GDPR Compliant Data Processing and Privacy Preserving Technologies: A Literature Review on Notable Horizon 2020 Projects

GDPR Compliant Data Processing and Privacy Preserving Technologies: A Literature Review on Notable Horizon 2020 Projects*

Orhan Gazi Yalcin[1][0000-0001-7990-6531]

University of Bologna (beneficiary)
Polytechnic University of Madrid
University of Turin
[1] Via Galliera, 3, 40121, Bologna, Italy
orhangazi.yalcin2@unibo.it
https://last-jd-rioe.eu/

Abstract. This paper presents a practical literature review focusing on privacy preserving technologies and organizational measures developed and proposed for GDPR-compliant data processing. Based on the selected Horizon 2020 projects, it identifies the substantial data processing and big data challenges relevant to data protection and privacy. Then, it visits the prominent privacy preserving technologies and organizational measures addressing these challenges. Finally, it analyzes the focus areas of the selected projects, identifies the solution they propose, draws quantitative conclusions, and asserts recommendations for future projects.

Keywords: Privacy Preserving Technologies, Data Protection, Information Ethics, GDPR, Tech and Ethics, Technical and Organizational Measures

1 Introduction

The 20th century witnessed the foundations of data-oriented research fields such as data science, artificial neural networks, deep learning, and machine learning. However, data-oriented technologies gained tremendous significance in the decision-making process only in the 21st century. Although data collection practices were poorly regulated initially, soon, the governments realized the dangers these practices may cause, and data protection regulations started to emerge. This regulatory policy resulted in a new field to surface: Data Protection. The main document regulating the data protection field in the European Union is General Data Protection Regulation (GDPR) 2016/679, which supersedes the Data Protection Directive 95/46/EC. The fundamental principles dictating GDPR are based on the European ethical principles and fundamental rights and freedoms. Several articles of GDPR, particularly Art. 32, lay out the obligations of data controllers and processors regarding data processing systems, which can be summarized as "technical and organizational measures"- including the utilization of privacy

* This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie ITN EJD "Law, Science and Technology Rights of Internet of Everything" grant agreement No 814177.

preserving technologies (PPTs). This paper reviews the latest developments in privacy-preserving technologies and organizational measures that address the challenges that surfaced out of the clash between the GDPR obligations & ethical principles against technical & economic goals.

To detect the advancements in PPT solutions and organizational measures, this paper reviews the notable Horizon2020 projects that aim to develop PPT solutions and ethical & technical standards for GDPR compliant data processing. After analyzing these projects, we will uncover the existing trends & patterns and detect which areas require additional attention.

2 Methodology

This paper refers to the common themes established by legal and ethical frameworks in the European Union to review the notable Horizon 2020 projects. The final goal of this research is to shed light on the common themes & missing pieces and contribute to the direction of the state of the art of data processing and AI applications. The overall methodology consists of three pillars:

- **An Interdisciplinary Approach** is adopted to cover all the issues in different fields related to AI and data processing practices. The field of data protection is deeply connected with information technology, data protection law, and information ethics. Therefore, to have a thorough analysis of data protection issues, the principles and methodologies in these fields will be used.
- **A Pragmatic Approach** based on effective GDPR rules is followed without disregarding ethical principles. Another essential component of the research methodology is the pragmatic approach, as there will be recommendations for future projects based on the existing projects.
- **A Comparative Approach** is the third pillar of the methodology adopted to compare the existing solutions. The comparative approach allows us to see the existing solutions' relative competencies and what they lack in terms of GDPR compliance and ethical principles.

3 Big Data Challenges and Proposed Solutions

After a detailed literature review, this paper identifies ten critical big data and data protection challenges at the intersection of information technology, data protection regulations, and principles of information ethics. These challenges are addressed by the Horizon 2020 projects with 15 technical and organizational measures.

3.1 Challenges to security and privacy in Big Data

In the European Union, GDPR regulates and heavily limits the processing of personal data. Although the introduction of this regulatory practice is easily justified with ethical and societal concerns, regulation inevitably creates challenges for innovation.

Therefore, **the contradiction between big data innovation and data protection** is one of the main challenges in this sphere [1]. For example, Article 9 of GDPR defines the special categories of personal data. The personal data that belongs to one of these categories is often referred to as sensitive data. Processing sensitive personal data is subject to even stricter rules. However, sensitive data can be instrumental in providing more personalized service and improving the life quality of the data subject (e.g., predicting and preventing a future disease by using the health data of a data subject). In addition to complying with the additional legal restrictions for processing sensitive personal data, the data subject must also be given additional privacy warranties to give consent for processing his sensitive personal data. Achieving a satisfactory level of privacy and security for **processing sensitive personal data** is one of the challenges of big data solutions [2]. Big data technologies can be used to limit free will and to manipulate data subjects with unethical profiling practices. They can create a systematic unfairness, especially in sensitive areas, which may damage ethical values such as equality, non-discrimination, digital inclusion. Especially when automatic decision making is utilized in these areas, explainability mechanisms must be integrated into the systems to guarantee transparency, accountability, and trustworthiness. Designing big data solutions that address **the societal and ethical implications of big data technologies** regarding human welfare, autonomy, non-maleficence, justice, accountability, trustworthiness, privacy, dignity, solidarity is an important and difficult challenge that developers face [3.1, p. 19].

In the current state of the art, once the data is shared with a third party, there is almost no guarantee for the secrecy of the data. Therefore, most data collectors prefer to keep their datasets strictly private, which hampers the efforts to create a data market and economy. To create a functional data market, technical standards and organizational measures based on legal norms & ethical principles must be adopted, and new solutions allowing **secure and trusted personal data sharing** must be developed. Technical solutions developed for this purpose should enable secure data transfer, and privacy-aware analytics with inexpensive computational costs and integration with the existing systems must be guaranteed [4]. Although there are many PPT solutions for anonymization and pseudonymization, most of these solutions have exploitable imperfections. Big data system designers must warrant the irreversibility of these solutions. While a rigid anonymization method may damage the value of the data, a light one may not be reliable since attackers can deanonymize the reversibly anonymized data. Therefore, another big data challenge is **to overcome the limits of anonymization and pseudonymization** [5.1, p. 66]. Accumulation of a high volume of data and careful analysis provides opportunities for knowledge and value creation. As the volume of data grows, the number of data sources also increases, which creates another challenge: **Dealing with multiple data sources and untrusted parties**. These sources must be integrated, and the data processors and controllers should be encouraged and provided with the correct solutions to share data across organizations for better services. However, efficiency and security issues are some of the obstacles which discourage data sharing across organizations. PPT-enabled analytics solutions and multiparty computation techniques must create a secure and reliable ecosystem to make data available for encrypted processing as an alternative to unprotected data sharing [6].

The development of PPT solutions enabling data subjects to set the limits for the collection, processing, and sharing of their personal data strengthens the enforceability of data protection regulations. These technologies should also facilitate the exercise of fundamental rights, such as the right to be forgotten. While designing these mechanisms, developers must ensure that they follow **a general, easy-to-use, and enforceable data protection approach**. In addition to these mechanisms, technical measures must be introduced to examine the conformity of the data processing systems to the data subjects' consents and the limits regarding their personal data [5.2, p. 66]. Besides keeping the services efficient and reliable, implementing these solutions cause another related challenge: Scalability of the solutions. Big data system designers must find efficient and scalable ways to implement PPT solutions such as encryption and anonymization. Therefore, another critical challenge that data processing systems face is to **ensure the reliability and efficiency of the services while maintaining robust data privacy**. [5.3, p. 66]. According to GDPR, one of the responsibilities of the data controller and processor is to ensure the safety of the processed data and continuously assess the risk of security issues. Therefore, another challenge in big data processing is **calibrating data controllers' obligations with a risk-based approach**. Especially when dealing with multiple datasets and data from multiple sources, the risks associated with data processing reaches an even higher level where data controllers must approach the risks with the utmost attention. Therefore, developing and utilizing tools to assess and prevent these risks constantly is one of the challenges that the data processing and AI community faces [5.4, p. 67]. GDPR requires data controllers to utilize appropriate technical and organizational measures for end-to-end data protection. These measures include (i) technical measures such as secure hardware enclaves, secure multiparty computation, encryption, & anonymization and (ii) organizational measures such as IT awareness training, auditing, and certification. **Combining different techniques for end-to-end data protection** might be very costly and cause performance overheads. Therefore, achieving end-to-end data protection is a challenge, which can only be overcome with careful planning and efficient optimization [7.1, p. 7-8].

3.2 Privacy-Preserving Technologies (PPT) and Organizational Measures for the Challenges in Big Data

Many potential PPT solutions and organizational measures are proposed in the literature to address the identified data privacy and protection challenges.

Explainable AI is a critical PPT solution that can be used to address the ethical and societal implications of big data technologies by enhancing transparency and fairness [7.2, p. 11] and help with GDPR compliance, particularly with regards to the right to explanation (*see GDPR Article 15.1.h*). Explainable AI aims to develop techniques to provide local and global explanations [8.1, p. 5]. Model agnostic (i.e., post-hoc) explainability techniques can be applied to any machine learning model. Ante-hoc techniques try to utilize specific model architectures to provide explanations [8.2, p. 5-7].

Secure multiparty computation (MPC) allows multiple parties to compute a joint function without having access to each other's data. Sensitive information is distributed among the parties, and individual shares do not reveal the actual sensitive information

by themselves. The parties complete the calculation on sensitive information without revealing each other this information. MPC can address several challenges, including maintaining robust data privacy with utility guarantees, securing trusted personal data sharing, and dealing with multiple data sources and untrusted parties [9.1, p. 285-286].

Distributed ledger technologies (e.g., blockchain) can achieve overcome some of the challenges identified in this report, such as developing generic, easy to use and enforceable data protection solutions, processing sensitive data with enhanced privacy, and maintaining robust data privacy without damaging the utility of the personal data [5.5, p. 66]. **Self-sovereign identity (SSI) management** is another privacy-preserving solution based on the concept that the users should be the sole owners of their identity data. With self-sovereign identity, the data subjects do not have to rely on an intermediary to verify their identity on a digital platform and create their own verifiable credentials [10]. **Homomorphic encryption** is a form of encryption, which allows calculations on encrypted data. Self-sovereign identity management and scalable homomorphic encryption solutions can be helpful to overcome big data challenges such as secure and trusted personal data sharing [9.2, p. 286].

Public release of datasets carries risks of identification even though they are carefully anonymized [11]. One effective PPT solution against the risks of the public release of datasets is **differential privacy**. Differential privacy improves privacy by perturbing the values (N) in the dataset with added random noise (L) and maintain a balance between utility and privacy. **Document sanitization and document redaction** are two similar PPT solutions developed to ensure that only the intended information can be accessed from a document. While document redaction consists of removing or blacking out sensitive information in a text document (e.g., AIDS →****), document sanitization consists of replacing sensitive information with more generic information (e.g., AIDS → disease) [9.3, p. 290]. These PPTs can be used with anonymization methods and provide additional privacy against de-anonymization. Therefore, they shine out as complementary remedies against the limitations of anonymization and pseudonymization techniques.

The **federated learning** approach proposes the decentralized model training across multiple edge devices or servers using their local data samples without access to external data sources [8.3, p. 4-5]. Federated learning approaches usually take advantage of other PPT solutions such as homomorphic encryption [12], secure multiparty computation (MPC), and differential privacy [13] to enhance privacy.

With the **sticky policies**, machine-readable policies can be added to the released data in a standard format (e.g., XML or JSON) to improve privacy and terms of use. Sticky policies can regulate the formats that data can be accessed, the way it can be used throughout its life cycle, and the limitations of its use and share [14]. Sticky policies can provide assurance to the data processor regarding how the data they release is used, which would increase the circulation of the data in a privacy-enhanced fashion. Therefore, they can strengthen the pace of big data innovations by providing additional privacy properties.

Auditing is an effective measure to identify the legal and ethical risks that the big data systems might carry [15]. **Algorithmic auditing** is an effort to develop solutions that automatically evaluate these systems and identify the risks in a streamlined fashion

[16]. By automating and standardizing the auditing process, data controllers and processors may analyze their services based on ethical & legal standards. They can remedy the issues without breaching data subjects' rights as soon as these issues (e.g., algorithmic bias, illegitimate profiling, and discrimination) are detected [3.2, p. 19].

Risk assessment tools can shield the data processor from outstanding risks with an early detection mechanism. They are a perfect answer to the risk-based data protection principle since they can measure the level of compliance to data privacy regulations (e.g., GDPR), identify privacy and cybersecurity risks (e.g., the reversibility of the anonymization mechanisms), recommend mitigations against these risks (e.g., differential privacy), and reveal accountability [5.6, p. 67].

The feasibility and applicability of techno-regulation is a hotly debated issue. While the widespread adoption of techno-regulation is under question, some fields offer unquestionable opportunities for techno-regulation implementations [7.3, p. 14]. Therefore, especially for these fields, **automated compliance** can be a powerful solution to address some of the challenges we face today, such as measuring and reducing data controllers' obligations [7.4, p. 11]. Implementation of techno-regulations can streamline and automate the legitimacy verification of all the processes within the lifecycle of personal data [7.5, p.15].]. With the advancement in these technologies, several protective protocols can be integrated into data processing systems, which requires compliance with the existing laws and valid certificates for processing. The main issue related to techno-regulation and automated compliance is identifying the limitations of the machine-readable policies to protect the data subjects' rights [5.7, p.55].

Data governance can be defined as rules for accessing and sharing personal data by taking into account privacy and data protection concerns. Data governance deals with the standardization of these procedures for sharing metadata, defining terms between stakeholders, and providing guidance on the use of privacy preserving technologies [5.8, p.26], such as encryption, pseudonymization, and anonymization for better privacy protection [5.9, p.88]. Privacy is a property of high-quality big data, and one of the main tasks undertaken by data governance is to ensure good data privacy and protection procedures are in place [5.10, p.56].

In addition to PPT solutions, another essential component of a successful data protection policy is to set and follow **ethical and technical standards and guidelines**. By involving the complete value chain of big data stakeholders, organizations can agree on and adopt standards and guidelines that reflect common ethical and societal values. The values identified in existing AI guidelines (e.g., transparency, justice & fairness, non-maleficence, responsibility, privacy, beneficence, freedom & autonomy, trust, sustainability, dignity, solidarity) can be utilized and improved to enhance the protection of legal and ethical values [17].

One of the organizational measures that data processors must undertake is coordinating the implementation and **integrating technical data protection measures**, including, but not limited to, the relevant approaches, toolboxes, overviews, and repositories of privacy-preserving technologies. Data processors and controllers can achieve a higher level of end-to-end data protection by combining different techniques [7.6, p.11].

4 Evaluations of Existing Solutions for Responsible Data Processing and ICT Systems

This paper reviews some of the notable Horizon 2020 projects that address the data processing challenges with the PPT solutions and organizational measures identified above. There are other very successful and significant projects addressing the big data challenges all around the world. However, since the focus in this report is on GDPR and its application, we limit the geographical scope of the analysis with European projects. This paper focused on the projects funded as per the Horizon 2020 program to narrow its scope further. To prepare this paper, 21 notable projects that adopt at least one of the solutions identified above are identified and analyzed. Table 1 lists these projects, the challenges they address, and the privacy-preserving technologies and organizational measures they utilize:

Table 1. Notable European PPT and Organizational Measure Projects with Their Main Focus Areas and the Challenges They Address

Project Abbreviation	The Addressed Challenges	PPTs and Org. Measures
BOOST 4.0	C6, C7	DLT, ETS
A4CLOUD	C2, C9	RAT, DG, AC
GenoMed4ALL	C1, C4, C6	XAI, FL
TRUSTS	C2, C3, C5	MPC, HE, ETS
RESTASSURED	C1, C3, C5, C9, C10	HE, SP, RAT, INT
DECODE	C4, C7, C10	DLT, INT
MUSKETEER	C1, C3, C7, C8, C10	FL, MPC, HE
SPECIAL	C1, C8	SP
MHMD	C3, C4, C10	DLT, INT
AEGIS	C3, C6	DLT
BPR4GDPR	C5, C7, C8, C9, C10	AC, AA, RAT, INT
PAPAYA	C1, C3, C6, C7, C8	MPC, AA
SMOOTH	C2, C7, C10	AA, RAT, INT
PDP4E	C10	AA, RAT, DG, INT
DEFeND	C7, C10	DG, ETS, RAT, INT
MOSAICrOWN	C5, C6, C10	DG, DSR, INT
SODA	C3, C4, C6	DP, MPC
PoSeID-on	C2, C9	DLT, RAT
E-SIDES	C2	ETS
LINDDUN	C7, C10	RAT, ETS, INT
XAI	C2	XAI, ETS

Privacy-Preserving Technologies and Organizational Measures: XAI: Explainable AI • MPC: Secure Multiparty Computation • SSI: Self-sovereign Identity (SSI) Management • HE: Homomorphic Encryption • DP: Differential Privacy • DSR: Document Sanitization and Redaction • FL: Federated Learning Approaches • DLT: Distributed Ledger Technologies and Blockchain • SP: Sticky Policies • AA: Algorithmic Auditing • RAT: Risk Assessment Tools • AC: Automated Compliance • DG: Data Governance • ETS: Ethical and Technical Standards, Guidelines, Laws, and Codes of Conduct • INT: Integration of Approaches, Toolboxes, Overviews, and Repositories of Privacy-Preserving Technologies

The Challenges Identified in the Previous Section: C1: Contradiction between Big Data innovation and data protection • C2: Societal and ethical implications of big data technologies • C3: Secure and trusted personal data sharing • C4: Processing sensitive data • C5: Limits of anonymization and pseudonymization •

C6: Dealing with multiple data sources and untrusted parties • C7: A general, easy to use and enforceable data protection approach • C8: Maintaining robust data privacy with utility guarantees • C9: Risk-based approaches calibrating data controllers' obligations • C10: Combining different techniques for end-to-end data protection

The projects examined under this section have a wide range of starting dates. While the earliest of them all, A4CLOUD, started in October 2012, the most recent project, GenoMed4ALL, started in January 2021. The median starting date for the projects is January 2018. The starting date of the project is a significant indicator of its focus area. Among these 21 projects, while the initial theme was the adoption of blockchain technology until 2018, after 2018, we see a redistribution of the themes from risk assessment tools to encryption techniques. Finally, the most recent project, GenoMed4ALL, priorities explainable AI and federated learning, which might be the starting point of a new trend. The PPT and organizational measure solutions and the total number of projects that use these solutions are shared below:

Table 2. The Frequency Table of the Adoption of the PPT Solutions and Organizational Measures in the Notable Horizon 2020 Projects

Abb.	PPT or Org. Measure	#Projects
INT	Integration of Approaches, Toolboxes, Overviews, and Repositories of PPT	9
RAT	Risk Assessment Tools	8
ETS	Ethical and Technical Standards, Guidelines, Laws, and Codes of Conduct	6
DLT	Distributed Ledger Technologies and Blockchain	5
MPC	Secure Multiparty Computation	4
AA	Algorithmic Auditing	4
DG	Data Governance	4
HE	Homomorphic Encryption	3
FL	Federated Learning Approaches	2
SP	Sticky Policies	2
AC	Automated Compliance	2
XAI	Explainable AI	2
DP	Differential Privacy	1
DSR	Document Sanitization and Redaction	1
SSI	Self-sovereign Identity (SSI) Management	0

Table 2 shows that the Integration of Approaches, Toolboxes, Overviews, and Repositories of PPTs is the most preferred solution for big data issues, as nine of the 21 projects offer relevant solutions. Therefore, most projects choose to integrate existing privacy-preserving technologies and offer data privacy and protection solutions in a particular field such as medicine [18] or SMEs [19.1]. Apart from integrating existing technologies, eight projects develop risk assessment tools for GDPR compliance. Additionally, we see a high adoption ratio of DLT and Blockchain technologies, especially in relatively older projects. Although there are five projects developing solutions in "ethical and technical standards, guidelines, laws, and codes of conduct," this category does not seem saturated relative to its broad scope. Finally, although some PPT solutions such as explainable AI, differential privacy, document sanitization and redaction, federated learning, and multiparty computation can be helpful to address the limitations of anonymization, pseudonymization, and encryption techniques, **these promising**

PPT solutions are mostly disregarded. One explanation for this mismatch can be the infancy of these technologies.

Table 3. The Frequency Table of the Coverage of the Data Privacy and Data Protection Challenges in the Selected Projects

Abb.	Challenge	#Projects
C10	Combining different techniques for end-to-end data protection	10
C7	A general, easy to use and enforceable data protection approach	8
C3	Secure and trusted personal data sharing	7
C6	Dealing with multiple data sources and untrusted parties	6
C2	Societal and ethical implications of big data technologies	6
C1	The contradiction between Big Data innovation and data protection	5
C4	Processing sensitive data	4
C5	Limits of anonymization and pseudonymization	4
C8	Maintaining robust data privacy with utility guarantees	4
C9	Risk-based approaches calibrating data controllers' obligations	4

For the challenge categories covered in the notable data protection projects list, in parallel with the integration efforts, Table 3 shows that most projects aim to combine different techniques for end-to-end protection. This observation is in line with the trend that **most data privacy and data protection projects aim to combine existing PPT solutions and offer a platform for businesses in a particular field**. The second most popularly addressed challenge (i.e., adopting a general, easy to use and enforceable data protection approach) also supports this thesis. Since most companies do not have in-house data protection expertise, most projects try to standardize and facilitate the adoption of GDPR-compliant data processing platforms. While some projects aim to integrate PPT solutions into existing infrastructures already in the market [20.1], others develop their own platforms by combining PPT solutions with other technologies [21].

After these two closely related challenge categories, secure and trusted personal data sharing and dealing with multiple data sources and untrusted parties are the following two critical challenges. After identifying that there is ever-increasing data flow, data generated, and the data sources, around a third of the Horizon 2020 projects aim to address the secure transfer of data and dealing with multiple data sources. While multiparty computation and homomorphic encryption support secure data transfer and enable privacy-enhanced data analytics and machine learning among untrusted parties, sticky policies, integration efforts, and ethical & technical standards help to deal with multiple data sources.

Examination of the notable projects showed that **the least addressed challenges are the limits of anonymization and pseudonymization, maintaining robust data privacy with utility guarantees, and risk-based approaches calibrating data controllers' obligations**. Although many projects offer secondary solutions to these challenges, most projects do not prioritize them. Although identifiability is a significant indicator of personal data, and most anonymized and pseudonymized data can be reversed for identification, there is not enough emphasis on addressing this problem. Additionally, even though there are projects aiming to develop dynamic consent and sticky policy solutions, the number of solutions in this field is also minimal.

Another significant finding is that six projects actively aim to address big data technologies' adverse societal and ethical implicants. Considering the wide variety of societal and ethical implications of big data technologies, this number seems very small. On the one hand, each project at least mentions partially cover ethical and legal frameworks in their deliverable documents. However, partial mentions are only a determinant for being included in this review paper, not an indicator that these projects address this challenge in a comprehensive manner. **The number of projects that particularly emphasize the ethical and societal implications of big data systems is minimal, and their number should increase**.

One of the most frequently mentioned issues in the field of data protection is the user-centered data protection approach. This issue is usually covered as **giving control of the data back to its owner**. In many projects, we see platform proposals in which the data subject can dynamically edit the data processing consent. The SPECIAL project utilizes sticky policies attached to data in metadata form so that the data receivers would always know to what extent they can use and share the personal data [1.2]. The MOSAICrOWN project also provides deployable solutions that allow data owners to maintain control of the data sharing process [22].

There has been a discussion on whether it is healthy to automate regulatory actions or hardcoding laws is a dangerous practice because the legal field is argumentative and dynamic [7.7, p.14]. While this theoretical discussion continues, many projects try to achieve **automation in data privacy compliance, auditing, and risk assessment**. While SPECIAL aims to achieve automated compliance by utilizing sticky policies [1.3], BPR4GDPR [20.2], SMOOTH [19.2], and PDP4E [23] projects propose automated auditing with risk assessment tools and data governance frameworks.

As we approach the mass adoption of AI in sensitive fields, one of the hotly debated data privacy issues is right to explanation [24]. Utilizing AI systems in sensitive fields may cause the violation of several ethical principles such as non-discrimination, transparency, and accountability [25]. Although explaining the decisions of AI systems is an essential issue for the ethical and societal implications of big data technologies, **explainability techniques are hardly mentioned in notable projects**. Out of 21 examined projects, only two projects, GenoMed4ALL [26] and XAI [27], offer explainable AI solutions.

5 Conclusion

Designing and implementing GDPR-compliant solutions for big data processing is a challenging task that requires expertise in legal, ethical, and technical domains. This paper identifies the leading data processing and big data challenges relevant to data protection and privacy. After identifying the legal obligations of the data processors and controllers and the challenges, the privacy-preserving technologies (PPTs) that address these obligations and issues are explained. To have practical and useful results, 21 notable Horizon 2020 projects are analyzed and examined. After analyzing these projects, findings of the quantitative analysis of the challenges and the solutions are

shared with an emphasis on the issues that require additional attention. While the integration of existing tools is selected by most of the projects as the primary goal, the development of novel PPT solutions is usually disregarded.

The most important conclusion of the review is that there is a clear need for interdisciplinary research projects that focus on developing novel PPT solutions (e.g., explainable AI techniques, sticky policies, homomorphic encryption) for enhanced data protection practices in line with legal norms & ethical principles.

References

1. SPECIAL. Home. Retrieved January 30, 2021, from https://www.specialprivacy.eu/
2. Rizzo, A. (2017). MHMD Project Presentation. In My Health My Data, 4. http://www.myhealthmydata.eu/deliverables/D11.2-MHMD-Project-Presentation.pdf.
3. Custers, B., La Fors, K., Jozwiak, M., Esther, K., Bachlechner, D., Friedewald, M., & Aguzzi, S. (2018). Lists of Ethical, Legal, Societal and Economic Issues of Big Data Technologies. *SSRN Electronic Journal*, 19. https://doi.org/10.2139/ssrn.3091018
4. Markopoulos, I. (2020). Industry specific requirements analysis, definition of the vertical E2E data marketplace functionality and use cases definition I, 11. https://trusts-data.eu/
5. European Big Data Value Association. (2017). Strategic Research and Innovation Agenda. European Big Data Value, 4(October), 66. https://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf
6. Veeningen, M. (2020). SODA - Scalable Oblivious Data Analytics. SODA Project. https://soda-project.eu/
7. Timan, T. & Z. Á. M. (eds). (2019). Data protection in the era of artificial intelligence. Trends, existing solutions and recommendations for privacy-preserving technologies, 7-8.
8. Budig, T., Herrmann, S., Dietz, A., Pandl Supervisor, K., & Sunyaev, A. (n.d.). Trade-offs between Privacy-Preserving and Explainable Machine Learning in Healthcare, 5. Retrieved February 1, 2021, from www.kit.edu
9. Domingo-Ferrer, J., & Blanco-Justicia, A. (2020). Privacy-Preserving Technologies. In International Library of Ethics, Law and Technology (Vol. 21, pp. 279–297), 285-286. Springer Science and Business Media B.V. https://doi.org/10.1007/978-3-030-29053-5_14
10. Allen, C. (2016). The Path to Self-Sovereign Identity. Life With Alacrity. http://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html
11. Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. Proceedings - IEEE Symposium on Security and Privacy, 111–125. https://doi.org/10.1109/SP.2008.33
12. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated Machine Learning: Concept and Applications. In ACM Trans. Intell. Syst. Technol (Vol. 10), 12:4. https://doi.org/
13. Truex, S., Steinke, T., Baracaldo, N., Ludwig, H., Zhou, Y., Anwar, A., & Zhang, R. (2019). A hybrid approach to privacy-preserving federated learning. Proceedings of the ACM Conference on Computer and Communications Security, 1–11, 1. https://doi.org/10.1145/3338501.3357370
14. Pearson, S., & Casassa-Mont, M. (2011). Sticky policies: An approach for managing privacy across multiple parties. Computer, 44(9), 60–68, 60. https://doi.org/10.1109/MC.2011.225
15. Deborah Raji, I., Smart, A., White Google Margaret Mitchell Google Timnit Gebru Google Ben Hutchinson Google Jamila Smith-Loud Google Daniel Theron Google Parker Barnes Google, R. N., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron,

D., & Barnes, P. (2020). Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing ACM Reference Format, 1. https://doi.org/10.1145/3351095.3372873
16 .Kassir, S. (n.d.). Algorithmic Auditing: The Key to Making Machine Learning in the Public Interest. The Business of Government, 1–4. http://www.businessofgovernment.com/sites/default/files/Algorithmic Auditing.pdf
17. Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence: the global landscape of ethics guidelines, 7. In arXiv.
18. MHMD. (2019). My Health My Data. In My Health My Data. http://www.myhealthmydata.eu/
19. SMOOTH. (n.d.). About Smooth Project. Retrieved March 20, 2021, from https://smoothplatform.eu/about-smooth-project/
20. BPR4GDPR. (n.d.). Innovation Proposal. 48. Retrieved March 20, 2021, from https://www.bpr4gdpr.eu/about/research-description/
21. DEFeND. (n.d.). What is the Defend Project - Defend Project. Retrieved March 20, 2021, from https://www.defendproject.eu/
22. MOSAICrOWN. (n.d.). Homepage. Retrieved March 21, 2021, from https://mosaicrown.eu/
23. Yod, S. M. (2019). PDP4E - D 2.4 Overall system requirements. https://www.pdp4e-project.eu/deliverables/
24. Sartor, G. (European U. I. of F. (2020). The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. Panel for the Future of Science and Technology (STOA), 1st, 76-79. https://doi.org/10.2861/293
25. Yalçın, O. G. (2020). Examination of current AI systems within the scope of right to explanation and designing explainable AI systems, 2-3. CEUR Workshop Proceedings, 2598. https://www.academia.edu/44158508/Examination_of_Current_AI_Systems_within_the_Scope_of_Right_to_Explanation_and_Designing_Explainable_AI_Systems
26. GenoMed4All. About. 2021. Retrieved March 20, 2021, from http://genomed4all.eu/about/
27. XAI. Research lines. 2021. Retrieved May 30, 2021, from https://xai-project.eu/research-lines.html/