining of Journal Articles

The sources for the mining process of journal articles are not known at this point as isn't the involved technical process. Therefore, the following section will provide a very general assessment. This section will be further expanded as more information is available. In the context of mining journal articles (without knowing the sources nor the exact technical process) copyright law, database protection law and data protection law might be relevant. Should these rights apply, consent from the right holders might be necessary. First, a general overview of these rights will be given, and second, exceptions from these rights discussed. Should the rights but no exception apply, consent of the right holder is necessary.

2.5.1 Copyright Law

2.5.1.1 Usage of rights

Reproduction

If the mining process involves the extraction of the data and duplicates are created or pdf files are converted into XML files, and those data are protected by copyright this constitutes a reproduction according to Art 2 InfoSoc-RL, which requires consent of the rights holder, if no exception applies.[32] One might say, since through crawling only a few words are copied and not the whole text, the right to reproduce is not affected at all, but the ECJ has clearly stated in the 'Infopaq' decision, that the reproduction of even eleven words could fall under this right.[33]

One could also argue that there are mining techniques that do not copy at all through crawling. So they do not involve acts of reproduction. Reproductions made under the mandatory exception for temporary acts of reproduction provided for in Article 5(1) of Directive 2001/29/EC, still benefit from this exception, as long as the text and data mining techniques do not involve the making of copies beyond the scope of that exception.[34]

Publication

The right to publish is regulated in Art. 3 InfoSoc-directive and is generally not applicable to text mining processes, because the original work itself is not published. Also, the research results obtained by the mining process of the assembled data, do not contain parts of the original work, but represent interpretations of the knowledge gained from the mining software.[35] If the original corpus is published, which will contain parts of the copied works (for example to verify research results from the mining process) this right could apply.[36]

[32] Spindler, Text und Data Mining – urheber- und datenschutzrechtliche Fragen, GRUR 2016, 1113.
[33] Infopaq Stelle zitieren, dass 11 Wörter reichen.
[34] Recital 9, Directive (EU) 2019/790.
[35] Spindler, Text und Data Mining – urheber- und datenschutzrechtliche Fragen, GRUR 2016, 1113.
[36] Spindler, Text und Data Mining – urheber- und datenschutzrechtliche Fragen, GRUR 2016, 1113.

Modifications

Apart from the extraction and copying of potentially protected works, the mining process also usually includes the transformation of the collected data. This transformation phase could constitute an edition. Editions are not governed by the InfoSoc-directive. According to § 5 Austrian Copyright Act, translations and other editions are protected as original works, as far as they are an intellectual creation of the editor, regardless to the copyright existing on the edited work. The use of one work in the creation of another does not make it an edition, if it represents an independent new work in comparison to the work used.[37] To evaluate whether an edition applies or not one has to check whether the transformation of the collected data (which are not whole texts generally, but parts of the original work) into a different format or/and the enriching of this data with metadata constitute an edition.[38] Spindler (2016) does not see an edition in these two steps, since, on the one hand, the parts of the work are copied identically into another format (so no other work is created) and the metadata does not touch the actual work, but rather only enriches it and systematizes it according to its own scheme.[39]

Relevant exceptions from Copyright Law

a) EU exception: This is discussed above.

2.5.2 Database Protection Law

Should the collected journal articles be part of a database, the copyright of these database as well as the sui generis protection of databases has to be observed. Both rights are explained above as well as the exceptions form database protection law.

2.5.3 Data Protection Law

Should the journal articles that are mined contain any personal data, the GDPR might be applicable. The consequences of that scenario are discussed further below.

2.6 Mining of Social Media Posts

Since neither the sources nor the technical processes are fully known yet, this chapter will provide a general overview. With regard to copyright and database protection law, read further above.

2.6.1 Personal Data Protection

Personal data protection is an area of the law that could also be affected by the mining process.[40] The GDPR[41] applies to the processing of personal data.[42] Processing means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means such as “collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination restriction, erasure or destruction”.[43] Personal data could be “a name, a photo, an email address, bank details, posts on social networking websites, medical information, or a computer IP

[37] Translated via Google Translate.
[38] Spindler, Text und Data Mining – urheber- und datenschutzrechtliche Fragen, GRUR 2016, 1114.
[39] Spindler, Text und Data Mining – urheber- und datenschutzrechtliche Fragen, GRUR 2016, 1114.
[40] J.-P. Triaille, J. de Meeûs d’Argenteuil, A. de Francquen (2014). Study of the legal framework of text and data mining (TDM), prepared for the European Commission, 91.
[41] Regulation (EU) 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 Directive 95/46/EC (General Data Protection Regulation), Official Journal of the European Union L119/1.
[42] Art 2 GDPR.
[43] Art 4 (2) GDPR.

address".[44] The mining process as such can be subsumed as processing according to the regulation, as it is a set of operations that are performed on (possibly) personal data. Therefore, the mining process falls under this definition.[45]

Conclusion

Should the project aim to collect data from posts on social networking websites to make use of the Text Mining function of sentiment analysis this amounts to the processing of personal data and the applicability of the GDPR.

2.6.2 Principles Art 5 GDPR

Lawfulness, fairness and transparency

Personal data shall be processed lawfully, fairly and in a transparent manner in relation to the data subject.

Purpose limitation

Data must be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes. Further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes is, in accordance with Art 89 (1) (see below), not considered to be incompatible with the initial purposes.

Data minimisation

Data collection is adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.

Data accuracy

Data are accurate and, where necessary, kept up to date. Every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay.

Storage limitation

Storage of the data in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed. Personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89 (1) (*see below*). This is subject to the implementation of the appropriate technical and organisational measures required by the GDPR.

Integrity and Confidentiality

Data are processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures.

2.6.3 Legality of the Mining Process

According to Art 6 any processing of personal data must be based on a legal basis and this applies without limitation also to data processing for scientific research purposes. Processing of personal data must therefore comply with one of the

[44] European Commission, "Data protection reform: Frequently asked questions", MEMO/12/41, 25 January 2012, http://europa.eu/rapid/press-release_MEMO-12-41_en.htm?locale=en (accessed 20.5.2019).
[45] *Triaille J-P, de Meeûs d'Argenteuil J, de Francquen A* (2014) Study of the legal framework of text and data mining (TDM), prepared for the European Commission, 91.

conditions set out in Article 6 (1) (a) to (f) in order to be lawful. Consent or a legal provision are the most likely legal grounds for the processing. According to Art 6, the processing of personal is unlawful, unless at least one of the following applies:

Consent

(a) The data subject has given consent to the processing of his or her personal data for one or more specific purposes. The conditions for consent are laid down in Art 7 of the regulation. According to Art 4 (11) consent of a data subject means a "freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her". Fulfillment of all of these requirements may often be difficult to achieve when for example consenting to data processing for scientific research purposes, since the exact purposes of data processing are not yet clear at the beginning of a research project.[46] Recital 33 GDPR covers the so-called broad consent, where it is not possible to fully identify the purpose of processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose.

Legal reasons

(b) Processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;

(c) Processing is necessary for compliance with a legal obligation to which the controller is subject;

(d) Processing is necessary in order to protect the vital interests of the data subject or of another natural person;

(e) Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

(f) Processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

Point (f) of the first subparagraph shall not apply to processing carried out by public authorities in the performance of their tasks. In addition to the consent of data subjects, data processing in the field of scientific research could rely on legal provisions. The GDPR does not regulate the legal provisions on which the processing could be based, but stipulates opening clauses, the member states could make use of.

Conclusion

According to Art 6 any processing of personal data must be based on a legal basis and this applies without limitation also to data processing for scientific research purposes. In addition to the consent of data subjects, data processing in the field of scientific research could rely on legal provisions.

2.6.4 Rights of the Data Subject

2.6.4.1 Art 15 Right of Access by the Data Subject

This right is an addition to the information obligations in Art 13 and 14 (*below*). The data subject has the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed and, where that is the case, access to the personal data and the following information: the purposes of the processing, the categories

[46]Knotzer, Wissenschaftliche Forschung und Datenschutz: Eine kritische Analyse ausgewählter Aspekte der österreichischen Rechtslage, ZTR 2018, 205.

of personal data concerned (should allow the data subject a quick assessment[47]), the recipients or categories of recipient to whom the personal data have been or will be disclosed (the controller should store the information regarding which recipient received which data[48]), in particular recipients in third countries or international organisations, where possible, the envisaged period for which the personal data will be stored, or, if not possible, the criteria used to determine that period, the existence of the right to request from the controller rectification or erasure of personal data or restriction of processing of personal data concerning the data subject or to object to such processing, the right to lodge a complaint with a supervisory authority, where the personal data are not collected from the data subject, any available information as to their source, the existence of automated decision-making, including profiling, and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject. Where personal data are transferred to a third country or to an international organisation, the data subject shall have the right to be informed of the appropriate safeguards pursuant to Article 46 relating to the transfer.

As opposed to Art 13 and 14, these information obligations in Art 15 have to be provided for by the controller only when the data subject requests it.[49] The request does not need any special form or content and is free of charge.[50] Where the controller processes a large quantity of information concerning the data subject, the controller should be able to request that, before the information is delivered, the data subject specify the information or processing activities to which the request relates.

Further, the data subject can request a complete copy of the personal data undergoing processing. This right is connected to the above mentioned information rights but they do not rule each other out.[51] The controller has to provide a copy with the personal data as he has them stored, he does not need to prepare them in any special condition.[52] For any further copies requested by the data subject, the controller may charge a reasonable fee based on administrative costs. Where the data subject makes the request by electronic means, and unless otherwise requested by the data subject, the information shall be provided in a commonly used electronic form. Where possible, the controller should be able to provide remote access to a secure system which would provide the data subject with direct access to his or her personal data.[53]

The right to obtain a copy referred to above shall not adversely affect the rights and freedoms of others (for example copyright rights). The right to obtain a copy is limited in case the rights and freedoms of others are affected. The information obligations according to Art 15 (1) and (2) could be restricted by the member states.

Conclusion

The data subject has the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed and, where that is the case, access to the personal data and related information.

2.6.4.2 Art 16 Right to Rectification

According to Art 16, the data subject shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her. For the right to rectify, it does not matter, whether the controller made a mistake or the data subject entered false data.[54] Taking into account the purposes of the processing, the data subject shall have the right to have incomplete personal data completed, including by means of providing a supplementary statement.

The data subject has to make an application with the controller to assert the right to rectification. The controller should also provide means for requests to be made electronically, especially where personal data are processed by electronic means.[55] It is advisable for the data subject to provide information or proof of their identity, should the controller have

[47] Bäcker, Art 15 DSGVO Rz 14.
[48] Bäcker, Art 15 DSGVO Rz 18.
[49] Bäcker, Art 15 DSGVO Rz 1.
[50] Bäcker, Art 15 DSGVO Rz 30.
[51] Bäcker, Art 15 DSGVO Rz 39.
[52] Bäcker, Art 15 DSGVO Rz 40.
[53] ErwG 63 GDPR.
[54] Herbst, Art 16 DSGVO Rz 14.
[55] ErwG 59 GDPR.

justifiable doubt about the identity he can demand additional information regarding the identity.[56] The controller is obliged to respond to requests from the data subject without undue delay and at the latest within one month and to give reasons if the controller does not intend to comply with any such requests.[57] Should the controller need some time to verify the right to rectification from the data subject, the data subject has in the meantime the right to demand the limitation of the processing of these personal data (Art 18 (1) lit a). Should the controller refuse to rectify the personal data, he has to justify this decision (Art 12 (4) and point to the right to lodge a complaint with a supervisory authority or court. Should the controller have shared the personal data, he has to inform the recipients (Art 19 GDPR). Should the right to rectify and the right to erasure (Art 17 below) both apply, the data subject has the choice which right they assert.[58]

Conclusion

Even if the data corpus of the project never will be published, the data subject has the right according to Art 15 to know and control what data has been processed. As a result, the data subject has the right to rectify inaccurate personal data concerning him or herself. This has to be provided for both technically and organisationally. This rectification has to be free of charge according to Art 12 (5).

2.6.4.3 Art 17 Right to Erasure

The data subject has the right to erasure where one of the following grounds apply: the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed (fulfills the principle of data minimization), the data subject withdraws consent on which the processing is based and where there is no other legal ground for the processing, the data subject objects to the processing Art 21 (1) and (2) (below) and there are no overriding legitimate grounds for the processing, the personal data have been unlawfully processed, the personal data have to be erased for compliance with a legal obligation to which the controller is subject, the personal data of children have been collected in relation to the offer of information society services. If one of these legal grounds apply, the data subject has a right to erasure.

Also regulated in Art 17 (1) is the obligation to erasure that falls to the controller and is completely independent of the right to erasure of the data subject. So, independent of the assertion of rights through the data subject, the controller is obliged to erase personal data that falls in the above mentioned criteria[59]: The personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed (which makes sense, the data subject does not has this knowledge usually[60]), the data subject withdraws consent on which the processing is based (the withdrawal of consent has therefore the erasure of the personal data as a consequence) and where there is no other legal ground for the processing, the data subject objects to the processing Art 21 (1) and (2) (below) (the objection of the data subject has therefore the erasure of the personal data as a consequence) and there are no overriding legitimate grounds for the processing, the personal data have been unlawfully processed, the personal data have to be erased for compliance with a legal obligation to which the controller is subject, the personal data of children have been collected in relation to the offer of information society services.

There is no definition of “erasure” in the GDPR. Important is the result of the erasure: The impossibility to detect the information of the deleted data.[61] After the erasure nobody should be able to detect the information of the deleted data without disproportionate effort.[62] This has to technically ensured- for example with special erasure software, or by deleting links or codes that are necessary to detect the information.[63] Insufficient are simple organisational measure like the marking of the data as erased data.[64] The obligation does not cover copies third parties have created. In this case, the controller has

[56] Herbst, Art 16 DSGVO Rz 33.
[57] ErwG 59 GDPR.
[58] Herbst, Art 16 DSGVO Rz 17.
[59] Herbst, Art 17 DSGVO Rz 2.
[60] Herbst, Art 17 DSGVO Rz 9.
[61] Herbst, Art 17 DSGVO Rz 37.
[62] Herbst, Art 17 DSGVO Rz 37.
[63] Herbst, Art 17 DSGVO Rz 38.
[64] Herbst, Art 17 DSGVO Rz 40.

to inform the recipients according to Art 19. Should the controller have any backup copies, these have to be deleted as well.[65] According to Art 12 (5) the erasure has to be free of cost. The controller has at the latest one month until he erases the data and should he refuse to fulfill this obligation he has to give the reasons.[66] Regarding the erasure obligation of the controller, the controller should plan for an erasure concept (at what time and at what intervals are the data controlled?).[67]

Where the controller has made the personal data public and is obliged to erase the personal data, the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data. This regulation is the more specific norm as opposed to Art 19.

Conclusion

For ManyLaws relevant is Art 17 (3) lit d, that regulates that the erasure obligations do not apply, to the extent that processing is necessary for scientific research purposes or statistical purposes in accordance with Article 89 (1) in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing. The reference to Art 89 (1) ensures, that the exception obliges the controller to ensure appropriate safeguards for the rights and freedoms of the data subject.[68]

The exception for scientific research requires the right referred to in paragraph 1 to render impossible or to seriously impair the achievement of the objectives of the processing. This includes the right of the data subject to erasure as well as the (independently of the data subjects right to erasure) obligation to erasure that falls to the controller.[69] Whether the obligation referred to in paragraph 1 is likely to at least seriously impair the achievement of the objectives of the processing is part of a prognosis.[70] The objectives of the processing could at least be seriously impaired by the fact that the research depends on a complete data set.[71] The impairment has in all cases to be at least seriously, a minor impairment is not enough.[72]

2.6.4.4 Art 18 Right to Restriction of Processing

The right to restriction is closely connected to the right to erasure (Art 17).[73] The aim of this regulation is to not erase the data but in the same time to not process the data.[74] The restriction of processing means the marking of stored personal data with the aim of limiting their processing in the future (Art 4(3)). Art 18 regulates cases, where the right and the obligation of the controller to erase the data is clear, but the data subject opposes the erasure because it would conflict with his or her interests, or cases in which the assessment takes time.[75]

The data subject shall have the right to obtain from the controller restriction of processing where one of the following applies: The accuracy of the personal data is contested by the data subject for a period enabling the controller to verify the accuracy of the personal data, the processing is unlawful and the data subject opposes the erasure of the personal data and requests the restriction of their use instead, the controller no longer needs the personal data for the purposes of the processing but they are required by the data subject for the establishment, exercise or defence of legal claims, the data subject has objected to processing pursuant to Article 21(1) pending the verification whether the legitimate grounds of the controller override those of the data subject. Where processing has been restricted under paragraph 1, such personal data shall, with the exception of storage, only be processed with the data subject's consent or for the establishment, exercise or

[65] Herbst, Art 17 DSGVO Rz 42.
[66] Herbst, Art 17 DSGVO Rz 46.
[67] Herbst, Art 17 DSGVO Rz 47 and for example: http://din-66398.de/.
[68] Herbst, Art 17 DSGVO Rz 81.
[69] Herbst, Art 17 DSGVO Rz 82.
[70] Herbst, Art 17 DSGVO Rz 82.
[71] Herbst, Art 17 DSGVO Rz 82.
[72] Herbst, Art 17 DSGVO Rz 82.
[73] Herbst, Art 18 DSGVO Rz 1.
[74] Herbst, Art 17 DSGVO Rz 1.
[75] Herbst, Art 17 DSGVO Rz 1.

defence of legal claims or for the protection of the rights of another natural or legal person or for reasons of important public interest of the Union or of a Member State. A data subject who has obtained restriction of processing pursuant to Paragraph 1 shall be informed by the controller before the restriction of processing is lifted.

Technical measures have to ensure, that the marked data are solely used for the purposes according to Art 18 (2).[76] These technical measures need to include the backup files. Methods by which to restrict the processing of personal data could include temporarily moving the selected data to another processing system, making the selected personal data unavailable to users, or temporarily removing published data from a website. In automated filing systems, the restriction of processing should in principle be ensured by technical means in such a manner that the personal data are not subject to further processing operations and cannot be changed. The fact that the processing of personal data is restricted should be clearly indicated in the system.[77]

2.6.4.5 Art 20 Right to Data Portability

The purpose of the regulation is to facilitate the change of provider, for example from one social network to another.[78] The aim is to prevent lock-in-effects.[79] Art 20 GDPR states, that the data subject must have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided, where: (a) the processing is based on consent or on a contract and (b) the processing is carried out by automated means.

The personal data have to concern the data subject, it is not enough that the data subject simply provided the data.[80] According to recital 68: "*Where, in a certain set of personal data, more than one data subject is concerned, the right to receive the personal data should be without prejudice to the rights and freedoms of other data subjects".* The fact that the data set includes personal data from other data subjects is therefore no reason to exclude the right to data portability.[81]

In exercising his or her right to data portability, the data subject must have the right to have the personal data transmitted directly from one controller to another, where technically feasible. That right shall not apply to processing necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller. According to Art 68: "*By its very nature, that right should not be exercised against controllers processing personal data in the exercise of their public duties. It should therefore not apply where the processing of the personal data is necessary for compliance with a legal obligation to which the controller is subject or for the performance of a task carried out in the public interest or in the exercise of an official authority vested in the controller*."

Conclusion

In exercising his or her right to data portability, the data subject must have the right to have the personal data transmitted directly from one controller to another, where technically feasible.

2.6.4.6 Art 21 Right to Object

Art 21 regulates the Right to Object against processing that is necessary for the performance of a task carried out in the public interest or necessary for the purposes of the legitimate interests pursued by the controller or by a third party, including profiling based on those provisions. Further it regulates the right to object against processing for marketing reasons, which includes profiling to the extent that it is related to such direct marketing. It also regulates the right to object where processing is done for direct marketing purposes and for scientific research purposes.

[76] Herbst, Art 17 DSGVO Rz 29.
[77] ErwG 67.
[78] Herbst, Art 20 DSGVO Rz 1.
[79] Herbst, Art 20 DSGVO Rz 2.
[80] Herbst, Art 20 DSGVO Rz 9.
[81] Herbst, Art 20 DSGVO Rz 10.

Conclusion

Art 21 (6) states that, where personal data are processed for scientific or historical research purposes or statistical purposes pursuant to Article 89(1), the data subject, on grounds relating to his or her particular situation, shall have the right to object to processing of personal data concerning him or her, unless the processing is necessary for the performance of a task carried out for reasons of public interest.

As regulated in Art 20, the right to object concerns personal data that concern the data subject, not simply data that was provided for by the data subject.[82] To claim the right to object, the data subject has to lie down the grounds relating to his or her particular situation.[83] Where a data subject claims the right to object, where personal data are processed for scientific research, and does so while stating the grounds relating to his or her particular situation, the controller is obliged to end the processing of these personal data (and delete them according to Art 17), unless the processing is necessary for the performance of a task carried out for reasons of public interest.[84] This right to object has at the latest at the time of the first communication with the data subject, be explicitly brought to the attention of the data subject and must be presented clearly and separately from any other information. Similar information obligations are regulated in Art 13 and 14. In the context of the use of information society services (and notwithstanding Directive 2002/58/EC), the data subject may exercise his or her right to object by automated means using technical specifications (for example do-not-track-settings).[85]

2.6.4.7 Art 22 Automated Individual Decision-making including Profiling

The title of this article might be confusing, since profiling is a form of processing whereas automated individual decision-making builds on the act of processing and concludes decisions during this act but does not include any processing on its own.[86] As a result, Art 22 does not regulate a form of processing as much, but the handling of results of that processing.[87]

According to Recital 71: "*The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention. Such processing includes 'profiling' that consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her.*"

Therefore, Art 22 (1) regulates the right of the data subject to not be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her. "Which produces legal effects" means a change of the legal position for the data subject: entering a contract, the concrete form and arrangement of the contract etc.[88] "Or similarly significantly affects" means a serious disruption in the economical or personal development of the data subject: termination of a loan, not entering a contract, etc.

Recital 71 further states: "*However, decision-making based on such processing, including profiling, should be allowed where expressly authorised by Union or Member State law to which the controller is subject, including for fraud and tax-evasion monitoring and prevention purposes conducted in accordance with the regulations, standards and recommendations of Union institutions or national oversight bodies and to ensure the security and reliability of a service provided by the controller, or necessary for the entering or performance of a contract between the data subject and a controller, or when the data subject has given his or her explicit consent.*" Therefore, Art 22 (2) states that Paragraph 1 shall not apply if the decision:

(a) is necessary for entering into, or performance of, a contract between the data subject and a data controller;

[82] Herbst, Art 21 DSGVO Rz 48.
[83] Herbst, Art 21 DSGVO Rz 49.
[84] Herbst, Art 21 DSGVO Rz 52.
[85] Herbst, Art 21 DSGVO Rz 43.
[86] Buchner, Art 22 DSGVO Rz 3.
[87] Buchner, Art 22 DSGVO Rz 11.
[88] Buchner, Art 22 DSGVO Rz 24.

(b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or

(c) is based on the data subject's explicit consent.

Regarding special safeguards in case the automated individual decision-making is allowed Recital 71 states: "*In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision. Such measure should not concern a child. In order to ensure fair and transparent processing in respect of the data subject, taking into account the specific circumstances and context in which the personal data are processed, the controller should use appropriate mathematical or statistical procedures for the profiling, implement technical and organisational measures appropriate to ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised, secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject and that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect. Automated decision-making and profiling based on special categories of personal data should be allowed only under specific conditions.*"

Art 22, therefore, further regulates that in the cases referred to in points (a) and (c) of Paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision. And, decisions referred to in Paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place. Further information obligations in this case are regulated in Art 13 (2) lit f, Art 14 (2) lit g and Art 15 (1) lit h. Furthermore according to Art 35 (3) lit a a data impact assessment has to be carried out.[89]

Conclusion

If ManyLaws aims to use the results of the processing of personal data in a way that produces legal effects or similarly significantly affects for the data subject, and the exceptions of Art 22 (2) do not apply, the data subject has the right to not be subject to such a decision.

2.6.5 Obligations

2.6.5.1 Art 24 Responsibility of the Controller

Taking into account the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons, the controller has to implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with the GDPR. Risk should be evaluated on the basis of an objective assessment, by which it is established whether data processing operations involve a risk or a high risk.[90] These measures could also include the implementation of appropriate data protection policies. Approved codes of conduct or approved certification mechanisms could also be used to demonstrate compliance with the obligations. Art 24 serves, as a general rule, that is concretised in Art 25, 32 and 35.[91]

2.6.5.2 Art 25 Data Protection by Design and by Default

Data Protection by Design

Taking into account the state-of-the-art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the

[89] Buchner, Art 22 DSGVO Rz 3.
[90] Recital 76 GDPR.
[91] Hartung, Art 24 DSGVO Rz 1.

processing, the controller has to implement appropriate technical and organisational measures and to integrate the necessary safeguards into the processing.

Such measures could consist of minimising the processing of personal data, pseudonymising personal data as soon as possible, transparency with regard to the functions and processing of personal data, enabling the data subject to monitor the data processing, enabling the controller to create and improve security features.[92] These measures should be designed to implement data-protection principles (Art 5 (1)), such as data minimisation, in an effective manner. As already shown by Art 24, appropriate technical and organisational measures have to be implemented. The intention of these measures in Art 24 relate to security, while Art 25 demands these measures regarding the data protection principles such as data minimisation.[93] The controller has to implement these technical and organisational measures both at the time of the determination of the means for processing and at the time of the processing itself.

Conclusion

A concrete guidance, methodology or instruments for the implementation of Data Protection by Design are not offered in the GDPR. However, there are generally accepted models and methods regarding Data Protection by Design.

Data Protection by Default

The controller has to implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. Therefore it is not sufficient to simply offer options to the data subject, on the contrary, those options have to be preset in a data protection friendly way.[94] If and how the preset options could be changed by the data subject are not regulated by the regulation.[95] That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility.

2.6.5.3 Art 32 Security of Processing

Taking into account the state of the art (all technical measures that are marketable[96]), the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security (data security and IT-security[97]) appropriate to the risk, including:

(a) The pseudonymisation and encryption of personal data;

(b) The ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;

(c) The ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;

(d) A process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

In assessing the appropriate level of security, the risks have to be considered that are presented by processing, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed. The controller and processor have to take steps to ensure that any person acting under their authority, who has access to personal data does not process them except on instructions from them.

[92] Recital 78 GDPR.
[93] Hartung, Art 25 DSGVO Rz 15.
[94] Hartung, Art 25 DSGVO Rz 24.
[95] Hartung, Art 25 DSGVO Rz 26.
[96] Jandt, Art 32 DSGVO Rz 10.
[97] Jandt, Art 32 DSGVO Rz 3.

2.6.5.4 Art 33 Notification of a Personal Data Breach to the Supervisory Authority

In the case of a personal data breach within the scope of the controller (so not in case the data subject fell prey to phishing for example[98]), the controller has to notify the personal data breach to the competent supervisory authority (Art 55: In general the supervisory authority where the data was processed) within 72 hours after having become aware of it unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons. The processor also has to notify the controller without undue delay in case of a personal data breach. This notification to the controller has to at least contain:

(a) Describe the nature of the personal data breach including where possible, the categories and approximate number of data subjects concerned and the categories and approximate number of personal data records concerned;

(b) Communicate the name and contact details of the data protection officer or other contact point where more information can be obtained;

(c) Describe the likely consequences of the personal data breach;

(d) Describe the measures taken or proposed to be taken by the controller to address the personal data breach, including, where appropriate, measures to mitigate its possible adverse effects.

- Where, and in so far as, it is not possible to provide the information at the same time, the information may be provided in phases without undue further delay.
- The controller shall document any personal data breaches, comprising the facts relating to the personal data breach, its effects and the remedial action taken. That documentation shall enable the supervisory authority to verify compliance with this Article.

2.6.5.5 Art 34 Communication of a Personal Data Breach to the Data Subject

Art 33 regulates the notification of a personal data breach to the supervisory authority and Art 34 the communication to the data subject when the data breach is likely to result in a high risk to the rights and freedom of natural persons. The communication to the data subjects aims at making the situation more transparent.[99] Paragraph 1 regulates the conditions for the communication to the data subject: When the personal data breach is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall communicate the personal data breach to the data subject without undue delay. Whether or not the data breach is likely to result in a high risk to the rights and freedom of natural persons, the controller has to decide via a prognosis.[100] This prognosis could be verified by the supervisory authority.[101]

Paragraph 2 describes the requirements of the communication: The communication to the data subject must describe in clear and plain language the nature of the personal data breach and contain at least the information and measures referred to in points (b), (c) and (d) of Article 33 (3) (name and contact details of the data protection officer or other contact point where more information can be obtained, description of the likely consequences of the personal data breach, description of the measures taken or proposed to be taken by the controller to address the personal data breach, including, where appropriate, measures to mitigate its possible adverse effects). Paragraph 3 regulates the exceptions from the communication: The communication to the data subject referred to in Paragraph 1 shall not be required if any of the following conditions are met:

(a) The controller has implemented appropriate technical and organisational protection measures, and those measures were applied to the personal data affected by the personal data breach, in particular those that render the personal data unintelligible to any person who is not authorised to access it, such as encryption;

(b) The controller has taken subsequent measures which ensure that the high risk to the rights and freedoms of data subjects referred to in Paragraph 1 is no longer likely to materialise;

[98] Jandt, Art 32 DSGVO Rz 8.
[99] Jandt, Art 34 DSGVO Rz 1.
[100] Jandt, Art 34 DSGVO Rz 5.
[101] Jandt, Art 34 DSGVO Rz 5.

(c) It would involve disproportionate effort. In such a case, there shall instead be a public communication or similar measure whereby the data subjects are informed in an equally effective manner.

Paragraph 4 regulates the notice of the supervisory authority: If the controller has not already communicated the personal data breach to the data subject, the supervisory authority, having considered the likelihood of the personal data breach resulting in a high risk, may require it to do so or may decide that any of the conditions referred to in Paragraph 3 are met.

Conclusion

In case of a personal data breach and when the data breach is likely to result in a high risk to the rights and freedom of natural persons not only the supervisory authority has to be notified but the incident must potentially also be communicated to the affected data subject.

To make use of the exceptions, it would be advisable, to oblige to certain rules, before the breach such as stated above: Implement appropriate technical and organisational protection measures to the personal data, in particular those that render the personal data unintelligible to any person who is not authorised to access it, such as encryption (even if it were exactly those safety measures that failed[102]), or to ensure that subsequent measures are planned for which ensure that the high risk to the rights and freedoms of data subjects is no longer likely to materialise.

2.6.5.6 Information to be Provided Where Personal Data are Collected from the Data Subject

Art 13 regulates the information obligations in case of personal data collection from the data subject, whereas Art 14 regulates the information obligations, where these data are not obtained from the data subject. The personal data are not obtained from the data subject if these data are obtained from a third source. This also applies, if the data subject publishes personal data (via Social Media).[103] Even then, there is no personal contact between the controller and the data subject and Art 13 does not apply (but Art 14).[104]

Should data be collected from the data subject, the following information have to be provided: the identity and the contact details of the controller, the identity and contact details of the data protection officer (if there is one), the purposes of the processing for which the personal data are intended (complete and detailed to a degree, that the data subject has an idea of the data processing) as well as the legal basis for the processing, the foreseeable recipients or categories of recipients of the personal data (for example subsidiary companies), where applicable, the fact that the controller intends to transfer personal data to a third country or international organisation and the existence or absence of an adequacy decision by the Commission, the period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period, the existence of the right to request from the controller access to and rectification or erasure of personal data or restriction of processing concerning the data subject or to object to processing as well as the right to data portability, where the processing is based on consent, the existence of the right to withdraw consent at any time, without affecting the lawfulness of processing based on consent before its withdrawal, the right to lodge a complaint with a supervisory authority, whether the provision of personal data is a statutory or contractual requirement, or a requirement necessary to enter into a contract, as well as whether the data subject is obliged to provide the personal data and of the possible consequences of failure to provide such data, the existence of automated decision-making, including profiling, and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.[105]

This information will have to be delivered prior to the processing. Should the purposes of the processing change at a later time (for example the stored data will be used for further yet not defined processing), the controller is prior to that further processing obliged to provide the data subject with information on that other purpose and with any relevant further

[102] Jandt, Art 34 DSGVO Rz 14.
[103] Bäcker, Art 12 DSGVO Rz 16.
[104] Bäcker, Art 12 DSGVO Rz 16.
[105] Art 13 GDPR; Bäcker, Art 12 DSGVO Rz 16-55.

information.[106] Those information obligations do not exist where and insofar as the data subject already has the information.[107]

Conclusion

Form and display of the information obligation: According to Art 12 (1), (7) and (8) the information have to be easily accessible without media breaks. It is not enough to simply passively provide the information on the webpage, the information has to be provided actively to the data subject.[108] It is sufficient to point out that the information on the website to the data subject, as long as the data subject has the secure possibility to access that information before the processing.[109] There is no research exception from these obligations, but according to Art 23 (4) the member states have the possibility to limit these information obligations.

2.6.5.7 Art 14 Information to be Provided Where Personal Data Have Not Been Obtained from the Data Subject

Art 14 regulates the information obligations, where these data are not obtained from the data subject. Should data be collected from the data subject, as regulated in Art 13 the following information have to be provided: the identity and the contact details of the controller, the identity and contact details of the data protection officer (if there is one), the purposes of the processing for which the personal data are intended (complete and detailed to a degree, that the data subject has an idea of the data processing) as well as the legal basis for the processing, the categories of personal data concerned (not the processed data themselves, just the data categories[110]), the foreseeable recipients or categories of recipients of the personal data (for example subsidiary companies), where applicable, the fact that the controller intends to transfer personal data to a third country or international organisation and the existence or absence of an adequacy decision by the Commission, the period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period, the existence of the right to request from the controller access to and rectification or erasure of personal data or restriction of processing concerning the data subject or to object to processing as well as the right to data portability, where the processing is based on consent, the existence of the right to withdraw consent at any time, without affecting the lawfulness of processing based on consent before its withdrawal, the right to lodge a complaint with a supervisory authority and the existence of automated decision-making, including profiling, and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.[111]

Art 14 further regulates the following information obligations: From which source the personal data originate, and if applicable, whether it came from publicly accessible sources. As a general rule, the controller shall provide the information referred to above within a reasonable period after obtaining the personal data, but at the latest within one month, having regard to the specific circumstances in which the personal data are processed. If the personal data processing has its legal ground in the consent of the data subject, those information have to be provided for before consent is given.[112] The controller has to provide this information actively and in a manner, that the data subject has effective and timely access to the information. Where the controller intends to further process the personal data for a purpose other than that for which the personal data were obtained, the controller shall provide the data subject prior to that further processing with information on that other purpose and with any relevant further information as referred to above.

These information obligations do not apply: Where and insofar as the data subject already has the information. Further, these information obligations do not apply, where the provision of such information proves impossible. For example if the controller doesn't know the data subject, the controller is not obliged to research contact information of the data subject.[113] On the other hand, if the controller identifies the data subject, the fact that he has to research the contact details alones is

[106] Art 13 (3) GDPR.
[107] Art 13 (4).
[108] Bäcker, Art 12 DSGVO Rz 59.
[109] Bäcker, Art 12 DSGVO Rz 59.
[110] Bäcker, Art 14 DSGVO Rz 58.
[111] Art 14 GDPR.
[112] Bäcker, Art 14 DSGVO Rz 32.
[113] Bäcker, Art 14 DSGVO Rz 54.

not enough to render the provision of such information impossible.[114] Further, these information obligations do not apply, where this would involve a disproportionate effort. For all of these Art 14 exceptions from the information obligations, the controller has to take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available. Making the information publicly available is often an appropriate alternative, especially if individually informing the data subjects is impossible.[115]

Conclusion

These information obligations do not apply for processing for scientific research purposes, subject to the conditions and safeguards referred to in Article 89(1). According to Art 23 member states have the possibility to limit these information obligations.

2.6.5.8 Notification Obligation Regarding Rectification or Erasure of Personal Data or Restriction of Processing.

If the controller is obliged to rectify, erase, or restrict the processing of personal data, and should the controller share the personal data with other recipients[116] (for example through dissemination[117]), he has to communicate any rectification or erasure of personal data or restriction of processing to each recipient to whom the personal data have been disclosed, unless this proves impossible or involves disproportionate effort (the amount of effort for the notification and the amount of the data and their value for the data subject are to be taken into consideration[118]). This allows the recipient to fulfil their duty and the data subject to see their rights through.[119] Those notifications have to be free of charge according to Art 12 (5) GDPR. The controller must inform the data subject about the recipients if the data subject requests it.

Conclusion

Should the controller of the personal data have shared those with other recipients and a data subject assert their rights, the controller has to notify the recipients of this fact.

2.6.6 Research Privileges

Since the definition of 'scientific research' has a tremendous impact on the extent to which personal data may be processed, this chapter will start with the definition of 'scientific research'. 'Scientific research' is not defined in the GDPR. However Recital 159 states: *"Where personal data are processed for scientific research purposes, this Regulation should also apply to that processing. For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner [...]"*[120]. Scientific research should, therefore, be interpreted in a broad manner. The Article 29 Working Party (WP29), on the other hand, argues that the term should not be stretched further than its common meaning and "*understands that 'scientific research' in this context means a research project set up in accordance with relevant sector-related methodological and ethical standards, in conformity with good practice.*"[121] According to Recital 159 this is to be understood to be "*including for example technological development and demonstration, fundamental research, applied research and privately funded research.*"[122] It does not depend on whether a university or a private research organisation undertakes the research[123].

---

[114] Bäcker, Art 14 DSGVO Rz 54.
[115] Bäcker, Art 14 DSGVO Rz 62.
[116] According to Art 4 (9) GDPR a recipient "means a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not."
[117] *Herbst*, Art 19 DSGVO, in *Kühling/Buchner* (Hg), Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO/BDSG. Kommentar (2017) Rz 6.
[118] Herbst, Art 19 DSGVO Rz 9.
[119] Herbst, Art 19 DSGVO Rz 1 and 12.
[120] Recital 159 GDPR.
[121] Article 29 Working Party Guidelines on consent under Regulation 2016/679, adopted on 28 November 2017 as last revised and adopted on 10 April 2018, WP 259 rev. 01, 28.
[122] Recital 159 GDPR.
[123] Feiler/Forgó, EU-DSGVO. Kurzkommentar (2017) Art 89 Rz 3.

Since ManyLaws is a research project set up in accordance with the relevant sectors related standards and concerns itself with not only technological development, the mining activities of the ManyLaws project are to subsumed as 'scientific research'. As a consequence, the GDPR privileges for research apply to the ManyLaws project. The fact that the processing of personal data is part of a research project, does not make the processing per se legal.[124] Furthermore, the controller is responsible for and must be able to demonstrate compliance with the general GDPR principles according to Art 5 (2).[125]

The research privileges include in short[126]:

- regarding the purpose limitation (Art 5 (1) lit b) "f*urther processing for [...] scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes[...]*"[127];

- regarding the storage limitations (Art 5 (1) lit e) "*personal data may be stored for longer periods insofar as the personal data will be processed solely for [...] scientific or historical research purposes or statistical purposes in accordance with Article 89(1) [...]*";

- regarding the processing of sensible data (Art 9 (2) lit j) "*processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.*";

- regarding the information obligations (Art 14 (5) lit b) *"the provision of such information proves impossible or would involve a disproportionate effort, in particular for processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, subject to the conditions and safeguards referred to in Article 89(1) or in so far as the obligation referred to in paragraph 1 of this Article is likely to render impossible or seriously impair the achievement of the objectives of that processing. In such cases the controller shall take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available*";

- regarding the Right to Erasure (Art 17 (3) lit d) "*for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing*";

- Art 89 (2): "*Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.*";

- Art 89 (3): "*Where personal data are processed for archiving purposes in the public interest, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18, 19, 20 and 21 subject to the conditions and*

[124] Buchner/Tinnefeld, Art 89 DSGVO, in *Kühling/Buchner* (Hg), Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO/BDSG. Kommentar (2017) Rz 1-2.
[125] According to Art 4 (7) 'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law; According to Art 4 (8) processor' means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.
[126] Buchner/Tinnefeld, Art 89 DSGVO, in *Kühling/Buchner* (Hg), Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO/BDSG. Kommentar (2017) Rz 2.
[127] Art 5 GDPR.

safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes."

According to Art 89 (1) all of these above mentioned privileges are subject "*to appropriate safeguards for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner*". However, Art 89 (1) provides no new regulations since they repeat those which have been stated (e.g. data minimisation or privacy by design).[128]

Conclusion

Anonymization

According to Art 89, for technical and organisational measures one has to check first whether or not the purposes of the processing can be fulfilled with anonymized data.[129] According to Recital 26, the principles of data protection do not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. Regarding anonymized data, one has to make sure that re-identification in combination with other data sets are not a possibility.[130]

Pseudonymisation

If anonymization is not the case, the safeguards and measures in Art 89 (1) have to be fulfilled. This could include pseudonymisation[131] if the purposes can be fulfilled in that manner. According to Art 4 (5) GDPR pseudonymisation means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. Art 89 aims to ensure the principle of data minimization also in scientific research. Further it wants to ensure the principles regarding the purpose and integrity and confidentiality (Art 5 (1) lit b and f) which could be provided for by encryption or controlled access to the data.[132]

2.6.7 Data Protection Impact Assessment

Where a type of processing in particular using new technologies, and taking into account the nature, scope, context and purposes of the processing, is likely to result in a high risk to the rights and freedoms of natural persons, the controller has

[128] Buchner/Tinnefeld, Art 89 DSGVO, in *Kühling/Buchner* (Hg), Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO/BDSG. Kommentar (2017) Rz 3.
[129] Buchner/Tinnefeld, Art 89 DSGVO Rz 17.
[130] Buchner/Tinnefeld, Art 89 DSGVO Rz 17.
[131] According to Art 4 (5) GDPR pseudonymisation' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.
[132] Buchner/Tinnefeld, Art 89 DSGVO Rz 19.

to, prior to the processing, carry out an assessment of the impact of the envisaged processing operations on the protection of personal data. A single assessment may address a set of similar processing operations that present similar high risks.

A data protection impact assessment has to be carried out in the case of:

(a) A systematic and extensive evaluation of personal aspects relating to natural persons which is based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;

(b) Processing on a large scale of special categories of data, or of personal data relating to criminal convictions and offences referred to in Article 10; or

(c) A systematic monitoring of a publicly accessible area on a large scale.

The assessment has to contain at least:

(a) A systematic description of the envisaged processing operations and the purposes of the processing, including, where applicable, the legitimate interest pursued by the controller;

(b) An assessment of the necessity and proportionality of the processing operations in relation to the purposes;

(c) An assessment of the risks to the rights and freedoms of data subjects and the measures envisaged to address the risks, including safeguards, security measures and mechanisms to ensure the protection of personal data.

Compliance with approved codes of conduct by the relevant controllers or processors are taken into account in assessing the impact of the processing operations performed by such controllers or processors.

3. MANYLAWS PORTAL GDPR COMPLIANCE

The ManyLaws portal will be hosted on UAEGEAN servers which follows specific policies that will be described below. UAEGEAN collects and processes legal open data legally, honestly and transparently, in accordance with the General Data Protection Regulation (GDPR). UAEGEAN limits collection of personal data to what is strictly necessary, in accordance with the principle of data minimisation. The ManyLaws portal will also use cookies that enhance the website use by remembering things, so a cookies policy will be provided to the users as well. In addition, ManyLaws portal will use GDPR ckan extension (ckanext-gdpr) in order to block the access to anonymous user information.

3.1 User Profiling (required/optional fields for logging users based on type)

The ManyLaws portal will provide personalised services based on user type for all the registered users. For that reason, the required fields in the signup form may vary between the different user types. In any case, the ManyLaws portal will mark only the necessary data as required.

More specifically, the required personal data for each user type have formed as follows:

1. Citizen - Username, e-mail address, password and country will be enough for the citizens.
2. Lawyer - Username, e-mail address, password, country and occupation will be required by the lawyers.
3. Business Person - Username, e-mail address, password, country, occupation and business sector will be required by the business people.
4. Legal Administrator - Username, e-mail address, password, country and occupation will be required by the Legal Administrators.
5. Parliamentary Administrator - Username, e-mail address, password, country and occupation will be required by the Parliamentary Administrators.

The country is a required field for all the registered users, because the system has to match every user with at least one national legal framework. Of course, the user will be able to select any of the available national legal frameworks when using the portal.

3.2 Privacy Policy

The ManyLaws portal will be hosted on UAEGEAN servers which follows the following privacy policy:

3.2.1 Information and Rights of Data Subjects

UAEGEAN hereby informs you clearly about how it processes personal data as part of its business activity, how data are collected, used and protected. UAEGEAN is available to provide data subjects any information about processing carried out as part of the Project. For any request or complaint about the processing of personal data, it is possible to contact UAEGEAN at this e-mail address: manylaws@gmail.com.

In particular, any data subject has the right to ask UAEGEAN: for access to the personal data supplied; to correct the data; to object to the processing, when such processing is based on UAEGEAN's legitimate interest and given the particular situation of the data subject; or to exercise their right to portability of their information.

On the Right to Portability, UAEGEAN offers you the option to return all the personal data about a subject, at their express request. The data subject is thus guaranteed better control of their data and retains the possibility of reusing them. These data will be supplied in an open and easily reusable format, directly into the hands of the other data controller when desired and when technically possible. For it to be accepted, the request message must be accompanied by proof of identity.

Any person receiving the newsletter has the option to unsubscribe, unless this person is bound to receive the aforementioned newsletter under its obligations to a Partner. UAEGEAN ensures an effective unsubscribe link is provided in the Project newsletter.

The Privacy Policy may be modified by the UAEGEAN at any time, particularly to comply with regulatory, case law, editorial or technical developments. Before browsing, the data subject should refer to the latest online version on the Site or sent electronically.

3.2.2 Data Collected and Purposes of Processing

As part of carrying out the project, the Partners transfer personal data to UAEGEAN making it possible to identify and contact (first & last name, business e-mail address, photograph) their employees due to their job titles or third-parties involved in the Project, such as experts (hereafter designated the 'Partners Data'). In this case, the Partner remains responsible for supplying the legal information to the people involved in the processing operations prior to or when the data are collected.

The purpose of processing Partners' Data is: to compile files on members of the Consortium and people likely to contribute to the Project due to their job titles or expertise; UAEGEAN manages, monitors and guides the Project in fulfilment of its obligations to the Consortium; communication on the Project; sending a newsletter and information about events related to the Project; or compiling statistics related to the Project.

Further, as part of using the Site, UAEGEAN may collect the following data categories about Users of the Site directly from data subjects (hereafter designated the 'Users Data'):

1. Identification data (first & last name, telephone number, e-mail address), relating to professional life (organisation, job title), the purpose of which is: To respond appropriately to persons wishing to join UAEGEAN by completing the Site contact form; Sending newsletters, as long as the User ticks the box provided to express their acceptance. Any data collection form states the objectives of gathering these data (purposes) and if these data are compulsory or optional to administer the request. This Confidentiality Policy is freely accessible to the User, who should read it before sending their data to UAEGEAN.
2. UAEGEAN follows a standard procedure of using log files. These files log visitors when they visit websites. The information collected by log files includes internet protocol (IP) addresses, browser type, Internet Service Provider (ISP), date and time stamp, referring/exit pages, and possibly the number of clicks.

The information is used for analysing trends, administering the site, tracking users' movement on the website, and gathering demographic information.

3.2.3 Recipients of Data

UAEGEAN complies with the legal rules that could prevent, limit or govern the distribution or processing of information or data, and particularly the GDPR. Personal data are collected by UAEGEAN for internal use only. Under no circumstances will these data be sold, transferred or communicated to third parties under conditions not specified herein.

Based on legal obligations, the personal data of data subjects may be disclosed pursuant to a law, a regulation or in accordance with a decision by a competent regulatory or legal authority. The personal data collected may be added to UAEGEAN's database. It may be passed to third parties after being anonymised, solely for statistical purposes. If your personal data are communicated to a third party, UAEGEAN will ensure that the third party is bound to apply conditions of confidentiality at least identical to those herein.

The Website uses Social Plugins (Plugins) of different social networks like Twitter/LinkedIn/YouTube/Google+, which are marked by their logos. When you open a site with these plugins, browser data are transferred to the owners of those social networks. Thus, they are informed that you have opened the respective site. If the User is logged in to one or several of these social networks, these data can be linked to his/her account. The IP address may be logged without being a member of a social network or logging into a network.

Please note that data processing takes place outside the European Union. We have no influence over and take no responsibility for the volume of data transferred through social plugins. For more information, please look up the data security policies of the respective network:

http://www.twitter.com/privacy

http://www.google.com/intl/de/policies/privacy/

https://www.linkedin.com/legal/privacy-policy

https://www.youtube.com/static?template=privacy_guidelines

By logging out of networks before surfing or using privacy protecting software, you can limit the amount of data transferred.

3.2.4 Storage of Data

Personal data processed are not stored longer than necessary to fulfil the obligations defined by the Consortium. Beyond this period, the data will be anonymised and stored solely for statistical purposes and will not be used in any other way whatsoever.

More generally, data purging procedures are implemented to plan their effective deletion as soon as the storage or archiving duration necessary to fulfilling the predetermined or required purposes is reached. Any person not actively involved in relation to the processing purposes described over a three-year period will have their data deleted.

Cookies are stored for a maximum period of 13 months after they are first saved on the user's computer, corresponding to the validity period of User consent to use these Cookies. The lifetime of cookies is not extended by each visit. The user's consent must, therefore, be renewed at the end of this period.

3.2.5 Procedures In Case of Security Breach Detected by UAEGEAN

UAEGEAN undertakes to implement all appropriate technical and organisational measures using physical and logistical security resources to guarantee an adequate security level to meet risks of accidental unauthorised or illegal access, disclosure, alteration, loss or destruction of the personal data collected. However, the UAEGEAN cannot guarantee that all risks of data misuse are eliminated. In the event were the UAEGEAN become aware of illegal access to personal data stored on its servers or those of its contractors, or unauthorised access resulting in the risks identified above being realised, UAEGEAN undertakes to:

• Notify the person affected by the incident as quickly as possible, if this meets a legal requirement;
• Investigate the causes of the incident;
• Take the necessary reasonable measures to reduce the negative effects and harm that could arise from the aforesaid incident.

Under no circumstances may the commitments defined above relating to notification in the event of a security breach be considered as any form of acknowledgement by UAEGEAN of fault or liability for the occurrence of the incident in question.

3.3 ManyLaws Cookies Policy

3.3.1 What Are Cookies?

A cookie is a small text file that a website saves on your computer or mobile device when you visit the site. It enables the website to remember your actions and preferences (such as login, language, font size and other display preferences) over a period of time, so you don't have to keep re-entering them whenever you come back to the site or browse from one page to another.

3.3.2 Our Cookies Policy

Our website uses cookies to manage sessions, provide personalized web pages and adapt advertising and other content to reflect your particular needs and interests. Further, we may use cookies to generate anonymous, cumulative statistics which enable us to understand how the public uses our website and help us improve their structure and content. The information collected does not allow us to identify you. You can change your browser settings to reject some or all cookies, except cookies which are absolutely necessary. You need to know that certain features are available only if cookies are used. Choosing to reject cookies may make these features unavailable.

3.3.3 Absolutely Necessary Cookies

Absolutely necessary cookies are critical for the proper functioning of your website, allow you to browse and use its features, such as access to secure areas (see Table 1). These cookies do not identify you personally. Without these cookies, we cannot provide an efficiently functioning website.

For this particular category of cookies we do not ask for your specific consent. According to the relevant legal framework (paragraph 5 of Article 4 of Law 3471/2006), you cannot choose whether to reject the installation of these cookies, as without them it would not be technically possible to provide services from the ManyLaws website.

Cookie Name	Source	Usability
Session cookie	manylaws.eu	This cookie is saved only for authenticated users after they login and is deleted when they logout.
Persistent login cookie	manylaws.eu	These cookies remain on your hard drive until you erase them or they expire. How long a cookie remains on your browser depends on how long the visited website has programmed the cookie to last (more on persistent cookies).
cookie-agreed	manylaws.eu	These cookies are used to remember a user's choice about cookies on www.manylaws.eu. Where users have previously indicated a preference, that user's preference will be stored in these cookies. They have a duration of 100 days.
cookie-extras-agreed	manylaws.eu	
has-js	manylaws.eu	This cookie is saved for all users of www.manylaws.eu for the duration of their session. It lasts until the user closes the tab or the browser and allows the webpage to know whether the user's browser supports the programming language Javascript.

Table 1: Absolutely Necessary Cookies

3.3.4 Performance Cookies (Third Party)

Performance cookies are small files sent from our web server to your web browser that ask for permission to be placed on your hard drive to gather and remember information about your browsing preferences. The information gathered is anonymous. The third party performance cookies we use is 'Google Analytics'. Google Analytics cookies determine the number of unique visitors to our Website, provide consistent experiences for our users, and optimise web usage. Google Analytics cookies may differ from our own Website cookies in the duration for which they collect information.

Cookie Name	Source	Expiration Time	Description
_ga	Google	2 years	Used to distinguish users.
_gid	Google	24 hours	Used to distinguish users.
_gat	Google	1 minute	Used to throttle request rate. If Google Analytics is deployed via Google Tag Manager, this cookie will be named _dc_gtm_<property-id>.

Table 2: Performance Cookies (Third Party)

3.3.5 Google Analytics

This website uses Google Analytics, a web analytics service provided by Google, Inc. ("Google"). Google Analytics uses "cookies", which are text files placed on your computer to help the website analyse how visitors use the site. The information generated by the cookie about your use of the website (including your IP address) will be transmitted to and stored by Google on servers in the United States. Google will use this information for the purpose of evaluating your use of the website, compiling reports on website activity for website operators and providing other services relating to website activity and internet usage. Google may also transfer this information to third parties where required to do so by law, or where such third parties process the information on Google's behalf. Google will not associate your IP address with any other data held by Google. You may refuse the use of cookies by selecting the appropriate settings on your browser, however please note that if you do this you may not be able to use the full functionality of this website. By using this website, you consent to the processing of data about you by Google in the manner and for the purposes set out above.

You can prevent Google's collection and use of data (cookies and IP address) by downloading and installing the browser plug-in available under https://tools.google.com/dlpage/gaoptout?hl=en.

Please note that this website initializes Google Analytics with the setting anonymizeIp. This guarantees anonymized data collection by masking the last part of your IP address.

More information about Google Analytics' terms and conditions of use and data privacy.

3.3.6 Google Maps Cookies

The embedded Google Map on our site sets three cookies known as NID, PREF and khcookie. These cookies allow Google Maps to remember which browser you are using and what preferences you have set when you view maps.

For more information about Google Maps terms and conditions of use and data privacy visit https://cloud.google.com/maps-platform/terms/.

3.3.7 YouTube Cookies

We embed videos from our official YouTube channel using YouTube's privacy-enhanced mode. This mode may set cookies on your computer once you click on the YouTube video player, but YouTube will not store personally-identifiable cookie information for playbacks of embedded videos using the privacy-enhanced mode. To find out more please visit YouTube's embedding videos information page.

More information about YouTube terms and conditions of use and data privacy visit https://support.google.com/youtube/answer/171780?hl=el-GR.

3.3.8 Automatically Collected Information (Log Files)

Information that is automatically collected as you navigate through our site (log files) includes Internet Protocol (IP) address, geolocation information, unique device identifiers, browser type and language, time and date of access, referring/exit pages, time spent on pages, keywords used to find our site via search engines and other information of this nature. This information is anonymous and cannot be directly linked to individual users. We use this information to diagnose problems with our server and in conjunction with the Google Analytics website statistics package to analyse trends in how our website is accessed and utilized.

3.3.9 How Do I Change My Cookie Settings?

You can change your cookie preferences at any time by navigating to http://www.manylaws.eu. You can then adjust the available sliders to 'On' or 'Off', then clicking 'Save my cookie preferences'. You may need to refresh your page for your settings to take effect. Alternatively, most web browsers allow some control of most cookies through the browser settings. Find out how to manage cookies on popular browsers:

- Google Chrome
- Microsoft Edge
- Mozilla Firefox
- Microsoft Internet Explorer
- Opera
- Apple Safari

3.4 Terms of Use

3.4.1 Disclaimer

The content of the website is maintained by the UAEGEAN Research Unit (INEA-CEF grant agreement No 156047). The views expressed on this website reflect the views of the editors. The INEA is not liable for its content and the use that may be made of the information contained herein.

Due to the general informational character of the website, no liability can be assumed for the topicality, correctness, completeness, quality or constant availability of the information provided. The editors accept no liability for any damage caused by the use of the information provided.

3.4.2 Publisher/ Responsible Editor

University of the Aegean – Research Unit

UAEGEAN

Administration Building, University Hill, Mytilene,

GR-81100, North Aegean, Greece

3.4.3 Hosting Company

University of the Aegean – Research Unit

UAEGEAN

Administration Building, University Hill, Mytilene,

GR-81100, North Aegean, Greece

4. ETHICAL CONSIDERATIONS OF ADVANCED COMPUTER SYSTEM DESIGN AND DEVELOPMENT

Over the past three decades, computer hardware and software have become increasingly integral parts of every aspect of human activity. During this exponential growth in the size, quantity, and sophistication of technological artefacts, while significant attention has been focused on the consequences of technology use – good and bad – by actors within society, due diligence has not been paid to the impacts of human actions undertaken, and decisions made during the design and deployment phases of these systems, within complex human environments. This is also true for the field of Computer Ethics, which has until recently been mostly concerned with moral issues relating to the use, regulation, and social implications of new information technology; and has largely ignored the design of computer systems[133]. Pierce & Henry define Computer Ethics as referring to "*...a set of rules or principles used for moral decision making regarding computer technology and computer use.*"[134] It is now increasingly recognised that software design and development can have subtle yet crucial impacts on people and their social, cultural and organisational contexts[135]. Increased emphasis is being placed on the values and ethics associated with digital technologies located within complex socio-technical systems as software designers and programmers become aware of the potential negative by-products of their work in the form of misinformation campaigns, online harassment, exclusionary tools, and biased algorithms[136]. Dutton (2014) argues that impeccable technical visions will not lead inexorably to successful public and private infrastructures that automatically support an increased quality of everyday life; indeed, core values as privacy, equality, trust and individual choice could be gravely undermined if technological networks - including the much-vaunted Internet of Things (IoT) - are not designed, implemented and governed in appropriate ways[137].

In this regard, the proposed ManyLaws system may be considered as an *emerging* technology, in contrast to an existing or *entrenched* one[138]; characterised by the application of advanced and innovative techniques, and being still under development with its larger socio-economic impact being as yet largely unknown. This makes the ethical dilemmas associated with its design and deployment a function of *uncertainty*, and consequently difficult to ascertain, according to Sollie (2007) when the final outcome of opaque and multi-agency Research & Development trajectories and finished products of a technology are not fully known, as neither are their uses and broader consequences[139]. Ethicists working on emerging technologies often distinguish between extrinsic and intrinsic concerns regarding them[140]. According to Sandler (2014), *extrinsic concerns* pertain to possible problematic outcomes or consequences associated with a technology[141]; including (i) environment, health and safety, (ii) justice, access and equality, (iii) individual rights and liberties, (iv) autonomy, authenticity and identity, and (v) dual use[142]. *Intrinsic concerns*, on the other hand, involve relate to objections made to a technology itself, regardless of its outcomes[143]; including (i) 'playing God' or extending human agency to activities

[133] P. Brey (2000) Method in Computer Ethics: Towards a multi-level interdisciplinary approach. *Ethics and Information Technology*, vol. 2, no. 2, p. 125

[134] M.A Pierce & J. W. Henry (1996). Computer Ethics: The Role of Personal, Informal and Formal Codes. *Journal of Business Ethics*, vol. 5, iss. 1, p. 425

[135] A.J. Thomson & D.L. Schmoldt (2001) Ethics in computer software design and development. *Computers and Electronics in Agriculture*, vol. 30, iss 1-3 p. 86

[136] K. Shilton (2018). Values and Ethics in Human-Computer Interaction. Foundations and Trends® Human–Computer Interaction, vol. 12 no. 2, pp 107-109.

[137] Cf. W. H. Dutton (2014). Putting things to work: social and policy challenges for the Internet of things. *Info: the Journal of Policy, Regulation and Strategy for Telecommunications, Information and Media* vol.16 no. 3, 1-21.

[138] P. A. E. Brey (2012) Anticipatory ethics for emerging technologies. *NanoEthics,* vol. 6, no. 1, pp. 1-2

[139] P. Sollie (2007). Ethics, technology development and uncertainty: an outline for any future ethics of technology. *Journal of Information, Communication and Ethics in Society*, vol. 5, no. 4, p. 294

[140] R. L. Sandler (2014). Introduction. In R. L. Sandler (Ed.). *Ethics and emerging technologies*. Palgrave MacMillan, Basingstoke, UK, p. 12

[141] Ibid.

[142] Sandler (2014), op. Cit., pp. 12-15.

[143] Sandler (2014), op. Cit., p. 12

that are inappropriate, (ii) 'hubris' or the overestimation of technological capabilities, and (iii) respecting nature and/or the natural[144].

4.1 Assessing the Ethical Implications of Emerging Technologies

Lucivero et. al. (2011) argue that an assessment of the ethical implications of emerging technologies begins with the evaluation of the 'promises, expectations and visions'[145] – in particular the plausibility of developer claims focusing on the technological feasibility, societal usability, and overall desirability of the expected technology[146]. In adopting a rhetorical approach to analyse the nature of expectations, the authors identify three interrelated claims that generally underlie this type of statement made about emerging technologies[147]: a) claims about the characteristics and functioning of the technology; b) claims about how the technology will be adopted by the intended users and integrated in current practice; c) claims about how the technology will address a social problem or need. In practice, therefore, these need to be identified and considered before a new solution or technology goes to market.

When considered from a policy point of view, the European Union has recently been positioning itself as a global leader in the debate surrounding advanced technology and governance[148]. This ambition has recently been enshrined in the General Data Protection Regulation (GDPR), whose provisions encompass the ethical deployment of technology within European organisations. Of direct interest for a project making use of advanced high performance computing infrastructure like ManyLaws are the provisions contained in Section 5 of the GDPR on the Right to Object (Article 21)[149], concerned with the processing of personal data and the express right of the data subject to object to its perceived unauthorised use; and Automated Individual Decision-Making Including Profiling (Article 22)[150], concerned with the right of any data subject to not be bound by any decision based solely on the outcome of automated processing. Significant debate has arisen in scholarship as to what is meant by wording of these two provisions, and whether the provisions adequately address the ethical concerns that arise from the processes dealt with in practice[151]. Austria in particular has emerged as frontrunner in this context, demonstrating a strong interest in the GDPR and similar collaborative European initiatives[152].

Until the late twentieth century, technology was commonly regarded as value neutral, being considered an artefact that facilitated rather than influenced human endeavours[153]. Langdon Winner (1980) was one of the first prominent contemporary critics of the natural view of technology, arguing that technology contributes not only to productivity and efficiency, but also exhibits moral and political choices that shape power and authority within a society[154]. In recognising the political and moral dimensions of technology, moral philosophers following Winner's example have come to explore not only the direct consequences of technology use, but also how advanced computer systems influence the morality of human action – in particular, decision making. Bynum (2008) identifies two major strands of thinking as being seminal to the way we currently understand the relationship between society and so-called 'intelligent' machines, and relevant to the

[144] Sandler (2014), op. Cit., pp. 15-16

[145] F. Lucivero, T. Swierstra, & M. Boenink (2011). Assessing expectations: towards a toolbox for an ethics of emerging technologies. *NanoEthics*, vol. 5 iss. 2, p. 129.

[146] Lucivero et. al. (2011), op. Cit.pp. 133-138

[147] Lucivero et. al. (2011), op. Cit., p. 132

[148] A. Daly, T. Hagendorff, L. Hui, M. Mann, V. Marda, B. Wagner, W. Wang, & S. Witteborn (2019) Artificial Intelligence - Governance and Ethics: Global Perspectives. Report, 28 June 2019. pp.11-13

[149] General Data Protection Regulation (2018) Art. 21 GDPR Right to Object. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e2803-1-1

[150] General Data Protection Regulation (2018) Art. 22 GDPR Automated Individual Decision-making, including profiling, Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e2803-1-1

[151] Daly et. al. cite L. Edwards & M. Veale (2017). Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. Duke Law & Technology Review, vol. 16 iss. 1, pp. 18-84 and S. Wachter, B. Mittelstadt, & L. Floridi, (2017). Why a right to explanation of automated decisionmaking does not exist in the General Data Protection Regulation. International Data Privacy Law, vol. 7 iss. 2, pp. 76-99 as two seminal works in this respect.

[152] A. Daly, T. Hagendorff, L. Hui, M. Mann, V. Marda, B. Wagner, W. Wang, & S. Witteborn (2019) Artificial Intelligence - Governance and Ethics: Global Perspectives. Report, 28 June 2019. p.14

[153] N. Manders-Huits (2011). What values in design? The challenge of incorporating moral values into design. *Science and engineering ethics*, *17*(2), p. 272.

[154] L. Winner (1980). Do artifacts have politics?. *Daedalus*, vol. 109, no.1, pp. 121-136.

discussion brought forward here[155]. The first of these, propounded by James H. Moor (1998; 1985), takes an extremely practical approach to computer ethics[156]. Moor notes that although computer technology is both 'logically malleable' (i.e. they may be manipulated syntactically and semantically to perform any activity characterised in terms of inputs/outputs/and logical operations), as well as being 'informationally enriching' (i.e. they may be put to use to enhance human capabilities and performance), there is no immediate moral imperative or obligation for their use in day-to-day activities. Indeed, further ethical dilemmas arise, especially in the case of disruptive technologies, when there are no policies or explicit legislation in place to govern either their design or any computer-mediated activity[157].

The second influential way of considering the role of computer technology in society is advanced by Luciano Floridi (1999), whose theory of 'information ethics' is considered by Bynum to be foundational for modern thinking on computer ethics[158]. For Floridi, any object or structure that either preserves or increases information may be considered to possess at least minimal ethical worth[159]. As a consequence of having minimal ethical value being attributed to these so-called 'informational entities'[160], their use by human agents results in a process or action that "…may be morally good or bad, irrespective of its [pleasure and pain] consequences, motives, universality, or virtuous nature." Thus, the design of computer technology, and the detailed sequence of operations arising from the algorithms that run it, are subject to ethical standards, and their use to ethical dilemmas requiring resolution[161].

Moor (1985) discusses 'invisibility' in computer design and operations as having significant ethical ramifications. He specifically mentions three manifestations of invisibility - *invisible abuse*, *invisible programming values*, *invisible calculations* - that cause ethical dilemmas[162]. The first kind is 'invisible abuse', or the intentional but invisible appropriation of computer operations to cause harm. An action perpetrated under this condition may result in theft of property, an invasion of privacy, or a breach in personal security. The second type of invisibility pertains to the presence of hidden programming values. Here, software architects are compelled to take numerous design and deployment decisions that may or may not be directly specified at the outset. For this they fall back on their own intrinsic value systems, which become embedded into the programme's eventual design. The third variety of invisibility is concerned with the complex calculations beyond the sphere of normal human capacity or comprehension that advanced computer systems make in order to produce particular results. In all three cases, moral responsibility is ascribed to software professionals as the moral agents performing the actions and taking the decisions that have wider socio-economic impact.

4.2 Ethical Decision-making in Advanced Computer System Design and Development

This essay will focus on the second type of invisibility factor, invisible programming values, as the potential cause of ethical tension during the development of the ManyLaws platform. Thomson and Schmoldt (2001) enumerate a list of ethical considerations that system developers need to be mindful of during the various stages of system development; i.e. as a part of data source selection, and during data fusion and storage, data provision, and the formulation and presentation of alternatives[163]. These ethical issues include *privacy*, *accuracy*, *property*, *accessibility*, *use of knowledge*, and *quality of life* or *societal wellbeing*[164]. What makes the approach taken in this report new? On the one hand, technology-related concerns involving privacy, property, accessibility, quality of life, etc. can still be explored as expressions of traditional ethical notions

[155] T. W. Bynum (2008). Norbert Wiener and the Rise of Information Ethics. In .J. Van der Hoven & J. Weckert (eds.) *Information Technology and Moral Philosophy*. Cambridge, UK: Cambridge University Press.pp. 20 - 21

[156] Cf. J.H.Moor (1998) Reason, Relativity and Responsibility. *Computer Ethics. Computers and Society*, vol. 28 no. 1, p. 15; and J. H. Moor (1985) What is Computer Ethics? *MetaPhilosophy*, vol. 6, no. 4, pp 272 - 275.

[157] J.H.Moor (1998) Reason, Relativity and Responsibility in Computer Ethics. *Computers and Society*, vol. 28 no. 1, p. 15

[158] Bynum (2008). Op. cit. p.21.

[159] L. Floridi (2002) On the intrinsic value of information objects and the infosphere. *Ethics and Information Technology*, vol. 4. Iss. 4, pp. 290 - 292

[160] L. Floridi and J.W. Sanders (2004) The Foundationalist Debate in Computer Ethics. In R. A. Spinello & H.T. Tavani (eds.) *Readings in CyberEthics Second Edn*., Jones and Bartlett Publishers, Sudbury: M.A., p.93

[161] L. Floridi (1999) Information ethics: On the philosophical foundation of computer ethics. *Ethics and Information Technology*. vol. 1, no. 1, p.38.

[162] J. H. Moor (1985) What is Computer Ethics? *MetaPhilosophy*, vol. 6, no. 4, pp 272 - 275

[163] Thomson & Schmoldt (2001), Op. cit., p. 86

[164] Ibid.

like autonomy, fairness, justice, responsibility, and respect for persons. However, the speed, complexity, and multi-functionality afforded by newer technologies – including systems like the proposed ManyLaws architecture based on high performance computing infrastructure – and their deployment as decision support systems that not only complement but have a direct impact on political, economic and societal outcomes, have necessitated the closer examination of hitherto neglected moral questions associated with the actions of their creators, developers and promoters, especially before the technology becomes publicly available. As society becomes more and more reliant on information created, collected, collated, and communicated through the use of technology, maintaining the integrity of information becomes increasingly central to its computer-mediated provision and retrieval. Robust and reliable information is also thus an important consideration for the ethical development and operation of any computer-based information retrieval system, and it becomes imperative that those individuals responsible for the planning, development and operation of these systems are aware of their moral obligation to guarantee information integrity and universal access[165].

This is all the more important because computer system involvement in decision making leads to fundamental shifts in traditional moral issues of right and wrong, honesty, reliability, responsibility, trust, accountability and fairness[166]. Ethical practice in information system development, argues Rogerson et. al. (2019), is the outcome of a combination of *process* (activities undertaken by information system practitioners, pertaining to whether the conduct of professionals is considered virtuous), and *product* (the outcome of information system endeavour, pertaining to whether a developed system might be considered ethically viable)[167]. Different approaches to system design and deployment emphasise different aspects of the surrounding environment within which these processes occur, and into which products are launched, and can either hinder or facilitate the addressing of dilemmas and conflict in an ethical manner[168]. Here, much like in the case of the ethical use of computer technologies[169], ethical decisions related to advanced software system development may be located within a matrix of three primary influences: (1) the individual developer's own personal code of conduct; (2) any informal code or guidelines of ethical behaviour that exists in the workplace or within the larger professional community; and (3) a developer's exposure to formal or institutionalised codes of ethics. This interplay of influences may be illustrated in Figure 3, below.

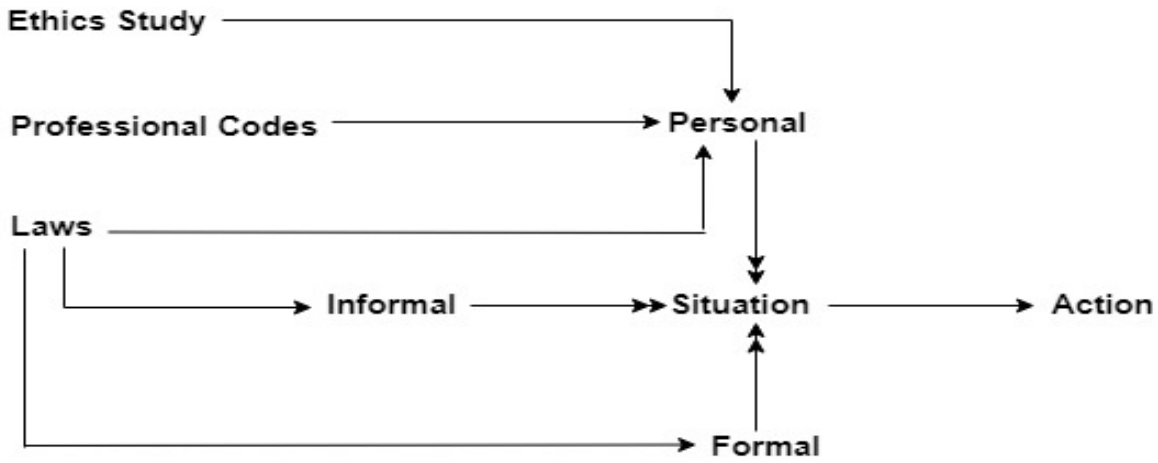

Figure 3: Model of Ethical Decision making Related to Computer Technology (Adapted from Pierce & Henry, 1996)

From the model described it becomes apparent that Pierce & Henry (1996) propose, in addition to the three primary codes of ethics that influence people when confronted with an ethical decision related to computer technology or use, three other factors that also have an impact on the process[170]. The formal study of ethics, the use of professional codes, and the

[165] Cf. S. Rogerson (2011). Ethics and ICT. In R. Galliers and W. Currie, W. (Eds), *The Oxford Handbook on Management Information Systems: Critical Perspectives and New Directions*, Oxford: Oxford University Press, pp. 601-622.
[166] Thomson & Schmoldt (2001) Op. cit, p. 99
[167] S. Rogerson, K. W. Miller, J. S. Winter, & D. Larson (2019). Information systems ethics - challenges and opportunities. *Journal of Information, Communication and Ethics in Society*, vol. 17, iss. 1, p. 88
[168] Rogerson et. al. (2019), Op. cit, pp. 91-92
[169] Pierce & Henry (1996), Op. cit., pp. 245 - 246
[170] Pierce & Henry (1996), Op. cit., p. 426

application of the law are all considered to affect an individual's personal ethical decisions. The law in particular is seen to significantly influence formal and informal codes of conduct as it contains references to specific prohibitions that circumscribe behaviour. Two concepts integral to ethical decision making, and deemed essential to being mentioned here, are *judgement* and *choice*. Judgement and choice refer to two distinct responses necessitated by the circumstances of a decision [ref]; wherein judgement is concerned with the explicit evaluation of a set of alternatives along a continuum or multilevel scale, and choice requires the selection or rejection of one alternative over the rest[171]. This distinction is first and foremost important because the response required on a task in each case is likely to impact the decision processes and the final decision taken[172] [ibid]. When considered in the context of the new digital or disruptive technologies, understanding this distinction, and the options available to software professionals when designing or deploying complex computer systems becomes of paramount importance, as Billings and Scherer (1988) note that there is virtually no empirical research on even the effects of the application of choice versus judgment *per se* on more generic decision tasks[173].

Based on this line of reasoning, the central argument of this section runs as follows: at each stage of the design process or during the deployment of a technology system, within each layer of the high-level architecture, certain fundamental decisions that change the nature and/or direction of the workflow have to be made and subsequently effectuated by certain key people. Each decision represents a point of ethical contention. This argument implies two distinct constituent elements: *desi*gn or process decisions, and the people who take them – and the interplay between the two. As human beings are subject to 'bounded rationality'[174], the extent to which a decision taken by a systems developer is 'ethical' or morally justifiable can be impacted by a whole range of biases - including hindsight bias, preference for certainty, and loss vs. benefit[175]. Five main areas of enquiry present themselves, that each warrant a closer look in this context: (i) the design of the computer system itself; (ii) the selection of data sources; (iii) the processing and/or cleaning of metadata; (iv) the publication of metadata, and (v) the electronic monitoring and collection of end-user personal information. It follows that it becomes imperative, therefore, to ascertain who takes core process decisions, based on what criteria, at what stage in the design or deployment workflow of a system, and how these decisions are then implemented.

Consequently, it becomes important during each step of the design and deployment process to identify correctly and understand the major stakeholders involved - users, co-developers, the general public, commercial enterprises, and government agencies. Developing an understanding about the information used by the technology, the major sources of data, and the output generated by the system is also fundamental. From these, may be derived a list of *ethically preferable actions*, *ethically relevant machine features*, *recommendations for the ethical deployment of sophisticated computer systems*, and the identification of *further points of potential conflict/ethical dilemmas*.

4.3 A Moral Compass to Guide Advanced Computer System Design

It may be argued, therefore, that the most important factor in the effective ethical deployment of advanced computer technology is the development of a better understanding of the people who comprise the professional community responsible for the creation and maintenance of a software system - their attitudes, their values, their actions, and their sense of right and wrong. Software developers are the driving force behind the achievement of some of the world's biggest milestones in recent times. They write the code and train the algorithms that make increasingly important decisions about ordinary people's lives. They determine what information ordinary people can or cannot obtain, see first or do not see at all. They calibrate the difficulty level of each and every advanced software system that is designed, thereby controlling access and usability. In short, they are an integral part of the modern-day financial sector, retail, manufacturing, education, law enforcement, and national security critical infrastructure. This new order of human action requires, according to Jonas

[171] R. S. Billings & L.L.Scherer (1988). The effects of response mode and importance on decision-making strategies: Judgment versus choice. *Organizational Behavior and Human Decision Processes*, vol.41 no.1, p. 2
[172] Ibid.
[173] Billings & Scherer (1988), Op. cit., p. 1
[174] Cf. H.A. Simon (1979) Rational Decision making in Business Organisations. *The American Economic Review*, vol.69, no. 4, pp. 493-513.
[175] Thomson & Schmoldt (2001), Op. cit., p. 92

(2016), an equivalent degree of ethical foresight and responsibility[176]. A final ethical consideration is thus concerned with equipping of programmers with a moral compass, defined for the purposes of this report as "…an internalized set of values and objectives that guide a person with regard to ethical behaviour and decision-making", to inform their professional activities[177]. Based on either an explicit or an unwritten framework of guiding ethical principles[178], the role of the moral compass is to guide complex decision-making while taking into consideration the relevant contextual factors surrounding the given decision point[179]. Bowden and Green (2014) argue that although the ideal goal of a moral compass is to lead an individual to taking the "right" decision, more often than not this finality is difficult to ascertain and attain. Instead, a moral compass is likely to either influence what comes to be perceived as the "better" decision to make within a given set of circumstances, or lead to actions undertaken by an individual to try to change the circumstances[180].

It may be further postulated that a moral compass is necessary for information systems professionals in general, and advanced technology systems developers in particular, given that these systems are black-boxes functioning within complex socio-technical environments, and most end-users are non-experts possessing little knowledge of how systems architecture is constructed or of what the basis of a given code is[181]. In this respect, the average user has to trust the information systems professional to develop a system that respects personal privacy, protects individual freedoms, delivers accurate results (thereby reducing harm or personal risk), and delivers the maximum amount of good to the maximum number. For developers of a legal information retrieval system this issue is compounded thrice over: firstly, on the one hand, while the large-scale mining of legal information might not itself violate privacy or curtail individual freedoms, however, this type of information forms an important basis for accurate decision making, and its mishandling or misuse can result in losses to the people and groups that rely on it.

Secondly, the information retrieval landscape is already fraught with discussions concerning the moral and political implications of technology and the potential of algorithms to amplify in-built bias and opacity[182], particularly when it comes to large-scale corporations like Google - the search engine's domination of the global marketplace for information, and their proprietary rights over algorithms like Page Rank[183]. This is of concern if one considers information as the basis for knowledge and subsequently power. Moral issues involved in such debates include degrees of *fairness*, *trust*, and *accountability*; discussed particularly in terms of the principle of universal access and its denial to the larger public. For ManyLaws, the moral issues involved in the compilation of databases, where choices are made about data sources and data classification schemes, are of great import[184]. This brings us to the third issue, wherein the presence of bias in legal search engine algorithms and code can magnify losses caused by the misuse of legal data, increasing power and other inequalities within society. When applied to the context of government users especially, bad decisions taken as a result of algorithmic bias or the retrieval of faulty information can seriously disrupt society itself.

A fourth related consideration arises from the electronic monitoring of user activity for marketing and e-commerce purposes, which may result in an (real or perceived) invasion of privacy, property, and autonomy; ergo a violation of those

---

[176] H. Jonas (2016). Technology and Responsibility: Reflections on the New Tasks of Ethics. In R. L. Sandler (Ed.). *Ethics and emerging technologies*. Palgrave MacMillan, Basingstoke, UK, p. 42

[177] Dictionary.com (2019). Moral Compass - Definition. https://www.dictionary.com/browse/moral-compass; last accessed 05.08.2019

[178] L. J. Thompson (2004) Moral leadership in a postmodern world. *Journal of Leadership & Organizational Studies* vol. 11, no. 1, pp. 33-34

[179] A. Bowden & P. Green (2014) A Moral Compass Framework for Resolution of Wicked Problems in Doctoral Education and Supervision. *Quality Assurance in Education*, vol. 22, no.4, p. 361

[180] Bowden & Green (2014), Op. cit., p. 364

[181] Cf. L. Lessig (2009) *Code and other laws of cyberspace*. New York, N.Y, Basic Books.

[182] Cf. T. Gillespie (2014). The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society*, *167*, p. 167; L. D. Introna & H. F. Nissenbaum (2000). Shaping the Web: Why the politics of search engines matters. *The Information Society*, vol 16 iss.3, 169–185.

[183] Cf. M. Zimmer (2008). Privacy on planet Google: Using the theory of ''Contextual Integrity'' to expose the privacy threats of Google's quest for the perfect search engine. *Journal of Business & Technology Law*, vol 3, iss. 1, pp. 109–126.; and B. Pan, H. Hembrooke, T. Joachims, L. Lorigo, G. Gay, & L. Granka (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of computer-mediated communication*, vol. 12, iss. 3, pp. 801-823.

[184] Cf. G.C.Bowker & S.L. Star (1999). *Sorting things out: Classification and its consequences*. Cambridge, M.A.: MIT Press.

moral rights[185]. These concerns hold particularly true when code, in the form of cookies, for example, regularly collects personally identifying information from the end-user and applies it to profiling them without either their full knowledge of the process or their complete consent[186]. The ethical use of cookies as a marketing practice has been observed by Palmer (2005) to operate on two levels: on the one hand, their use *per se* to profile users and/or customise user experience is subject to ethical evaluation, and on the other, there are ethical concerns underlying the practices involved in engaging potential customers of the product or service for which information gleaned is applied[187]. It may be argued from a utilitarian perspective that electronic monitoring is ethically justifiable if the benefits outweigh the harm[188], however with business interests involved a moral compass for concerned stakeholders is herein deemed essential.

A so-called moral compass for developers, in this case a professional code of ethics, it is postulated, is also useful in creating an awareness within system developers of the implications that the usability decisions they take during the design and deployment phases of a project can have on a wider community of stakeholders, and possibly the most vulnerable sections of their target user population. This may appear to be a no-brainer, but even simple choices such as the colour scheme of the home page or the placement of buttons can have profound ethical implications in this context. This is because, the notion of 'usability' directly implies accessibility, which is a moral right. Features that make a system usable or appealing directly impact the level of technology acceptance, which in turn determines who can and will access a particular service or functionality. At another level, usability is also a direct function of the diversity embodied within the group of the users of a system like ManyLaws. Not all users are technical experts or legal experts, and non-experts are likely to be the most dependent on the system in both cases. There is a moral imperative for the system to be designed to be user-friendly and bias-free, therefore. What does it mean, therefore, to be an ethical information systems professional? Rogerson (2004) articulates six social responsibility principles that may be used to establish an ethos of ethical professionalism with information systems development[189]. Overall, he argues that these six axes constitute the attainment of a balance between rights and justice, care and empathy, and lead to a concrete definition of ethical responsibilities within the professional community[190].

[185] D. Charters (2002). Electronic monitoring and privacy issues in business-marketing: The ethics of the doubleclick experience. *Journal of business ethics*, vol 35, iss. 4, p. 244

[186] Cf. G. Elmer (2004). *Profiling machines: Mapping the personal information economy*. Cambridge, M.A.: MIT Press.

[187] D. E. Palmer (2005). Pop-ups, cookies, and spam: Toward a deeper analysis of the ethical significance of Internet marketing practices. *Journal of business ethics*, vol. 58, iss. 1-3, p. 272 .

[188] Charters (2002). Op. cit., pp. 248 - 249

[189] Cf. S. Rogerson (2004). Aspects of Social Responsibility in the Information Society. In G.I. Doukidis, N.A. Mylonopoulos and A. Pouloudi (Eds), *Social and Economic Transformation in the Digital Era*, Hershey P.A.: IGI Global Inc.,pp.31-46.

[190] Ibid.

5. MANYLAWS EXPLORATORY WORKSHOP ON LEGAL AND ETHICAL ASPECTS

5.1 Examining the Technical, Legal and Ethical Implications of Improved Access to Legal Information Using Supercomputing Technology: The ManyLaws Project

The ManyLaws project team organised a half-day workshop under the auspices of JURIX 2018: The 31st International Conference on Legal Knowledge and Information Systems, held on 12 December 2018 in Groningen, The Netherlands. The workshop, titled *Examining the Technical, Legal and Ethical Implications of Improved Access to Legal Information Using Supercomputing Technology: The ManyLaws Project*, aimed to engage its audience in a critical exploration of the factors contributing to the effective delivery of the services proposed by the project, together with the legal and ethical implications associated with the application of advanced computing technologies to the acquisition, storage, and processing of legal information. In particular, the workshop sought to concentrate on the need for a legal and ethical framework concerning the application of legal text mining, and the possible form that such a framework might take. Particular emphasis was placed on data protection and copyright laws, and attempts made to find possible solutions for those concerns.

I. Motivation for the Workshop

Although society is overwhelmed with an overload of legal information, only legal experts are able to follow the latest legislation and case law produced by parliaments and courts on a national and on a European level. This creates something of a paradox, as accurate, target-orientated, and timely information is needed not only by lawyers and other legal professionals but also by EU institutions, national governments, local administrations, businesses and citizens at almost every stage of the decision making process. The issue is further compounded by the fact that this type of information is increasingly embedded as large amounts of textual data available on the Internet. Furthermore, the large amount of information concerning laws that apply in the EU countries currently remains fragmented across multiple national databases, and inaccessible systems.

Due to the sheer volume of the data, the manual extraction of the relevant information contained therein is nearly impossible. The application of text mining and analysis tools becomes necessary to contend with issues related to quantity, quality, and the appropriateness of data. The information processing stage of any legal information retrieval system, such as ManyLaws, should make use of parallel high performance computing tools, balancing the load between batch and real-time service modes. Two stages have to be supported: (i) *Pre-processing:* this stage includes data reading and initial cleansing, anonymization if needed, semantic annotation and formulation for processing; and (ii) *Mining:* this stage includes processing tasks based on text mining tools and algorithms relying on a super-computing infrastructure, in order to produce service-oriented intermediate results.

As legal information is generally considered at the core of the open data movement and a major part of public sector information[191], there are implications to be considered about the way text mining tools are applied to such datasets, and used to produce results. Within this context, it is envisaged that the operative services of ManyLaws will facilitate not only decision-making processes but also facilitate access to legal information to a number of different actors across the European Union. Van Wel & Royakkers (2004) argue that data mining as a process is not in itself ethically problematic; instead ethical dilemmas arise when the data to be mined is of a personal nature. Individual consent (or lack thereof) to data being collected and used is a further ethical consideration[192]. In this context, Wahlstrom et. al. (2006) identify four ethical issues

[191] European Parliament, Revised DIRECTIVE 2003/98/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 November 2003 on the re-use of public sector information Directive 2003/98/EC on the re-use of public sector information. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32003L0098&from=EN

[192] Cf. L. Van Wel, L. Royakkers (2004). Ethical issues in web data mining. Ethics and Information Technology vol. 6, iss. 2, pp. 129-140.

associated with data mining that warrant further investigation: *privacy*, *data accuracy*, *database security*, and *stereotyping*[193]. The legal and ethical considerations pertaining to the application and use of data mining tools and methods must hence not be overlooked.

II. Main Research Questions Addressed During the Workshop

The main goals of the workshop were to identify the factors that enable the effective delivery of legal data-related services, and to discuss the key legal and ethical considerations associated with the application of information processing tools to legal text mining and database compilation. The following questions were used as a starting point to guide group discussions:

RQ1. What are the criteria against which existing projects concerned with the publication and access to legal information can be assessed?

RQ2. What are some of the key lessons from current legal text mining projects (eg. OpenLaws) that can be taken as starting points for ManyLaws?

RQ3. Who are the key users of legal text mining services? What are their immediate needs and how can a project like ManyLaws fulfil these?

RQ4. What are the central legal and ethical considerations when developing a legal text mining service based on AI and supercomputing?

RQ5. What are possible solutions for the legal and ethical challenges and how could they be implemented?

III. Structure of the Workshop

The workshop was planned to run for half a day (or for a maximum 3 hours as per standard international conference format). The primary aim was to bring together legal scholars, practitioners, and policymakers within a stimulating environment to discuss current developments, opportunities and challenges relating to the application of text mining, the compilation of databases, and the use of AI and supercomputing to facilitate the publication of and access to legal information in the European context. The workshop was divided into two sessions, each comprising of three interrelated activities: 1) Introductory short presentations by the organisers 2) Group breakout sessions focusing on key issues, and 3) Interactive discussions to generate conclusions. The final structure of the workshop is outlined below in Table 3. Ideas and insights generated through discussions during this workshop were recorded by the organisers and will be shared with workshop participants and the general public at a future date. Ten participants from the field of legal informatics, representing either academia or industry, participated in both sessions. A larger audience from the TeReCom - Technologies for Regulatory Compliance workshop attended the first session as invitees to learn more about the ManyLaws project and some of the initial work being done by the project team.

[193] Cf. K. Wahlstrom, J.F. Roddick, R. Sarre, V.Estivill-Castro, D. deVries (2006). On the ethical and legal implications of data mining. Technical Report SIE-06-001, School of Informatics and Engineering, Flinders University, Adelaide, Australia.

Activity	Duration
Opening Remarks & Introductions Moderators: Dr. Shefali Virkar (Danube University Krems, Austria), Mag. Anna-Sophie Novak (Danube University Krems, Austria)	10 minutes
Session 1: Legal Text Mining – Existing Tools and Infrastructures	
Presentation 1: Many Laws - Technical Requirements and Proposed Infrastructure Presenter: Dr. Shefali Virkar (Danube University Krems, Austria)	10 minutes
Presentation 2: Legal Text Mining - Existing Projects, Tools and Infrastructures Presenter: Dr. Shefali Virkar (Danube University Krems, Austria)	10 minutes
Invited Presentation: Lynx – Compliance Made Easy. Legal Knowledge Graph for Multilingual Compliance Services. Presenter: Prof. Elena Montiel-Ponsoda (Universidad Politécnica de Madrid, Spain)	20 minutes
Breakout Session 1: The Value of Applying Information Processing Tools to Enable Access to Legal Information	40 minutes
Session 2: Legal, Privacy, and Ethical Implications of Legal Text Mining	
Presentation 3: Exploring the Legal, Ethical and Privacy Implications of Legal Text Mining Presenter: Mag. Anna-Sophie Novak (Danube University Krems, Austria)	20 minutes
Breakout Session 2: Benefits of and Challenges to the Use of Text Mining within a Legal Context and Associated Ethical Implications	30 minutes
Group Discussion and Final Evaluation	30 minutes
Formulation of Conclusions	10 minutes
Closing Remarks	

Table 3: Structure of Workshop on Legal and Ethical Aspects

IV. Key Results/Outcomes

Breakout Session 1: The Value of Applying Information Processing Tools to Enable Access to Legal Information

During the joint workshop session, immediately following the invited presentation, the two project teams from Lynx and ManyLaws discussed the commonalities and differences between the two projects, the legal and technical challenges encountered thus far, and the possibility for future collaboration in terms of technical development and research exchanges. The audience, comprising of experts from the technical and legal fields, was also invited to share their experiences with legal information retrieval systems, and to comment on the progress made by each project, as well as the prospect of future collaboration between the two. The discussion incorporated the five guiding questions prepared the ManyLaws moderators, namely: (i) *Are there any other existing solutions that already facilitate access to legal information?*; (ii) *Are CEF DSIs appropriate software building blocks for an action like ManyLaws?*; (iii) *Are there any European data standards that could impact the project?*; (iv) *What makes ManyLaws stand out from the myriad of research projects with similar objectives?*; and, (v) *Which user group would particularly benefit from a project like ManyLaws?*

It emerged that, similar to ManyLaws, Lynx proposes to harness high performance computing, big data technologies and machine translation to facilitate access to legal information across Europe. The projects differ in terms of target user group: while ManyLaws currently focuses on facilitating public sector actors' access to legal information of various sorts, Lynx proposes offer European business entities - particularly SMEs - a legal knowledge and information one-stop shop service for the provision and management of regulatory compliance documents. The fundamental component of the Lynx system architecture is the legal knowledge graph (or LKG) that integrates, aggregates, and links heterogeneous compliance data sources from different jurisdictions, languages and orders - including legislation, case law, standards, industry norms, best practices, and other private contracts - through a collection of advanced services.

Breakout Session 2: Benefits of and Challenges to the Use of Text Mining within a Legal Context and Associated Ethical Implications

The second breakout session of the workshop focused the audience on the legal implications of text mining techniques and tools. Participants were challenged to explore whether there were any further legal implications for developers deploying these tools not already covered during the moderator's presentation, and how issues and complications arising might be resolved. The potential benefits and challenges associated with the application of HPC and text mining to extract useful information from legislation were also examined in this context. Questions concerning the ethical implications of large-scale text/data mining and legal database compilation were raised; in particular, both the allocation of responsibility for the undertaking of various steps within the process, as well as who is to held liable in case of perceived moral failure, were critically examined.

Group Discussion and Final Evaluation

For the final session of the workshop, the audience was divided into two groups, each curated by one of the workshop moderators. Group 1 discussed potential legal barriers to text mining in Europe, while Group 2 focused on the potential ethical barriers to the application of these techniques. The resulting insights are presented here.

Group 1: The participants in this group discussed the applicability of different laws – copyright law, database law, database protection law – to the problem of big open legal data aggregation within the context of the ManyLaws project. Legal provisions in the Austrian case were used as a central reference point against which different national legal frameworks were compared and the specificities discussed. To begin with, Group 1 explored in some detail the general legal implications involved in the large-scale mining of different types of legal information – legislation, case law, explanatory material, and social media posts. The legal implications of data handling in this context were then examined; especially, the nature of the actions involved, and the exceptions arising in different country cases. Particular attention was subsequently paid to the legal exceptions applicable in different national contexts. Here, participants upheld copyright law in Germany as a prime example, and its use by German public administration as a convenient excuse for not opening public data. The availability of versions of legislation in Germany across time, and its similar prevalence in Austria, were also examined in this respect. Copyright law in Greece was also touched upon, and compared to the German and Austrian contexts. The discussion then moved to copyright exceptions, and the fair-use principle in the USA was explored as a cornerstone provision.

Group 1 also took a close look at legal publishers; in particular, the market activities of LexisNexis and the tendency for bigger players to invest in startups as against in academic research initiatives. The nature of the legal information retrieval system market, and the prevalence of high subscription fees as a barrier to unfettered access, was also looked at. Projects similar to ManyLaws, and providing access to European legislation and court judgements, were mentioned. Ease-of-access to legal information was further explored by participants through an analysis of government-developed APIs and cross-references within laws during their framing. The legal ramifications of privacy were also touched upon briefly. A final point of discussion was concerned with methods of due diligence employed by law firms. Participants were of the opinion that private law firms, in their quest for quality legal information, expected projects such as ManyLaws to meet stringent quality standards, and consequently invested large amounts of money in training their human resources in due diligence.

Group 2: Amongst the members of Group 2, there was immediate consensus that the predisposition of disruptive technologies towards raising ethical questions has a tendency to complicate fundamental technical problems and challenges found in theoretical informatics literature. In practice, said some participants, this becomes even more messy, as human actors operate these systems, and understanding users and their vested interests becomes central to the identification and resolution of ethical dilemmas. The discussion then moved on to types of data that ManyLaws proposed to publish, particularly court decisions, that involve an almost inseparable combination of facts and people. Parallels were drawn between the re-publication and storage of these certain types of legal data and medical data. This led to questions being asked by various members about the processes involved in the creation of a big open legal database. One participant remarked that during the compilation of a massive corpus of legal data to train legal algorithms, developers first need to verify whether the raw data available can actually be used. Another discussion centred around whether it was indeed 'ethical' to re-publish data in one standardised format, after it has first been published under different standards? This is questionable, argued one participant, because of the national differences in languages and formats across different EU member states.

One participant wished to clarify whether the ManyLaws project would create one single meta-database, and whether this would be decentralised. This was important, he said, as coupled with existing differences in language and storage formats discussed previously it could lead to so-called 'differential privacy' or the ad hoc removal of records from a database during the cleaning of data, following decisions taken by developers that were either opaque or ethically questionable. Would there be further, he queried, a distinction made in the treatment and handling of different types of law? Not that this would resolve an arising ethical dilemma, countered another participant, for in fact it would merely raise more questions concerning who does the anonymising of data, who is responsible for its processing, and what their vested interests might be. Agile system development here mirrors the so-called Evolution-of-Things, as it produces something that works in a robust way by starting with the basics – hardware, software, and people – and then adapting the development of the system as user requirements change over time.

Looking at the bigger picture, the discussion moved to what the next steps could possibly be for ManyLaws. There exist two potential situations or trade-offs emerging, namely, on the one hand, the threat posed to individuals and their fundamental rights and privacy, contrasted against the benefits that can accrue to society through, for instance, the long-term storage of rich legal data. One participant pointed out that facilitating access to legal information, both in literal terms and in terms of its comprehension leads to citizens circumventing access barriers to knowledge; wherein providing complex information in the layman language of citizens results in a level of simplification that becomes key to highlighting ethical dilemmas and resolving them. Organisations such as think-tanks, the media, and in particular the investigative press, all play an important role in reducing complexity. The demystification of the legal process is almost certainly to result public sector transparency that in turn fosters governmental accountability. In other words, by allowing citizens to understand legislation and policy, asymmetries of knowledge are reduced, and government can be held to account.

Ethical training for a project like ManyLaws is a must, it was decided, as developers are manipulating something – legal data - that is not a neutral artefact, but is instead inherently biased despite all best efforts. A significant part of this training should involve instilling within developers a form of Moral Compass, i.e. developers need awareness training to ensure that the benefits of the database that they are building, and the system that they are creating, do not help malicious people or hostile governments with intrusive surveillance or violations of individual freedoms.

6. CONCLUSION

The aim of the ManyLaws project is to deliver a novel set of services for citizens, businesses and administrations of the European Union, built upon text mining, advanced processing and semantic analysis of legal information. The Action will attempt to build the proper environment and vision of semantically annotated Big Open Legal Data (BOLD), easily searchable and exploitable with proper visualization techniques. The ultimate objective is to provide the technical foundation and the tools for making legal information available to everyone, in a customizable, structured and easy to handle way. To this end, big legal data will need to be accessed, which is currently produced and published in multiple national or EU public databases (e.g. RIS, EUR-Lex) or privately-owned legal databases (e.g. NOMOS). Those datasets will need to be extracted, linked and transformed into a structured relational, open database to prepare them for the mining process. The project also includes the development of services that also involve the deployment of large-scale technical processes facilitated by supercomputing technologies. It is for this reason that exploring the legal and ethical implications of using legal data and of the technical processes involved is both relevant and timely.

Legal Implications of Text and Data Mining

This report aims to, at the outset, give the reader a comprehensive overview of the potential legal risks relating to text and data mining techniques. On the one hand, the project focuses on the European, Greek and Austrian legal frameworks. To be more specific, ManyLaws will use five sources of legal information. The identified Greek legal sources are the Greek National Printing office and the Hellenic Parliament Portal while the Austrian are the RIS and the Austrian National Parliament. On the other, the legal fields that could be affected by the mining process have been identified herein as *copyright law*, *database protection law* and *data protection law*. Since the types of data - legislation, case law, social media posts and journal articles – required for the project are already known, they have also been considered in relation to the informatics techniques discussed, where found to be relevant. It has been found that, according to the nature of the legal data extracted using these methods, the data providers who control access to the data sources, and ways in which data is handled by project systems developers, different combinations of laws might be applicable at any particular moment. Exceptions or limitations to a particular law may or may not apply. This, in turn, bears influence on the rights and obligations of the various project stakeholders. Other factors, including time and monetary resources expended, together with technological considerations also play key roles in the overall assessment. Although this report does not purport to offer official legal advice, we maintain that the legal uncertainty concerning text and data mining should be addressed through explicit provisions for universities and other research organisations that encourage collaboration with the private sector within the framework of pan-European public-private partnerships.

Analysing GDPR Compliance

The General Data Protection Regulation (GDPR) is a regulation in EU law on data protection and privacy for all individual citizens of the European Union (EU) and the European Economic Area (EEA). It also addresses the transfer of personal data outside the EU and EEA areas. The General Data Protection Regulation aims primarily to give control to individuals over their personal data and to simplify the regulatory environment for international business by unifying the regulation within the EU. As all systems have to fully comply with this regulation, ManyLaws portal guarantees full compliance with the GDPR and full protection of the users' personal data it collects. The collection of personal data is limited to the information necessary for the functionality of the ManyLaws portal. All data are gathered in total transparency in accordance with the (GDPR). All data collection processes and the ManyLaws *Terms of Use* will be described in detail inside the portal so users can be fully informed about their personal data protection process.

The Importance of the Ethical Dimension

In acknowledging that politics and morality are integral to the design of advanced computer systems it becomes highly desirable in the case of ManyLaws to understand how the actions and decisions at the heart of these processes may be ethically evaluated and justified. Similarly, an increased awareness of the possible ethical implications of overall system

design and design choices made is expected to contribute to an increased ability on the part of system developers to control and influence this process.

Morally responsible behaviour on the part of a systems architect begins with the adoption of an ethical outlook in one's day-to-day professional activities. One technique to do so is the conscious avoidance of activities, laws, and policies that cause obvious harm to others, particularly through the acquisition of a significant grasp of current legislation and societal context, and is a good beginning towards ethically responsible conduct. For instance, leaving a "cookie" on a user's hard drive without informing them constitutes both a breach in data protection law itself, and also a deviance from core moral values of privacy, freedom to choose, informed consent, and user security. On the other hand, giving the same user the ability to accept or decline a "cookie", and the information required to take this decision, guarantees freedom of action and upholds the right to knowledge.

A conscious awareness of core moral values that leads, in turn, to the development and adoption of permanent or quasi-permanent professional standards that contribute to an increase in ethically responsible behaviour amongst members of the professional community is another way in which ethically responsible behaviour amongst system developers can be encouraged. For instance, if universal access to information is upheld as a fundamental human right, and is embodied in writing as part of a professional code of conduct, then these steps may lead directly to the creation of a platform that is not merely user-friendly, but accessible by people with disabilities.

A third technique to make system design and development decisions morally justifiable is for software engineers to consciously adopt an impartial stance while planning for and in evaluating the outcomes of their actions against certain fundamental ethical values. This implies both a concrete knowledge of core ethical values and principles, and an active engagement with the moral implications of the advanced computer systems that they create and deploy. This means that programmers too need to be actively aware of their own failings and potential for bias.

7. REFERENCES

[1] Art 3 Text and Data Mining, Amendments adopted by the European Parliament on 12 September 2018 on the proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)), europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2018-0337+0+DOC+XML+V0//EN; last accessed 05.01.2019

[2] Article L122-5 10° Code de la propriété intellectuelle, legifrance.gouv.fr/affichCodeArticle.do;jsessionid=446EB38B67F366F6A76125EF7EDFB179.tplgfr29s_2?idArticle=LEGIARTI000037388886&cidTexte=LEGITEXT000006069414&dateTexte=20190106; last accessed 06.01.2019

[3] Article L342-3 5° Code de la propriété intellectuelle, legifrance.gouv.fr/affichCodeArticle.do;jsessionid=446EB38B67F366F6A76125EF7EDFB179.tplgfr29s_2?idArticle=LEGIARTI000033219347&cidTexte=LEGITEXT000006069414&dateTexte=20190106; downloaded 06.01.2019

[4] M. Baroni, R. Bernardi, R. Zamparelli.2014. Frege in space: A program for compositional distributional semantics. Linguistic Issues in Language Technology, 9.

[5] Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979) (Authentic text), wipolex.wipo.int/en/text/283693; last accessed 20.12.2018.

[6] R. S. Billings, L. L. Scherer.1988. The effects of response mode and importance on decision-making strategies: Judgment versus choice. *Organizational Behavior and Human Decision Processes 41*(1), 1-19.

[7] A. Bowden, P. Green. 2014. A Moral Compass Framework for Resolution of Wicked Problems in Doctoral Education and Supervision. *Quality Assurance in Education 22(*4), 355-369.

[8] G. C. Bowker, S.L. Star.1999. *Sorting things out: Classification and its consequences*. Cambridge, M.A.: MIT Press.

[9] P.A. Brey.2012. Anticipatory ethics for emerging technologies. *NanoEthics 6*(1), 1-13.

[10] P.A. Brey. 2000. Method in Computer Ethics: Towards a multi-level interdisciplinary approach. *Ethics and Information Technology 2*(2), 125 - 129.

[11] Briefing requested by the JURI committee, European Parliament, europarl.europa.eu/RegData/etudes/BRIE/2018/604942/IPOL_BRI(2018)604942_EN.pdf; last accessed 02.01.2019

[12] C. Brunschwig. 2001. Visualisierung von Rechtsnormen, Legal Design, Zürcher Studien zur Rechtsgeschichte, hg. von M. T. Fögen [u.a.], Zürich: Schulthess Juristische Medien.

[13] Buchner, Tinnefeld. 2017. Art 89 DSGVO. In J. Kühling, B. Buchner (Hg), Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO/BDSG. Kommentar (2017) Rz 1-2.

[14] Bundesgesetz über das Urheberrecht an Werken der Literatur und der Kunst und über verwandte Schutzrechte (Urheberrechtsgesetz), BGBl. I 2018/63.

[15] T.W. Bynum. 2008. Norbert Wiener and the Rise of Information Ethics. In .J. Van der Hoven & J. Weckert (eds.) *Information Technology and Moral Philosophy* (pp. 8-25). Cambridge, UK: Cambridge University Press.

[16] D. Charters. 2002. Electronic monitoring and privacy issues in business-marketing: The ethics of the doubleclick experience. *Journal of business ethics 35*(4), 234 – 254.

[17] Code de la propriété intellectuelle, legifrance.gouv.fr/affichCodeArticle.do;jsessionid=446EB38B67F366F6A76125EF7EDFB179.tplgfr29s_2?idArticle=LEGIARTI000037388886&cidTexte=LEGITEXT000006069414&dateTexte=20190106; last accessed 06.01.2019

[18] A. Daly, T. Hagendorff, L. Hui, M. Mann, V. Marda, B. Wagner, W. Wang, S. Witteborn. 2019. Artificial Intelligence - Governance and Ethics: Global Perspectives. Report, 28 June 2019.

[19] Dictionary.com. 2019. Moral Compass - Definition. https://www.dictionary.com/browse/moral-compass; last accessed 05.08.2019

[20] W. Dillenz, D. Gutmann. 2004. Praxiskommentar zum Urheberrecht. Österreichisches Urheberrechtsgesetz & Verwertungsgesellschaftsgesetz 2, Springer Verlag, Wien 2004.

[21] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, Official Journal of the European Communities 1996 L 77/20.

[22] W. H. Dutton. 2014. Putting things to work: social and policy challenges for the Internet of things. *Info: the Journal of Policy, Regulation and Strategy for Telecommunications, Information and Media 16*(3), 1-21.

[23] G. Elmer. 2004. Profiling machines: Mapping the personal information economy. Cambridge, M.A.: MIT Press.

[24] EuGH 9.11.2004, Rs C-203/02, British Horseracing Ltd, Slg 2004, I-10415, Rz 5.

[25] European Parliament (2003) Revised DIRECTIVE 2003/98/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 November 2003 on the re-use of public sector information Directive 2003/98/EC sector information. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32003L0098&from=EN; last accessed 12.07.2019

[26] L. Feiler, N. Forgó. 2017. EU-DSGVO. Kurzkommentar (2017) Art 89.

[27] L. Floridi. 2002. On the intrinsic value of information objects and the infosphere. Ethics and *Information Technology 4*(4), 287- 304.

[28] L. Floridi. 1999. Information ethics: On the philosophical foundation of computer ethics. *Ethics and Information Technology 1*(1), 33-52.

[29] L. Floridi, J.W. Sanders. 2004. The Foundationalist Debate in Computer Ethics. In R. A. Spinello & H.T. Tavani (eds.) *Readings in CyberEthics Second Edn.* (pp. 81-95). Sudbury: M.A.: Jones and Bartlett Publishers.

[30] M. V. França, G. Zaverucha, A.S.D.A Garcez. 2014. Fast relational learning using bottom clause propositionalization with artificial neural networks. *Machine learning 94*(1), 81-104.

[31] General Data Protection Regulation. 2018.
https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN, last accessed 26.08.2019

[32] German Copyright Act, Text and Data Mining, dejure.org/gesetze/UrhG/60d.html; last accessed on 6.1.2019

[33] T. Gillespie. 2014. The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society*, *167*.

[34] I. Hargreaves. 2011. Digital Opportunity. A Review of Intellectual Property and Growth, assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/32563/ipreview-finalreport.pdf; last accessed 05.01.2019

[35] Herbst. 2017. Art 19 DSGVO. In J. Kühling, B. Buchner (Hg), Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO/BDSG. Kommentar (2017).

[36] G. Heyer, U. Quasthoff, T. Wittig. 2006. Text Mining: Wissensrohstoff Text: Konzepte, Algorithmen, Ergebnisse, W3L AG 2006.

[37] L.D. Introna, H.F. Nissenbaum. 2000. Shaping the Web: Why the politics of search engines matters. *The Information Society 16*(3), 169–185.

[38] H. Jonas. 2016. Technology and Responsibility: Reflections on the New Tasks of Ethics. In R. L. Sandler (Ed.). *Ethics and emerging technologies* (pp. 37-47). Palgrave MacMillan, Basingstoke, UK.

[39] P. Katzenberger. 1972. Die Frage des urheberrechtlichen Schutzes amtlicher Werke, GRUR.

[40] S. Knotzer. 2018. Wissenschaftliche Forschung und Datenschutz: Eine kritische Analyse ausgewählter Aspekte der österreichischen Rechtslage, ZTR 2018.

[41] M.C. Lacity, M. A. Janson. 1994. Understanding qualitative data: A framework of text analysis methods. *Journal of Management Information Systems 11*(2), 137-155.

[42] N. Lagos, M. Gallé, A. Chernov. 2017. U.S. Patent Application No. 14/850,060.

[43] L. Lessig. 2009. *Code and other laws of cyberspace*. New York, N.Y, Basic Books.

[44] F. Lucivero, T. Swierstra, M. Boenink. 2011. Assessing expectations: towards a toolbox for an ethics of emerging technologies. *NanoEthics*, *5*(2), 129-141.

[45] N. Manders-Huits. 2011. What values in design? The challenge of incorporating moral values into design. *Science and Engineering Ethics 17*(2), 271-287.

[46] J.H. Moor. 1985. What is Computer Ethics? *MetaPhilosophy 6*(4), 272-275.

[47] J. H. Moor. 1998. Reason, Relativity and Responsibility in Computer Ethics. *Computers and Society 28*(1), 14-21.

[48] A.-S. Novak, C. Udokwu, C. Alexopoulos, M. A. Loutsaris, S. Virkar. 2019. Mining Legislation: An Analysis of Legal and Technical Implications. In: E. Schweighofer, F. Kummer, A. Saarenpää (eds.) Internet of Things – Proceedings of the 22nd International Legal Informatics Symposium IRIS 2019 (pp. 631-640). Bern, Editions Weblaw.

[49] OGH 12. Juni 2007, 4 Ob 11/07g, ÖBl 2007/65 S 291 (Dittrich) - ÖBl 2007,291 (Dittrich) = ecolex 2007/332 S 783 (Schumacher) - ecolex 2007,783 (Schumacher) = MR 2007,384 = Burgstaller, MR 2008,15 = jusIT 2008/43 S 94 (Mader) - jusIT 2008,94 (Mader) = RdW 2008/109 S 147 - RdW 2008,147 = SZ 2007/95 = Thiede/Schacherreiter, JBl 2015,287.

[50] D.E. Palmer. 2005. Pop-ups, cookies, and spam: Toward a deeper analysis of the ethical significance of Internet marketing practices. *Journal of Business Ethics 58*(1-3), 271-280.

[51] B. Pan, H. Hembrooke, T. Joachims, L. Lorigo, G. Gay, L. Granka. 2007. In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication 12*(3), 801-823.

[52] M.A. Pierce, J.W. Henry. 1996. Computer Ethics: The Role of Personal, Informal and Formal Codes. *Journal of Business Ethics 5*(1), 425-437.

[53] Provision Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules. Part B Exceptions and Limitations Chapter 1 Text and Data Mining,

ip.mpg.de/fileadmin/ipmpg/content/stellungnahmen/MPI_Position_Statement_Part_B_Chapter_1_Update23022017.pdf; last accessed 06.01.2018

[54] S. Rogerson. 2004. Aspects of Social Responsibility in the Information Society. In G.I. Doukidis, N.A. Mylonopoulos, and A. Pouloudi (eds) *Social and Economic Transformation in the Digital Era* (pp.31-46), Hershey P.A.: IGI Global Inc.

[55] S. Rogerson. 2011. Ethics and ICT. In R. Galliers and W. Currie (eds), *The Oxford Handbook on Management Information Systems: Critical Perspectives and New Directions* (pp.601-622), Oxford UK, Oxford University Press.

[56] S. Rogerson, K. W. Miller, J. S., Winter, D. Larson. 2019. Information systems ethics - challenges and opportunities. *Journal of Information, Communication and Ethics in Society 17*(1), 87-97.

[57] K.F. Röhl, S. Ulbrich. 2007. Recht anschaulich: Visualisierung in der Juristenausbildung (Vol. 3). Herbert von Halem Verlag.

[58] P. Sollie. 2007. Ethics, technology development and uncertainty: an outline for any future ethics of technology. *Journal of Information, Communication and Ethics in Society 5*(4), 293-306.

[59] K. Shilton. 2018. Values and Ethics in Human-Computer Interaction. *Foundations and Trends® Human–Computer Interaction 12*(2), 107-171.

[60] H.A. Simon. 1979. Rational Decision making in Business Organisations. T*he American Economic Review 69*(4), 493-513.

[61] G. Spindler. 2016. Text und Data Mining – urheber- und datenschutzrechtliche Fragen. GRUR 2016 (pp. 1112-1120), https://beck-online.beck.de/Dokument?vpath=bibdata%2Fzeits%2Fgrur%2F2016%2Fcont%2Fgrur.2016.1112.1.htm&anchor=Y-300-Z-GRUR-B-2016-S-1112-N-1

[62] L.J. Thompson. 2004. Moral leadership in a postmodern world. *Journal of Leadership & Organizational Studies 11*(1), 27-37

[63] A.J. Thomson, D.L. Schmoldt. 2001. Ethics in computer software design and development. *Computers and Electronics in Agriculture 30* (1-3), 85-102.

[64] M. Truyens, P. Van Eecke. 2014. Legal aspects of text mining. *Computer Law & Security Review 30*(2), 153-170.

[65] L. Van Wel, L. Royakkers. 2004. Ethical issues in web data mining. *Ethics and Information Technology 6*(2), 129-140.

[66] K. Wahlstrom, J. F. Roddick, R. Sarre, V. Estivill-Castro, D. de Vries. 2006. On the ethical and legal implications of data mining, Technical Report SIE-06-001, School of Informatics and Engineering, Flinders University, Adelaide, Australia, 2006.

[67] L. Winner. 1980. Do artifacts have politics?. *Daedalus 109*(1), 121-136.

[68] A. Zemann. 2018. In Kuscko, G. & Handig, C. (eds), urheber.recht. systematischer kommentar zum urheberrechtsgesetz2, Manz Verlag, Wien 2018.

[69] M. Zimmer. 2008. Privacy on planet Google: Using the theory of ‘‘Contextual Integrity’’ to expose the privacy threats of Google’s quest for the perfect search engine. *Journal of Business & Technology Law 3*(1), 109–126.

ANNEX A: IRIS 2019 PUBLISHED CONFERENCE PAPER

MINING LEGISLATION: AN ANALYSIS OF LEGAL AND TECHNICAL IMPLICATIONS

Anna-Sophie Novak / Chibuzor Udokwu / Charalampos Alexopoulos / Michalis Loutsaris / Shefali Vrikar

Research Associate, Danube University Krems, Department for E-Governance and Administration
Dr. Karl-Dorrek-Straße 30, 3500 Krems, AT
anna-sophie.novak@donau-uni.ac.at

Research Associate, Danube University Krems, Department for E-Governance and Administration
Dr. Karl-Dorrek-Straße 30, 3500 Krems, AT
cjobuzor@gmail.com

Postdoctoral Researcher, University of the Aegean, Department of Information and Communication Systems Engineering
Voulgaroktonou 30, 11472 Athens, GRC
alexop@aegean.gr

PhD Candidate, University of the Aegean, Department of Information and Communication Systems Engineering
Karlovassi 832 00, Samos, GRC
mloutsaris@aegean.gr

Research Associate, Danube University Krems, Department for E-Governance and Administration
Dr. Karl-Dorrek-Straße 30, 3500 Krems, AT
shefali.virkar@donau-uni.ac.at

Keywords: ***Text Mining, Legislation, Legislative Data, Copyright Law, Database Protection Law***

Abstract: *Text mining is the process by which information and connections are obtained from large amounts of text using algorithms. ManyLaws is a project that seeks to increase the accessibility of legal information by offering innovative services. Although several text mining techniques for extracting and storing information exist, an in-depth analysis of the legal and technical issues regarding text mining of legislative information systems is still missing. This paper identifies and examines the legal and technical implications associated with the mining of legislative data, with a focus on copyright and database law. Possible solutions to mitigate the identified issues are also discussed.*

1. Introduction

The aim of the ManyLaws project is to deliver a novel set of services for citizens, businesses and administrations of the European Union, built upon text mining, advanced processing and semantic analysis of legal information. To that end big legal data will need to be accessed, which is currently produced and published in multiple national or EU public databases (e.g. RIS, EUR-Lex). Those datasets will need to be extracted, linked and transformed into a structured relational, open database to prepare them for the mining process.

With regard to the legal field possible applications include the research through legal corpora, analyzing the alignment of national legislation with EU legislation, comparing national laws which target the same life events, analyzing the references to European legislation by national laws, analyzing related laws within the same Member State, timeline analysis for all legal acts, visualization of the progress and current status of a specific national or European piece of legislation and sentiment analysis towards new legislation.[1]

[1] Those are the services proposed by the ManyLaws project. The University of the Aegean, the Hellenic Parliament, Intrasoft, Danube University Krems and the Austrian Parliament are the Consortium of the ManyLaws project. The project runs for two years and is financed under the Connecting Europe Facility (CEF). The aim of this project is to deliver a set of novel services for citizens,

For the proposed applications, the standard Text Analysis pipeline performs several levels of analysis: morphological, syntactic, semantic, and discourse (Lacity/Janson 1994). The morphological and syntactic analysis is usually performed with a syntactic parser (Lagos et al., 2017), which recognizes the syntactic word classes such as nouns and verbs, and the syntactic dependency structure of the constituents of a sentence (main verb, subject, object, etc.). In general language processing, recognizing the basic semantic roles of a sentence constituents, i.e., the «who», «does what», «where», «when», and «how» constituents, is a well-established task for English. Co-reference resolution is identifying when two mentions of an entity or event refer to the same underlying person, place, thing or event in the real world. A layered approach supporting the data flow, from source data to visualized outputs will handle the large volumes of data. When it comes to the information processing layer, various text mining algorithms are applied in different processing tasks, relying on a supercomputing infrastructure. Service-oriented intermediate results are: Creation of reverse indexing, occurrence and frequency tables for millions of words, creation of various n-grams for the identification of important terms or phrases, semantic comparison of different law sets (e.g. EU Directive against national legal framework) performing full word-level and document-to-document comparison for billions of pages. Therefore, the power of thousands of processors is needed.

Although several methods and techniques for executing text mining processes exist, a detailed analysis on the implications and limitations of text mining, especially on legislative information systems, is still missing. Legislative information is usually stored in national databases, whose contents and outputs vary. Therefore, there is a gap in identifying and addressing legal and technical issues in executing text mining on legislative databases. To address this gap, this paper is guided by the main research question: What are the legal and technical implications in mining legislative information? Sub-questions, further derived from the main research question, are as follows: What are the existing information systems for storing legal information? What are the legal and technical issues in mining these information systems? What technical approaches can be used in addressing the identified issues?

The remainder of this paper is structured as follows: Section 2 discusses existing legislative databases within the context of the ManyLaws project and the selection of the databases that are relevant for this paper. Section 3 describes how the text mining process of legislative information is designed. Section 4 identifies the legal and technical implications of text mining and Section 5 provides a conclusion of the study.

2. Existing Legislative Databases

To identify the existing databases for storing legislative information, the following steps were completed. First, we performed a systematic search to identify databases that store legal information including public and private databases. Once this was completed, we consulted with legal experts to identify the main private databases. However, only legislative databases of the European Union, Austria and Greece were selected for this study since those are the pilot cases for the ManyLaws project. The results can be seen in Table 1.

businesses and administrations of the European Union, built upon text mining, advanced processing and semantic analysis of legal information. To that end we will access big legal data.

Platform	Country	Output type	Content type	Access
RIS (Legal Information System Austria)	Austria	HTML, PDF, RTF	Federal and State Legislation and Case Law	Free
Austrian Parliament	Austria	rss, xml	Government Bills, Legislative Initiatives, Explanatory Materials o	Free
RDB	Austria		Commentary, Doctrine, Legislation	Full access is not free
Lexis-Nexis	Austria		Commentary, Doctrine, Legislation	Full access is not free
Hellenic Parliament	Greece		Bills, Proposals, Legislation	Free
Greek National Printing House	Greece	Mainly PDF	Laws, Presidential Decrees, regulatory acts "	Free
NOMOS	Greece		Doctrine and Commentary, Law, Regulations	Full access is not free
EUR-Lex	Europe	HTML	Legislation, Case-Law, Treaties	Free

In the following sections, the mining process of the governmental legislative databases will be described theoretically. The analyses performed in these sections focus on governmental legislative databases as they provide the necessary data via open data portals. In addition, the legal and technical discussions focus on the Austrian databases.

3. The Text Mining Process

In order to design the appropriate text mining process, it is deemed necessary to first identify the service layer. Services are to be provided towards citizens, businesses and administrations, based on the most common needs of each user type. Through a user interface supporting simplification or advanced usage, these are some of the services to be provided at real time by the ManyLaws project: Research through EU member-state legal corpora, analyzing the alignment of national legislation with EU legislation, comparing national laws which target the same life events, analyzing the references to European legislation by national laws, analyzing related laws within the same Member State or with different EU-States, timeline analysis for all legal acts, visualization of the progress and current status of a specific national or European piece of legislation, visualizations of correlations, dependencies and conflicts between different laws and sentiment analysis towards new legislation.

Based on the above services and sub-products, a variety of add-on services can be developed after capturing new requirements from citizens, businesses and administrations. The proposed system consists of three service components:

User Generated Queries: The proposed system infrastructure stores search queries made by users for analysis and optimization purposes. The terms and structure that make up the user query are feeding into the semantic search engine to enhance the relevance of the results based on inferred concepts and semantic annotations.

Search relevant content based on queries: The search engine is used for searching through the system's triple store using a scalable Solr-based semantic search engine. Furthermore, the search engine is taking into account semantic relations between search terms and stored entities (e.g. synonyms). Best practices such as faceted search are also used to present the user with more search options, relevant to the search terms by semantic association.

Search results: This component is responsible for retrieving and presenting to the user the search results in an efficient and user-friendly manner.

3.1. Design of the Text Mining Process

The information processing layer of the architecture deals with the text mining process of the identified legislative databases. These are the four necessary steps:

Data Preparation and Translation Services: In this stage, data from the identified legislative databases is acquired and prepared for the text mining tools to follow. This stage includes data reading and initial cleansing, semantic annotation and formulation for processing. Due to the diversified origin of the texts acquired, a large amount of effort and computational power is devoted to Optical Character Recognition (OCR) and translation

in English. Translation is based on EuroVoc and automated translation services, both for the complete texts but also for the various indexes and n-grams to be created at the processing stage.

Text Mining: Various algorithms are being applied in different processing tasks, relying on a super-computing infrastructure, in order to produce service-oriented intermediate results as described above.

Structured Data: This component represents the information collected from various sources and adhering to a common model and format that can then be used more effectively by the Visual Analytics Service. The data stored include any harvested and derived information that is necessary to realize the project's use cases.

Visual Analytics Service: The Visual Analytics Service provides the ability to access the entire data transformation pipeline from raw or semantic data to interactive visual representations. The main goal is to enable user-centred and comprehensible solutions for getting insights and knowledge about the entire domain.

4. Legal and Technical Issues in Mining Legislative Databases

After we have discussed the text mining approach in the sections above, we will continue below to identify the technical and legal issues in mining the governmental databases.

4.1. Technical Implications of the Text Mining Process

Table 1 shows that there are seven databases that store legislative data in the locations that were chosen for the ManyLaws pilot project. We consider the following as the technical issues when mining legislative text in Austrian databases. These include availability and accessibility of data, content type and content output type.

4.1.1. Accessibility and Availability of Data

The accessibility of information from the data sources depends on the data output of the data sources. Our analysis shows that, although most of the legislative data sources provide full access to the output data, this requires payment of a subscription service to access this information. For the other legislative databases that do not provide full access to their output information, it is therefore not possible to collect complete information from such data sources. The availability of text to be mined depends on the service level agreements entered with the data providers.

4.1.2. Content Type

The content type refers to the type of legal information contained in the output data from the data sources. As shown in table 1, our study shows that some of that data sources output data contain both legislative data and case law, while the others contain only legislative data. It is necessary to identify the type of legal information stored in each data source before a meaningful data analysis can be carried out on such data. To analyze data, it is necessary to process the aggregated texts into structured data by classifying and organizing them into various groups and columns. In this case, if the content types are not properly identified in the processing stage, information generated from such analyses will be incorrect.

4.1.3. Content Output Type

As shown in table 1, all the databases analysed use different formats for displaying legal information to the users. There are technical implications related to the format of the law texts under examination in the different databases. For instance, in the process of transforming of legal documents to plain text (this applies mainly for information stored in the PDF format), the following issues can be experienced. Identifications of columns in the law text, images included in the document, dashes used in continuing a word to the next row, types of article separation, information not related to law content, and character replacement step necessary for eliminating problems of using English language characters similar to other languages.

4.1.4. Mitigation Methods for the identified Issues

To reduce loss of information when converting PDF documents to text, the original XML files can be requested from the database publishers before converting to PDF. This is a possible solution because of the limited number of databases involved. Another way to reduce the loss of information and maintain data integrity is to properly outline the contexts of the legislation documents and incorporate natural language processing capabilities of artificial intelligence (AI) tools. For the content types, it is necessary to properly structure the information to be mined by properly categorizing the text outputs into various groups and columns before performing data mining analyses. The issue with the accessibility of data can easily be addressed by mining only databases that provide full access to the content in order to guarantee the authenticity of information generated.

4.2. Legal Implications of the Text Mining Process

Austria currently has no provisions specifically for the text mining process. Other countries have introduced explicit laws to ensure legal certainty. Germany regulates in section 60d German Copyright Act[2] that copying and publishing of protected works is admissible for the text mining process performed for non-commercial research purposes. The corpus may be published and shared with a distinguishable group of persons for joint research efforts or with third parties in order to review the research quality.

In 2010, then Prime Minister David Cameron commissioned a review of UK's intellectual property rights fearing those might not be suitable to enable innovation and growth.[3] One of the results was an exception from copyright law in Section 29A Copyright, Designs and Patents Act 1988, which states that copying of a work for text mining purposes is no infringement of copyright as long as the person has lawful access to the work. Furthermore, the sole intention of copying has to be research for non-commercial purposes and lastly it is to be accompanied, if possible, by a sufficient acknowledgement. Furthermore, a contractual exclusion of this right is unenforceable and there is no provision for remuneration. The copyright holders are entitled to adopt technical provisions, since the text mining process could overload their servers.

In 2016, France introduced two exceptions for text mining from its Intellectual Property Code (Code de la propriété intellectuelle). Article L122-5 paragraph 10[4] states that the author cannot prohibit the copying for research purposes from sources a person has lawfully access to, if the work has been rightfully published. Commercial research is excluded from this exception as well. Article L342-3 paragraph 5[5] covers the exception for databases. It states that the entitled person cannot prohibit the copying of the database for non-commercial research purposes by a person whose access to the database is lawful, should the database have been rightfully published. Any contractual clause contrary to this Article is void. In the course of the application of this exception, the normal operation of the database must not be interfered with.

Since Austria currently has no specific text mining exception, the extraction of legislative information from ris.bka.gv.at and parlament.gv.at will be investigated concentrating on copyright law and the legal protection of databases.

---

2 § 60d German Copyright Act, Text and Data Mining, dejure.org/gesetze/UrhG/60d.html (accessed on 6 January 2019).

3 HARGREAVES, Digital Opportunity. A Review of Intellectual Property and Growth, assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/32563/ipreview-finalreport.pdf (accessed on 5 January 2019), 2011.

4 Article L122-5 10° Code de la propriété intellectuelle, legifrance.gouv.fr/affichCodeArticle.do;jsessionid=446EB38B67F366F6A76125EF7EDFB179.tplgfr29s_2?idArticle=LEGIARTI000037388886&cidTexte=LEGITEXT000006069414&dateTexte=20190106 (accessed on 6 January 2019).

5 Article L342-3 5° Code de la propriété intellectuelle, legifrance.gouv.fr/affichCodeArticle.do;jsessionid=446EB38B67F366F6A76125EF7EDFB179.tplgfr29s_2?idArticle=LEGIARTI000033219347&cidTexte=LEGITEXT000006069414&dateTexte=20190106 (downloaded on 6 January 2019).

4.2.1. Copyright Law

The Austrian Copyright Act[6] protects individual and intellectual creations in the fields of literature, sound art, fine arts and film art. In the context of the text mining process, it depends firstly on the text mining technique used and secondly on the selected texts whether there is an infringement of copyrights of third parties. Most text mining techniques, as does the proposed technique above, rely heavily on copying the texts for them to be analyzed and annotated. The act of copying protected texts is covered by section 15 of the Copyright Act. It states that solely the author has the right to copy his or her work. The right to copy is understood broadly and includes even the technically required automated copying.[7] Therefore, the described mining process would be unlawful, if the used texts are protected and no exception applies. There are exceptions to Copyright Law that could find application to the text mining process. These exceptions are discussed as follows.

The **research exception**[8] includes solely non-commercial research. Further, the researcher has to attribute the source of the used data, if this is feasible. It is questionable whether a company's in-house research is also included in this exception.[9] Organizational structure and finances are not relevant for determining whether the research is commercial or non-commercial. Relevant is the research purpose.[10]

The **temporary copies exception**[11] allows the copying if it is an essential and integral part of a technical process and itself not of economic importance. As the name of the exception suggests, the copies are allowed to be saved temporarily as long as the technical process requires them.[12] The establishment of a permanent corpus with the original elements is therefore not possible under this exception.[13]

The **text mining exception** adopted by the European Parliament in the directive on copyright[14], states an exception to the reproduction right (copyright law) and the extraction right (*sui generis* database law) of the author if those were made for scientific research purposes, by research organizations that had lawful access to the texts. Unlike the exceptions mentioned above, the EU-exception does not require non-commercial research. However, further specifics on how to interpret the term «research organizations» are missing for now. According to a briefing requested by the JURI committee: «By ‹research organizations›, it is in fact intended universities, research institutes, non-profit or public interest research-intensive organizations.»[15] Public-Private partnerships are not left out, as long as the commercial interests are not the decisive influence and control over the research organization.[16] Member states of the EU are required to implement this exception into their legislation and contracts that contradict the exception are unenforceable. Some criticize the unequal treatment of

---

6 Bundesgesetz über das Urheberrecht an Werken der Literatur und der Kunst und über verwandte Schutzrechte (Urheberrechtsgesetz), BGBl. I 2018/63.

7 Dillenz/Gutmann, Praxiskommentar zum Urheberrecht. Österreichisches Urheberrechtsgesetz & Verwertungsgesellschaftsgesetz[2], Springer Verlag, Wien 2004, § 14 Rz 22.

8 § 42 (2) Urheberrechtsgesetz.

9 Zemann, in: Kuscko/Handig, urheber.recht. systematischer kommentar zum urheberrechtsgesetz[2], Manz Verlag, Wien 2018, § 42 Rz 23.

10 Dillenz/Gutmann, UrhG und VerwGesG[2], § 42 Rz 10.

11 § 41a Urheberrechtsgesetz.

12 Truyens/Van Eecke, Legal aspects of text mining, lrec-conf.org/proceedings/lrec2014/pdf/452_Paper.pdf. (accessed 2 January 2019).

13 Truyens/Van Eecke, Legal aspects of text mining.

14 Art 3 Text and Data Mining, Amendments adopted by the European Parliament on 12 September 2018 on the proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)), europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2018-0337+0+DOC+XML+V0//EN (accessed 5 January 2019).

15 Briefing requested by the JURI committee, European Parliament, europarl.europa.eu/RegData/etudes/BRIE/2018/604942/IPOL_BRI(2018)604942_EN.pdf (accessed 2 January 2019), p. 8.

16 Briefing requested by the JURI committee, European Parliament, p. 8.

non-research organizations who have lawful access to texts.[17] Their text mining processes are not covered by the EU-exception.

As we have now discussed the general application of copyright law and the relevant exceptions on the text mining process, the following section will give an assessment of the applicable laws when mining Austrian legislative databases.

Article 2 paragraph 4 Berne Convention[18] states that «It shall be a matter for legislation in the countries of the Union to determine the protection to be granted to official texts of a legislative, administrative and legal nature, and to official translations of such texts.» According to section 7 paragraph 1 Austrian Copy Rights Act, laws, regulations, official decrees, notices and court judgements do not fall under the Copyright Act. Since the copying of legislative data falls under this exception, copyright law is not applicable to the extraction of legislative information. Concerning the explanatory materials from the website of the Austrian parliament one has to make sure that those are also covered by section 7 paragraph 1 Copyright Act. Explanatory Materials are of considerable importance for interpreting the law. Furthermore, they are attributable to the parliament, an authority with public authority tasks. As a result, copyright law is not applicable to explanatory materials, which can be subsumed as official notices.[19]

4.2.2. Database Law

Should the texts for the mining process be part of a database, copyright law as well as the *sui generis* database law have to be considered. Those two fields of law are independent of each other and can apply to the same database. According to the 17th recital of the directive on the legal protection of databases, the term «database» should be understood to include literary, artistic, musical or other collections of works or collections of other material such as texts, sound, images, numbers, facts, and data, collections of independent works, data or other materials which are systematically or methodically arranged and can be individually accessed. According to the European Court of Justice, databases «in any form», whether they are in electronic or non-electronic format, are covered by the directive.[20] Below we will discuss general database protection and then go on to our specific goal of mining governmental legislative databases.

Copyright Protection of Databases

The copyright protection of databases does not extend to their contents. The selection or arrangement of the contents mark the author's own intellectual creation and are protected as such by copyright. No temporary or permanent reproduction by any means and in any form, in whole or in part can be made without the consent of the author[21]. Art 6 paragraph 2 of the directive allows the member states to make certain exceptions. In Austria, section 42 paragraph 2 and 40 h paragraph 2 Copyright Act state that the reproduction for non-commercial research purposes is admissible. It is doubtful that the text mining process includes the extraction of the protected selection and arrangement of the contents, as the mining process concentrates on the content.

Sui Generis Protection of Databases

Section 76c of the Austrian Copyright Act protects investments made in databases. According to Article 7 of the directive, the maker of a database who has made qualitatively and/or quantitatively a substantial investment

---

17 Provision Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules. Part B Exceptions and Limitations Chapter 1 Text and Data Mining, ip.mpg.de/fileadmin/ipmpg/content/stellungnahmen/MPI_Position_Statement_Part_B_Chapter_1_Update23022017.pdf (accessed 6 January 2018), p. 4.

18 Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979) (Authentic text), wipolex.wipo.int/en/text/283693 (accessed 20 December 2018), p. 4.

19 Katzenberger, Die Frage des urheberrechtlichen Schutzes amtlicher Werke, GRUR 1972, p. 692.

20 Decision of ECJ of 9 November 2004 C-203/02 *British Horseracing Ltd*, Slg 2004, I-10415, Rz 5.

21 Art 5 lit a Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, Official Journal of the European Communities 1996 L 77/20.

in either the obtaining, verification or presentation of the contents has the right to prevent extraction and/or re-utilization of the whole or of a substantial part of the contents of that database.[22] The extraction and re-utilization of an insubstantial part of such a database is therefore allowed, as long as they are not executed repeatedly and systematically. According to § 76d paragraph 3 number 2 Copyright Act, the extraction and re-utilization of a substantial part of a publicly available database is lawful for non-commercial research purposes, to a justifiable extent, as long as the source is named.

As we have now discussed the general application of database law on the text mining process, the following section will give an assessment of the applicable laws when mining Austrian legislative databases.

As neither the selection nor the arrangement of the contents of the Austrian legislative databases mark an intellectual creation, those are not protected as such by copyright protection of databases. Nonetheless, in the imprints on the website ris.bka.gv.at one will find a copyright notice in favor of the Federal Ministry of Digitalization and Economy and on the website parlament.gv.at one will find a copyright notice in favor of the Austrian Parliament. Concerning the *sui generis* protection of databases, one has to check if a substantial investment in either the obtaining, verification or presentation of the contents was made.[23] A qualitatively and quantitatively substantial investment was made especially in the presentation of the contents of the Austrian legislative databases. According to the CJEU, the investment going into the creation of the single entries is not taken into account.[24] According to the Austrian Supreme Court, the definition of an investment does not depend on whether the data is given to the maker or if law prescribes the presentation of the data.[25] Rather those expenses that run into the presentation and the updating of the database content are to be considered as investment.[26] Therefore, the maker of the websites ris.bka.gv.at and parlament.gv.at have the right to prevent extraction and/or re-utilization of the whole or of a substantial part of the contents of that database via the website. An application of section 7 Copyright Act on protected databases by analogy was denied by the Austrian Supreme Court.[27] Since both websites offer some of their data on the national open data portal (data.gv.at) this right does not constitute an insurmountable barrier for the mining process.

5. Conclusion

In this paper, we analyzed and discussed legal and technical implications experienced in mining legislative information. The legislative databases in Greece, Austria and Europe were chosen, as those are the pilots for the ManyLaws project. The ManyLaws project seeks to provide innovative services to increase accessibility of legal information. To address the gap identified by the main research question of this study, we systematically identified databases that store legal information and described text mining procedures for mining them. Then we analyzed the properties of these databases and discussed technical and legal implications that could arise in mining these databases.

The main limitation of this paper is that the study focuses only on the technical and legal implications of text mining legislative databases in relation to the ManyLaws project. This limits the scope of the paper. Also, as

[22] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, Official Journal of the European Communities 1996 L 77/20.

[23] § 76c Urheberrechtsgesetz.

[24] Decision of ECJ of 9 November 2004 C-203/02 *British Horseracing Ltd*, Slg 2004, I-10415, Rz 40 «However, such prior checks are made at the stage of creating the list for the race in question. They thus constitute investment in the creation of data and not in the verification of the contents of the database.».

[25] OGH 12 June 2007, 4 Ob 11/07g, ÖBl 2007/65 S 291 (Dittrich) - ÖBl 2007,291 (Dittrich) = ecolex 2007/332 S 783 (Schumacher) - ecolex 2007,783 (Schumacher) = MR 2007,384 = Burgstaller, MR 2008,15 = jusIT 2008/43 S 94 (Mader) - jusIT 2008,94 (Mader) = RdW 2008/109 S 147 - RdW 2008,147 = SZ 2007/95 = Thiede/Schacherreiter, JBl 2015,287.

[26] OGH 12 June 2007, 4 Ob 11/07g.

[27] OGH 9 April 2002, 4 Ob 17/02g.

a result of the page limitation of this paper, the study does not fully describe the necessary steps in addressing the technical issues identified in this paper.

6. References

Heyer, Gerhard/Quasthoff, Uwe/Wittig, Thomas, Text Mining: Wissensrohstoff Text: Konzepte, Algorithmen, Ergebnisse, W3L AG 2006.

German Copyright Act, Text and Data Mining, dejure.org/gesetze/UrhG/60d.html (accessed on 6 January 2019).

Hargreaves, Ian, Digital Opportunity. A Review of Intellectual Property and Growth, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/32563/ipreview-finalreport.pdf (accessed on 5 January 2019), 2011.

Code de la propriété intellectuelle, legifrance.gouv.fr/affichCodeArticle.do;jsessionid=446EB38B67F366F6A76125EF7EDFB179.tplgfr29s_2?idArticle=LEGIARTI000037388886&cidTexte=LEGITEXT000006069414&dateTexte=20190106 (accessed on 6 January 2019).

Bundesgesetz über das Urheberrecht an Werken der Literatur und der Kunst und über verwandte Schutzrechte (Urheberrechtsgesetz), BGBl. I 2018/63.

Dillenz, Walter/Gutmann, Daniel, Praxiskommentar zum Urheberrecht. Österreichisches Urheberrechtsgesetz & Verwertungsgesellschaftsgesetz^2, Springer Verlag, Wien 2004.

Zemann, Adolf, in: Kuscko, Guido/Handig, Christian, urheber.recht. systematischer kommentar zum urheberrechtsgesetz^2, Manz Verlag, Wien 2018.

Truyens, Maarten/Van Ecke, Patrick, Legal aspects of text mining, http://lrec-conf.org/proceedings/lrec2014/pdf/452_Paper.pdf (accessed 2 January 2019).

Art 3 Text and Data Mining, Amendments adopted by the European Parliament on 12 September 2018 on the proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)), http://europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2018-0337+0+DOC+XML+V0//EN (accessed 5 January 2019).

Briefing requested by the JURI committee, European Parliament, http://europarl.europa.eu/RegData/etudes/BRIE/2018/604942/IPOL_BRI(2018)604942_EN.pdf (accessed 2 January 2019).

Provision Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules. Part B Exceptions and Limitations Chapter 1 Text and Data Mining, http://ip.mpg.de/fileadmin/ipmpg/content/stellungnahmen/MPI_Position_Statement_Part_B_Chapter_1_Update23022017.pdf (accessed 6 January 2018).

Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979) (Authentic text), http://wipolex.wipo.int/en/text/283693 (accessed 20 December 2018).

Katzenberger, Paul., Die Frage des urheberrechtlichen Schutzes amtlicher Werke, GRUR 1972.

Lagos, N., Gallé, M., & Chernov, A. (2017). U.S. Patent Application No. 14/850,060.

Lacity, M. C., & Janson, M. A. (1994). Understanding qualitative data: A framework of text analysis methods. Journal of Management Information Systems, 11(2), 137–155.

Baroni, M., Bernardi, R., and Zamparelli, R. (2014). Frege in space: A program for compositional distributional semantics. Linguistic Issues in Language Technology, 9.

França, M. V., Zaverucha, G., & Garcez, A. S. D. A. (2014). Fast relational learning using bottom clause propositionalization with artificial neural networks. Machine learning, 94(1), 81–104. Brunschwig, Colette (2001): Visualisierung von Rechtsnormen, Legal Design, Zürcher Studien zur Rechtsgeschichte, hg. von M. T. Fögen [u.a.], Zürich: Schulthess Juristische Medien.

Röhl, K. F., & Ulbrich, S. (2007). Recht anschaulich: Visualisierung in der Juristenausbildung (Vol. 3). Herbert von Halem Verlag.

ANNEX B: JURIX 2018 PUBLISHED WORKSHOP DESCRIPTION

Examining the Technical, Legal and Ethical Implications of Improved Access to Legal Information Using Supercomputing Technology: The ManyLaws Project

Shefali VIRKAR[a,1] Anna-Sophie NOVAK[b,2] Charalampos ALEXOPOULOS[c] Yannis CHARALABIDIS[d] and Michalis LOUTSARIS[e]
[a] *Danube University Krems, Austria*
[b] *Danube University Krems, Austria*
[c] *University of the Aegean, Greece*
[d] *University of the Aegean, Greece*
[e] *University of the Aegean, Greece*

Abstract. Legal information is an asset for decision making not only by EU institutions but also by member states, local administrations, businesses and citizens. Accurate, target-orientated, and timely information could enhance the digitisation of decision-making processes. Data such as legislation acts, bills, case laws, resolutions and decisions, published in each Member States' language, as well as administration and citizen-generated content is increasingly embedded in large amounts of textual data available on the Internet. The vision of ManyLaws is to produce semantically annotated Big Legal Open Data, easily searchable and exploitable based on text mining tools and algorithms offered through proper visualization techniques. The proposed workshop will focus on the factors that can contribute to the effective delivery of the new services, together with the legal and ethical implications associated with the application of advanced computing technologies to the acquisition, storage, and processing of legal information.

Keywords. Big Legal Open Data, Text Mining, Supercomputing, Ethical and Legal Implications

1. Motivation for the Workshop

Although society is overwhelmed with an overload of legal information, only legal experts can follow the latest legislation and case law produced by parliaments and courts on a national and on a European level. Accurate, target-orientated, and timely information is needed not only by EU institutions but also by member states, local

[1] Shefali Virkar, Department for E-Governance and Administration, Danube University Krems, Dr.-Karl-Dorrek-Straße 30, 3500 Krems an der Donau, Austria, E-mail: shefali.virkar@donau-uni.ac.at.
[2] Anna-Sophie Novak, Department for E-Governance and Administration, Danube University Krems, Dr.-Karl-Dorrek-Straße 30, 3500 Krems an der Donau, Austria, E-Mail: anna-sophie.novak@donau.uni.ac.at.

administrations, businesses and citizens at almost every stage of the decision making process.

Such, by policy makers required and/or produced data, is increasingly embedded in large amounts of textual data available on the Internet. Furthermore, the large amount of information concerning laws that apply in the EU countries currently remains fragmented across multiple national databases or inaccessible systems, mainly consisting of documents (legislation acts, bills, case laws, resolutions, decisions) published in each Member States' language. In addition, administration-generated content (e.g. local communications, regulations), citizen-generated relevant content (e.g. blogs, newsletters, social media posts) and news published in EU member states concerning legal events (e.g. law publication, draft law deliberation, EU directive publication) could be considered of major importance in every-day or in mid- and long-term decision making. It is estimated the above database will contain more than 1 trillion words in 21 different languages, corresponding to about 10 million "volumes" of classical books, when another 5,000 such "volumes" will be added for study on a daily basis.

Due to the sheer volume of data, the manual extraction of the relevant data it contains is nearly impossible. Text mining and analysis tools become necessary to address the problem of volume, of currentness, and in order to provide the right information in the proper format. The information processing stage of such an infrastructure should make use of massively parallel computing tools, balancing the load between batch and real-time service modes. Two stages should be supported: (i) *Pre-processing*: This stage includes data reading and initial cleansing, anonymization if needed, semantic annotation and formulation for processing; (ii) *Mining*: This stage includes processing tasks based on text mining tools and algorithms relying on a super-computing infrastructure, in order to produce service – oriented intermediate results.

The vision of ManyLaws is to produce and build the proper environment of semantically annotated Big Legal Open Data, one that is easily searchable and exploitable with proper visualization techniques. The ultimate objective is to provide the technical foundation and the tools for making legal information available to everybody, in a customizable, structured and easy to handle way.

The developed services will ensure real time provision towards citizens, businesses and administrations based on the most common needs of each user type. The current envisioned services provide the following: parallel search in many EU member-state legal frameworks using simple keywords (through parallel translation of search terms), assessment of the degree of transposition of an EU directive in a national legal framework, indicating relevant national legislation and monitoring the status of transpositions, analysis of references to the european legislation by national laws, comparative analysis of equivalent or relevant laws from different EU member states, comparative analysis of connected laws from the same member state, timeline analysis for all legal elements, visualising the progress and current status of a specific national or European legislation (after amendment/extensions) over time including preparatory acts and agreements, interrelation of laws and news or social media posts, including sentiment analysis, various geo-related visualisations (e.g. EU maps indicating different parameters), various text-related visualisations (e.g. wordle, sentiment graphs, interrelation maps, etc.) and other common visual aids (e.g. graphs, charts, tables, etc.), visualizations of correlations, dependencies and conflicts between different laws and decision support services (e.g. impact assessment) within legal procedures.

Although legal information is generally considered to be at the core of the open data movement and a major part of public sector information [1], there are implications in

considering the way text mining tools are used to produce results. Within this context operative services of Many Laws will facilitate not only the decision making processes but also enable access to legal information across the European Union. The legal and ethical considerations pertaining to the application and use of data mining tools and methods must not be overlooked. Wahlstrom et. al. [2] argue that data mining as a process is not in itself ethically problematic; instead ethical dilemmas arise when the data to be mined is of a personal nature. Individual consent (or lack thereof) to data being collected and used is a further ethical consideration [3]. In this context, [2] identify four ethical issues associated with data mining that warrant further investigation: privacy, data accuracy, database security, and stereotyping.

In this workshop, we will enquire into the need for a legal and ethical framework concerning legal text mining and, if so, the possible versions of this framework. Particularly we will determine the concerned data protection, copyright laws and find possible solutions for those concerns.

1.1. Relevance of the Workshop to the JURIX Conference

The Many Laws project aims to apply information processing and text mining tools and methods to big legal data in order to develop novel legal information processing services for citizens, businesses and public administrations within the European Union. The proposed workshop will focus on the factors that can contribute to the effective delivery of the new services, together with the legal and ethical implications associated with the application of advanced computing technologies to the acquisition, storage, and processing of legal information.

1.2. Questions to be Addressed During the Workshop

The main goals of this workshop are to identify the factors that enable the effective delivery of legal data-related services, and to discuss the key legal and ethical considerations associated with the application of information processing tools to legal text mining and database compilation. The following questions will be used as a starting point to guide group discussions:

- What are the factors against which existing projects concerned with the publication and access to legal information can be assessed?
- What are some of the key lessons from current legal text mining projects (eg. OpenLaws) that can be taken as starting points for Many Laws?
- Who are the key users of legal text mining services? What are their immediate needs and how can a project like Many Laws fulfil these?
- What are the central legal and ethical considerations when developing a legal text mining service based on AI and supercomputing?
- What are possible solutions for the legal and ethical challenges arising, and how could they be implemented?

2. Format of the Workshop

This half day workshop (3 hours) aims to bring together legal scholars, practitioners, and policymakers within a stimulating environment to discuss current developments, opportunities and challenges relating to the application of text mining, the compilation of databases, and the use of AI and supercomputing to facilitate the publication of and access to legal information in the European context. The workshop will be divided into two sessions, each comprising of three interrelated activities: 1) Introductory short presentations by the organisers 2) Group breakout sessions focusing on key issues, and 3) Interactive discussions to generate conclusions.

Ideas and insights generated through discussions during this workshop will be recorded by the organisers and shared with workshop participants and the general public at a future date. The structure of the workshop is outlined below in Table 1.

Table 1. Workshop Agenda

Activity	Time
Open Remarks and Introductions	10 minutes
Session 1: Legal Text Mining – Existing Tools and Infrastructure	
Presentation I: ManyLaws – Technical requirements and Proposed Infrastructure	10 minutes
Presentation II: Legal Text Mining – Existing Projects, Tools and Infrastructures	10 minutes
Breakout Session: The Value of Applying Information Processing Tools to Enable Access to Legal Information	30 minutes
Group Feedback and Discussion	30 minutes
Session 2: Legal, Privacy, and Ethical Implications of Legal Text Mining	
Presentation 3: Exploring the Legal, Ethical and Privacy Implications of Legal Text Mining	10 minutes
Breakout Session: Benefits of and Challenges to the Use of Text Mining within a Legal Context	40 minutes
Group Discussion and Final Evaluation	30 minutes
Formulation of Conclusions	10 minutes

References

[1] European Parliament, Revised DIRECTIVE 2003/98/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 November 2003 on the re-use of public sector information Directive 2003/98/EC sector information.

[2] K. Wahlstrom, J.F. Roddick, R. Sarre, V.Estivill-Castro, D. deVries, On the ethical and legal implications of data mining, Technical Report SIE-06-001, School of Informatics and Engineering, Flinders University, Adelaide, Australia, 2006.

[3] L. Van Wel, L. Royakkers, Ethical issues in web data mining, Ethics and Information Technology 6, 2 (2004), 129-140.

Presenters' Biographies

Dr. Shefali Virkar is a Senior Researcher at the Department for E-Governance and Public Administration at the Danube University Krems (DUK), Austria, specialising in the theory and practice of electronic government. In particular, she studies the political, social, and economic implications of the new Information and Communications Technologies (ICT), and how these developments impact traditional forms of work and governance structures. She is has been involved on a variety of Austrian and European projects, and is a member of the EGOV-CeDEM-ePart conference organising committee. Shefali holds a D.Phil. (Ph.D) degree from the University of Oxford (UK), where her doctoral research focused on the impact that the perceptions and behaviour of political actors involved with strategic ICT projects in bureaucracies have on the ultimate outcome of such initiatives through an in-depth examination of a case study based in India. Prior to this, she was awarded a M.A. in Globalisation and Development from the University of Warwick (UK).

Mag. Anna-Sophie Novak is a Research Assistant at the Department for E-Governance and Public Administration at the Danube University Krems (DUK). Anna-Sophie holds a master degree in law (University of Vienna) and is currently a PhD candidate at the University of Vienna.

Dr. Charalampos (Harris) Alexopoulos is a postdoc researcher at the Department of Information and Communications Systems Engineering, University of the Aegean. He is a Project Manager at the Information Systems Laboratory of the same department, working on European and National funded research and pilot application projects for governments and enterprises. Harris also serves as Programme and Organisation Committee Member, Track and Minitrack chair for Samos Summit, HICSS, MCIS. He is also a course manager of two summer schools and he is teaching e-government and business management at pre-graduate and postgraduate level. His research interests lie on the fields of Decision Support Systems, Open Data, e-government and interoperability. Harris is a computer science graduate from the University of Peloponnese with an MSc in Management Information Systems from the University of the Aegean. In 2015, Harris was ranked as one of the most prolific researchers in open data research worldwide by Hossain, Dwivedi and Rana (2015).

Dr. Yannis Charalabidis is Associate Professor in the Department of Information and Communication Systems Engineering of the University of Aegean. In parallel, he serves as Director of the Innovation and Entrepreneurship Unit of the University, designing and managing youth entrepreneurship activities, and Head of Information Systems Laboratory, coordinating policy making, research and pilot application projects for governments and enterprises worldwide. He has more than 20 years of experience in designing, implementing, managing and applying complex information systems as project manager, in Greece and Europe. He has been employed for 8 years as an executive director in SingularLogic Group, leading software development and company expansion in Greece, Eastern Europe, India and the US. He has published more than 200 papers in international journals and conferences, while actively participating in

international standardisation committees and scientific bodies. In 2016 he was nominated as the 8th most productive writer in the world, among 9500 scholars in the Electronic Government domain, according to the Washington University survey. He is Best Paper Award winner in the International IFIP e-Government Conference (2008,2012, 2016), winner of the first prize in OMG / Business Process Modelling contest (2009) and 2nd prize winner in the European eGovernment Awards (2009). As of August 2018, Yannis is among the 100 most influential people in Digital Governance worldwide, according to the apolitica.co list.

Mr. Michalis Avgerinos Loutsaris (male) is PhD candidate in the University of the Aegean. Michalis holds a Bachelor Degree and a Master of Science in "Technologies and Management of Information and Communication Systems" from University of the Aegean, Department of Information and Communication Systems Engineering. His research interests involve around Entrepreneurship, Artificial Intelligence and Text Mining. He is Lab Assistant and has some experience in teaching lab courses as assistant. He is member of the Information Systems Laboratory (ISLab) and is working in several lab projects.