Compliance by Design Methodologies in the Legal Governance Schemes of European Data Spaces

Kossay Talmoudi, Khalid Choukri, Isabelle Gavanon
ELDA, DELCADE
9 Rue des Cordelières, 75013 Paris, 19 Rue du Colisée, 75008 Paris
{Kossay, Khalid}@elda.org, Igavanon@delcade.fr

Abstract

Creating novel ways of sharing data to boost the digital economy has been one of the growing priorities of the European Union. In order to realise a set of data-sharing modalities, the European Union funds several projects that aim to put in place Common Data Spaces. These infrastructures are set to be a catalyser for the data economy. However, many hurdles face their implementation. Legal compliance is still one of the major ambiguities of European Common Data Spaces and many initiatives intend to proactively integrate legal compliance schemes in the architecture of sectoral Data Spaces. The various initiatives must navigate a complex web of cross-cutting legal frameworks, including contract law, data protection, intellectual property, protection of trade secrets, competition law, European sovereignty, and cybersecurity obligations. As the conceptualisation of Data Spaces evolves and shows signs of differentiation from one sector to another, it is important to showcase the legal repercussions of the options of centralisation and decentralisation that can be observed in different Data Spaces. This paper will thus delve into their legal requirements and attempt to sketch out a stepping stone for understanding legal governance in data spaces.

Keywords: Data space, compliance, legal governance

1. Introduction

As the data economy is developing at an unprecedented pace, major regulatory changes have operated on the European level to maximise the value of data while regulating how many actors collect, use, and share it.

In this sense, the European Commission introduced a high-level European Data Strategy in February 2020 that introduced, among others, a landscape compiled of European Common Data Spaces that harness the data potential in various key sectors such as health, agriculture, and language.

Data Spaces are set to be one of the catalysers of data innovation. However, these Data Spaces require a set of legal standards that need to be taken into account as early as possible in the process of their conception, thus ensuring their sustainability and legal compliance in the long term.

Some of the major legal texts regulating data exchanges in the European Union are the Data Governance Act ("**DGA**") [1], the Data Act ("**DA**")[2], and the Open Data Directive[3]. These texts refer to the importance of complying with horizontal obligations stemming, among others, from the General Data Protection Regulation ("**GDPR**")[4], the E-Privacy Directive[5], the Copyright in the Digital Market Directive[6], and the proposed AI Act. Compliance with these texts requires a "by design" methodology that needs to guide the conceptualisation of Data Spaces as it is indispensable in the mission of fostering data potential responsibly.

Organisational modalities need to be put in place in order to unlock the data potential while respecting the obligations laid down in the different legal texts. There is thus an important need for a practice-facing methodology to allow dialogue between technicians building the Data Spaces and lawyers who are expected to proactively accompany the building of Data Space. Therefore, the legal obligations and frameworks governing the transfer of data need to be translated into actionable tools.

This paper tries to answer the question as to how a practical organisation of compliance be implemented in the framework of Common European Data Spaces. In fact, the data collected needs to be compliant among others, with the GDPR, trade secrets, and copyright law, among other obligations. In order to analyse this, it is important to identify the scope of the compliance with applicable law before identifying debtors of such warranties.

The legal obligations need to be laid down, not only in the documentation but also in the processes that allow for data transfers thus reflecting the operationalisation of these obligations and allowing for going beyond the "box-checking exercise" to rather dynamic compliance aligned with state-of-the-art technologies.

Legal rules would apply differently depending on the range of services offered by the Data Space and answers to operational questions are a prerequisite to the analysis. These questions include the creation of value in data spaces, their economic model, their definition as a marketplace or as an intermediation service. The

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A52020PC0767
[2] https://eur-lex.europa.eu/eli/reg/2023/2854
[3] https://eur-lex.europa.eu/eli/dir/2019/1024/oj
[4] https://eur-lex.europa.eu/eli/reg/2016/679/oj
[5] https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32002L0058
[6] https://eur-lex.europa.eu/eli/dir/2019/790/oj

questions regarding the architecture are also of great relevance. In this sense, it is important to know whether the data space is set to centralise all data or merely hold a catalogue of metadata, and to understand the services it would offer.

This paper intends to focus mainly on data governance and not on data space governance, which shall be addressed in a second step, once the legal constraints applicable to the data are identified.

2. Impact of the various applicable rules

2.1. Commitments to respect the third parties' rights to data in compliance with the requirements of the Data Act and the Data Governance Act

- **Reconciling the rights of third parties with the uses envisaged**

The DGA presents chapters covering specific types of data, such as data from public sector institutions that are not covered by the Open Data directive obligations in accordance to them containing rights of third persons, including copyright, trade secret, and personal data. The Data Act in its Chapter sets out standards of data sharing in business-to-business and consumer-to-business framework. It notably presents obligations regarding data created through the use of Internet of Things (IoT) devices.

In its explanatory memorandum of the DGA, the European Commission states that the general objectives of such legislation is: "(*1) making public sector data available for re-use in situations where such data is subject to rights of others (such as privacy rights, IP rights, trade secrets, or other commercially sensitive information); (2) sharing data among businesses, against remuneration in any form; (3) allowing personal data to be used with the help of a personal data sharing intermediary that safeguards data subjects' rights under the GDPR; and (4) allowing data use on altruistic grounds*"

According to article 3 DGA, rules on re-use apply to *"data held by public sector bodies which are protected on grounds of: (a) commercial confidentiality, including business, professional and company secrets;(b) statistical confidentiality; (c) the protection of intellectual property rights of third parties; or (d)the protection of personal data"*

The obligations to respect third-party rights are horizontal in the new legislative landscape pertaining to data in the EU. These obligations are considered as a cornerstone of responsible data sharing and data services. This covers novel types of entities created by the DGA such as data intermediaries and data altruism organisations whose scope and role in the various data spaces is in the process of being defined.

- **Contractual fairness in the Data Act**

Article 8 of the Data Act states that: *"Where a data holder is obliged to make data available to a data recipient under Article 5 or under other Union law or national shall do so under fair, reasonable and non-discriminatory terms and in a transparent manner"*

The data act thus prohibits clauses that are deemed unfair in a business-to-business contract when the object of the contractual relationship is access and use of data. Clauses are deemed unfair if they limit liability for intentional acts or gross negligence that may bring prejudice to the contracting party. Some clauses can also be deemed unfair if a dominant party to the contract puts in place limitations on the other party regarding use or reuse of data during the period of the contract.

These limitations need to be integrated in the functioning of the data space through the standardisation of contract templates.

2.2. Data not covered by the DA and DGA

- **Data collected through text and data mining**

The Text and Data Mining exception incorporated in the 2019 Copyright in the Digital Market Directive aims to create a possibility of using copyrighted material that is accessed lawfully, either through subscriptions, open licences, or online availability, to extract new data while balancing the interests of rights holders.

A permissive Article 3 is limited to research organisations and cultural heritage institutions, while Article 4 is broader, allowing TDM by any entity. However, from the latter usage, rights holders can signal an opt-out that makes it impossible to mine the data.

The EU's soon-to-be AI act is noteworthy in its explicit connection between the use of copyrighted works for training AI models and the text and data mining exception outlined in Article 4 of the 2019 Copyright Directive. This entails an obligation of compliance with the opt-out option. The exceptions thus represent an important step in adapting EU copyright law to the development of AI technologies. It is therefore important that the data transactions in the data space are subject to due diligence when it comes to respecting opt-out options.

- **All other data**

Data and data sets can otherwise be protected under copyright or the database sui generis right

stemming from the EU Database Directive 96/9/EC[7] when it comes to intellectual property mechanisms and GDPR when it comes to data that contains personal data. As well as data that is covered by trade secrets, in this sense, templates of confidentiality agreements can be proposed in order to align with the rights of third parties.

3. Warranties of data compliance with applicable laws

The various definitions of a data space allow a level of freedom of interpretation of its scope as they can be different depending on the sectoral needs.In this sense, the definition given by the Data Spaces Support Center (DSSC) of a Common European data space is “A sectoral/domain-specific data spaces established in the European single market with a clear EU-wide scope that adheres to European rules and values.” Data spaces can therefore have different architectures and can centralise the data or not.

3.1 Warranting compliance in a decentralised data space architecture

One of the fundamental repercussions of the architecture is the different levels of responsibility that data spaces have in regard to data made available. In the framework of a decentralised data space, the data space operator does not centralise the data that transits through it. Rather, the data space acts as an enabler of transactions, generally via a central catalogue of metadata.

The data space operator does not assume this responsibility on behalf of the participants. Each participant in the data space must take full responsibility for their compliance. However, the data space operator can intervene in the framework of decentralised data spaces. This cannot be conceived as the entity that would be liable if non-compliant data transmits through it, rather it would mean that the data space operator offers help towards compliance as a service. In this sense, this service can be done through the provision of templates of compliance documentation and templates related to contractual transactions that can be adopted through data audit services that can be plugged into the data space. It is to be highlighted that in such cases where the data is not held and, by principle, not accessed by the data space operator, it is the data space participant who would define what data is exchanged, and in the case of data containing personal data, how the data is processed and for which objectives; thus coinciding with the “*means and purposes*” of GDPR.

In its guidelines 4/2019 on Article 25 Data Protection by Design and by Default (DPbDD)[8], the EDPB states that: *”Although not directly addressed in Article 25, processors and producers are also recognized as key enablers for DPbDD, they should be aware that controllers are required to only process personal data with systems and technologies that have built-in data protection.”* The fact that the decentralised data space has no control over the data therefore does not exempt it from the expectations of complying with data protection by design and by default considerations as it enables eventual data processing activities.

Data governance arrangements should be in place at data spaces entry and exit points, based on the "compliance by design" principle, to optimise compliance with data (personal, non-personal, copyrighted…) protection laws, reduce the cost of compliance (barriers to entry) and gain efficiency.

For example, this is done via the verification that no data transfers allowed by the data space include personal data that has not gone through due diligence by the data participants. This is to allow the free flow of personal and non-personal data while respecting European values.

By clearly delineating the roles and responsibilities, the ecosystem ensures that all participants are aware of their obligations and can make informed decisions about their involvement in the data space, particularly about handling personal data.

The data space operator may put in place rules that close the space to non-compliant data. Using a code of conduct, it is possible to state that only data assessed by ex-ante against data protection, copyright, and trade secret considerations can transit via the data space It is notable that this does not necessarily imply a legal obligation of compliance from the data space operator’s side when it is a decentralised architecture; however, it is evident that providing obligations of compliance to the exchanged data constitutes an added value for reusers who can be aware that the data was collected lawfully.

In fact, the data space operator plays a crucial role in facilitating compliance for data providers, enabling them to focus on the core aspects of data sharing and utilisation within the data space ecosystem. By offering standardised tools, the data space operator can significantly reduce the burden on data providers when it comes to implementing compliance measures. This can encourage more data providers to participate in the data space, as the administrative overhead is minimised. As the data spaces are a soft infrastructure, they can also create value via the provision of compliance services.

[7]https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31996L0009

[8]https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf

A similar analysis should be made regarding compliance with other bodies of rules such as copyright law or protection of trade secrets.

3.2 Warranting compliance in a centralised data space architecture

In this sense, although more and more data space initiatives are taking up that route, data spaces are not necessarily decentralised. This corresponds to a central platform that aggregates the data brought to the table by participants. Where the data space is an entity that centralises the data that will be the subject of transactions, may they be in the framework of selling, lending, or sharing on an altruistic basis, responsibilities in regard to compliance with relevant legislation can differ.

It is to be noted here that compliance with the rights of third parties becomes, among others, the responsibility, not only of the data space participant but also of the data space operator who centralises the data. In this sense, it becomes adequate to conceptualise the data space participants as controllers of data, where they can be individual controllers or joint-controllers on the basis of Article 26 GDPR. In this architecture, the data space operator would be generally conceived as a data processor that would therefore have specific legal responsibilities.

It is also important to assess the centrality of the data space against the risks of having unfair advantages towards data participants. As a centralised data space is naturally more closed and less penetrable by eventual data participants, this may result in unfairness that is proscribed by the Data Act.

4. Implementing mechanisms to automate compliance warranty in data spaces

4.1 Metadata interoperability

Data sharing requires incentives for data holders to develop metadata to increase certainty about rights and obligations in relation to data; indeed, efficient metadata management, with the development of standards for semantic and technical interoperability, is of utmost importance. Therefore, Article 28 of the Data Act provides that "*Operators of data spaces shall comply with, the following essential requirements to facilitate interoperability of data, data sharing mechanisms and services: a) the dataset content, use restrictions, licences, data collection methodology, data quality and uncertainty shall be sufficiently described to allow the recipient to find, access and use the data;[…]*"

Special focus on these matters shall be made for an optimal allocation of data to the benefit of society. These cross-sectoral interoperability standards, will be one of the responsibilities of the European Data Innovation Board (EDIB) as one of its missions is proposing guidelines for common European data spaces pursuant to Article 30 of the DGA.

4.2 Integration of Distributed Ledger Technology and Smart Contracts

The integration of Distributed Ledger Technology (DLT) into data spaces can have many benefits for data space participants. By integrating DLT into the data space the ecosystem can foster greater trust, transparency, and efficiency in the exchange and utilisation of valuable data assets.

DLT, mostly reflected in blockchain, empowers data providers as it significantly enhances the value proposition within the data space ecosystem. By creating a permanent, tamper-proof record of data contributions, blockchain can ensure that sellers' efforts are irrefutably acknowledged and their intellectual property rights are protected.

Similarly, smart contracts enabled by DLT can automate the process of compensating data providers upon the fulfilment of predefined conditions, ensuring timely and fair remuneration for their data assets.

DLT's transparency and immutability allow data providers to conclusively prove the origin and ownership of their data, while also guaranteeing that the information has not been altered from its original form, preserving its quality and reliability. As data transactions are recorded on the distributed ledger, providers can establish a verifiable track record of their contributions, which can lead to increased business opportunities and enhanced pricing power over time. Data users can easily verify the complete history of a data asset, including its origins, ownership changes, and any relevant licensing terms, reducing administrative overhead.

Smart contracts can streamline the process of obtaining the necessary rights to use the data, further enhancing efficiency and reducing the risk of licensing disputes. The decentralised architecture of DLT makes it significantly more challenging to engage in fraudulent activities, as altering recorded data would require consensus across the entire network.

Moreover, DLT's inherent security features can provide an additional layer of protection for the storage and transaction of sensitive language data, mitigating the risks of unauthorised access or tampering.

5 Conclusion

By aligning best practices and harmonising regulatory requirements across different data space initiatives, a cohesive and interoperable compliance framework emerges thus resulting in

data spaces where data can flow freely and securely.

As data spaces emerge as infrastructure enabling the exchange and utilisation of valuable information assets, the need for a comprehensive compliance framework becomes paramount. This is translated through the integration of compliance services into the foundation of the data space infrastructure

At the heart of this compliance-driven approach lies the imperative to safeguard compliance of the data to applicable laws.

As the data space ecosystem continues to evolve, compliance-driven approaches need to mutate and be flexible enough to respond to the needs of data space participants.

6 Bibliographical references

- Duisberg, A., "Legal Aspects of IDS: Data Sovereignty-What Does It Imply?." *Designing Data Spaces* 61 (2022).
- Ruohonen, J., Mickelsson, S. Reflections on the Data Governance Act. *DISO* **2**, 10 (2023).
- Margoni, T., Ducuing, C., Schirru, L., Data Property, Data Governance and Common European Data Spaces. Computerrecht: Tijdschrift voor Informatica, Telecommunicatie en Recht, (2023)