ISSN 1831-9424

European Commission

JRC SCIENCE FOR POLICY REPORT

EUROPEAN DATA SPACES

Scientific insights into data sharing and utilisation at scale

2023

Farrell, Eimear; Minghini, Marco; Kotsev, Alexander; Soler-Garrido, Josep; Tapsall, Brooke; Micheli, Marina; Posada, Monica; Signorelli, Serena; Tartaro, Alessio; Bernal, Jaime; Vespe, Michele; Di Leo, Margherita; Carballa-Smichowski, Bruno; Smith, Robin; Schade, Sven; Pogorzelska, Katarzyna; Gabrielli, Lorenzo; De Marchi, Davide

EUR 31499 EN

Joint Research Centre

Contact information
Name: Alexander Kotsev
Address: via E. Fermi, 2749 – TP 263 - 21027 Ispra (VA), Italy
Email: alexander.kotsev@ec.europa.eu / jrc-data-spaces-kb@ec.europa.eu
Tel.: +39 0332 78 9069

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC129900

EUR 31499 EN

Print	ISBN 978-92-68-03166-7	ISSN 1018-5593	doi:10.2760/301609	KJ-NA-31-449-EN-C
PDF	ISBN 978-92-76-53522-5	ISSN 1831-9424	doi:10.2760/400188	KJ-NA-31-449-EN-N

Luxembourg: Publications Office of the European Union, 2023

How to cite this report: Farrell, E.; Minghini, M.;Kotsev, A.; Soler-Garrido, J.; Tapsall, B.; Micheli, M.; Posada, M.; Signorelli, S.; Tartaro, A.; Bernal, J.; Vespe, M.; Di Leo, M.; Carballa-Smichowski, B.; Smith, R.; Schade, S.; Pogorzelska K.; Gabrielli, L.; De Marchi, D., *European Data Spaces: Scientific insights into data sharing and utilisation at scale*, Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/400188, JRC129900.

Contents

Abstract ... 4

Foreword ... 5

Acknowledgements ... 6

Executive summary ... 7

1 Introduction ... 10

1.1 Policy context: a Europe fit for the digital age ... 10

1.1.1 What is a common European data space? ... 12

1.1.2 Principles, requirements and features of a data space ... 13

1.1.3 Sector- and domain-specific common European data spaces ... 14

1.1.4 Key actors ... 17

1.2 A scientific perspective on setting up, operating and maintaining common European data spaces ... 18

1.2.1 Scope and intended audience ... 18

1.2.2 Synergies with other relevant initiatives ... 18

1.2.3 Components of the report ... 18

2 Identification of requirements for common European data spaces ... 20

2.1 High-level requirement identification ... 21

2.2 Technical perspective ... 23

2.3 Governance perspective ... 26

2.3.1 Goals of Governance: guidance and guardrails ... 26

2.3.2 Governance: principles, practices and processes ... 29

2.3.3 Data Governance ... 29

2.3.4 Governance within and between common European data spaces ... 31

2.4 Relevant data sharing initiatives and component reuse ... 33

3 Data sharing How-to guides ... 35

4 JRC dashboard on European data spaces ... 71

5 Discussion and conclusions ... 73

List of abbreviations ... 78

Abstract

This Science for Policy report distils technical and organisational lessons learned from the scientific work of the Joint Research Centre (JRC) that can inform the scoping and implementation of common European data spaces as envisioned by the European strategy for data. Those lessons stem from the JRC's long term commitment and relevant research on topics relating to data sharing and use, as well as on digital infrastructures and emerging technologies. The audience targeted by this report comprises policy officers and domain experts that are currently working on, or expect to work on, the scoping and implementation of common European data spaces. The report introduces the policy context, and positions the JRC's evidence as an instrument for translating policy conceptualisations and theoretical understandings of data spaces into a practical framework for implementation. This is followed by a synthesis of the requirements for common European data spaces, derived from relevant EU policy documents. A set of data sharing How-to guides is provided, in addition to a description of the dedicated JRC Knowledge Base that corresponds to the functional and non-functional requirements of common European data spaces. In the investigation of the technical and governance aspects of common European data spaces and acknowledging sectoral specificities, it becomes clear that there is no single approach that can be applied for their establishment. However, in addition to horizontal principles and requirements, a minimum stack of protocols and specifications can be employed in a decentralised and federated manner. Therefore a community-based approach through co-creation and co-design of common European data spaces, which considers the domain-specific context is the only feasible way forward that would ensure buy-in by a broad spectrum of stakeholders. Finally, scientific evidence should continue to play a central role in the conceptualisation and operationalisation of common European data spaces.

Foreword

The 'Europe fit for the digital age' priority that President von der Leyen's Commission has embarked on is ambitious. Our challenge is to leverage the huge opportunities that modern digital technologies provide without putting our fundamental rights and freedoms at risk or making our environment pay. At the same time, this human-centred digital transformation, "Made-in-EU", should not be based on dependencies that hamper the competitiveness of our Digital Single Market.

Clearly, in order to achieve such ambitious objectives, access to data is of paramount importance. The European strategy for data aims to make Europe a leader in a data-driven society. If data are not available and easily reusable across borders and in different application domains, the agenda for the digital transformation cannot be rolled-out successfully. We need to facilitate the reuse of data shared by businesses, citizens, public authorities, academia and other organisations, and build common European data spaces with an inclusive approach. This will lead to trust, guarantee fairness and improve relationships among actors of the data economy.

In my view, we can offer a lot. As the scientific arm of the European Commission, at the Joint Research Centre (JRC) we are committed to providing our expertise and research findings to implement and sustain the common European data spaces. As a centre of excellence, JRC has been deeply involved in data sharing since its inception. Our expertise covers multidisciplinary topics such as data harmonisation, interoperability, standardisation, artificial intelligence, economic modelling and social impacts of digital transformation, and data governance. Our open data portal alone contains more than 3200 readily available and well documented datasets. In addition, the JRC publications repository holds more than 500 articles that deal explicitly with various technical, social, economic or organisational aspects of data sharing and use.

This Science for Policy report brings together these dispersed assets and targeted scientific evidence, embedding our long-term commitment to data-driven innovation, which is highly relevant for the development of common European data spaces.

I would also like to emphasise that in addition to research on data sharing, our scientific and evidence-based policy work will tremendously benefit from the pan-European single market for data. In the JRC, together with our partners in the other European Commission services, we are committed to making the common European data spaces initiative a success.

Bernard Magenhann

Deputy Director General
in charge of Digital Transformation
and Data, European Commission
Joint Research Centre (JRC)

Acknowledgements

The authors express their sincere gratitude to the following for their advice and an early reading of the report: Maria Rosaria Coduti, Johan Bodenkamp, Federico Milani, Andrea Halmos, Seth van Hooland, Alexandra Balahur, Carlos Torrecilla Salinas, Michael Lutz, Eva Martinez Rodriguez, Jordi Escriu, Andrea Musumeci, Pierpaolo Cira, Loizos Bailas, Karen Fullerton, Miriam Giubilei and Francesca Siciliano.

The authors would also like to thank Koen Jonkers, Editor in Chief of the JRC Editorial Review Board, Jacopo Grazzini and Charles Heard, whose review of the report greatly helped the authors improve the content. We also wish to thank Sabrina Cochepin and Csaba Mosonyi (Publications Office of the EU) who designed the cover page and created the final layout design of this report.

The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission. All errors remain the sole responsibility of the authors.

Authors

European Commission,
DG Joint Research Centre,
Unit T.1 Digital Economy

FARRELL, Eimear; MINGHINI, Marco; KOTSEV, Alexander; SOLER-GARRIDO, Josep; TAPSALL, Brooke (ARHS DEVELOPMENTS SA); MICHELI, Marina; POSADA, Monica; VESPE, Michele; CARBALLA-SMICHOWSKI, Bruno; SCHADE, Sven; POGORZELSKA, Katarzyna; SMITH, Robin (EURORA SUPERNOVA); TARTARO, Alessio[1]; BERRNAL, Jaime[2].

European Commission,
DG Joint Research Centre,
Unit S.4 Scientific Development Programmes

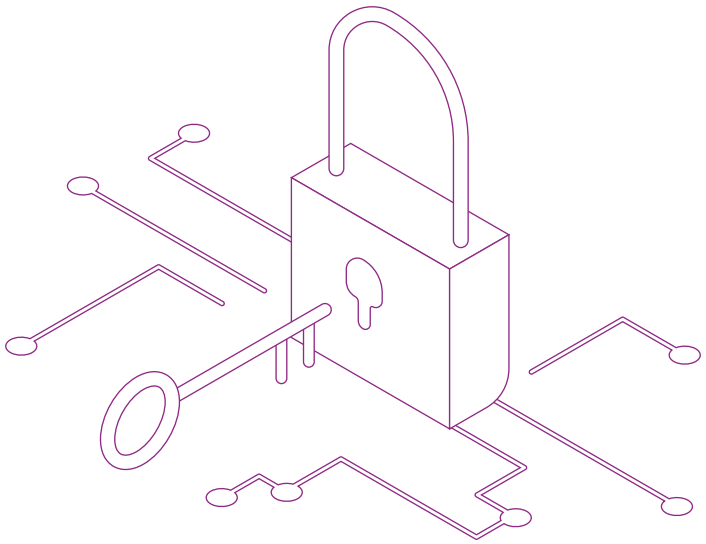
SIGNORELLI, Serena; GABRIELLI, Lorenzo

European Commission,
DG Joint Research Centre,
Unit T.4 Data Governance and Services

DE MARCHI, Davide

Arcadia SIT Srl

DI LEO, Margherita



1 JRC trainee between 16 March - 15 August 2022.
2 JRC trainee between 1 May - 30 September 2022.

Executive summary

→ Policy context

Ensuring fair and trustworthy data sharing is at the core of the European Commission's policy agenda for the period 2019-2024 under the priority "A Europe fit for the digital age". Published in 2020, the European strategy for data set the ambitious vision to establish common European data spaces in all strategic societal sectors and domains of public interest. Data spaces are envisioned as sovereign, trustworthy and interoperable data sharing environments where data can flow within and across sectors, in full respect of European ambitions, rules and values.

To support the establishment of data spaces while meeting the Digital Decade objectives for Europe's digital transformation by 2030, several **cross-sectoral legislative instruments** have been adopted or proposed following the publication of the European strategy for data. These include: the **Data Governance Act**, introducing a set of horizontal measures to boost trustworthy data sharing in the EU; the **Data Act**, aiming to make more business data available for reuse through the definition of rules on who can access and use what data and for which purposes; the **Implementing Act on High Value Datasets**, implementing the Open Data Directive by specifying a list of datasets that public sector bodies shall make available for free and under open access licenses, as well as making them accessible in machine-readable formats via application programming interfaces (APIs); and the **Digital Markets Act**, which sets out some measures on data access and portability to regulate the "gatekeeper power" of digital companies prone to unfair business practices.

→ Key conclusions

There is **no single technical or organisational approach** that can be applied for the establishment of common European data spaces. Along those lines, it is not possible to come up with a centralised governance model for data sharing. Each sector and thematic domain has its own specificity in terms of data types, data flows, business models and stakeholder needs. Therefore, a community-based approach through co-creation and co-design of data spaces that considers the domain-specific context is the only feasible way forward that would ensure buy-in by a broad spectrum of stakeholders.

Similarly, from a technical perspective, **a single architecture or stack of technologies and standards cannot be universally applied**. However, a minimum stack of protocols and specifications that can be used in a decentralised and federated manner, ideally through the means of the web, is highly desirable. The forthcoming European Data Innovation Board, defined by the Data Governance Act and supported by the Data Spaces Support Centre, should play a central role in the choice of such technologies and standards. Within that context, scientific evidence should continue to play a central role in the conceptualisation and implementation of European data spaces.

→ Main findings

The Staff Working Document on Common European Data Spaces describes these as bringing together relevant data infrastructures and governance frameworks in order to facilitate data pooling and sharing. From the descriptions contained in relevant EU policy documents, as well as our analysis of the scientific literature, we can state that common European data spaces aim to overcome legal, organisational, semantic and technical barriers to data sharing between diverse actors by combining the necessary tools and infrastructures, and by way of common rules and standards, always in compliance with the applicable laws. They will thus connect currently fragmented and dispersed data from various ecosystems in order to support EU priorities, create new business opportunities and facilitate the use of data for innovation.

The high-level and overarching **vision** defined for common European data spaces is to establish a single market for data in the EU, promoting trustworthy data sharing practices that are aligned with common societal values and existing legal frameworks, and that will bring about many economic, social and environmental benefits. The report thus distils several **key principles** that are important to respect in common European data spaces, based on the afore-mentioned EU policy and legislative documents. We assert that the combination of these principles constitutes the **mind set for sharing data in Europe** and they should be propagated throughout a data value chain. In accordance with this overarching vision, relevant stakeholders of the data value chain should be engaged in the creation, maintenance and governance of common European data spaces.

In addition, and derived from the same policy sources, the report identifies and synthesises a set of high-level **requirements** for common European data spaces, classifying these into both functional and non-functional categories. Such requirements can be of an organisational or governance type (e.g., data governance and transparency), or of a technical nature (e.g., that data spaces should support data transfer and exchange between participant infrastructures). The list of requirements presented is not exhaustive and could also form the basis for a more granular requirement elicitation in the future. Furthermore, it is our contention that common European data spaces should also include **features** (or properties) that implement some, or all, of these requirements (e.g., a secure and privacy-preserving infrastructure to pool, access, share, process and use data). At the same time, the requirements and features of a particular data space instance may need to be adapted to the specific needs of discrete ecosystem, depending on the given requirements of a domain, sector, or territory.

In providing a scientific perspective on data sharing and utilisation at scale, the JRC aims to provide useful evidence, insights and guidance to stakeholders involved in setting up, operating and maintaining common European data spaces, based on its own substantive experience and expertise in this area. To help address some of the key questions that may arise for stakeholders when conceptualising and operationalising a common European data space, the report includes a compilation of dedicated **How-to guides**, which elaborate on selected technical, governance and organisational aspects of sharing, accessing and using data. For example, these How-to's elaborate on topics such as, 'How to choose the best software stack?', 'How can stakeholders benefit from synthetic data in a common European data space?', 'How to facilitate the discovery of data in a common European data space?', etc.

In addition to the How-to series that is presented, the report includes an interactive **dashboard** that is designed as a tool to explore the **JRC Knowledge Base** on data spaces. This allows those interested to search by keyword and retrieve relevant JRC publications mapping to the principles, requirements and identified features of common European data spaces, without having to perform single queries on the JRC Publications Repository. Moreover, this visual tool offers the possibility to restrict or enlarge the search to other terms, and consequently to retrieve additional publications. There is also an accompanying **Wiki page**, where the JRC Knowledge Base can be further updated and enriched based on real-world implementation and related feedback from European data space actors.

As it should be clear from the above, and as it emerges from the report findings, **implementing a common European data space requires attention not only to technical aspects**, but also to governance-related elements. Governance typically includes business, legal, sociological, political and organisational aspects. It can span from contractual, organisational and operational agreements; to technical standards and tools; and from goals and principles to laws and regulations. Governance is defined by the Data Spaces Support Centre as the system of rules, practices and processes to direct and manage an organsiation and the mechanisms by which it, and its stakeholders are held to account.

Governance is essential to configure and coordinate the necessary actions and interactions by different stakeholders that make a data space work and meet its objectives. Emphasising this wider governance perspective, the report observes that different levels of governance may be needed within a data space and that data governance cannot be easily separated from the governance of organisations and individuals. At a high level, a broad governance structure is necessary to set out the strategy and values for a common European data space and to ensure an open participation; whereas a narrower governance structure may be needed for more technical and operational aspects, such as standards-setting and the establishment of processes and procedures. Interoperability governance is also important as a cross-cutting activity - while the realisation of interoperable European data spaces presents a coordination challenge, an approach to systematically address and govern this challenge is provided by the European Interoperability Framework (EIF).

As the exchange of data is one of the core functions of a common European data space, this report identifies **trustworthy and effective data governance as a primary concern**. Common European data spaces should put in place an appropriate governance structure to ensure fair, transparent, proportionate and non-discriminatory access to, sharing and use of data. Data governance should outline clear duties, standards and responsibilities, and ensure that data is appropriately protected according to EU laws and standards, while also supporting data sharing, and openness to data mobility and portability. For this purpose, it is recommended to look at existing frameworks and leverage good practices – with this in mind, the How-to series contained within the report addresses some pertinent data governance-related questions, e.g., 'How can Governments access data of public interest in a common European data space?' or 'How to leverage data sharing in a common European data space?'

While the EU Data Governance Act provides an overarching horizontal governance framework for common European data spaces, they will also need to operate according to other applicable EU laws and regulations, such as those relating to data protection, cybersecurity, competition, intellectual property, fundamental rights, safety, trade, etc., as well as complying with relevant sectoral-related legislation. Therefore, as underlined in this report, **it is essential to ask which legal elements are relevant when setting up a particular data space**, as well as to consider the appropriate data and organisational model.

Furthermore, **specific governance systems are beginning to be elaborated** for some of the sectoral European data spaces. While there is no single technical or organisational approach that can be applied for the establishment of common European data spaces, it is nevertheless important that any initiatives for implementing sectoral European data spaces should be aligned with the general set of horizontal rules that is applicable to all common European data spaces. Governance and trust become especially important when working across sectoral boundaries and where there is a need for multiple commons, communities and actors to interact together, for example in use cases across thematic European data spaces, such as the future Green Deal Data Space.

→ Related and future JRC works

Due to its specific mandate within the European Commission, the **JRC has been actively working on topics relating to data sharing since its establishment**. This is evident from the more that 3000 datasets and numerous scientific publications explicitly addressing topics such as interoperability, harmonisation, standardisation and data governance. The experience and expertise of the JRC in this domain is one of the main drivers behind writing this report. From our perspective, the JRC knowledge fits well within the concept of common European data spaces, and the synthesis of some of our lessons learned can inform the establishment of these data spaces. In addition, the scientific work performed by the JRC in support of all the different stages of the policy cycle can tremendously benefit from the improved access to data that will be provided by common European data spaces and the emerging single market for data. From that perspective, our Directorate is also user and end beneficiary of common European data spaces. Considering this duality, there is a strong interest in supporting the activities related to the establishment of data spaces, of course providing a distinctly scientific perspective.

With the high level of ambition of the policy initiatives defined by the European strategy for data, **targeted scientific work should be conducted that covers the technical, organisational, legal and economic aspects related to data sharing** in the EU. As the outcomes of such research should inform the policy-making process, it would make sense if the topics to be analysed are identified together with the policymakers in the domains where common European data spaces are being developed. This would ensure that the scientific findings of the JRC can be operationalised and scaled. Specific topics worth further investigation are already mentioned in the different sections of this report. From a technical point of view, those revolve around the definition and testing of architectural patterns, standards and protocols that can interconnect fragmented infrastructures in a user centric manner. In addition, some of the governance elements to be explored include data governance, legal issues, interoperability and organisational governance aspects. Finally, economic evidence around the business models that can sustain and scale sector-specific European data spaces and contribute to Europe's technological sovereignty remain equally important.

→ Quick guide

This Science for Policy report is subdivided into five interdependent chapters.

1 As an introduction to the report, **Chapter 1** provides an overview of the EU policy context defined within the overall frame of the 'Europe Fit for the Digital Age' priority of the EU, and in particular the European strategy for data. The introduction also describes the concept of a common European data space, describes their key actors and elaborates briefly on the principles, requirements and features of these data spaces. It subsequently justifies the need for scientific evidence in support of data sharing and presents a brief overview of the intended readership.

2 **Chapter 2** is dedicated to the '**Identification of requirements for common European data spaces**'. This is exclusively based on the most prominent EU policy documents relating to this topic, and in particular the European strategy for data.

3 Following the identification of the functional and non-functional requirements, and the equally important governance aspects related to European data spaces, **Chapter 3 contains a series of data sharing How-to guides**. These do not have a binding nature, and aim to serve as a guide to addressing some of the key questions that may arise for stakeholders when conceptualising and operationalising a European data space.

4 This is followed by an overview of the dedicated **JRC database on European data spaces** (**Chapter 4**) that exposes some of the JRC knowledge mapped to the requirements for data spaces.

5 Finally, the **Discussions and Conclusions** (**Chapter 5**) summarises the work done and includes the key findings, together with an outline of possible future work strands.

1 Introduction

Data-driven innovation plays a key role in the digital transformation of our society and organisations (Granell et al., 2022). That is why fair and trustworthy data sharing that can support the development of innovative products and services is at the core of the European Commission's agenda, which focuses on the twin green and digital transition. It is increasingly necessary to access heterogeneous data at scale to harness its full value for the benefit of the European economy and society, and to enable new technologies such as artificial intelligence (AI). An ambitious policy agenda was set by the **European strategy for data** in 2020 (European Commission, 2020a), with the objective to achieve the full potential of data-driven innovation in the European Union (EU), which is currently hampered by issues including fragmentation of both infrastructures and data sharing practices. This Strategy envisages the establishment of **common European data spaces** in strategic sectors and domains of public interest, aligned with European ambitions, rules and values. As described in the Strategy, common European data spaces are decentralised infrastructures where diverse actors can share, access and use data in a secure, reliable and trustworthy manner, following common governance, organisational, regulatory and technical mechanisms. They will interconnect various data ecosystems in a demand-driven process, which is key to achieving data exchange across societal actors. Common European data spaces will thus ensure that more data becomes available for use in the economy and society, while companies, organisations and individuals who generate the data retain control.

1.1 Policy context: a Europe fit for the digital age

The focus on digital technologies presented under the priority "*A Europe fit for the digital age*" lies at the core of the Commission's policy agenda for the period of 2019-2024 (European Commission, 2020b). The Commission has presented a vision for Europe's digital transformation by 2030, which is emphasised by the title "Digital Decade" and has set ambitious targets aimed at strengthening digital sovereignty through specific actions on data, technology and infrastructures (European Commission, 2021a). In addition, the Annual Single Market report published in 2023, which marks the 30th anniversary of the Single Market, highlights the ambition to create a single EU data economy through a data-driven Single Market where interoperability within and across data spaces is ensured (European Commission, 2023a).

Foreseen by the European strategy for data (European Commission, 2020a), several cross-sectoral legislative instruments have been adopted or proposed in order to create the necessary overarching governance framework for a data-agile economy, and to address common data sharing issues between different sectors, actors and domains (see **Figure 1**). Several of these legislative instruments are elaborated on below, as being particularly relevant in the context of common European data spaces.

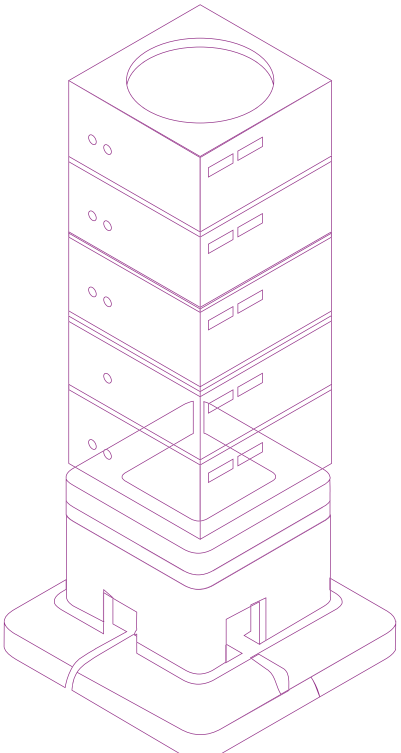
More specifically, the **Data Governance Act (DGA)** introduces a set of horizontal measures to boost trustworthy data sharing, thereby also supporting the establishment and development of common European data spaces by: (i) facilitating the reuse of specific public sector data that cannot be made available as open data (such as health data); (ii) regulating the role of data intermediaries within the common European data spaces; (iii) making it easier for businesses and citizens to make their data available for the benefit of society; and (iv) facilitating data sharing across sectors and borders and enabling suitable targeted applications. The DGA, proposed in 2020, entered into force in June 2022 and will be applicable from September 2023 following a 15-month grace period (European Commission, 2020c).

The **Data Act**, proposed by the Commission in February 2022 and expected to be adopted in 2023, includes specific rules on who can access and use data generated in the EU across all economic sectors. This should make more data accessible for, and within, the sectoral data spaces, stimulating a competitive data market and boosting data-driven innovation. It includes measures to: (i) allow users of connected devices (electronic devices that may be connected to other devices or networks) to gain access to data they generate and share these data with third parties; (ii) allow public sector institutions to access and use data held by private companies under particular circumstances (such as emergencies); (iii) rebalance negotiation power for small and medium-size enterprises (SMEs) by preventing abuse of contractual imbalances in data sharing contracts; and (iv) allow customers to switch between different providers of cloud data processing services (European Commission, 2022a).

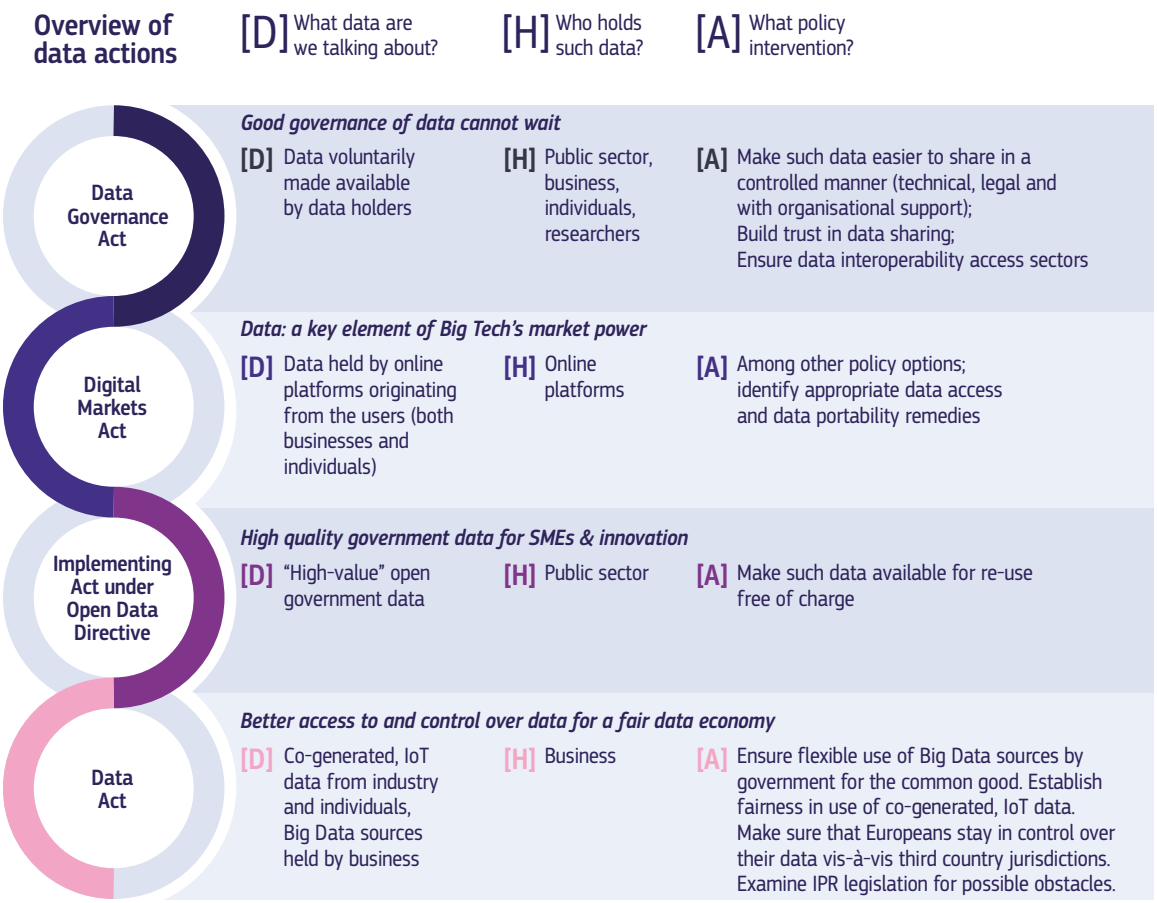
Originating from the Open Data Directive (European Union, 2019a), the third relevant legal instrument is the recently published **Implementing Act on High Value Datasets** (European Commission, 2023b). The concept of high-value datasets was introduced by the Open Data Directive to describe datasets held by the public sector, the reuse of which holds the potential to generate important benefits for society, the economy and the environment. The Open Data Directive requires that such datasets are made available for free, under open access licenses, and that they are accessible in machine-readable formats via suitable application programming interfaces (APIs) and, where relevant, as a bulk download. While the Directive only defined six categories of high-value datasets (Geospatial, Earth Observation and Environment, Meteorological, Statistics, Companies and Company Ownership, and Mobility), the Implementing Act will detail the actual datasets belonging to each category, together with the requirements for their publication, e.g., formats, granularity, key attributes and licenses.

Finally, the **Digital Markets Act,** which was also published recently (European Union, 2022), establishes a set of measures to identify appropriate data access and portability remedies, comprehensively regulating the "gatekeeper power" of digital companies, which are more prone to unfair business practices.

The horizontal legislative framework mentioned above may be complemented by sector-specific legislation when required by a sector's specificities, for example as in the case of the proposal for a Regulation on the European Health Data Space (European Commission, 2022b).



→ **Figure 1. Overview of the legislative framework for the implementation of the European strategy for data.**



Source: adapted from European Commission[3].

1.1.1 What is a common European data space?

The European strategy for data sets out a **vision for a "single European data space"**, which it describes as "*a genuine single market for data – open to data from across the world – where personal and non-personal data, including sensitive business data, are secure and businesses have easy access to high-quality industrial data, boosting growth and creating value*" (European Commission, 2020a). It emphasises that horizontal actions towards a European data space need to be accompanied by the development of sectoral or domain-specific data spaces in strategic areas such as manufacturing, agriculture, health and mobility.

In February 2022, the Commission, at the request of the EU Council, published a Staff Working Document (SWD) providing an overview of the **state of play of the ten common European data spaces** initially listed in the European strategy for data, together with several additional ones (European Commission, 2022c). It describes these data spaces as "*bringing together relevant data infrastructures and governance frameworks in order to facilitate data pooling and sharing*".

According to this SWD, common European data spaces should include *inter alia*:

i) the deployment of **data-sharing tools and services** for the pooling, processing and sharing of data by an open number of organisations, as well as federated energy-efficient and trustworthy cloud capacities and related services;

(ii) **data governance structures**, compatible with relevant EU legislation, which determine, in a transparent and fair way, the rights concerning access to and processing of data;

(iii) an **improved availability, quality and interoperability of data**, both in domain-specific settings and across sectors.

3 https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020SC0295&rid=6

From these descriptions, we can state that common European data spaces aim to overcome legal, organisational, semantic and technical barriers to data sharing between diverse actors by combining the necessary tools and infrastructures, and by way of common rules and standards, always in compliance with the applicable laws. A recent Digital Europe call for proposals refers to a data space as "data infrastructure with tailored governance mechanisms that will enable secure and cross-border access to key datasets in the targeted thematic area" (DIGITAL, 2022). Similarly, the Data Spaces Support Centre explains a data space as a "Decentralised, governed and standard-based structure to enable trustworthy data sharing between the data space participants on a voluntary basis" (Data Spaces Support Centre, 2022).

Common European data spaces will thus serve as repositories connecting currently fragmented and dispersed data from various ecosystems in order to support EU priorities, create new business opportunities and facilitate the use of data for innovation. They will offer an interoperable, trusted information technology (IT) environment for data processing, and a set of rules of a legislative, administrative and contractual nature that determine the rights of access to and use of data. They will thus enable data reuse and secondary use within and across sectors, fully respecting EU values and contributing to European economic and social development. It is also important to mention that in various policy documents the term "data space" is sometimes used alongside terms such as a "data ecosystem", a "data framework" and a "data infrastructure" as a way of describing European data sharing practices (e.g., European Commission, 2022c).

What is a common European data space?
- Availability of large pools of data
- Infrastructure to use and exchange data
- Appropriate governance mechanisms

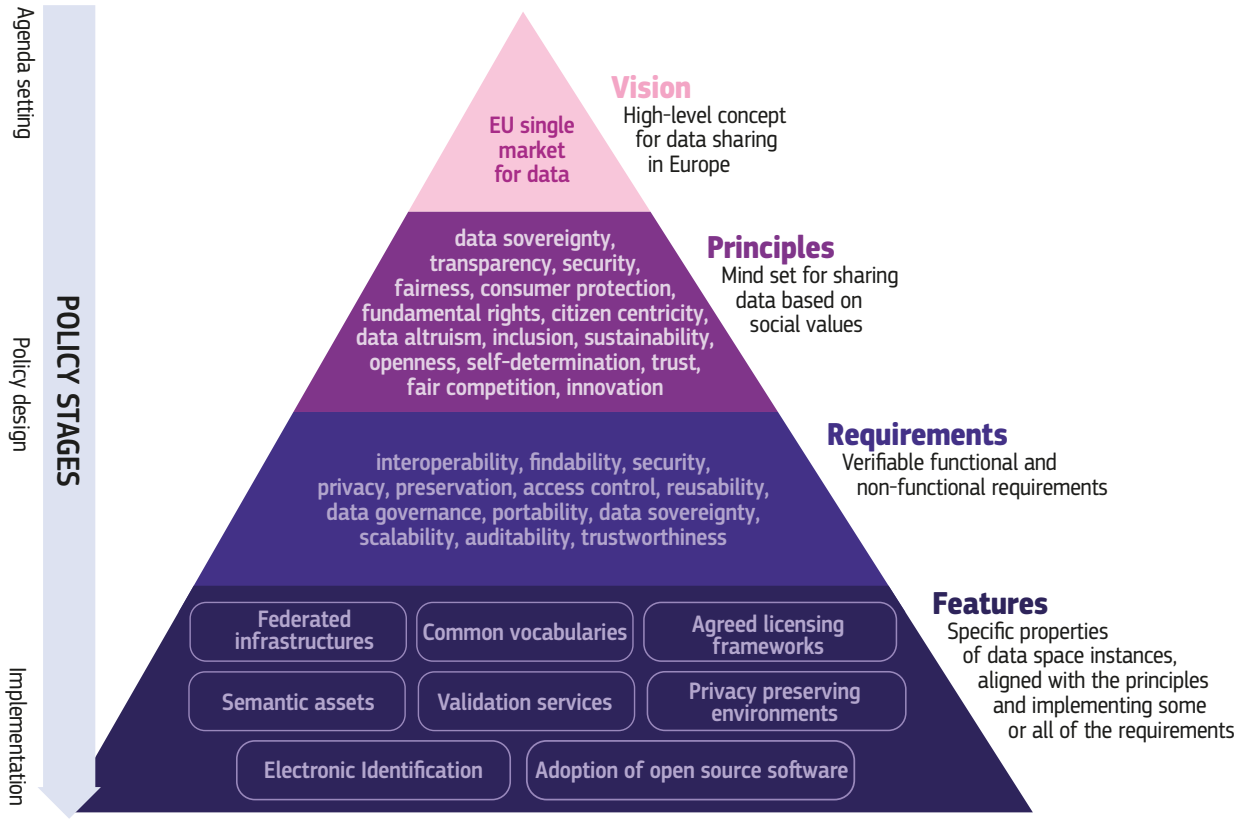
1.1.2 Principles, requirements and features of a data space

The high-level and overarching **vision** defined for common European data spaces is to establish a single market for data in the EU, thus bringing about many economic, societal and environmental benefits. In addition, the EU policy initiatives dedicated to data-driven innovation strongly signal that data sharing practices should be ethical, trustworthy and based on common societal values. For this to be achieved, the rules for accessing, using and sharing data must be fair, clear and practicable.

Accordingly, several key principles for data sharing in European data spaces, that are relevant for common European data spaces, are listed in **Figure 2** ('Selected characteristics of common European data spaces'), as these have been set out in relevant EU policy and legislative documents (see e.g., European Union, 2019a; European Commission, 2020a; European Commission, 2020c; European Commission, 2021b; European Commission, 2022a). Similar principles are also outlined in the "European declaration on digital rights and principles" that was published in 2022. Combined, such principles constitute the mind-set for sharing data in Europe based on common societal values and should be propagated throughout a data value chain (i.e., the full data lifecycle from collection to analysis and usage).

In addition, and at the policy design stage, a set of **requirements** for common European data spaces can be distilled from the existing EU policy and legislative documents relating to the digital sphere, as mentioned in the sections above. Further elaborated in **Section 2**, these requirements can be classified into functional and non-functional categories and include requirements of an organisational type (e.g., data governance, data protection, open participation, transparency, sustainability), as well as requirements of a technical nature (e.g., data transfer and exchange, identification, data findability and reusability, privacy preservation, authentication and access control, and interoperability). It is important to emphasise that additional principles and requirements may further derive from the specificity of the domains within the framework of the European strategy for data. Finally, during the implementation stage of the policy cycle, depending again on the specific application domain, a particular instance of a common European data space should include **features** (or properties) that implement some, or all, of these requirements. For instance, the SWD on Data Spaces (European Commission, 2022c) outlines some key features such as a secure and privacy-preserving infrastructure to pool, access, share, process and use data. **Figure 2** provides a graphical representation of these main characteristics of common European data spaces that stem from the overarching vision or high-level concept for data sharing in Europe, i.e., principles, requirements and features, seen in correspondence to the different policymaking stages, from the initial agenda setting to design and implementation.

→ **Figure 2. Selected characteristics of common European data spaces.**



Source: JRC's own elaboration based on existing EU policy documents.

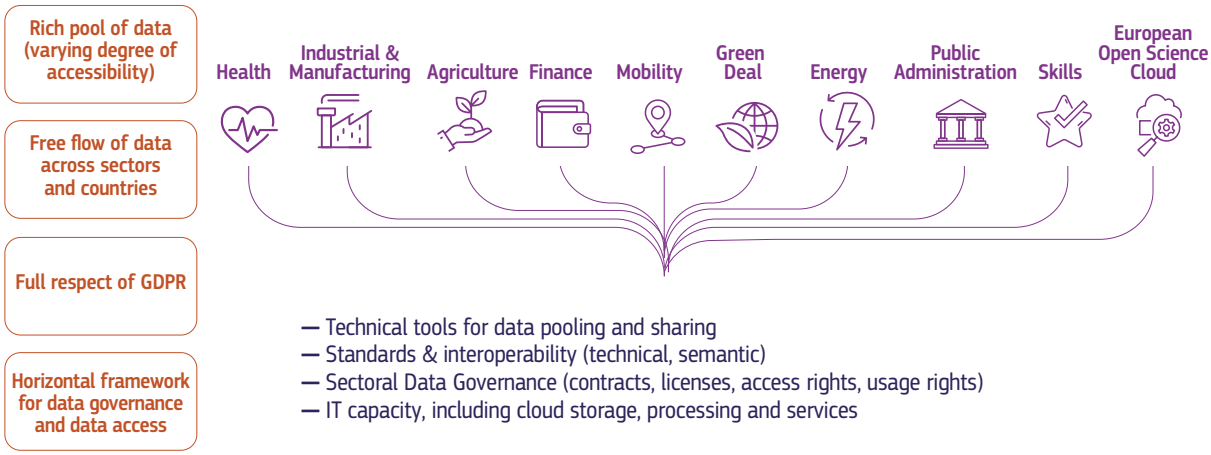
Following the logic outlined in **Figure 2**, it is possible to provide an example of characteristics in the case of the European Green Deal Data Space (see **Section 1.1.3**), which aims to exploit the full potential of data in support of the Green Deal priority actions on climate change, circular economy, zero-pollution, biodiversity, deforestation and compliance assurance. The **vision** of this data space is to ensure the effective and efficient reuse of heterogeneous data in support of all stages of environmental policies, to meet the objectives of the European Green Deal. This vision is guided by the same data sharing **principles** outlined earlier in this section. The **requirements** for the European Green Deal Data Space include, among others, harmonisation, documentation and exposition of geospatial environmental data in accordance with legal provisions (European Union, 2007), to preserve privacy when reusing citizen-generated data on the environment, and to provide licensing information alongside the data to ensure their reusability. The **features** that implement such requirements may include, e.g., an infrastructure for federating European data sharing infrastructures, aligned with European values and existing legal frameworks; a validation service to ensure adherence of data encoding and sharing provisions to reference/legal standards; and a common licensing framework to standardise data access and use conditions.

1.1.3 Sector- and domain-specific common European data spaces

In the European strategy for data, the Commission undertakes to promote the development of EU-wide common, interoperable data spaces in strategic economic sectors and domains of public interest, and in collaboration with relevant stakeholders. A cross-cutting objective is to support the creation of data pools for big data analytics and machine learning, in a manner compliant with relevant legislation, and to support the twin green and digital transitions. The ten initial common European data spaces are listed as follows (European Commission, 2020a):

- **an Industrial/Manufacturing Data Space**, to support the competitiveness and performance of the EU's industry
- **a Green Deal Data Space**, to use the major potential of data in support of the Green Deal priority actions on issues such as climate change, circular economy, pollution, biodiversity and deforestation
- **a Mobility Data Space**, to position Europe at the forefront of the development of an intelligent transport system
- **a Health Data Space**, essential for advances in preventing, detecting and treating diseases as well as for informed, evidence-based decisions to improve the healthcare systems
- **a Financial Data Space**, to stimulate innovation, market transparency, sustainable finance, as well as access to finance for European businesses and a more integrated market
- **an Energy Data Space**, to promote a stronger availability and cross-sector sharing of data, in a customer-centric, secure and trustworthy manner
- **an Agriculture Data Space**, to enhance the sustainability performance and competitiveness of the agricultural sector through the processing and analysis of data. It will consider the experience of the stakeholder-led Code of Conduct on agricultural data sharing
- **data spaces for Public Administrations**, to improve transparency and accountability of public spending and spending quality, fighting corruption, both at EU and national level
- **a Skills Data Space**, to reduce the skills mismatches between the education and training systems and the labour market needs
- the **European Open Science Cloud (EOSC)**, which provides the basis for a science, research and innovation data space that will bring together data from research and deployment programmes and that will be connected and fully articulated with the other sectoral data spaces. It provides seamless access and reliable re-use of research data to European researchers, innovators, companies and citizens through a trusted and open distributed data environment and related services (European Commission, 2022d).

→ **Figure 3. Overview of the initial breakdown of sector-specific common European data spaces.**



Source: Adapted from European Commission, 2020a.

Subsequent initiatives to create common European data spaces in additional sectors have since emerged in response to the European strategy for data, such as in the areas of language, cultural heritage, media, construction and tourism (European Commission, 2022b). This list is not exhaustive and is expected to evolve over time as the Commission may consider launching (sequentially) further common European data spaces in other sectors. For example, the Update of the 2020 New Industrial Strategy (European Commission, 2021c) announces a follow-up to the European strategy for data in this regard, including the launch of common European data spaces in specific sectors and value chains to enable industrial ecosystems to tap into the full potential of data coming from multiple domains.

The European strategy for data acknowledges that while the sectoral data spaces may have some differences according to domain, they should also be interconnected in order to simplify the exchange of data across sectors and enable cross-exploitation and collaboration for innovation (this is particularly relevant in the case of the Green Deal Data Space). In fact, data sharing and exchange within specific domains and sectors is already taking place within some existing initiatives (see e.g., Granell et al., 2022), where stakeholders in need of accessing new sources of data have already been organising data spaces. However, many of these initiatives follow their own individual approach, and therefore are not interoperable. The European strategy for data aims to include, and build upon, these existing data sharing initiatives to improve their interoperability, by introducing a harmonised framework for data exchange and a horizontal governance framework. By recognising data spaces in strategic sectors as common European data spaces, the Commission will foster their deployment and adoption at EU level.

In terms of policy and legislative initiatives on particular sectoral data spaces, in May 2022, the Commission presented a proposal for a regulation establishing the European Health Data Space, which complements the Data Governance Act (European Commission, 2020c) with specific rules for the use of datain the health sector (European Commission, 2022b). The proposal establishes a set of rules, infrastructure and governance mechanisms to promote interoperability and both primary and secondary uses of electronic health data, while ensuring data protection and strengthening cybersecurity. The forthcoming EU Action Plan on "Digitalising the European Energy System" announces the launch of an energy data space by 2024. The legislative frameworks for implementing sectoral common European data spaces should strike the right balance between the general set of rules applicable to all common European data spaces on the one hand, and the sector-specific rules and pre-existing or planned investments, on the other hand.

Finally, the European Commission supports the creation of common European data spaces through funding programmes such as the **Digital Europe Programme (DIGITAL)** for digital deployment initiatives, the **Horizon Europe** programme for research and innovation, the **Connecting Europe Facility (CEF)** for digital infrastructures, and the **NextGenerationEU's Recovery and Resilience Facility** for reforms and investments in the digital domain (European Commission, 2020e). Furthermore, the Multi-Country Project on European Common Data Infrastructure and Services (European Commission, 2021d) will deliver a technical infrastructure for the deployment of data spaces. Through DIGITAL, the Commission will fund the creation of European data spaces in specific sectors or domains where the EU financial contribution will have an impact on their deployment as European digital infrastructures – these data spaces will be based on the common data infrastructure, including SIMPL[4] – the smart middleware for cloud-to-edge federations - procured by the Commission in order to assure interoperability across sectors. Funding programmes supporting such data spaces have already been adopted and the first calls launched in 2021. However, it should be noted that in order to be considered as a common European data space, it is not necessary to receive EU funding or to be officially recognised as such by the European Commission.

In addition, ad hoc financial instruments are defined, based on the level of maturity of the development of common European data spaces and their specific objectives, for instance the Coordination and Support Actions (CSAs) defined in the DIGITAL Programme.

4 https://digital-strategy.ec.europa.eu/en/news/simpl-cloud-edge-federations-and-data-spaces-made-simple

1.1.4 Key actors

In accordance with the overarching vision for common European data spaces (see **Section 1.1.2**), relevant stakeholders of the data value chain should be engaged in their creation, maintenance and governance. These stakeholders may include **actors from the private sector, public sector, academia and civil society, as well as individuals**. They may play different roles in the ecosystem (Micheli et al., 2020), such as data producers, data receivers, data rights holders, data users or consumers, data or data service providers, data intermediaries, and data space enablers, as well as technology partners and standardisation bodies. By bringing together different stakeholders, common European data spaces will ensure that more data becomes available for use in the economy and society, while data control is retained by those businesses, organisations and individuals who have a role in generating the data. In any data space, and common European data spaces are not an exception, the various actors are brought together by a soft infrastructure consisting of rules and agreements of a legal, functional, technical and operational nature. As stakeholders may act in multiple ecosystems at the same time, they are not limited to sharing data within a single data silo or data domain. For example, the same public sector body can be primarily a producer of environmental data, but at the same time also a consumer of transport data that has been produced by other organisations (from the public and/or private sector), thereby participating as an actor within multiple data spaces.

Another key actor wil be the new **European Data Innovation Board** (EDIB), defined in the Data Governance Act and to be set up by September 2023, a structure that will be formed by representatives of common European data spaces and that will support the Commission in issuing guidelines on how to facilitate the development of common European data spaces, as well as identifying the relevant standards and interoperability requirements for cross-sector data sharing (European Commission, 2020c). On the more operative level, there is also the **Data Spaces Support Centre** (DSSC)[5] funded by the DIGITAL programme and defined in its Coordination and Support Actions, with the aim to coordinate all relevant actions on sectoral data spaces and make available (blueprint) architectures, best practices and common data infrastructure requirements, including relevant technologies, processes, standards and toolboxes. Launched in October 2022, the DSSC includes a consortium of leading associations and knowledge centres in the domain of data spaces, with a broad membership of associated and collaboration partners. Initial outputs of the DSSC activities include a Starter Kit, which provides a checklist for the creation of data spaces structured into the business, legal, operational, functional and technical dimensions (Data Spaces Support Centre, 2023a) and the DSSC Glossary of key terms that are relevant in the context of data spaces (Data Spaces Support Centre, 2023b).

The common European data spaces also have logical links to other regional, national and transnational initiatives for data infrastructures. For example, large Europe-wide research and business networks such as the Big Data Value Association (BDVA)[6], the Data Spaces Business Alliance (DSBA)[7] and the International Data Spaces Association (IDSA)[8] have been actively engaged in recent years in the development of data sharing initiatives in support of use cases, policies and digital innovation. The European Alliance for industrial data, edge and cloud (comprising both public and private actors) will also be connected with the common European sdata spaces to foster innovative data sharing environments based on open, interoperable, secure and resource-efficient cloud and edge solutions (European Commission, 2021e).

Other initiatives such as the industry-driven GAIA-X[9], or the Digital Transport and Logistics Forum (DTLF)[10], which aim to create a federated data infrastructure based on European values regarding data and cloud sovereignty, are also relevant for the development of specific aspects of the data spaces. Furthermore, the network of European Digital Innovation Hubs (EDIHs) will also work closely with the data spaces, together with European AI Testing and Experimentation Facilities (TEFs) and will promote the use of these infrastructures to SMEs.

5 https://dssc.eu
6 https://www.bdva.eu
7 https://data-spaces-business-alliance.eu
8 https://internationaldataspaces.org
9 https://gaia-x.eu
10 https://transport.ec.europa.eu/transport-themes/digital-transport-and-logistics-forum-dtlf_en

1.2 A scientific perspective on setting up, operating and maintaining common European data spaces

As data-driven innovation is a cross-cutting topic it can be investigated from multiple interrelated perspectives such as technical, organisational, legal, social and economic. Within that complex multidisciplinary context, scientific evidence can inform the scoping and implementation of data-sharing policies, but also their consequent monitoring and calibration.

As the in-house European Commission science and knowledge service, the Joint Research Centre (JRC) is well positioned to support the establishment of common European data spaces through the provision of independent scientific insights and recommendations. The JRC and the Digital Economy Unit in particular, have many years of experience in supporting data-related policies led by a number of Directorate Generals (DGs) of the European Commission, including CNECT, DIGIT, ENV, AGRI, ESTAT, JUST and GROW. The work presented here builds upon the outcomes and lessons learned from such experiences. We have tried to organise some of the knowledge and insights gained in the form of a practical guide, supporting the set-up, operation and maintenance of common European data spaces, based on the JRC's scientific perspective. With this report, we want to contribute to the process of **translating policy conceptualisations and theoretical understandings of data spaces into a practical framework for implementation**. **Accordingly, should you wish to get in touch with the authors, please contact the following mailbox: jrc-data-spaces-kb@ec.europa.eu**

1.2.1 Scope and intended audience

As outlined, many different stakeholders and actors have a role and an interest in the establishment, sustaining and scaling of the planned common European data spaces. They may include users and providers of data, standardisation bodies, open source communities, data intermediaries, and organisations that offer data integration and value-added services. The main audience targeted by this report comprises policy officers and domain experts that are currently working on, or expect to work on, the scoping and implementation of common European data spaces, particularly those which are still at an early stage in the process and require evidence and guidance for making their first steps. As the JRC's perspective on data sharing that is presented here includes both technical and organisational aspects (i.e., non-technical), the content could also be of interest to the broader spectrum of data space stakeholders and actors.

1.2.2 Synergies with other relevant initiatives

The objective of the work presented in this report is to organise and share some of the JRC's work that relates explicitly to data sharing (see **Section 4** on the JRC Knowledge Base), as well as to provide our scientific perspective on certain emerging questions that may be raised on organisational and technical topics relating to data-driven innovation. Clearly, **it is not our intention to cover all aspects related to data sharing in Europe.** This report also aims to inform readers of the work planned by (i) the European Data Innovation Board, and (ii) the Data Spaces Support Centre. In addition to this report, we have created an accompanying Wiki[11], which is intended to serve as a living document that can be incrementally updated with additional content and new scientific insights that can complement the other emerging (and bigger) initiatives dedicated to facilitating the sharing and use of data in Europe (see **Section 1.1.4**).

1.2.3 Components of the report

The remainder of this report includes three components that, when combined, can provide insights on data sharing and utilisation: (i) synthesis of the requirements for common European data spaces, based on existing EU policy documents and addressing organisational, governance-related, legal and technical aspects; (ii) data sharing How-to guides (also covering organisational, legal, governance and technical aspects); and (iii) a description of the JRC Knowledge Base relevant to the establishment of data spaces. These are briefly introduced below.

A) Synthesis of the requirements for common European data spaces

To understand and distil a set of requirements for common European data spaces, it is necessary to reference the European strategy for data (European Commission, 2020a) and the Staff Working Document on Data Spaces

11 https://wikis.ec.europa.eu/display/jrcdataspaceswiki/JRC+Data+Spaces+Knowledge+Base

(European Commission, 2022c), together with the various horizontal legal initiatives being developed within the context of that Strategy (see **Section 1.1**). A synthesis of these requirements, as we have derived them from the relevant policy sources, is presented in **Section 2**, taking into account the broader context of the vision and key principles for European data sharing. The requirements cover a range of organisational, governance and technical aspects.

B) Data sharing How-to guides

While the concept of a data space is relatively new, the topic of data ecosystems is well studied, and a solid body of evidence exists around most aspects of data sharing and exchange. However, this knowledge, which is also highly relevant in the context of data spaces, often remains fragmented and dispersed. Therefore, and in order to help address some of the key questions that may arise for stakeholders when conceptualising and operationalising a common European data space, we have prepared dedicated data sharing information sheets or guides called How-to's, which aim to bring some of this knowledge together and elaborate on selected technical, governance and organisational aspects of sharing, accessing and using data in data spaces (see **Section 3**). When considering the scope of the How-to guides, we tried to imagine the type of questions that most commonly might arise, some of which may also relate to the data space requirements and features we identified. The criteria we used for identifying these questions included: (i) co-creation and internal peer review aligned with the authors' research evidence and interests; (ii) dialogue with prominent stakeholders, including policy officers in European Commission services, national administrations and scientific peers; and (iii) the existing body of knowledge of the JRC.

For example, the questions covered in the How to's address topics such as: 'How to Choose the best software stack for a data space?'; 'How to facilitate the discovery of data in a data space?'; 'What is the best way to provide access to data in a data space?'; 'Which actors are providing what types of data within scope of a data space?', and 'How can voluntary data sharing be leveraged in a data space?'

Structurally, the How-to's outline a high-level problem statement or scenario, followed by a proposed approach which may be applied in order to address it. Where possible, concrete examples are included to explain the problem and/or proposed solution. Furthermore, references to additional resources and materials are provided, including, but not limited to those produced by the JRC. Many of the questions covered by the How-to's are complex and we acknowledge that some may be subject to debate, or that different approaches can be adopted. It is not our aim to be fully comprehensive; nonetheless, we consider that the JRC's perspective on these questions, based on our knowledge and research in the area of data sharing, might prove of value to the target audience. It is clear that some of the How-to's may be of particular interest and relevance to specific stakeholders and actors, while others may be of a more general interest.

C) JRC dashboard on European data spaces

The JRC has a long-term commitment to topics such as data sharing, data interoperability and standardisation, digital infrastructures and emerging technologies, but also conducts sector-specific work with, and on data that spans across multiple disciplines and domains. The JRC's research related to data sharing is made available through academic papers and dedicated reports, organised in different categories based on the topic and target audience (including, e.g., technical, science for policy, foresight, policy briefs and workshop proceedings). In addition, JRC datasets and their metadata are made available in a structured and standardised manner through the JRC data catalogue[12]. At the time of writing this report, a search through the JRC's publications repository for publications including "data" AND "sharing" in the title or abstract[13] returns a total of 545 results. Furthermore, the JRC data catalogue contains (as of February 2023) 3,220 datasets organised in 10 science areas. These JRC research outputs on data sharing standards and infrastructures, together with its experience in collecting, harmonising and combining heterogeneous data in support of EU policies, uniquely position the JRC to contribute its knowledge and perspectives on these topics. Using all of the above as input, the JRC Knowledge Base on data spaces includes a online dashboard, serving as an interactive discovery tool that allows those interested to identify JRC publications that map to the principles, requirements and identified features of European data spaces. This tool, presented in more detail in **Section 4** below, can be accessed upon registration.

12 https://data.jrc.ec.europa.eu/
13 https://publications.jrc.ec.europa.eu/repository/search?query=%22data+sharing%22

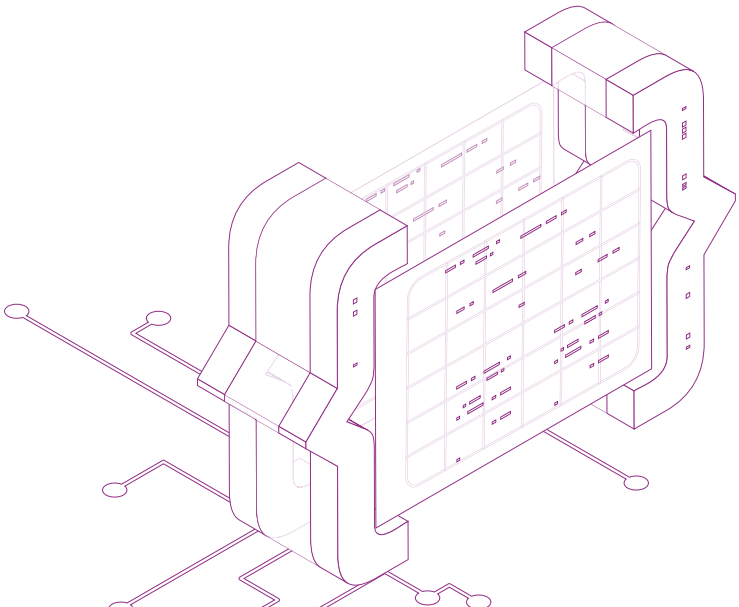
2 Identification of requirements for common European data spaces

In this section we summarise the requirements for common European data spaces (as defined in **Section 1.1.2**), based on relevant EU policy documents. We do this by first identifying high-level requirements, from which to derive an implementable design that satisfies the concerns of key stakeholders, in this case those of the participants in common European data spaces. Identification of requirements should assume that common European data spaces are not to be seen only as technical infrastructures, but as inclusive initiatives that cater for a balanced set of technical, organisational and legal or governance-related aspects. Some particularities of such data spaces within that context are the following:

- Many instantiations of common European data spaces are expected, corresponding to the different sectors of operation. These are anticipated to have a common base but will inevitably present slightly different priorities and trade-offs in their design, according to the nature of the data assets managed, as well as the characteristics of their participants and the applications hosted.
- They should provide functionality for a broad range of societal stakeholders. These include not only policymakers and public administrations, but also private sector organisations of all sizes, the scientific community and the public at large.

While being a new initiative at the EU level, data spaces are not a fully greenfield project to be built from scratch. Indeed, many public and industry-led initiatives have already started to build and pilot data sharing ecosystems, and there is an expectation that while these may be heterogeneous, some of the know-how, lessons learned — and even the software building blocks — produced by these examples could be reused for the common European data spaces, which, while not representing a formal continuation of these initiatives, may share a similar vision.

These considerations and constraints will influence the process for defining the requirements of data spaces, requiring a tailored approach, and in this section, we provide an initial reflection on what this process may entail.



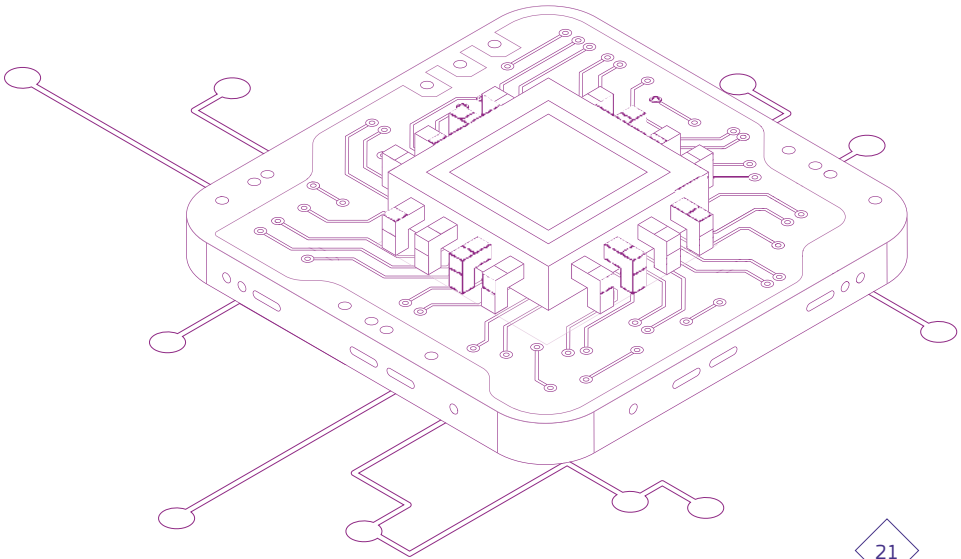
2.1 High-level requirement identification

As mentioned above, the identification of requirements represents the starting point for a common European data space as it provides the necessary context and scope, including the desired functionality, as well as the main quality attributes. Requirement elicitation is an iterative process performed in stages, with a gradually increasing level of detail. The initial set of requirements identified typically corresponds with high-level functionality, without which a data space cannot achieve its purpose. These, albeit high-level and often not fully precise, tend to be Architecturally Significant Requirements, defined as those that have a profound impact on the final design, for example, that common European data spaces should support data transfer and exchange between participant infrastructures. Once identified, top-level requirements are further broken down into more granular requirements, for example, which interfaces and standards to follow, building a hierarchical requirement tree gradually through the design lifecycle, and often iterating between architecture definition, implementation and further requirement elicitation steps.

In this analysis we are mostly concerned with top-level requirements. Two effective ways to identify these are by conducting stakeholder interviews, as well as by conducting an analysis of the business goals of the system, or its ultimate mission, e.g., through documentation review. As common European data spaces are a European policy initiative, with a clearly defined objective to increase the availability of data and make the most out of their use within the economic and social fabric of the EU, our top-level requirement analysis focuses on relevant EU policy sources such as the SWD on data spaces (European Commission, 2022c) as well as the DIGITAL work programme (European Commission, 2021b).

Since the EU policy documents and programmes are themselves based on extensive stakeholder consultation, the requirements extracted from these sources are already informed by the needs of data space users across all sectors of interest. We therefore scanned relevant policy documents in order to extract requirements for common European data spaces and we further classified these into functional and non-functional categories (see **Section 2.2**). While functional requirements define what data spaces must provide, i.e., they explain the concrete components of data spaces, non-functional requirements define how data spaces should work, i.e., their quality attributes.

Table 1 includes a collection of selected EU policy provisions on common European data spaces, based on the afore-mentioned policy sources. Such provisions are translated into groups of functional and non-functional requirements. These could form the basis for more granular requirement elicitation in the future, as well as for an analysis of the functionality provided by existing data sharing initiatives, as outlined in the next sub-section.



→ Table 1: Selected list of EU policy provisions translated into functional and non-functional requirements for common European data spaces

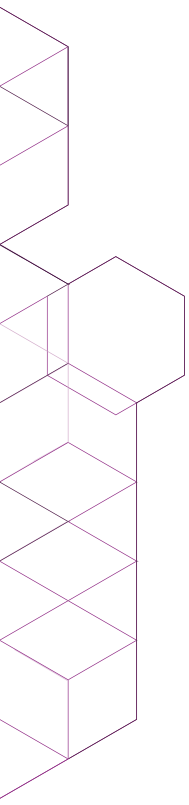
Policy provision	Functional requirements	Non-functional requirements
A secure and privacy-preserving IT infrastructure to pool, access, process, use and share data	Data transfer and exchange Data storage Data processing and analytics Data pooling and collaboration	Security Confidentiality
Data holders will have the possibility, in the data space, to grant access to or share certain personal or non-personal data under their control	Identity, authentication and access control Usage control policies	Confidentiality
… promote the development of tools to pool, access, use and share all types of data favouring the development of common open standards and findable, accessible, interoperable and reusable (FAIR) principles …. data holders could use these tools to ease the uploading of data into data spaces, to give or revoke their authorisation to data and to change access rights and specify new conditions of how their data can be accessed and reused over time	Data transfer and exchange Identity, authentication and access control Usage control policies	Interoperability Auditability
Data that is made available can be reused against compensation, including remuneration, or for free	Transaction metering and billing	
Participants […] use the common technical infrastructure and building blocks which will allow the data spaces to be built in an efficient and coordinated manner	—	Maintainability Variability
The common technical infrastructure will have to […] integrate the cybersecurity-by-design principle	—	Security
Participation of an open number of organisations/ individuals	Identity, authentication and access control	Scalability
Common European data spaces could be developed on international standards, INSPIRE (for spatial data) and FAIR principles to favour interoperability, …exploitation of data on EU computing infrastructures (e.g., cloud and HPC) and be interconnected and progressively made interoperable	Data interoperability features Data processing and analytics	Interoperability Performance
European rules and values, in particular personal data protection, consumer protection legislation and competition law, are fully respected	Compliance monitoring and auditing Data protection	Auditability
Enhance the development of new data-driven products and services in the EU and thereby create the core tissue of an interconnected and competitive European data economy	Data processing and analytics	Sovereignty
Data spaces middleware: provide a full cloud stack with basic services that can also be operated at the edge, while foreseeing the subsequent integration of high-performance computing and far edge computing	Multi-tier support, federation, orchestration	Portability Performance
Data spaces middleware: provide a technical baseline to be used by all EU common data spaces to avoid duplication of effort and overlaps and to ensure a proper alignment of the various implementation approaches	—	Maintainability Variability
Data spaces middleware: allow state-of-the art data management between cloud and edge, enabling seamless ultra-fast data workload balancing between them, and intelligent data porting between centralised and decentralised data infrastructures Ensure performance and quality of service in the execution of applications across multiple cloud and edge providers Provide a multi-cloud orchestration solution, with built-in identification and security management services	Data transfer and exchange Multi-tier support, federation, orchestration Identity, authentication and access Control	Portability Performance Security Maintainability
Data spaces middleware: provide data mapping services, data anonymisation and masking services	Privacy-preserving mechanisms Data interoperability features	Confidentiality Interoperability
Data spaces middleware: embed business intelligence services for multi-uses based on crosscutting, low power and software-enabled services	Data processing and analytics	—
Data spaces middleware: integrate an environmental tracking performance system to ensure services operate in a low power mode	—	Energy efficiency

Policy provision	Functional requirements	Non-functional requirements
Data spaces middleware: provide secure resource efficient data storage services	Data storage	—
Data spaces middleware: provide an "High Performance Computing as a service" connector to enable High Performance Computing resources to be accessible to users of the Cloud Federation	Multi-tier support, federation, orchestration	Performance Portability
Data spaces middleware: ensure that AI solutions [...] can operate over the middleware platform Support sustainable and ultra-low latency digital twins' business applications Allow the hosting of highly specialised tools for complex business activities simulation, forecasting and modelling	Data processing and analytics	Performance
Data spaces middleware: provide secured communication, productivity and collaboration services Provide workflow management services Facilitate the integration with [cloud-to-edge] services and [their] marketplace	Data pooling and collaboration	Portability
A common European data space brings together relevant data infrastructures and governance frameworks in order to facilitate data pooling and sharing A clear and practical structure for access to and use of data in a fair, transparent, proportionate and/non-discriminatory manner and clear and trustworthy data governance mechanisms	Data pooling and collaboration Usage control policies Privacy-preserving mechanisms Data protection Data governance	Inclusivity Fairness Sustainability Trustworthiness Transparency

2.2 Technical perspective

As described in the previous **Section 2.1** and the associated **Table 1**, the relevant EU policy sources (European Commission, 2021b; European Commission, 2022c) contain the following main categories of **functional requirements** for data spaces – many of these aspects are addressed in the How-to information sheets that are included as part of this Science for Policy report (see **Section 3**):

- **Data transfer and exchange.** The core functionality of data spaces, enabling participants to transfer data to other participants.
- **Data publication and discovery.** An effective mechanism for publication and discovery is expected to be a key functional requirement of data spaces, especially given the large amount of heterogeneous data expected to be made available in them.
- **Data Storage.** To support access to data, storage services can be either physical, i.e., based on independent copies of participants' data within the ecosystem, or virtual, providing access to data assets which are physically located in their owners' infrastructure.
- **Identity, authentication and access control.** These are key features upon which trust is built in the data sharing ecosystem, enabling participants to control who can access their data assets.
- **Data interoperability.** Features supporting the integration of heterogeneous data sources from the legal, organisational, technical and semantic perspectives.
- **Data processing and analytics.** The functionality of data spaces extends beyond making data available, and includes the utilisation of data for value-added applications, notably through data analytics and AI. Tools to streamline the development of AI solutions would be beneficial, especially if they target not only AI specialists but also domain-experts from the different sectors, e.g., through low-code, no-code, AutoML (automated machine learning methods and processes) and other approaches to make AI available for non-experts.
- **Data pooling and collaboration.** Collaboration tools are required to enable the joint development and exploitation of products and services by multiple participants in data spaces, possibly from different organisations and even economic sectors. Productivity and collaboration services could support and simplify the design, implementation and management of distributed processing workflows across ecosystem participants, ensuring an effective shared governance.



- **Multi-tier support, federation and orchestration.** Data spaces should provide development tools for multi-platform services that are supported by a wide range of underlying computing architectures, as well as interfaces for their orchestration – this is a key aspect of digital sovereignty.

- **Privacy-preserving mechanisms.** Ensuring data privacy is a key requirement for certain data spaces handling sensitive data (e.g., personally identifiable information or intellectual property). Data spaces should comply with the EU General Data Protection Regulation (European Union, 2016) and provide data privacy features, such as anonymisation and masking services – they may in the future incorporate more advanced privacy-enhancing technologies, such as federated learning, secure multi-party computation and homomorphic encryption.

- **Usage control policies.** Building on access control functionality, additional features should enable participants in data spaces to determine not only who is allowed to access their data, but also the manner in which these data can be used, providing effective monitoring and enforcement functionality.

- **Compliance and auditing.** This functional category encompasses features that enable participants in data spaces to attest and verify claims made by their peers regarding compliance with standards, regulations and general terms and conditions for using data and services. Such features include preconditions for making data available that are defined by their owners or by any other governing authorities.

- **Transaction metering and billing.** Features that enable participants in data spaces to monitor and monetise data flows, as well as the consumption of their services within the ecosystem.

- **Data governance.** Data governance can be defined as the set of rules, policies, relations, decision-making structures and processes established among different kinds of actors to collect, share and use data. In general terms, it is understood as the correct management and maintenance of data assets and related aspects, such as data rights, data privacy, and data security, among others. While being a functional requirement on its own, data governance is also an essential prerequisite for many other (e.g., technical) functional requirements of data spaces. And in turn, the technologies used in a European data space should meet the requirements of data and information governance.

- **Data protection.** Data spaces should protect the personal data of individuals that is shared within them, and comply with EU General Data Protection Regulation (GDPR) rules (European Union, 2016). The GDPR is a European law that establishes protections for privacy and security of personal data about individuals in European Economic Area (EEA)-based operations and certain non-EEA organizations that process personal data of individuals in the EEA. Privacy and data protection are also enshrined in the EU Treaties and in the EU Charter of Fundamental Rights.

It should be noted that not all categories of requirements may be necessary for each of the sector-specific common European data spaces. Therefore, as discussed in the following sections, a modular approach should be adopted. This approach should be based on the definition of fundamental functional building blocks that can be combined to form complete sectoral European data spaces.

Additionally, the following selected **non-functional requirements**, have been identified or inferred:

- **Security and confidentiality.** Security based on design principles is required in order to increase trust and eliminate barriers to sharing private data, including personal data as well as intellectual property.

- **Interoperability.** On the non-functional side, common vocabularies and standards are expected to provide a basis for interoperability of the different data and services built by ecosystem participants.

- **Maintainability.** The functionality of data spaces is expected to be developed gradually and to be in constant evolution as new technical developments emerge and the priorities of the different participants and domains shift.

- **Variability.** This quality supports the development of different variants of core building blocks to fulfil the requirements, in order to support the specific priorities of different sectoral data spaces, allowing for their tailoring and evolution in accordance with the characteristics of their respective users and data.

- **Scalability.** Data spaces should have the capacity to adapt and continue to function in the presence of potentially major changes in the amount and type of data available and exchanged, as well as in terms of the numbers of users and services deployed.

- **Performance.** Data spaces should support high-performance data analytics use cases, including those based on high-performance computing (HPC). Similarly, high-speed data transfer between infrastructures and participants is expected to be required.

- **Auditability.** Data spaces are expected to be transparent in their operation, allow public scrutiny while enabling users, governing bodies and regulators to monitor activities taking place within them and to enforce rules.

- **Portability.** A key quality attribute for the seamless transfer of services and data assets between different infrastructures, including across deployment tiers (field devices, edge, multi-cloud environments, including private and public components), effectively preventing vendor and technology lock-in.

- **Energy efficiency.** Data spaces should favour the use of digital infrastructures with an optimised energy demand and provide orchestration mechanisms involving low-power devices when feasible, effectively reducing the carbon footprint, notably when hosting intensive data processing workloads.

- **Inclusivity.** Data spaces should allow for open and broad stakeholder participation in decision-making processes concerning access, control, sharing and use of data.

- **Fairness.** Data spaces should maximise the value of data in the economy by ensuring that a wider range of stakeholders gain control over their data and that there is a greater and fairer flow of data in all sectors. They should thus enable fairness in the allocation of value from data among actors in the data economy and foster access to and use of data.

- **Sustainability.** Data sharing infrastructures should have (i) a viable business model, (ii) appropriate governance model, and (iii) appropriate rules for participation by the relevant actors that would ensure they can be sustained in the long term.

- **Trustworthiness.** Trust is a crucial factor in any data sharing arrangement. Participants and stakeholders in data spaces should feel confident that the data is governed responsibly, used with integrity and in accordance with the law, and that the technology being used is also trustworthy and secure. The idea of retaining control over one's own data and its use is an important basis for establishing trust amongst data space participants and is also in line with EU rules and values regarding data sovereignty. Trust can also be built through transparency, accessibility and participation.

- **Transparency.** A certain degree of transparency is needed for the proper functioning of a data space. This should include information about what is happening within the data space, including what data is provided or received, for what purpose and for what duration. The policies, rules, standards and participation modalities in a data space should be clearly and transparently defined - individuals should know how to contribute data, who has access to the data, where they are stored, how they are secured and protected, and how they can interact with the data.

- **Data sovereignty.** Data sovereignty involves enhancing control by organisations and individuals over data that they contribute to generating. It implies participation in data governance and allows individuals and organisations to self-determine how, when and at what price others may use their data across the value chain. It means that data holders can safeguard user data, and ensure that it is used only in accordance with strictly defined rules.

Non-functional requirements are in many ways even more relevant than functional requirements, as in most practical cases, they cannot be jointly optimised. It should be noted that the list provided above is not meant to be exhaustive. In fact, it mostly contains non-functional requirements that are either directly mentioned or that can be readily inferred, from EU policy documentation. More detailed analysis of policy priorities relating to common European data spaces, combined with stakeholder interviews, may lead to the identification of additional significant non-functional requirements. In fact, some non-functional requirements can be directly derived from the perspective (pertaining to the pure technical dimension) of data spaces as platforms, on top of which end-users can build solutions, whether these are data provision services or ready-made AI products. This implies a need to consider certain aspects of usability, notably creating an environment that effectively supports developer productivity or habitability, e.g., by

providing comprehensive tools such as software development kits (SDKs) and ensuring that user interfaces are clearly defined, meaningful and accessible from the developer perspective. Other implicit non-functional requirements extend beyond technical considerations and include governance-related aspects, e.g., the need for long-term sustainability of data spaces, as these digital infrastructures are expected to outlive the initial investment phase leading up to their development and operationalisation.

2.3 Governance perspective

This section provides an overview of the general goals and questions of governance at different levels in a common European data space. After providing a general introduction to the purpose and functions of governance in a data space, it outlines some key roles, as well underlining the importance of establishing core principles, practices and processes to address legal, business, technical and organisational dimensions. It describes the overarching governance context for common European data spaces (including data governance), as well as providing some examples of sectoral governance considerations. Furthermore, it emphasises the need for interoperability to enable cross-stakeholder, cross-border and cross-sectoral data sharing and interactions. **Figure 4** displays the various concepts, actors, policies and regulations that may shape the use and sharing of data sharing and use in a common European data space, while also recognising that governance can be a complex task, and that data spaces may involve a variety of different platforms and systems, fit adapted to different market designs and business processes.

2.3.1 Guidance and guardrails: goals of governance

As outlined by the Data Spaces Support Centre (DSSC, 2023a), common European data spaces can provide a clear and shared means (from a business, legal and governance, technical and operational perspective) to support data sharing within a data ecosystem. Other authors have noted that implementing a common European data space requires not only attention to technical aspects, but also to **stewardship, leadership and strategy for sharing resources** (e.g., Keller, 2021).

The SWD on Common European Data Spaces describes the vision behind common European data spaces as that of bringing together relevant data infrastructures and **governance frameworks** in order to facilitate trustworthy data pooling, processing and sharing (European Commission, 2022c).

The European strategy for data foresees that European data spaces will include the creation of **data governance structures**, compatible with relevant EU legislation, which determine, in a transparent and fair way, the rights concerning access to and processing of data (European Commission, 2020a). Furthermore, by setting requirements on the reuse and access of data, and creating a framework to increase trust between actors in the data market, for example by promoting the establishment of trustworthy data intermediation service providers (DISPs), the Data Governance Act supports the establishment of common European data spaces (European Commission, 2020c). It also states that sound governance is needed in which relevant stakeholders of a common European data space can participate and be represented.

Indeed, data governance cannot be easily separated from the **governance of organisations and individuals**. In this regard, Braun et al. (2022) suggest that governance of data spaces involves "the management of the way [the] data space is accessed, controlled, and used; the assessment and control of generating value from [the] data space, and of the redistribution of value between actors; reflection on and involvement in decision-making over [the] data space, who is able to make such decisions and to influence how data and the virtual space is accessed, controlled, used and benefited from".

Governance is therefore needed to **configure and coordinate the necessary actions** (and interactions) by different organisations that make a data space work and meet its objectives. The SWD on Common European Data Spaces states that the governance of a common European data space should include **adequate and non-discriminatory representation of relevant stakeholders** (European Commission, 2022c). Consequently, common European data spaces should be clear on the **roles different actors play** in the design, development and implementation of the data spaces. Governance mechanisms should guarantee the rights of the participants and a smooth exchange of data. Furthermore, data space users should also be seen as stakeholders, with the principles of user-centricity in technological design being considered. It is necessary to agree on standards and design principles for data spaces that are acceptable to all participants.

For example, regarding the common European Health Data Space, Kiseleva and de Hert (2021) have argued that the roles of all the stakeholders involved (such as patients, medical organisations and professionals, public authorities, private entities and research organisations) shall be clearly defined in terms of their rights and obligations with regards to adding data to the space, as well as having access to it.

These roles shall be based on common requirements to add data to the space and limitations imposed on specific data sets, but shall also take into consideration differences in roles, interests and applicable legal frameworks.[14]

Moreover, while common European data spaces are being established for those that recognise themselves as belonging to defined communities (such as health, mobility or environment), the users and beneficiaries of data spaces may engage with a range of different data sources and domains to tackle societal challenges that data space content would help address. As posited by Nagel and Lycklama (2021), the aim is that over time, European data spaces would systematically harmonise parts of their technical, operational, functional and legal aspects, leading to the emergence of a more uniform, de-facto 'soft infrastructure'. This soft infrastructure is described as a combination of design decisions, rules, agreements and processes concerning how to share, manage and use data, that would include common aspects for all data spaces, including (but not limited) to identity, metadata, consent, legal terms, and security. Such a soft infrastructure would also define the roles required for a data space to work effectively (data provider, data consumer, etc.). The harmonisation of common aspects should enable users (e.g., citizens, businesses, governments) to stay in control of their data even across different sectors and applications (i.e., across different data spaces).

Fritzenkötter et al. (2022) point out that different levels of governance may be required in a common European data space as part of a **governance continuum**. At a high-level, a broad governance structure is needed for strategy and values-setting and to bring in the participation of a wide range of those who can contribute, use or are affected by the relevant data and associated information systems. Narrower governance structures may be needed for more technical and operational aspects, such as standards-setting and process rules, as well as the management of interoperability assets, such as reference code lists. In between those levels, an oversight approach is needed to deliver transparency and accountability, founded in the broadly-agreed strategy and demonstrating to all that the governance of the data space is being realised fairly. This would include outlining the terms for participating in the data spaces as a provider or user, as well as the processes of recourse should any actor feel that those in charge of the data space had not respected their needs in a fair way.

While the realisation of interoperable common European data spaces may present a coordination challenge (International Data Spaces Association, 2021), an approach to systematically address and govern this challenge is provided by the European Interoperability Framework (EIF) developed by the European Commission, alongside its related frameworks and assets, where data aspects are also mentioned within its principles and recommendations. More specifically, the current form of the EIF considers "**Interoperability Governance**" as a cross-cutting activity alongside the main four levels of legal, organisational, semantic and technical interoperability. This aspect deals with interoperability framework decisions and institutional arrangements, organisational structures, roles and responsibilities, policies and agreements, as well as the monitoring of interoperability at national and EU levels. In addition, "Integrated public service governance" under the EIF considers how "European public services" should be governed to aid aspects such as integration, seamless execution, reuse of services and data and the development of new services and reusable 'building blocks' that can be deployed across IT infrastructures, such as eID. Importantly, the EIF will be reviewed as part of the currently proposed Interoperable Europe Act[15], with further details on the current EIF available in the JoinUp platform[16].

Drawing from all the above, **Figure 4** provides a visual representation of some of the main concepts, actors, policies, processes and regulations that may shape the overall governance context and set the rules of engagement in a common European data space. Recognising

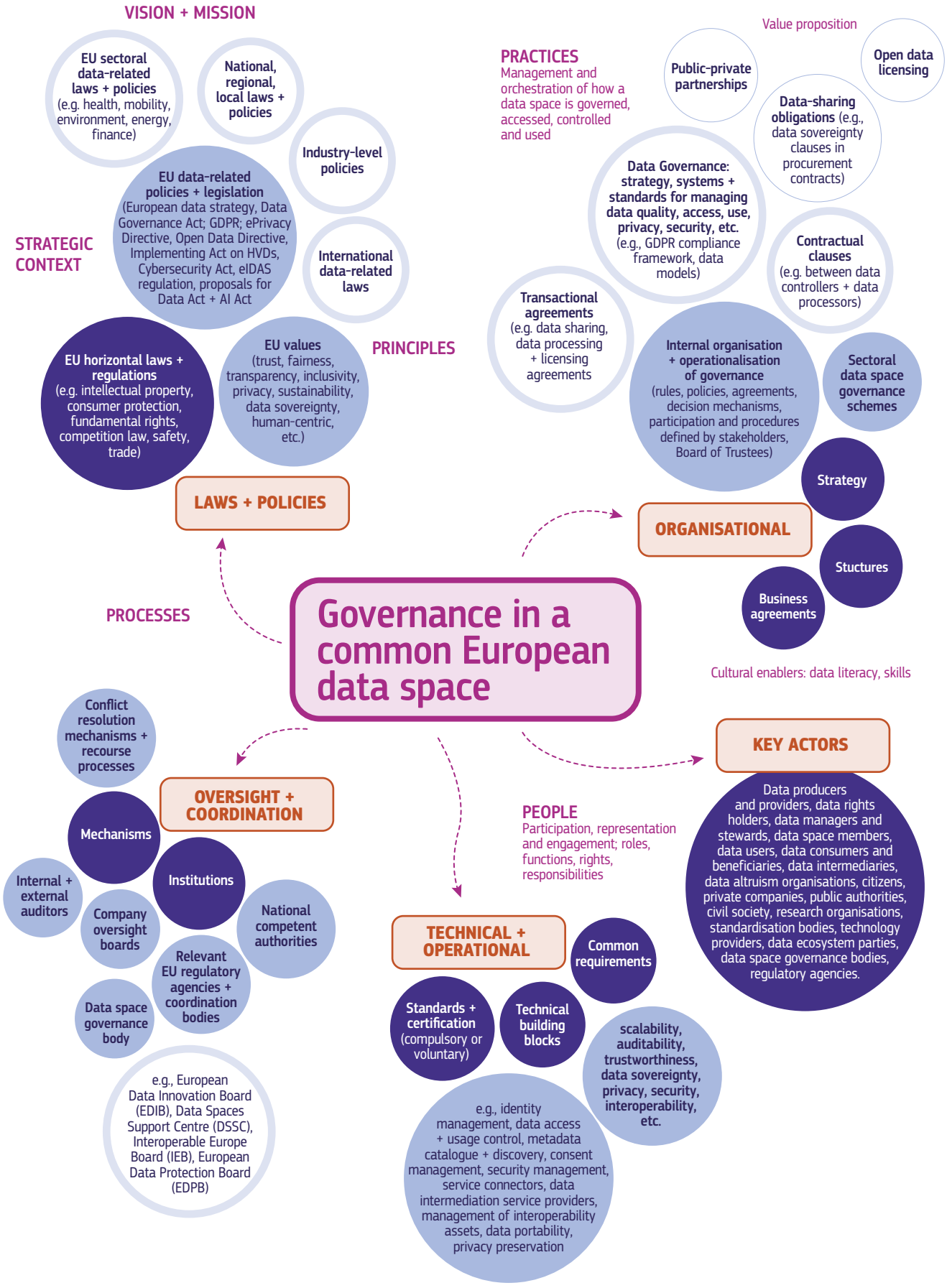
14 See for example the ARE3NA study on AAA for Data and Services that engaged with varying roles and access to geospatial data, **https://joinup.ec.europa.eu/collection/are3na/solution/are3na-study-aaa-data-and-services**

15 **https://commission.europa.eu/publications/interoperable-europe-act-proposal_en**

16 Displayed via the EIF Toolbox: **https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/solution/eif-toolbox/eif-toolbox**

that governance can be a complex task involving multiple dimensions, this diagram is not intended to be a fully comprehensive representation - nor does it attempt to capture the complex dynamics and interrelationships between the different aspects and actors. Data spaces may also involve a variety of platforms and systems, adapted to different market designs, sectoral needs and organisational or business processes.

→ **Figure 4. Governance in a common European data space.**



The following sub-sections elaborate in more detail the concepts and dimensions of governance, including data governance, and then seek to apply these within the context of the common European data spaces.

2.3.2 Governance – principles, practices and processes

Governance encompasses the system of rules, principles, policies, practices, and processes to direct and manage an organisation, as well as the mechanisms by which it, and its stakeholders are held to account (Data Spaces Support Centre, 2023a; 2023b). Fritzenkötter et al. (2022) define governance as the range of political, institutional and administrative principles, rules and practices; as well as formal and informal processes through which decisions are taken and implemented.

An effective governance framework must have a clear organising purpose that can inform the development of these principles, processes, and practices, and should ensure that data projects are pursued with consideration for the broader technical, social, political, and economic contexts within which the data is produced and consumed.

Governance also covers stakeholder interactions, how they articulate their interests and have their concerns considered. This understanding is informed by the notion of governance in political science, which acknowledges that a broad set of actors and institutions, beyond regulators, are involved in managing societies, including private sector, civil society and other non-governmental organisations (Kooiman, 2003), as well as the (democratic) influence of citizens.

Governance therefore includes business, legal (rights and obligations), sociological, political and organisational aspects. It is implemented with a set of overlapping legal, administrative, organisational, business, and technical measures and procedures that define the roles, functions, rights, obligations, attributions, and responsibilities of each participant (Torre-Bastida et al., 2022). **These measures and procedures may span from contractual, organisational and operational agreements; to technical standards and tools; and from goals and principles to laws and regulations**.

2.3.3 Data Governance

As the exchange of data is one of the core functions of a common European data space, trustworthy and responsible data governance is a primary concern. While there are multiple definitions of data governance, generally it is understood as the exercise of decision-making and authority for data-related matters and it consists of norms, principles, rules and processes concerning various types of data and their use. For Torre-Bastida et al., 2022, it entails defining, implementing and monitoring strategies, policies and shared decision making over the management, collection, analysis, sharing and use of data assets as well as related aspects, such as data rights, data privacy, and data security, among others. According to the Organisation for Economic Co-operation and Development (2022), data governance refers to "diverse arrangements, including technical, policy, regulatory or institutional provisions, that affect data and their creation, collection, storage, use, protection, access, sharing and deletion across policy domains and organisational and national borders." It also encompasses the social, economic, political, and cultural factors that explain how data is accessed and controlled by a multitude of actors (Liu, 2022). As explained by Davis (2022), good data governance efforts should seek to maximise the benefits from data, while addressing and minimising related risks and challenges, including to rights and interests.

Within the EU, a series of regulations have been introduced concerning the governance and protection of data, e.g., the GDPR (European Union, 2016), which regulates the processing and transfer of personal data, and the Open Data Directive (European Union, 2019a) which regulates the reuse of public sector data. Privacy and data protection are also enshrined in the Treaty on the Functioning of the European Union (European Union, 2012a) and in the EU Charter of Fundamental Rights (European Union, 2012b). More recently, and as mentioned in **Section 1.1**, several cross-sectoral legislative initiatives to govern the access, use and sharing of data have been launched under the European strategy for data. The Strategy recognises the need to develop governance structures for handling data and to increase pools of quality data, including across borders.

This European approach to data governance "envisions trustworthy data governance that reconciles responsible and human centric data governance, subject to full compliance with the EU's strict data protection rules, while enabling data governance to foster innovation, and to drive economic growth" (European Commission, 2020a).

Data governance within a common European data space should set out a framework outlining clear roles, duties, standards and responsibilities, to ensure that data is appropriately protected, while also supporting data sharing and openness to data mobility. Good data governance can facilitate a fair data economy, address the cross-border nature of data flows and data storage practices, as well as helping to foster public trust. It should include an assessment of different business models for data sharing and put in place contractual agreements to enable data flows between different stakeholders (to also be recognised as open data licensing, where appropriate).

Recognising the complex interplay between cross-cutting legal frameworks, organisational aspects, and contractual or transactional dimensions, the IDSA Rule Book (International Data Spaces Association, 2020; International Data Spaces Association, 2023) refers to "layers of data space governance". Similarly, Torre-Bastida et al. (2022) and Bodó et al. (2021), identify various levels of data governance in a data space – micro (intra-organisational), meso (inter-organisational) and macro (legislative framework). For example, at the meso or inter-organisational level, governance should provide the common principles, rules and requirements for orchestrating data access and management for participants according to the different roles they may play within the data space ecosystem. As legislation only provides the general, overarching framework for data sharing, the legal dimension of a data space also includes a contractual framework so that the different participants can agree on specified rules that fit their data sharing context. For example, the contractual clauses of a data space use case may state the availability of the product, the totality of participants or its publication under an open data license. The governance code of a data space should also enable the inclusion of information access and use policies in the corresponding agreements.

For this purpose, it can be useful to look at existing data governance models and good practices that can be leveraged in terms of sharing data between different types of stakeholders. This is anticipated by the Data Spaces Support Centre, which has announced the intention to publish a Data Governance Matrix, as well as a Legal Compass and a Catalogue of Contractual Models. The Data Governance Matrix will outline roles, business strategies and the legal environment, as well as recommending data governance models and mechanisms. Torre-Bastida et al. (2022) also provide some examples of principles that should usually be considered, as well as the type of rules and requirements that can be used in the governance of a data space. Similarly, SITRA's rulebook for a fair data economy provides a set of tools and contractual templates for building a data-sharing network, in addition to setting out relevant legal, business, technical, security, and administrative rules, as well as ethical guidelines to be observed (SITRA, 2022).

Linked to these models, resources and good practices, some relevant data governance questions are covered by the How-to series that is contained in **Section 3** of this report, e.g., '*How can Governments access private sector data of public interest in a common European data space?*', '*How can voluntary data sharing be leveraged in a common European data space?*' and '*Which actors are providing what types of data within the scope of a common European data space?*'

Finally, it is important also to recognise the trust-bringing role of third parties such as data intermediaries, such as those as regulated by the DGA, which can act as neutral facilitators and enablers of data reuse in a common European data space. These actors mediate between the suppliers of data, the data subjects, the data storage providers, and the data utilisers. Depending on specific needs, they can serve as matchmakers between supply and demand for data, offer services and relevant technologies to help share data more easily, and provide a guarantee that data will be handled in a trustworthy and legally compliant manner. For example, customer-facing companies can share data with data intermediaries that they may not be able to share directly with other businesses, often for reasons pertaining to privacy, competition or intellectual property rights. This data can then be reused by Business to Consumer (B2C) services that are connected to the data intermediary - in this way reliable and legal access is secured to the kind of data that the business requires in order to provide added value to its customers, all while retaining the individual's data sovereignty.

2.3.4 Governance within and between common European data spaces

Governance within a European data space should be based on commonly-held EU societal values and established principles relating to data sharing (see **Section 1.1.2**). As advocated by the Data Spaces Support Centre (2023a), these values and principles should be enshrined in concrete settings and brought to the level of technical implementation within European data spaces to ensure individual and collective control of data and safeguard a human-centric and fair approach. Establishing robust governance mechanisms to achieve this will help gain the trust of the stakeholders involved in a data space.

The SWD on Common European Data Spaces (European Commission, 2022c) outlines that data spaces should put in place an appropriate governance structure to ensure fair, transparent, proportionate and non-discriminatory access to, sharing and use of data. That structure should comply with existing provisions of horizontal data-related and Digital Single Market legislation, e.g., the GDPR (European Union, 2016), Free Flow of Non-Personal Data Regulation (European Union, 2018), ePrivacy Directive (European Union, 2002), Platform to Business Regulation (European Union, 2019b), as well as relevant sectoral EU data-related legislation. Common European data spaces will also need to comply with the applicable EU legal framework on cybersecurity, safety, intellectual property, fundamental rights, environmental protection, competition law, and other relevant rules of data services in the EU, such as international trade commitments under the World Trade Organization's (2022) General Agreement on Trade in Services and other trade agreements. In addition, adequate technical, legal and organisational measures should be put in place to prevent unauthorised access to personal and non-personal data.

As described in **Section 1.1**, an overarching governance context for European data spaces is formulated in the Data Governance Act (European Commission, 2020c), which offers a means for greater societal reuse and access to data, by setting the enabling conditions and providing the concrete arrangements in which trustworthy data sharing and/or data pooling can occur. The DGA comprises horizontal measures relevant for all common European data spaces, while leaving room for the application of sector-specific rules, where appropriate. This is complemented by the proposal for a Data Act (European Commission, 2022a), as mentioned earlier.

Furthermore, governance schemes are beginning to be elaborated for specific sectoral data spaces, for example the European Health Data Space will be established by a Regulation (European Commission, 2022b) and the forthcoming Commission Communication on the creation of the common European Mobility data space will describe the main features and objectives for the common European Mobility Data Space (EMDS), including a defined governance system[17]. Individual European data space initiatives will develop the governance framework that contains the rules and practices for that particular data space's governance, management and operations. At the same time, it is important that initiatives for implementing sectoral data spaces should be aligned with the general set of rules applicable to all common European data spaces.

From the above, it is clear that the various legal aspects affecting the creation and operation of common European data spaces are multi-fold, shaping their purposes, goals, business structures and technical features. In order to help navigate the legal landscape pertaining to data spaces, the Data Spaces Support Centre (2023a) has proposed consideration of legal issues across six different areas: (1) Substantive rights and obligations pertinent to data; (2) Data contracts; (3) Competition law; (4) Organizational aspects; (5) Techno-legal interoperability; and (6) Regulatory aspects. Therefore, and as outlined in the checklist contained in the DSSC's Data Spaces Start-up Kit, it is essential to ask which legal aspects are relevant when setting up a particular common European data space, as well as to consider the data and organisational governance model (Data Spaces Support Centre, 2023a). This is also one of the questions addressed by the How-to guides contained in **Section 3** of this report, i.e., '*Which legal aspects should be considered when creating, providing or using novel data-driven solutions in a data space?*'

For example, the European Commission's impact assessment study for the European Health Data Space - EHDS (European Commission, 2022e) clarifies that it will ensure coherence with relevant EU legislative frameworks, including the GDPR (European Union, 2016), the Data Governance Act (European Commission, 2020c), the proposed Artificial Intelligence Act (European Commission, 2021f), the cybersecurity regulatory framework (European Union, 2019c), the eIDAS (electronic IDentification, Authentication and Trust Services) regulation (European Union, 2014), the pharmaceutical regulatory framework[18], the Cross-border Healthcare Directive (European Union, 2011) and the Medical Devices Regulation

17 https://transport.ec.europa.eu/news-events/news/share-your-views-common-european-mobility-data-space-2022-11-14_en

18 https://www.ema.europa.eu/en/about-us/what-we-do/legal-framework

(European Union, 2017), as well as international trade commitments. Regarding the provision of a framework for health data reuse, it states that the EHDS will build upon the DGA. However, as a horizontal framework, the DGA cannot address the specificities of sensitive data, such as health or genetic data. Within the EHDS, it will therefore be necessary to consolidate the requirements and technical framework needed to achieve a functioning system in the field of primary and secondary use of health data, complementing the DGA rules with more detailed or more practical rules considering the specific nature of health data. The EHDS will also build upon the Data Act proposal and establish standards and specifications for data portability and interoperability, thus making the portability and access of data linked to devices (medical devices and wellness apps) both technically and practically possible.

As another example, in the case of the Green Deal Data Space, it will be important to consider the underpinning policies and sectoral data-related legislation, e.g., in the areas of financing, sustainability reporting, energy and mobility, etc., as well as forthcoming updates to the existing legal framework for environmental data sharing, and the Implementing Act on a list of High-Value Datasets.

As outlined by Nagel and Lycklama (2021), adapting data spaces to the specific needs of discrete ecosystems depends on the given requirements of a domain, sector, or territory. For instance, a data space for the energy industry should take into consideration the specific needs of the stakeholders of that domain (distribution system operators, transmission system operators, etc.), and it should also support the constraints of the smart-grid cyber-physical system (e.g., when building a digital twin, so that data exchange has an effect on grid operation).

Governance (and trust) will become especially important when working across sectoral boundaries and where there is a need for multiple commons, communities and actors to interact together, for example in use cases across thematic data spaces, such as the future Green Deal Data Space. Likewise, the EU Data Space for Smart Communities, to be interconnected with the Green Deal Data Space, should take into consideration the multitude and variety of possible cross-domain interactions (e.g., energy, mobility, health, water and waste management, etc.).[19] Or similarly, the EMDS may need to interact and link with, for example, the Tourism data space or the Industrial and Manufacturing data space. There will thus be a need to accommodate different data sources and services, and different layers of data curation, enrichment and analysis, which may be separated by legal, institutional, infrastructural and even funding silos (Islam et al., 2022). In this sense, common European data spaces should ensure inter-data space interoperability (including legal interoperability), allowing for cross-sectoral data exchange and synergies between multiple data space instances or data sharing initiatives, in order to pave the way towards the federation of interoperable data spaces envisaged by the European strategy for data.

According to Torre-Bastida et al. (2022), the governance in a data space sets the management and guides the process to achieve the vision of the community - that is, to create value for the data space community by facilitating the finding, access, interoperability, and reuse of data in a trusted and secure environment, irrespective of its physical location. For each participant of the community, this implies that the authority and decisions that affect its own data assets should be orchestrated also with the rights and interests of other collaborating participants of the data space.

A fit-for-purpose data space governance framework should create the conditions for more systematic, sustainable, and responsible data sharing and collaboration. It will thus be important to design a governance system that can incentivise participation and collaboration between diverse actors in a way that fosters trust and inclusion, Braun et al. (2022), for example, ask whether a hypothetical Urban data space could lead to an integration of stakeholders and offer a new sphere for creativity and participation, as well as novel formats for engaging those in the urban public sphere who were previously not involved. Similarly, in the context of the Green Deal Data Space, it will be interesting to explore opportunities and governance approaches for the integration of citizen-generated content, e.g., contributions of data via recognised data altruism organisations (RDAOs), as well as new citizen engagement practices.

19 See Data Space for Smart and Sustainable Cities and Communities (DS4SSCC), a preparatory action for the creation of a data space for smart communities under the DIGITAL-CSA DIGITAL Coordination and Support Actions, which aims to create a sustainable data governance scheme as well as a blueprint that connects existing local data ecosystems and EU systems and enables public and private stakeholders to develop cross-sector, cross-community, data services (**https://www.ds4sscc.eu**).

2.4 Relevant data sharing initiatives and component reuse

Non-functional requirements, such as those identified in the **Section 2.1**, are expected to be ranked and prioritised according to the specific needs and priorities of individual sectoral data spaces. Similarly, different sets of characteristics may be selected for sector-specific common European data spaces, as discussed in **Section 2.1**.

Indeed, as mentioned above, the construction of data spaces does not start from scratch. Many piloting activities have been already carried out internationally by different associations and organisations interested in data sharing (see **Table 2** for a sample of relevant initiatives) and there is also a substantial wealth of practical knowledge in Europe concerning the key ingredients and recipes for building data sharing ecosystems. Feasible architectural and governance patterns have already been identified in the context of pilots, and in some cases turned into technical or business specifications.

European data spaces should leverage these examples, supported by the adoption of a design process based on the principle of composing and assembling independently developed elements with clearly defined – and preferably standardised – interfaces. Besides favouring reusability, this strategy would substantially reduce development effort and time-to-market, as well as increase reliability by leveraging tried and tested components.

Considering this, the top-down requirement elicitation process described in **Section 2.1** should be complemented with a bottom-up analysis of existing alternatives. Some of these provide no technical specification, staying at the level of design principles. However, others provide architecture and interface specifications, and even open source implementations of key building blocks (see **Table 2**).

→ **Table 2: Sample of potentially relevant data sharing initiatives.**

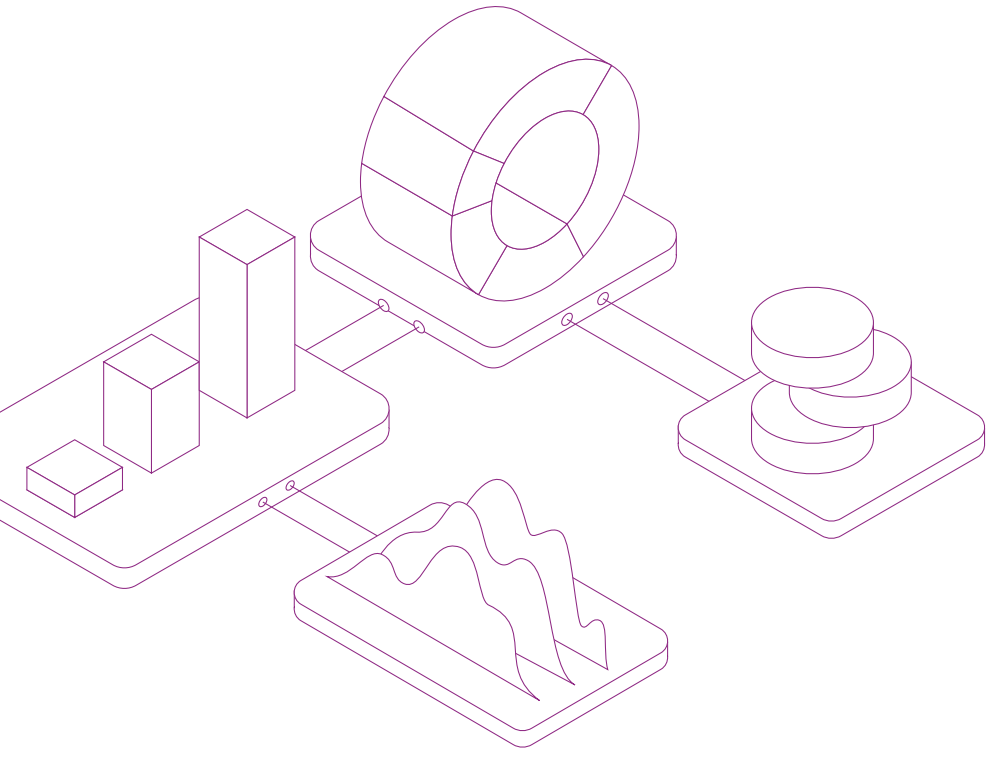
GAIA-X[20]	Specification of a federated open data infrastructure based on European values
IDSA Reference Architecture[21]	Reference architecture for international data spaces including a governance model and adoption strategy
FIWARE Catalogue[22]	Open source implementation of standards enabling the development of portable and interoperable smart solutions
S.O.L.I.D[23]	An open specification to enable participants to store their data securely in decentralised data stores
CEF Building Blocks[24]	Building blocks providing basic digital services developed in the context of the Connecting Europe Facility (CEF), an EU funding instrument
OCEAN Protocol[25]	Decentralised data exchange protocol enabling data owners and consumers to publish, discover and consume data in a secure, privacy-preserving fashion
INSPIRE[26]	EU-wide, standards-based geospatial data sharing infrastructure built on the legal requirements of the INSPIRE Directive, offering open source-based components
SIMPL[27]	Smart middleware developed within the DIGITAL programme that will enable cloud-to-edge federations and support data initiatives and the establishment of data spaces

20 https://gaia-x.eu
21 https://internationaldataspaces.org
22 https://www.fiware.org/
23 https://solidproject.org/
24 https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL
25 https://oceanprotocol.com/
26 https://inspire.ec.europa.eu
27 https://digital-strategy.ec.europa.eu/en/news/simpl-cloud-edge-federations-and-data-spaces-made-simple

The initiatives described in **Table 2** are varied, offering different sets of characteristics, implementing various management and governance mechanisms and prioritising different quality attributes. This represents a promising starting point for common European data spaces, as a wide range of conceptualisation styles could already be partially supported by existing implementations, including for instance (looking at the technical architecture):

- Centralised data catalogues offering high performance, auditability and productivity features;
- Federated data sharing ecosystems, following algorithm-to-data and edge-computing principles, offering stronger control, scalability and portability characteristics;
- Fully decentralised systems, e.g., blockchain-based for transparency and auditability, or personal information management systems, prioritising confidentiality, interoperability and portability qualities.

These are just some examples of the range of architectures already evaluated in the context of European and global data sharing initiatives. A next step could include the analysis of these initiatives against the functional and non-functional requirements presented earlier in this report. This would pave the way for the identification of promising building blocks for common European data spaces, and for the adoption of approaches from a set of field-tested options. It would also represent a step towards the definition of more concrete and measurable requirements, gradually breaking down high-level priorities into more granular ones. A systematic definition of requirements rooted in EU policy needs, combined with the analysis and reuse of lessons learned and building blocks from existing data sharing initiatives, would accelerate the design, development and deployment of European data spaces, promoting European competitiveness and digital sovereignty.



3 Data sharing How-to guides

As mentioned in **Section 1.2.3 B**, we have prepared dedicated data sharing information sheets or How-to guides, which aim to address some of the key questions that may arise for stakeholders conceptualising and operationalising a data space, including the common European data spaces.

These How-to's elaborate on selected technical and organisational (i.e., governance-related) aspects of sharing and accessing data in data spaces. Structurally, they outline a high-level problem statement or scenario, followed by a proposed approach which may be applied in order to address it. Where possible, concrete examples are included to explain the problem and/or proposed solution. Furthermore, references to additional resources and materials, including, but not limited to those produced by the JRC, are provided. The list of topics addressed by the How-to's is shown in Table 3.

→ **Table 3: List of technical and organisational How-to's.**

		How-to Information Sheets	Data Space Theme
TECHNICAL	1	How can stakeholders benefit from synthetic data in a data space?	**Synthetic data**
	2	How to choose the best software stack for a data space?	**Software stacks**
	3	How to ensure clear access and use conditions for a dataset in a data space?	**Licensing**
	4	How to ensure that datasets shared by different actors in a data space can be used together?	**Interoperability**
	5	How to ensure that technical requirements and standards are being followed?	**Data validation**
	6	How to facilitate the discovery of data in a data space?	**Data discoverability**
	7	How to select the most appropriate standards for a data space?	**Data standards**
	8	How to ensure that digital resources and data are uniquely referenced in a data space?	**Data registers**
	9	How to provide access to data in a data space?	**APIs for data access**
	10	How to preserve privacy and protect personal data and sensitive business data in a data space?	**Privacy enhancing technologies**
ORGANISATIONAL	1	Which actors are providing what types of data in scope of a data space?	**Data Actors**
	2	How to foster a people-centred approach to data in a data space?	**Citizen data**
	3	How can business benefit from sharing data in a data space?	**Benefits to business in data spaces**
	4	How can governments access private sector data of public interest?	**Accessing data (B2G)**
	5	How can data transparency for AI systems be increased in a data space?	**Transparency – AI data in data spaces**
	6	How to leverage voluntary data sharing in a data space?	**Voluntary data sharing**
	7	Which legal aspects should be considered when creating, providing or using novel data-driven solutions in data spaces?	**Legal**

TECHNICAL
PAGE 37 > 56

ORGANISATIONAL
PAGE 57 > 70

HOW CAN STAKEHOLDERS BENEFIT FROM SYNTHETIC DATA IN A DATA SPACE?

TECHNICAL

What is the problem?

Common European data spaces are conceptualised in a way that includes data from heterogeneous sources. This in turn may introduce different problems such as:

- Privacy constraints/threats to privacy
- Licensing issues
- Gaps in/lack of representativeness
- Data fragmentation/missing or insufficient data

Scenario

In real-world scenarios these problems can often be observed as a dataset constrained by privacy considerations that is fragmented and/or not representative.

A first scenario could involve the privacy of patient records. The medical community cannot share cancer patients' records for privacy reasons. At the same time, sharing these data would allow medical doctors and researchers to gain insights from many patients and build representative statistics. This would in turn help patients to receive improved treatment for their illness.

Another scenario details gaps in representation. Policymaking bases its activity on average demographic statistics. As such, minorities, people in situations of vulnerability and smaller communities, can share certain traits or patterns which may be generally under-represented in data. Conversely, Big Tech companies exploit advanced technologies to personalise the services and products offered to individual profiles.

A final scenario involves data augmentation. To train artificial intelligence (AI) algorithms for autonomous vehicles, data from sensors, cameras, laser and radar interferometry (LiDAR and RADAR) systems cannot cover all the different, and sometimes rare, scenarios that may be encountered by the vehicle, such as, e.g., a deer crossing the road. However, to ensure the safety of citizens, the algorithm should be trained to mitigate every type of event, from rare to general occurrences.

Proposed solution(s)

Synthetic data are data randomly and artificially generated from an original dataset and maintain a certain (given) similarity with it, while not including any of the original records but still preserving analytical value. Synthetic data can help overcome similar challenges in many fields of applications.

In the scenario where personal health data cannot be shared for privacy reasons, artificial data can replicate the characteristics of the original dataset, with no one-to-one mapping of single records. The synthetically generated dataset can be released with a pre-determined level of privacy that does not negatively impact its utility. Researchers can use synthetic cancer registry data to test hypotheses and gain insights without breaching patient privacy.

In the gap representation scenario, artificial profiles can be generated by exploiting technologies similar to those used by Big Tech companies for tailoring their advertising based on user preferences. Statistical personae are synthesised by crossing census data with additional data sources (e.g., behavioural surveys), and are used as input in model simulations to gain insights on habits, responses to policy scenarios, mobility patterns, contact tracing in epidemiological emergencies, etc.

Finally, in the AI scenario, many AI applications require varying levels of data augmentation. Artificial data can be generated by exploiting game engines and can entail any possible scenario.

It is important to point out that even synthetic data imply a certain risk of re-identification, which can, to a large extent, be controlled by design. As an example, imagine a certain rare trait in the population in sparsely populated areas. The closer the data are to the real population, the more valuable and valid they are for analytic purposes (utility), but the greater the risk of re-identification. Synthetic data that retain a high utility (hence most suited for analytical purposes) can also be costly to produce, but are generally more suitable to be re-used for more than one purpose. Therefore, it is important to assess the risk-benefit ratio of adopting synthetic data.

Recommendations

→ **Data providers/intermediaries**

- **Before releasing a dataset, analyse what level of risk of re-identification is acceptable on a case-by-case basis**
- **Consider the generation of synthetic datasets as an alternative to data anonymisation or other privacy enhancing technologies**

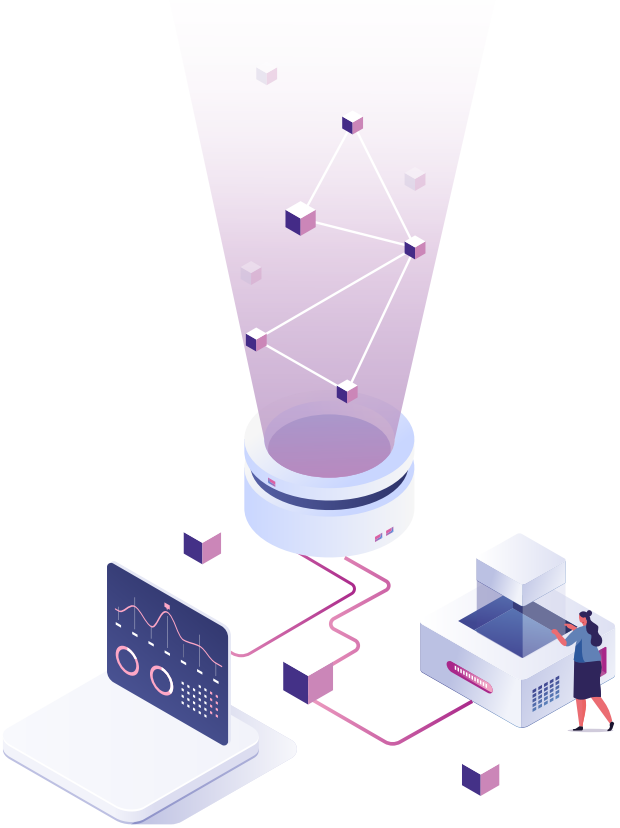


Illustration: Silvia/stock.adobe.com, icons: © iStock.com/Yuriy Bucharskiy, © iStock.com/da-vooda

Selected resources

→ Hradec, J., Craglia, M., Di Leo, M., De Nigris, S., Ostlaender, N. and Nicholson, N., **Multipurpose synthetic population for policy applications**, EUR 31116 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53478-5, doi:10.2760/50072, JRC128595.

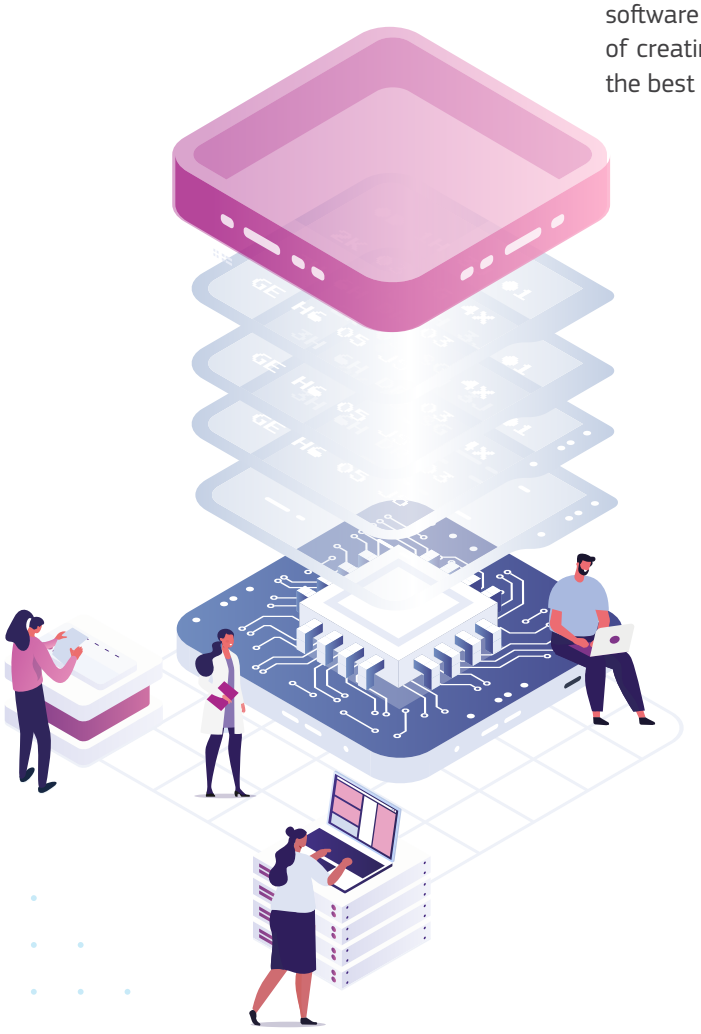
→ **Artificial cancer dataset created by Public Health England (PHE):** *https://simulacrum.healthdatainsight.org.uk*

HOW TO CHOOSE THE BEST SOFTWARE STACK FOR A DATA SPACE?

TECHNICAL

What is the problem?

A data-sharing infrastructure, such as a data space, is made up of a number of software components, collectively known as a "software stack". A software stack provides the functionalities required such as data storage, publication and discovery to authentication, data protection and access control, data validation, processing and analytics, orchestration and portability. A large number of software tools exist to address each of these functionalities, leading to the risk of creating fragmented infrastructures and to the problem of how to choose the best software stack for a specific data space.



Scenario

In the common European Green Deal Data Space, e.g., for an environmental impact assessment, several categories of data such as citizen-generated data, satellite imagery from Copernicus, public domain data and privately held sensors should be accessed, combined and processed through the use of AI algorithms.

This requires the use of versatile tools such as "extract, transform, load" (ETL) tools, as well as specific libraries for training and testing of AI algorithms. These tools can be used most effectively if the necessary data are made available in line with "findability, accessibility, interoperability and reusability" (FAIR) principles, which in turn require the use of additional instruments such as APIs, metadata editors and software implementations of international standards. Access control and privacy preservation mechanisms for citizen-generated data and sensitive private data should be also put in place. Finally, the re-users of the AI solutions, such as decision-makers and policymakers, can benefit from web and mobile applications for visualising and downloading the end results.

Proposed solution(s)

There is no universal solution to the problem of selecting the best software stack for a data space. The main classification is between proprietary software tools, which typically consist of monolithic applications (built/presented as a single unit) and offering several out-of-the-box functionalities, and open source tools, which are usually modular components that developers can then combine to build complex infrastructures. Aligned with the FAIR principles and from a European perspective (for example, see the Commission's Open Source Software Strategy 2020-2023), open source software represents a very good fit for the needs of data spaces and provides a viable alternative to proprietary solutions. Open source applications are (typically) available free of charge and provide transparent access to the underlying source code, which facilitates innovation and early adoption of new standards and technologies.

The forthcoming smart middleware platform (SIMPL) supported by the European Commission will provide a number of open source technical enablers for the establishment of data spaces.

Suggested criteria to be used in the choice of a software stack:

- Support of international standards.
- Existence and vitality of a dedicated software community. This implicitly indicates the maturity of a software and can be measured by, e.g., the existence and clarity of technical documentation, number of commits, issues and bug fixes on social coding platforms such as GitHub, organisation of community event, etc.
- The availability of clear contribution guidelines, codes of conduct and mechanisms for quality checking of the software codebase.
- Ease of maintenance and potential for technical evolution

In terms of software licenses, the European Union Public License (EUPL) is the Commission's reference for openly licensing software artefacts. While disregarding the question of compatibility of licenses, its use is recommended in case a license is to be attributed to a newly created software tool.

Recommendations

→ Infrastructure providers

- When possible, evaluate the use of open source software components
- Consider licensing new software/hardware products under open license, possibly the the EUPL
- Document technical solutions for reproducibility and auditability purposes

→ Data users

- Provide feedback on the usability of tools and data
- Be engaged in the co-creation and co-design of data-sharing practices and tools

→ Data intermediaries

- Identify existing software tools, or develop new ones, that can help interconnect existing fragmented infrastructures
- Promote good practices for the use of software tools for data-sharing
- Organise community events such as hackathons, datathons
- Facilitate the rapid testing of technology through dedicated sandboxes

Selected resources

→ **European Commission's open source software strategy 2020-2023**: *https://ec.europa.eu/info/departments/informatics/open-source-software-strategy_en*

→ **Introduction to the European Union Public License**: *https://joinup.ec.europa.eu/collection/eupl/introduction-eupl-licence*

→ Kotsev, A., Minghini, M., Cetl, V., Penninga, F., Robbrecht, J. and Lutz, M., **INSPIRE - A Public Sector Contribution to the European Green Deal Data Space**, EUR 30832 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-41564-0, doi:10.2760/8563, JRC126319.

→ **Simpl: cloud-to-edge federations and data spaces made simple:** *https://digital-strategy.ec.europa.eu/en/news/simpl-cloud-edge-federations-and-data-spaces-made-simple*

HOW TO ENSURE CLEAR ACCESS AND USE CONDITIONS FOR A DATASET IN A DATA SPACE?

TECHNICAL

What is the problem?

Very often datasets are made available with unclear or missing licensing information. This creates uncertainty for data users in a data space in terms of how data may or may not be used, leading to inefficient reuse of the data. This is exacerbated in cases where data space users want to combine existing datasets, and/or make use of artificial intelligence (AI) algorithms, in order to create a derivative dataset or service.

Scenario

A user wants to combine heterogeneous datasets produced by different organisations to perform a cross-border environmental risk assessment.

In cross-border use cases, the availability of licensing information in (different) national languages can make the user's task problematic, especially if the description of how the datasets may be used is either unclear or absent. Lack of clear information about licensing conditions can create challenges in carrying out the assessment without risking copyright infringement and potentially violating the access conditions to the data, including by generating a derivative product.

Proposed solution(s)

Licensing information should always be available for all datasets in a data space. Whenever possible, standard licenses should be used such as the Creative Commons (CC) family of licenses, which are very well known and have been adopted by a global community. Suggested CC licenses include CC0 or CC-BY, the latter only requiring that appropriate credit is granted to the licensor. The use of such licenses also makes it easy to determine the licenses attributed to derivative products and services. Finally, to facilitate the automated exchange and use of data, machine-readable licenses (i.e., licenses including code that allows automatic discovery by machines, e.g., search engines) should be provided.

Recommendations

→ **Data providers**

- Always provide licensing information together with datasets
- Consider changing custom licenses to standard-based licenses
- Adopt machine-readable licenses to facilitate the automated use of the data
- Consider open and non-restrictive licenses to maximise the benefits and opportunities for data reuse, such as CC licenses

→ **Data users**

- Carefully consider licensing information when using datasets and creating derivative products.

→ **Public authorities**

- Prescribe the use of licenses that would maximise the benefit generated from the reuse of the dataset
- Explicitly require that data licensing conditions are applied when procuring data

Selected resources

→ Granell, C., Mooney, P., Jirka, S., Rieke, M., Ostermann, F., Van Den Broecke, J., Sarretta, A., Verhulst, S., Dencik, L., Oost, H., Micheli, M., Minghini, M., Kotsev, A. and Schade, S., **Emerging approaches for data-driven innovation in Europe**, EUR 30969 EN, Publications Office of the European Union, Luxembourg, 2022, doi:10.2760/511775, JRC127730.

→ Kotsev, A., Minghini, M., Cetl, V., Penninga, F., Robbrecht, J. and Lutz, M., **INSPIRE – A Public Sector Contribution to the European Green Deal Data Space**, EUR 30832 EN, Publications Office of the European Union, Luxembourg, 2021, doi:10.2760/8563, JRC126319.

→ Minghini, M., Kotsev, A. and Lutz, M., **Comparing INSPIRE and OpenStreetMap data: how to make the most out of the two worlds**. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XLII-4, No W14, 2019, pp. 167–174. doi:10.5194/isprs-archives-XLII-4-W14-167-2019.

→ **Creative Commons (CC) licenses:** *https://creativecommons.org*

→ **CC0 license**: *https://creativecommons.org/share-your-work/public-domain/cc0*

→ **CC-BY license**: *https://creativecommons.org/licenses/by/4.0*

HOW TO ENSURE THAT DATASETS SHARED BY DIFFERENT ACTORS IN A DATA SPACE CAN BE USED TOGETHER?

TECHNICAL

What is the problem?

Interoperable data allow for easy creation, sharing and reuse of information between stakeholders. However, often datasets are very difficult or impossible to combine, and as such, have low interoperability. This is due to a complex combination of factors that may be legal, organisational, technical and semantic in nature. Datasets may be made available in different formats (e.g., open/proprietary, well/poorly supported); licensing information may be incompatible or missing; and data content from different sources may be semantically different. Combining data from multiple sources is a challenging task and can exacerbated by the factors mentioned above; therefore, achieving data interoperability is a topic of critical importance.

Scenario

A public administration is looking for multiple datasets in order to develop a new urban plan aimed at increasing walkable and green areas within a region. The public authority discovers and collects a number of relevant datasets provided by different actors, e.g.: multiple road network datasets published by several local transport agencies; number, size and location of green areas (multiple datasets provided by the national environmental ministry and by local agencies); the number and location of citizens' access points/ routes to the cities in the region, measured by sensors operated by different companies; and the popularity of city areas, determined from geotagged photos and comments posted on social networks.

The potential value generated from these datasets in developing a new, efficient urban plan, could be almost negated if one or more datasets are not provided in an interoperable manner. For example, it may not be possible to combine the local road network datasets into a single regional dataset if they are provided in non-standard formats featuring little or no software support, and/or if their data models are different, and/or if they are made available under different and non-compatible licenses.

Proposed solution(s)

Data interoperability is an integral part of digital solution design and is crucial for data reuse. Creating interoperable data can be achieved through the adoption of common specifications, such as those defined in open standards. Using commonly agreed, international, mature and well-supported standards is advisable for all aspects related to the documentation, modelling, encoding and sharing of data.

As interoperability specifications can quickly become complex, it is very important to decide on the desired level of complexity and interoperability that would best correspond to the envisioned use-cases, as well as to the capacities and incentives of the stakeholders involved. In addition, interoperability should be contextualised, and seen as a means for achieving specific objectives, i.e., as a tool, and not as the end objective of data sharing. A number of domain-specific interoperability initiatives already exist. A notable example is represented by Minimal Interoperability Mechanisms (MIMs) – a set of standards-based capabilities defined for smart cities.

Due to the critical importance of the topic, the proposed Data Act contains a dedicated Chapter with an emphasis on interoperability in the context of European Data spaces. The European Interoperability Framework (EIF) through its 12 principles and conceptual models, provides a solid foundation for improving the interoperability of data sharing practices around four dimensions: legal, organisational, semantic and technical. In addition, the European Commission has launched Interoperable Europe – an initiative committed to introducing a new cooperative interoperability policy for Europe that will transform public administrations and help them in their digital transformation.

Recommendations

→ **Data providers**

- Use open standards for encoding and sharing data.
- RESTful APIs, ideally documented by following the OpenAPI specifications, to facilitate data accessibility.
- Provide standard-based metadata for all datasets.
- Whenever possible, reuse existing interoperability assets instead of developing own ones.

→ **Data users**

- Consider the provision of feedback to data/service providers on datasets used with the aim to improve their usability and interoperability.
- Provide input, and whenever possible contribute to open source projects, interoperable solutions and standardisation initiatives.

→ **Public authorities**

- Consider data interoperability and the delivery of open source solutions as key requirements of procurements involving data and service provision.
- Contribute own developments to open source projects and standardisation initiatives.

Selected resources

→ Ulrich, P., Duch Brown, N., Kotsev, A., Minghini, M., Hernandez Quiros, L., Boguslawski, R. and Pignatelli, F., **Quantifying the Benefits of Location Interoperability in the European Union**, EUR 31004 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-48846-0, doi:10.2760/72064, JRC127330.

→ Boguslawski, R., Valayer, C., Mazzoli Kronborg, U. and Pignatelli, F., **European Union Location Framework Blueprint**, EUR 30872 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-42641-7, doi:10.2760/367025, JRC126653.

→ Kotsev, A., Minghini, M., Cetl, V., Penninga, F., Robbrecht, J. and Lutz, M., **INSPIRE - A Public Sector Contribution to the European Green Deal Data Space**, EUR 30832 EN, Publications Office of the European Union, Luxembourg, 2021, doi:10.2760/8563, JRC126319.

→ **European Interoperability Framework**: *https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/european-interoperability-framework-detail*

→ **Data Act: Proposal for a Regulation on harmonised rules on fair access to and use of data:** *https://digital-strategy.ec.europa.eu/en/library/data-act-proposal-regulation-harmonised-rules-fair-access-and-use-data*

→ **Interoperable Europe**: *https://joinup.ec.europa.eu/collection/interoperable-europe/interoperable-europe*

→ **Minimal Interoperability Mechanisms (MIMs) for smart cities**: *https://oascities.org/minimal-interoperability-mechanisms*

→ **European Commission's open source software strategy 2020-2023:** *https://ec.europa.eu/info/departments/informatics/open-source-software-strategy_en*

HOW TO ENSURE THAT TECHNICAL REQUIREMENTS AND STANDARDS ARE BEING FOLLOWED IN A DATA SPACE?

TECHNICAL

What is the problem?

If the technical requirements of the standards selected for data encoding and sharing are not closely followed, the opportunities to seamlessly combine different datasets, and to ensure an efficient data sharing within a data space, are sub-optimal. In addition, compliance with the provisions of specific legal acts is often based on certain preselected standards, and stakeholders need to verify that their technical implementations satisfy those legal requirements.

Scenario

Several European public sector data providers are mandated by an environmental EU Directive, and also by corresponding national legislation, to share data on the location of agricultural parcels. Technical guidelines and specifications based on the Directive, which in turn are built on top of international standards, make it possible for the data to be discovered, accessed and used across borders and in different application domains.

Following these guidelines and specifications, the data providers develop custom infrastructures to publish the relevant datasets. This is not enough to ensure that the national infrastructures are interoperable with each other, since they are all set up using different software tools and within different environments, and they serve different data (in terms of scope, amount, etc.)

Proposed solution(s)

Combining heterogeneous data sources is possible through the use of data specifications and approaches that are based on established international standards. The more closely the requirements of the standards are followed, the easier it becomes to combine and use together various data sources in an interoperable manner.

In addition, ensuring compliance with the technical specifications of data sharing standards as well as legal requirements improves the transparency and trustworthiness between the actors in the data economy. From experience, self-declaration of conformity to requirements and standards is not the optimal approach; automated techniques based on objective and quantifiable information are to be favoured. Validation of resources (data, metadata and services including APIs) is best implemented through the use of specialised validation tools (a valuable open-source example is ETF software) and following an iterative process where different errors are fixed and the tests are rerun until a satisfactory result is obtained.

Recommendations

→ **Data providers**

- Incorporate a data validation step in data production and data sharing workflows
- Develop own or reuse/extend existing Abstract and Executable Test Suites to ensure compliance with legal requirements and standards
- Ensure the availability of own data validation services/tools
- Liaise with standardisation bodies and software communities on the topic of data, metadata and service validation

→ **Data users**

- Consider the provision of feedback on used datasets with the aim to improve them

→ **Data intermediaries**

- Offer validation instances and guidelines to other stakeholders
- Develop certification schemes and labelling/quality stamps

Selected resources

→ Minghini, M., Cetl, V., Ziemba, L.W., Tomas, R., Francioli, D., Artasensi, D., Epure, E. and Vinci, F., **Establishing a new baseline for monitoring the status of EU Spatial Data Infrastructure**, EUR 30513 EN, Publications Office of the European Union, Luxembourg, 2020, doi:10.2760/296219, JRC122351.

→ Kotsev, A., Minghini, M., Cetl, V., Penninga, F., Robbrecht, J. and Lutz, M., **INSPIRE - A Public Sector Contribution to the European Green Deal Data Space**, EUR 30832 EN, Publications Office of the European Union, Luxembourg, 2021, doi:10.2760/8563, JRC126319.

→ **ETF validator framework**: *https://etf-validator.net*

HOW TO FACILITATE THE DISCOVERY OF DATA N A DATA SPACE?

TECHNICAL

What is the problem?

Often data, even when existent and made available through the web, are very difficult to find. Different natural languages and the various means that are used for discovering data – such as search engines, dedicated data catalogues and web Application Programming Interfaces (APIs) – each have their own specificities and can add complexities that make it even more complicated for users to locate and utilise data.

Scenario

A user needs to identify mobility data for city planning applications in several European cities, spread across multiple countries. She or he searches for existing datasets through multiple channels such as search engines and open data portals.

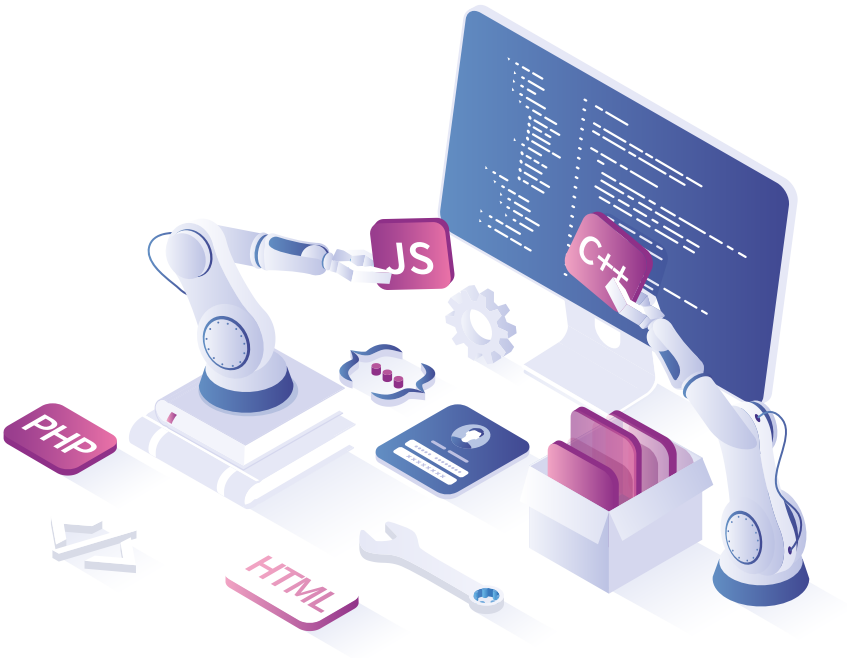
Those datasets, where documented properly through metadata, exposed through APIs, and indexed by search engines are easily discovered. On the other hand, other datasets, even though existing, cannot be found or do not include enough information allowing their reuse. This in turn limits the scope and potential benefits of the envisioned applications.

Proposed solution(s)

The provision of standardised metadata, in both machine and human-readable formats, as well as the sharing of data and metadata through Application Programming Interfaces (API), substantially improves the findability of those data.

In addition, providing data through multiple encodings, including web native formats such as HTML, would enable search engines to index the content.

Finally, annotating data using commonly agreed vocabularies such as those developed within the context of the SEMIC Action of the ISA2 programme, the Digital Europe programme, and collaborative industry-driven initiatives such as schema.org improves the findability of data on the web.



Recommendations

→ **Data providers**

- Compile metadata which are both machine and human readable, and based on international standards
- Ensure that metadata are up to date, self-explanatory and clear
- Use persistent identifiers for the datasets
- Use mainstream technology and standards and avoid custom metadata profiles and own standards where not explicitly needed
- Consider exposing the data through HTML representations
- Annotate existing datasets through the use of commonly agreed vocabularies

→ **Data users**

- Consider the provision of feedback on datasets and their metadata with the aim to improve them

→ **Public authorities**

- Devise strategies for aligning data-sharing practices to FAIR principles

Selected resources

→ **W3C Data on the Web Best practices**: *https://www.w3.org/TR/dwbp*

→ **Semantic Interoperability Community (SEMIC) platform**: *https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic*

→ **Data Catalogue Vocabulary (DCAT)**: *https://www.w3.org/TR/vocab-dcat-2*

→ **DCAT Application Profile for European Data Portals (DCAT-AP)**: *https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe*

→ **ISA² Programme**: *https://ec.europa.eu/isa2/home_en*

→ **Digital Europe Programme**: *https://digital-strategy.ec.europa.eu/en/activities/digital-programme*

→ **Schema.org**: *https://schema.org*

HOW TO SELECT THE MOST APPROPRIATE STANDARDS FOR A DATA SPACE?

TECHNICAL

What is the problem?

Although standards are widely recognised as key building blocks for interoperability and efficient data sharing in common European data spaces, the choice of which reference standard to use for concrete data and service provision, is often performed without the necessary considerations.

Incorrect standard selection may seriously hamper data-sharing processes and ultimately undermine the success of the data space as a whole. Non-optimal standards may be chosen for a set of different reasons such as the number of available standards, which, even across a single domain, could be very high. Standards with similar scope and provisions are often created by different organisations; in other cases, they overlap.

Also, some standards may exist that have not been fully adopted by a significant user base and/or there is a lack of mature software support for them. Finally, when standards remain at a draft stage for a long time yet are already used in practice, changes in the standard that is finally released can eventually break existing implementations.

Scenario

For a transport-related project, a business company needs to perform machine-learning analyses to evaluate the accessibility to the green areas located across districts within a city. When exploring and collecting available data, the company notices a high degree of fragmentation in terms of standards currently used for data encoding and sharing. The datasets found are encoded in different formats, including non-standard formats, as well as multiple standardised formats historically used by different communities, which highly overlap in scope but follow different data models; all of this makes "extract, transform, load" (ETL) conversion from one standard to the others impossible in some cases. In addition, one of the standards used for API-based data sharing is new and software tools, which are able to retrieve and consume the data directly, do not yet exist.

Proposed solution(s)

The selection of standards to be used for encoding and exchanging data within data spaces is crucial and requires significant attention. Selection should prioritise well-known standards adopted by global communities, that are developed and released by established, trusted international standards development organisations (SDOs) through a governance process involving several, diverse stakeholders, possibly including the public sector, the private sector and academia. A notable example from the geospatial domain is the Open Geospatial Consortium (OGC), an international standardisation body in existence since 1994 and currently numbering 500+ members from around the world. Cross-domain standardisation bodies include, on the other hand, the International Organization for Standardization (ISO) and the World Wide Web Consortium (W3C), the latter focusing on web-based data exchange.

The way in which standards are created should also be analysed in order to prioritise those developed in a participative, agile and collaborative way through the involvement of domain-specific experts and early adopters, (e.g., workshops/hackathons), and use of collaborative online platforms (e.g., based on Git) to co-create content. This ensures that, once adopted, the standards are implementable.

Finally, the maturity of standards should be considered, avoiding ones that are still in the draft stage or not yet published, giving preference to standards featuring a large user base with several implementations already available and a strong support for existing software tools. A notable exception to all these criteria is represented by so-called "de facto" standards, i.e., standards that have never been officially endorsed but whose usage and software support has become mainstream over several years. These should be given prominent consideration as a working alternative to formal standards endorsed and governed by SDOs.

Recommendations

→ **Data providers**

- Monitor the process of standards development and encourage appropriate, active participation.
- Consider the inclusion of standards in data-sharing practices "by design".
- Join technical committees of standardisation bodies.
- Modernise existing implementations that rely on standards which are outdated, inefficient, or unsupported by software implementations, in favour of modern, more efficient standards featuring high community and software support.
- Make use of validation services to test the actual compliance with reference standards.

→ **Data users**

- Contribute to the development/ evolution of standards.
- Test the implementation of standards in their own use cases.
- Provide feedback to standardisation bodies on the advantages and disadvantages in the adoption of standards, in particular informing about obstacles and bottlenecks.

→ **Data intermediaries**

- Consider participating in the funding/development of standards.

Selected resources

→ Kotsev, A., Minghini, M., Cetl, V., Penninga, F., Robbrecht, J. and Lutz, M., **INSPIRE - A Public Sector Contribution to the European Green Deal Data Space**, EUR 30832 EN, Publications Office of the European Union, Luxembourg, 2021, doi:10.2760/8563, JRC126319.

→ **International Organization for Standardization (ISO)**: *https://www.iso.org*

→ **World Wide Web Consortium (W3C)**: *https://www.w3.org*

→ **Open Geospatial Consortium (OGC)**: *https://www.ogc.org*

→ Guasch, C., Lodi, G. and Dooren, S. V. **Semantic Knowledge Graphs for Distributed Data Spaces: The Public Procurement Pilot Experience**. In: *Proceedings of the 21st International Semantic Web Conference*, 23-27 October 2022, 753-769. Cham, Springer, 2022, doi:10.1007/978-3-031-19433-7_43.

HOW TO ENSURE THAT DIGITAL RESOURCES AND DATA ARE UNIQUELY REFERENCED IN A DATA SPACE?

TECHNICAL

What is the problem?

A common challenge associated with the encoding and sharing of data, especially when different providers are involved (including those from different countries), is related with the appropriate use of terms, codelists, identifiers and other digital assets. This issue is further compounded if different natural languages are used, and the data are intended to be shared over the web.

Scenario

Two different data providers in neighbouring countries encode their data on buildings in a different data model and in their own language, without using commonly agreed vocabularies. When the need to combine the data arises, e.g., for flood risk assessment in a cross-border river basin, it becomes difficult to combine the data.

Proposed solution(s)

The use of commonly agreed data models, specifications and vocabularies will improve the interoperability of data. If those vocabularies are exposed on the web in a persistent manner through codelists, different providers can encode the data by providing the unique resource identifiers of codelist items independent of natural languages.

From a technical perspective, codelist items and their descriptions can be made available through different encodings such as RDF, JSON (incl. JSON-LD), CSV, XML. A notable solution is the "Re3gistry" open source software, which can be used to expose items that require clear descriptions that can be referenced through unique identifiers.

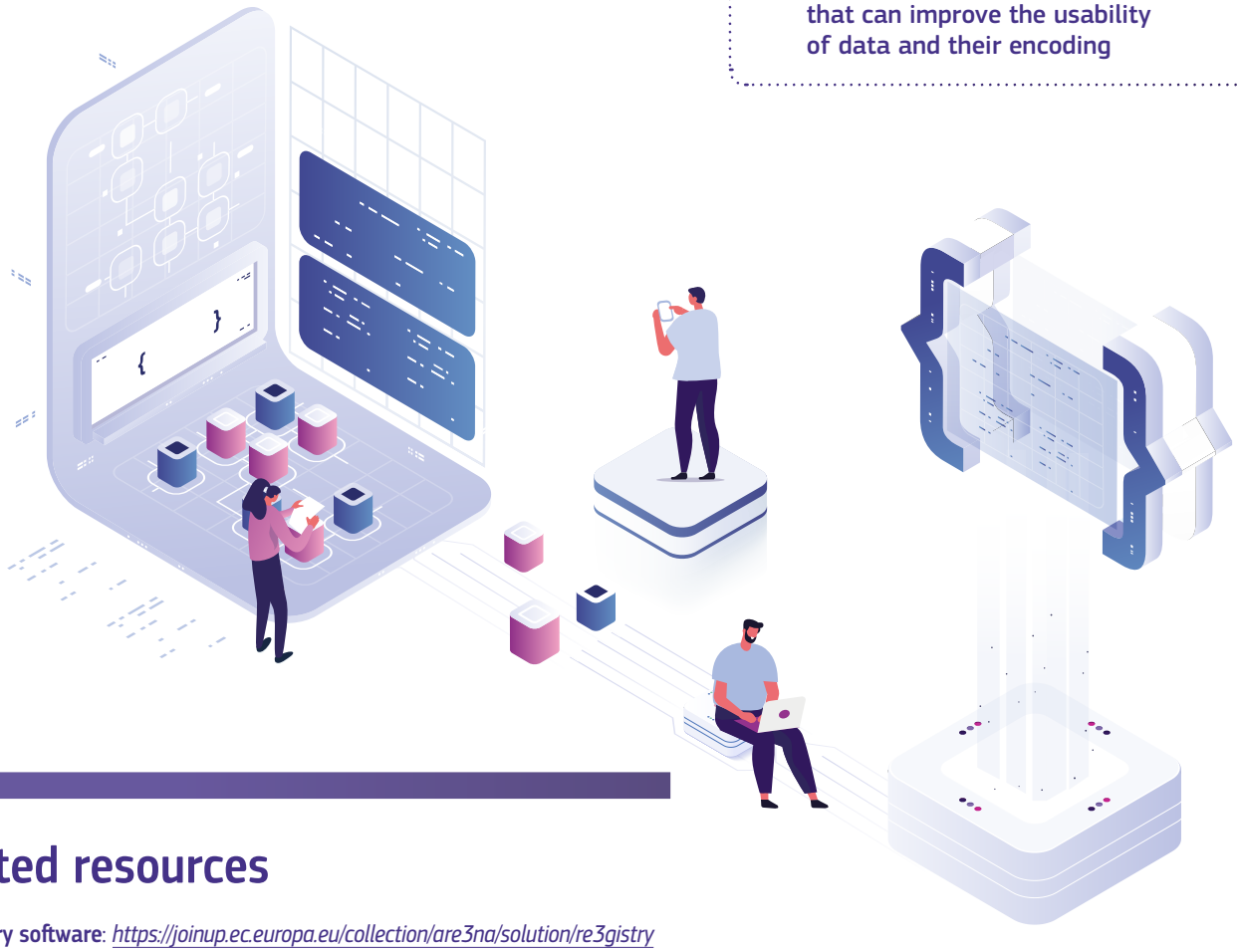
Recommendations

→ **Data providers**

- Use registers and codelists when encoding data
- Expose codelist content through APIs
- Provide unique identifiers of codelist items
- Offer different encodings of registers
- Participate in the co-creation of common data specifications
- Adopt standard classifications and vocabularies whenever possible

→ **Data users**

- Provide feedback to providers that can improve the usability of data and their encoding



Selected resources

→ **Re3gistry software**: *https://joinup.ec.europa.eu/collection/are3na/solution/re3gistry*

→ **Instance of the Re3gistry software used for the needs of the INSPIRE Directive**: *https://inspire.ec.europa.eu/registry*

→ Kotsev, A., Minghini, M., Cetl, V., Penninga, F., Robbrecht, J. and Lutz, M., **INSPIRE - A Public Sector Contribution to the European Green Deal Data Space**, EUR 30832 EN, Publications Office of the European Union, Luxembourg, 2021, doi:10.2760/8563, JRC126319.

→ Morales, L. G. and Orrell, T. **Data interoperability: A practitioner's guide to joining up data in the development sector**. United Nations World Data Forum (UNWDF): Dubai, United Arab Emirates, 2018.

HOW TO PROVIDE ACCESS TO DATA IN A DATA SPACE?

TECHNICAL

What is the problem?

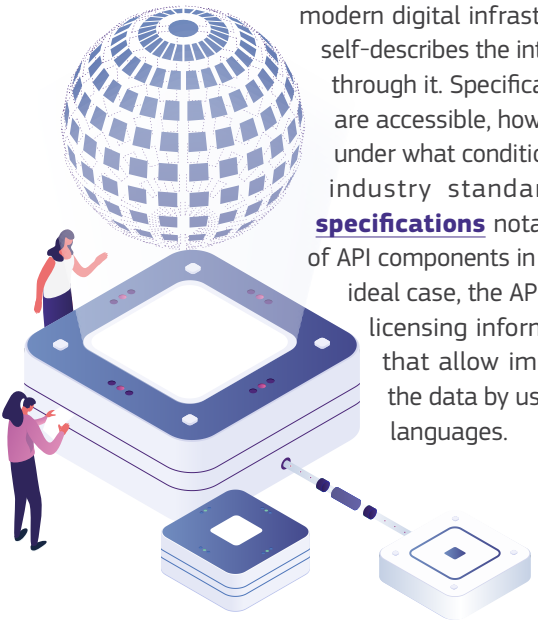
Data spaces aim to realise as much data value as possible by exploiting the complementarities of the different datasets located within them. Effective access to these datasets, along with the smooth integration of actors and systems in digital environments is often a challenge due to heterogeneity of technologies, standards and architectures.

Scenario

In order to build an application to monitor the exposure of citizens to air pollutants, an application developer could use data from air pollution sensors, traffic and road infrastructures – all of which are made available by different intermediate data providers. The envisioned application can only be implemented if simultaneous access is ensured to all these data sources.

Proposed solution(s)

Web-based *Application Programming Interfaces (APIs)* act as the "glue" to connect different actors and systems in digital environments. The use of APIs is the norm for connecting to modern digital infrastructures. An API interface self-describes the interactions that are possible through it. Specifically, it describes what data are accessible, how to access these data, and under what conditions. The use of widespread industry standards such as **OpenAPI specifications** notably eases the integration of API components in digital value chains. In the ideal case, the API specifications include the licensing information and code snippets that allow immediate interaction with the data by using different programming languages.



Recommendations

→ **Digital service providers across the full digital chain (e.g., app providers, digital service providers in the public sector)**

- ensure the robustness of the full digital chain, carefully considering API entry points
- ensure sustainability of digital interactions (technical and organisational coordination)
- ensure app compliance with the legal framework

→ **Intermediate service providers (e.g., high-value dataset API service providers in the public sector)**

- invest in facilitating discoverability and uptake of the API service by properly documenting and specifying correct use
- ensure compliance with technical and legal constraints from upstream components in the digital chain

→ **Intermediate service users (e.g., an organisation using external API components in their digital solutions)**

- ensure upstream compliance with technical and legal constraints
- ensure that the uses of the assets accessed are lawful

→ **Data providers**

- exhaustively describe licenses and conditions for use of data assets

→ **Data space orchestrators**

- stakeholder management and coordination (regarding conditions for APIs)
- ensure stable legal and technical digital interactions (encoded in the code and agreements for APIs)

Selected resources

→ Posada Sanchez, M., Pogorzelska, K. and Vespe, M., **The role of Application Programming Interfaces (APIs) in data governance and digital coordination**, European Commission, 2022, JRC128250.

→ Vaccari, L., Posada Sanchez, M., Boyd, M., Gattwinkel, D., Mavridis, D., Smith, R., Santoro, M., Nativi, S., Medjaoui, M., Reusa, I., Switzer, S. and Friis-Christensen, A., **Application Programming Interfaces in Governments: Why, what and how**, EUR 30227 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-18982-4, doi:10.2760/58129, JRC120429.

→ Santoro, M., Vaccari, L., Mavridis, D., Smith, R., Posada Sanchez, M. and Gattwinkel, D., **Web Application Programming Interfaces (APIs): general-purpose standards, terms and European Commission initiatives**, EUR 29984 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-13183-0, doi:10.2760/675, JRC118082.

→ Boyd, M., Vaccari, L., Posada Sanchez, M. and Gattwinkel, D., **An Application Programming Interface (API) framework for digital government**, EUR 30226 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-18980-0, doi:10.2760/772503, JRC120715.

→ Vaccari, L., **Tool: API for government framework - API strategy self assessment**, 2020, *https://joinup.ec.europa.eu/collection/api4dt/document/tool-api-government-framework-api-strategy-self-assessment*

→ Posada Sanchez, M., Vaccari, L. and Pogorzelska, K., **Unfolding opportunities from the use of APIs in Europe**, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-34203-8, doi:10.2760/074141, JRC124758.

HOW TO PRESERVE PRIVACY AND PROTECT PERSONAL DATA AND SENSITIVE BUSINESS DATA IN A DATA SPACE?

TECHNICAL

What is the problem?

Many datasets are not shared due to a multitude of reasons such as statistical confidentiality, privacy provisions defined by national and European legislation, and ethical constraints. Such data can be linked to individual entities, including citizens (e.g., personal health records) and businesses (e.g. business registers), or be a valuable digital asset deemed sensitive within private companies. Being able to utilise such datasets, while preserving their privacy, would bring multiple societal and economic benefits.

Scenario

Several hospitals hold data on a rare disease and are interested in developing a machine learning (ML) algorithm that can predict the occurrence of the disease among patients in the future. None of the hospitals has enough data records to obtain a robust result from the ML activity, nor do they have the necessary technical capacity. The existing data, if combined, would be sufficient for the desired prediction model; however, the patient data cannot be disclosed due to privacy considerations.

Proposed solution(s)

Several Privacy Enhancing Technologies (PETs) exist that would allow the reuse of sensitive business or citizen data (i) in a privacy preserving manner, and (ii) in line with the provisions of the General Data Protection Regulation (GDPR). These are, however, not a universal solution to privacy preservation, as there is a compromise to be made that is associated with their use (e.g., distorted output, additional computational resources to be involved, or the need to homogenise the input data).

In addition, the target of the privacy preservation is different according to the particular case. Some of these techniques (i) are capable of protecting the input, while (ii) others focus on preserving the privacy of the output, and (iii) a third group protect the algorithms involved (such as those based on AI). Also, depending on the architecture and the technique used, data can be protected at rest, in transit or during use.

The following most prominent PETs can be used:

- **Secure Multi-Party Computation (SMPC):** *techniques where a set of mutually distrusting data holders want to jointly compute a function without revealing anything apart from the output.*
- **Differential Privacy (DP):** *a technique in which noise is added to the input to avoid the identification of any individual record when this data is queried through any aggregation function, like mean, sum or variance.*
- **Homomorphic Encryption (HE):** *a set of encryption techniques aimed at performing operations on encrypted data instead of on raw data.*
- **Federated Learning (FL):** *a form of distributed computing where a model is trained through several nodes, each containing* only own data, without those data leaving where they are stored.
- **Zero Knowledge Proofs:** a cryptographic method that allows one party (the provider) to prove the validity of a statement to another party (the verifier) without revealing the secrets that make this statement true.
- **Trusted Execution Environments (TEEs):** a secure area inside a main processor where the code and data executed are safeguarded against unauthorised entities.
- **Pseudonymisation:** models aimed at protecting data using procedures that alter, delete or encode the identifiers that either directly reveal personal information or allow establishing a relation between the entities.

Each of the existing solutions comes at a certain "cost". While some of the PETs require that additional computational resources or hardware be used, others distort the output of data-driven analysis. As many of the available techniques are still being researched, it is important to consider the Technology Readiness Level (TRL) of each solution when selecting a particular solution.

Recommendations

→ **Data providers**

- Consider opening up sensitive data and plan for the use of PETs by design.
- Devise a strategy for choosing the most adequate privacy preserving technique considering their advantages, disadvantages and the desired outcome.
- Build capacity and prepare digital infrastructures that can be used for the development of PETs.
- Contribute to the development of technical building blocks (e.g., software libraries) for privacy preservation.
- Provide documentation that describes the approaches for using the data.

→ **Data intermediaries**

- Offer technical building blocks (e.g., infrastructures, harmonisation services, software tools) that would facilitate the utilisation of data in a privacy-preserving manner.
- Develop and curate good practices that showcase PETs and can be used as examples that build confidence and trust of data providers.

Selected resources

→ European Union, **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, 2016, pp. 1–88: https://eur-lex.europa.eu/eli/reg/2016/679/oj

→ Hansen, M., Hoepman, J. H. and Jensen, M. **Readiness analysis for the adoption and evolution of privacy enhancing technologies: Methodology, pilot assessment, and continuity plan**. Technical report, Version 1.0, European Union Agency For Network And Information Security, 2015.

WHICH ACTORS ARE PROVIDING WHAT TYPES OF DATA WITHIN SCOPE OF A DATA SPACE?

ORGANISATIONAL

What is the problem?

An enormous amount of data is produced on a daily basis by an increasing variety of stakeholders, ranging from public sector data providers and the academic community to the private sector (increasingly via sensors) as well as citizens. The data generated by all these actors are diverse in scope, granularity and quality, and if combined together, hold the potential to extract meaningful information that can benefit society.

To understand the full range and scope of data available, suitable time is required to search, explore and identify all datasets that may be relevant for a data space, as well as the actors that either produce or distribute those data. However, often only a subset of the available data is considered, or available, when building data sharing infrastructures. This occurs for a number of reasons, such as there is not enough knowledge or exploration of available data sources, or the data produced by certain actors are difficult to locate (e.g., data generated by privately held sensors). The result is that large quantities of potentially useful data remain unknown and therefore, unused, which may lead to missed opportunities for common European data spaces.

Scenario

An agency in an EU Member State is tasked to perform a climate change study evaluating the effects of land use/land cover transformations within the country over the past 10 years. Due to limited resources, the data exploration and collection phase is performed within a short time with the only available datasets being government land use/land cover digital maps available at the national level. These maps are only produced, or updated, every six years and are available at small scales with low spatial resolution.

The agency was either not aware of, or was aware of, yet did not use, additional data existing at the national level, such as, (i) a national university has already researched land use/cover changes in the Member State with a yearly frequency and produced openly-licensed change maps for each two-year period in the last 10 years; (ii) a private company has recently performed a survey of land use/land cover on a 10 km grid covering the whole country; (iii) an EU agency has released classified land use/land cover maps derived from satellite imagery; (iv) within a national citizen science project, volunteers have recently collected thousands of land use/land cover samples (geo-located photographs and descriptions) using a dedicated app. If these datasets were also used, they could have provided an essential contribution to complement, enrich and validate the government maps as well as to train and apply machine learning models assessing climate changes.

Proposed solution(s)

When conceptualising a common European data space and before starting to design its technological infrastructure, it is of primary importance to dedicate enough time to search, explore and identify all datasets that may be relevant for that data space, and the actors that either produce or distribute those data. As outlined in the European strategy for data, all relevant actors should be considered: e.g., the public sector, the private sector, the academic/research sector, citizens, data altruism organisations, and data intermediaries, all of which, depending on the specific data space, may have varying prominent roles. The identification of relevant data should then depend on the specific use cases defined for the data space.

Regarding possible challenges to access the identified data, it should be noted that: (i) Directive (EU) 2019/1024 (the Open Data Directive) and the related Implementing Act on a list of High-Value Datasets require free and open-access availability of several datasets from the public sector; (ii) scientists and academics are increasingly adhering to the Findable, Accessible, Interoperable and Reusable (FAIR) data sharing principles; and (iii) the Data Governance Act and the proposed Data Act aim to facilitate the access and use of privately held data as well as voluntary data produced by citizens.

Recommendations

→ Data providers

- Make the conditions for access and use of the datasets they produce as clear as possible
- Make the datasets produced as easy as possible to discover, access (e.g., through Application Programming Interfaces or APIs) and use (by relying on reference standards for data encoding and sharing to ensure interoperability)

→ Data users

- Consider using data from all possible sources and actors

→ Data intermediaries

- Facilitate and promote discovery of and access to data shared by all relevant actors, including through community engagement
- Identify existing software tools or develop new ones to facilitate and when possible automate access to and reuse of specific data sources

Selected resources

→ European Commission, **Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European Strategy for Data**, COM(2020) 66 final, 2020, *https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN*

→ European Union, **Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information**, *Official Journal of the European Union*, L 172, 2019, pp. 56–83, *https://eur-lex.europa.eu/eli/dir/2019/1024/oj*.

→ European Union, **Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)**, *Official Journal of the European Union*, L 152, 2022, pp. 1–44, *http://data.europa.eu/eli/reg/2022/868/oj*

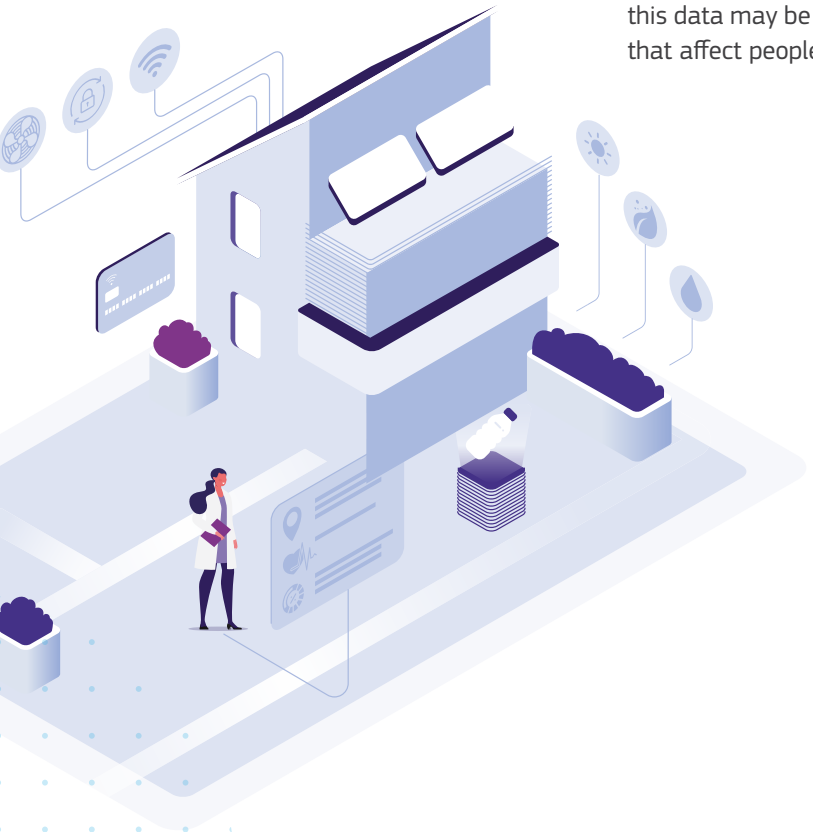
→ European Commission, **Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use**, *Official Journal of the European Union*, L 19, 2023b, pp. 43–75.

→ European Commission, **Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)**, COM(2022) 68 final, *https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022PC0068&from=EN*

→ Kotsev, A., Minghini, M., Cetl, V., Penninga, F., Robbrecht, J. and Lutz, M., **INSPIRE - A Public Sector Contribution to the European Green Deal Data Space**, EUR 30832 EN, Publications Office of the European Union, Luxembourg, 2021, doi:10.2760/8563, JRC126319.

HOW TO FOSTER A PEOPLE-CENTRED APPROACH TO DATA IN A DATA SPACE?

ORGANISATIONAL



What is the problem?

Ensuring a people-centred approach to data governance (with "data governance" defined broadly as the policies, relations, socio-technical arrangements and decision-making structures that define how to collect, share and use data) is an important element for the creation of a trustworthy environment in which data can be shared and exchanged within data spaces. Among other aspects, this involves including the perspectives and protecting the rights and interests of those who create the data, or who may be affected by decisions based on the data, as well of those who may not be adequately represented by the data due to marginalisation.

Large asymmetries often exist between those who collect data (who usually hold most of the decision-making power over how to share and use them), and other stakeholders, including the people who created the data in the first place. Yet this data may be used to build services and to develop products and AI systems that affect people's lives in significant ways.

Scenario

In the context of a digital innovation project, a neighbourhood digitises several public services, including public transport, analysis of waste collection and to monitoring of the energy consumption in social housing. This leads to greater data collection regarding the movement and activities of the residents, without adequately informing them. The city council establishes a partnership with a Big Tech corporation that offers its services to manage data on behalf of the administration. Notwithstanding the presence of an active neighbourhood council and high levels of digital literacy amongst the community, neighbourhood residents are neither involved nor consulted. As a result, they feel powerless and lose trust in the city council.

Proposed solution(s)

In the above example, a people-centred approach to data collection, sharing and use might be developed at different levels: (i) ensuring that people's voices and interests are well-represented within data governance; (ii) promoting digital literacy and skills amongst the general public and creating awareness of digital rights; (iii) enabling people to control how their data are accessed, shared and used; and (iv) providing people with opportunities to engage in the collection of data for public good and to influence decision-making involving data within their local area. Creating a trustworthy environment and a people-centred approach to data within a data space can help enhance data sharing and exchange, especially if the specific benefits for each actor are clear and tangible.

Selected resources

→ Bass, T., Sutherland, E. and Symons, T., **Reclaiming the smart city: Personal data, trust and the new commons**, 2018, *https://media.nesta.org.uk/documents/DECODE-2018_report-smart-cities.pdf*

→ Craglia, M., Scholten, H.J., Micheli, M., Hradec, J., Calzada, I., Luitjens, S., Ponti, M. and Boter, J., **Digitranscope: The governance of digitally-transformed society**, EUR 30590 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-30229-2, doi:10.2760/503546, JRC123362.

→ Micheli, M., Ponti, M., Craglia, M. and Berti Suman, A., **Emerging models of data governance in the age of datafication.** *Big Data & Society*, Vol. 7, No. 2, 2020, doi:10.1177/2053951720948087.

→ Ponti, M. and Craglia, M., **Citizen-generated data for public policy**, Publications Office of the European Union, Luxembourg, 2020, JRC120231.

→ Micheli, M., Delipetrev, B., Hupont, I. and Soler Garrido, J., **The landscape of data and AI documentation approaches in the context of new EU policies**, *Ethics and Information Technology*, in press.

→ **DECODE project:** *https://decodeproject.eu*

→ **Citizen Voices for Digital Rights (CVDR):** *https://www.demsoc.org/blog/launch-of-the-citizen-voices-for-digital-rights-cvdr-report*

→ **Digital City Alliance Berlin:** *https://digitalesberlin.info*

→ **Eurocities:** *https://eurocities.eu*

→ **Cities Coalition for Digital Rights:** *https://citiesfordigitalrights.org*

Recommendations

→ Public authorities and the private sector

- **Define clear roles and responsibilities for data governance**, emphasising a people-centred approach.
- **Adopt transparency and accountability measures:** Dataset documentation can help to build trust and communicate bias or limitations in datasets. For instance, due to digital divides and inequalities, many datasets collected as a by-product of using online services may be skewed towards representing younger, better resourced and more digitally literate populations. These, and any other biases, should be publicly acknowledged and communicated.
- **Establish/use the services of trustworthy data intermediaries:** In many cases it is not possible for members of the public to participate directly in data governance. Instead, they may have to rely on a data intermediary. Data intermediaries can allow data subjects to gain greater control over their personal data and to decide how it is managed , as well as for which purposes it may be used. There are different types of personal data intermediaries, which pursue a variety of goals, but all are intended to increase control by an individual or a group over their data and to unlock the value of that data to produce benefits for them, their community or society as a whole: e.g., personal information management systems (PIMS), data cooperatives, data trusts.
- **Implement privacy by design solutions and enhance individuals' control of data:** Local administrations can put in place initiatives for safeguarding citizens' rights to control and use their own personal data. For example, the city of Barcelona through project DECODE developed a platform to provide citizens with the ability to control the terms of sharing their data with public offices by selecting which information to share and the desired level of granularity.
- **Invite public engagement:** Engaging people from different socio-demographic, economic and cultural backgrounds in the governance of data, and as generators of data, can help ensure that the knowledge and value generated from data better represents their interests and avoids fostering inequalities. People can be involved (either by public authorities or civil society organisations) in data collection efforts to inform policy and address problems affecting their quality of life, e.g., by collecting data on air and water quality.

→ Civil Society

- **Establish networks and engage in civic advocacy efforts:** Within the European Union, there are numerous examples of local initiatives to promote and protect people's digital rights, as well as of civic groups aimed at engaging, equipping and empowering people with the skills, knowledge and tools to self-advocate for their digital rights (e.g., Citizen Voices for Digital Rights (CVDR), Digital City Alliance Berlin). There are also associations of cities promoting people-centred approaches to digital transformation, for example, Eurocities and the Cities Coalition for Digital Rights. The latter, which has the support of the United Nations Human Settlements Program (UN-Habitat), coordinates actions aimed at fostering human rights principles in the digital domain (such as privacy, freedom of expression, and democracy).

HOW CAN BUSINESS BENEFIT FROM SHARING DATA IN A DATA SPACE?

ORGANISATIONAL

What is the problem?

Organisations that foster trustworthy data sharing are set to outperform their peers on key business value metrics. Unfortunately, many are still unaware of the multiple benefits of data sharing. Common European data spaces offer a trustworthy and privacy-preserving infrastructure that allows businesses to securely share data. Aside from making profits by selling data, businesses can leverage data spaces to derive indirect financial and non-monetary compensation. For example, by combining complementary datasets in a data space, two parties can increase the size, quality and scope of the data.

Scenario

An electric car charging station collects data which is valuable for electric car manufacturers. The business owner of the charging station decides to advertise the data in the EU Mobility Data Space. Thanks to the efficacy of data spaces to coordinate supply and demand of data, an electric car manufacturer discovers the data and offers to purchase the whole dataset and enter into a business partnership.

This example shows how engaging in data sharing through common European data spaces can significantly spur data-driven innovation, facilitate coordination and generate profitable spillovers.

Proposed solution(s)

Sharing data with other businesses and accessing shared data in a data space can increase efficiency and reduce costs in multiple ways. Aside from remuneration, businesses can derive analytical insights, optimise supply chains, compare indicators, etc. Moreover, data spaces offer many marketing opportunities, allowing businesses to expand their market reach and improve their brand's reputation.

Recommendations

→ **Data providers**

- Consider cataloguing your data across the common European data spaces
- Share and exchange data with third parties
- Foster a "data sharing" culture and not a "data ownership" culture

→ **Data users**

- Seek opportunities to enrich data
- Leverage network opportunities and engage with data providers for partnerships
- Credit the contributions of data providers

→ **Public authorities**

- Facilitate the exchange of data with businesses through a European data space
- Fund the setting up of robust infrastructure and proper governance mechanisms
- Optimise the search engine to better coordinate the supply and demand of data

Selected resources

→ Carballa Smichowski, B. **The value of data: An analysis of closed-urban-data-based and open-data-based business models**, *CEPN Working Papers*, hal-01736484, HAL, 2018.

→ Gartner, **Data Sharing Is a Business Necessity to Accelerate Digital Business**, 2021, *https://www.gartner.com/smarterwithgartner/data-sharing-is-a-business-necessity-to-accelerate-digital-business*

HOW CAN GOVERNMENTS ACCESS PRIVATE SECTOR DATA OF PUBLIC INTEREST?

ORGANISATIONAL

What is the problem?

Businesses collect large amounts of data, some of which may be of public interest and beneficial for evidence-based public policymaking. For example, private sector data could be used to analyse and address health, societal and environmental challenges – this data relationship is also called Business-to-Government (B2G) data sharing. For instance, public authorities have sought to access location data collected by mobile phone operators in order to tackle the spread of the COVID-19 virus and monitor compliance with restrictions. Such data is also useful for better urban traffic management.

However, substantial asymmetries exist between governments, Big Tech corporations and other private sector entities concerning the ability to collect and access data. In this respect, public entities, especially local and regional governments, are often in a weaker position and may lack, or experience difficulties in finding sustainable instruments to access privately held data in a transparent and systematic manner. Another challenge is that businesses are often reluctant to share data with public sector organisations, mainly due to potential privacy risks and commercial sensitivity.

Scenario

In a local administration of a medium-sized city, several ridesharing providers are operating their services. While the city council manages some of these services, others are privately owned. This fragmented setup presents a challenge for the administration, as little is known about the overall uptake across the mobility services, and fleets of scooters and bikes are resources that are difficult to optimise and manage.

Access to comprehensive and integrated data from all operators would allow better urban planning and implementation of Mobility-as-a-Service (MaaS) by the city administration. Yet, private operators are often reluctant to share data collected as a by-product of delivering their services. The administration attempts to establish ad-hoc public-private data sharing agreements with each individual company, but these are suboptimal due to excessive costs and a low level of granularity in the data shared by the operators.

Proposed solution(s)

To improve access by public sector organisations to data of public interest collected by private companies, the following solutions could be implemented:

- Public sector bodies can adopt "**data sovereignty clauses**" in their procurement contracts requiring that data collected by a private company when providing public services (e.g., public transport, waste management and ridesharing) are made available and accessible to them, in a privacy-compliant way.
- Public sector bodies can establish **partnerships** with private sector companies willing to share at no cost and to analyse data for mutual interests (subject to networks and capacity).
- Governments can leverage **Corporate Social Responsibility initiatives around data sharing**, through which businesses may make their data available to organisations acting in the public interest or serving those in need. This sometimes occurs during public emergencies or for humanitarian purposes, and is subject to a company's willingness to share with a given organisation.*

** Future opportunities: The proposed EU Data Act would establish a requirement for companies to share data with governments free of charge in situations of emergency, and the possibility for companies to claim compensation for sharing data with public bodies in other exceptional situations.*

Selected resources

→ Galasso, G., Montino, C., Gori, M., Rasmussen, M., Roman, L., Mccolgan, O., Liva, G., Rebesco, E., Brynskov, M., Mulquin, M., Micheli, M., Schade, S., Smith, R. and Kotsev, A., **Sandboxing. How to use it to strengthen your local data ecosystem**, EUR 31285 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-58779-8, doi:10.2760/779684, JRC130555.

→ Micheli, M., **Public bodies' access to private sector data: The perspectives of twelve European local administrations**. *First Monday*, vol. 27, No. 2, 2022, doi:10.5210/fm.v27i2.11720.

→ Giovanni, L., Micheli, M., Schade, S., Kotsev, A., Gori, M. and Codagnone, C., **City data ecosystems between theory and practice: A qualitative exploratory study in seven European cities**, *Data & Policy*, Vol. 5, No. e17, 2023, pp. 1-16, doi:10.1017/dap.2023.13.

→ Alberti, V., Alonso Raposo, M., Attardo, C., Auteri, D., Ribeiro Barranco, R., Batista E Silva, F., Benczur, P., Bertoldi, P., Bono, F., Bussolari, I., Louro Caldeira, S., Carlsson, J., Christidis, P., Christodoulou, A., Ciuffo, B., Corrado, S., Fioretti, C., Galassi, M., Galbusera, L., Gawlik, B., Giusti, F., Gomez Prieto, J., Grosso, M., Martinho Guimaraes Pires Pereira, A., Jacobs, C., Kavalov, B., Kompil, M., Kucas, A., Kona, et al., **The Future of Cities**, Vandecasteele, I., Baranzelli, C., Siragusa, A. and Aurambout, J. editor(s), EUR 29752 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-03848-1, doi:10.2760/364135, JRC116711.

→ **Big Data Test Infrastructure:** *https://digital-strategy.ec.europa.eu/en/policies/bdti*

→ European Commission, **Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)**, COM(2022) 68 final, *https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022PC0068&from=E*

Recommendations

→ Public authorities

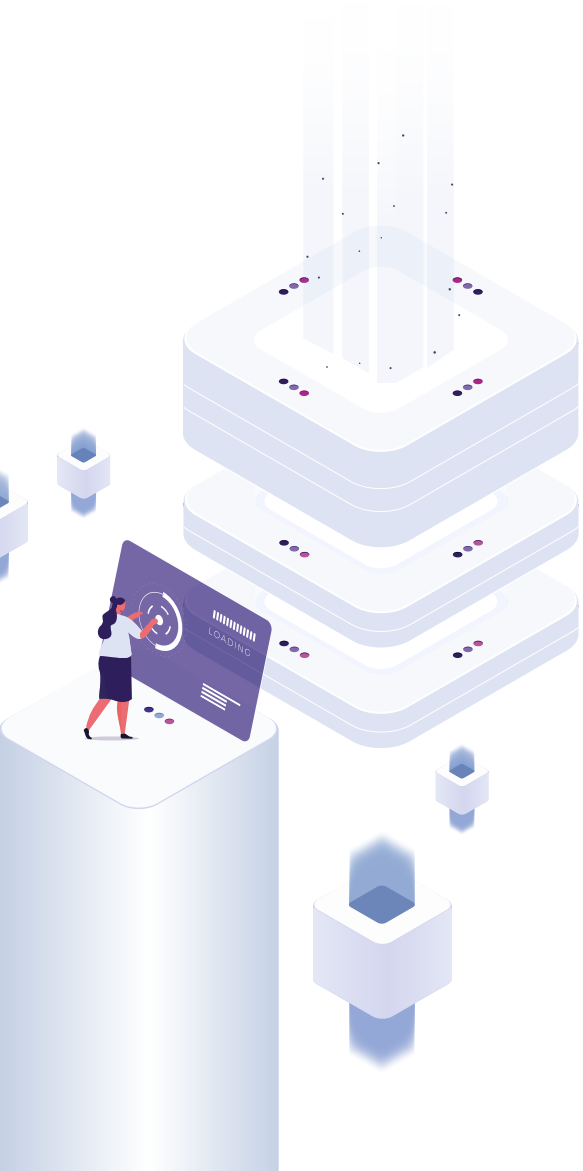
- **Join forces among public sector organisations to build collective bargaining power** and, at the same time, avoid duplication of requests to private companies. For example, cities can work within existing or new associations and networks of municipalities to develop joint and standard contractual frameworks for data sharing with private companies.
- **Establish a framework to map the demand side of privately held data,** helping to refine policy and research questions that can be addressed by public sector bodies through the access and use of such data.
- **Establish and recruit new professional roles and profiles, such as that of a data steward,** with the expertise to establish partnerships with the private sector for data access, sharing and management. Data scientists and analysts will also be needed within the public sector to process and interpret data; additionally, experts in ethics and data protection will be required to ensure fair and compliant data use, and IT experts to ensure security policies and safeguards for data access and use.
- **Engage in peer-learning** with organisations that have addressed similar challenges, and leverage existing prototypes of terms and conditions for data sharing or data partnerships.
- **Include a focus on organisational factors** – efforts and resources are needed to improve cooperation between government departments, as well as to increase internal data literacy and culture.

→ Public authorities and private sector

- **Engage in experimentation activities** to test the implementation of data sharing initiatives (e.g., sandboxing). These experiments should test solutions in a safe environment, addressing both technical and organisational data sharing challenges and promoting innovative approaches.

HOW CAN DATA TRANSPARENCY FOR AI SYSTEMS BE INCREASED IN A DATA SPACE?

ORGANISATIONAL



What is the problem?

Data in a data space can be used for the development of artificial intelligence (AI) systems. Choosing the right data and being aware of both the limitations of a dataset and the ethical implications of its use is important, as this can have a strong influence on the performance of the AI system and its outcomes. Data quality is a common concern in a wide range of domains, thus, before data are fed into downstream analytics and decision support tasks, it must be ensured that general quality criteria are met.

When using data for the development of AI systems, it is important to always use datasets that are appropriate for the intended task of the AI system. This is especially important in high-risk AI applications, where the unintentional misuse of data can amplify social bias and impact negatively on fundamental rights. Both policymakers and the general public increasingly expect transparency and accountability from organisations that share and manage data, especially if used for AI systems. This is reflected in the EU AI Act proposal, which lays down obligations and requirements concerning data quality, transparency and governance for high-risk AI systems.

Scenario

An organisation develops an AI-based fraud detection system that is used by a government welfare department. Due to bias against low-income individuals that is contained in the datasets used for training the algorithm, the AI system mainly targets this population. This can lead to stigmatisation and disproportionate targeting of poorer people, discrimination based on socio-economic status, and the violation of privacy rights, where one particular socio-demographic group is targeted more than others due to bias in the system. In addition, the results of using such biased data can cause reputational damage for the organisation that developed the AI system and, as a consequence, for the institution that uses it.

Proposed solution(s)

Documentation of datasets, AI models and algorithms according to best practices *(e.g., Datasheets for datasets, Dataset nutrition labels, Model cards* and *AI factsheets)* assists in the prevention of bias, discrimination and other possible risks to fundamental rights, as well as serving to increase trust in the organisations handling the data.

When AI developers have adequate knowledge about the provenance, scope, collection, preparation, correctness, representativeness and privacy of a dataset, they can choose the most appropriate one for the intended AI system/task. They also become aware of the limitations of the dataset and can put strategies in place to mitigate bias.

Such documentation of datasets and AI will be helpful not only for AI developers, but will also increase transparency and accountability among users of AI systems and the wider public.

Recommendations

→ **Data providers**

- Be responsible for documentation of datasets, AI models and algorithms according to best practices
- Include the following elements in the documentation of the dataset: provenance, scope, collection, preparation, correctness, representativeness and privacy of the data (see best pratices)
- Provide different documentation in different formats depending on the intended audience: technical (AI developers) or non-technical (end-users of AI systems, societal groups)

→ **AI developers**

- Use the documentation to understand the main features and limitations of datasets and choose the most appropriate one for a task

→ **AI developers, users of AI system, general public**

- Read the documentation to increase self-awareness about the potential ethical and social consequences of the use of the dataset/AI system

→ **Public authorities**

- Consider providing guidance on best practices for datasets documentation
- Transparently communicate (share documentation) about data sources used in AI systems

Selected resources

→ Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., III, H.D. and Crawford, K., **Datasheets for datasets**. *Communications of the ACM*, Vol. 64, No. 12, 2021, pp. 86–92, doi:10.1145/3458723.

→ Holland, S., Hosny, A., Newman, S., Joseph, J. and Chmielinski, K., **The dataset nutrition label**. *Data Protection and Privacy*, Vol. 12, No. 12, 2020, pp. 1–25.

→ Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, A., Raji, I.D. and Gebru, T., **Model cards for model reporting**. In *Proceedings of the conference on fairness, accountability, and transparency*, 2029, pp. 220–229, doi:10.1145/3287560.3287596.

→ Arnold, M., Bellamy, R.K., Hind, M., Houde, S., Mehta, S., Mojsilović, A., Nair, R., Ramamurthy, N., Olteanu, A., Piorkowski, D., Reimer, D., Richards, J., Tsay, J. and Varshney, K.R., **FactSheets: Increasing trust in AI services through supplier's declarations of conformity**. *IBM Journal of Research and Development*, Vol. 63, No. 4/5, 2019, pp. 6:1–6:13, doi:10.1147/JRD.2019.2942288.

→ Hupont, I., Micheli, M., Delipetrev, B., Gómez, E. and Soler Garrido, J., **Documenting high-risk AI: an European regulatory perspective**. *TechRxiv. Preprint*, 2022, doi:10.36227/techrxiv.20291046.v1.

→ Micheli, M., Delipetrev, B., Hupont, I. and Soler Garrido, J., **The landscape of data and AI documentation approaches in the European policy context**, *Ethics and Information Society*, in press.

HOW TO LEVERAGE VOLUNTARY DATA SHARING IN A DATA SPACE?

ORGANISATIONAL

What is the problem?

Increased data sharing is required to deal with today's societal, economic and environmental challenges. But despite the non-rivalrous nature of data (meaning that at a technological level, data are infinitely usable by multiple entities simultaneously), many key actors may lack access to data because there are limited venues for secure, voluntary data sharing. Privacy issues and fears of losing competitive advantage are some of the most common concerns cited by those who would otherwise be willing to voluntarily share their data.

Scenario

The study of rare diseases would benefit from health data voluntarily released by patients and hospitals. The combination of data from various sources can generate new insights into rare diseases and speed up innovation in the field. In this context, a trustworthy actor is needed such as a recognised data altruism organisation that can collect, manage, process and redistribute data made available for the public good by the parties involved.

Proposed solution(s)

Against this background, common European data spaces offer a protected environment in which to exchange data for the public benefit. To enable new ways of sharing data among a wider range of societal actors, the Data Governance Act includes a focus on servic4es that are based on data altruism and introduces a new category of "data altruism organisations recognised in the EU" (RDAOs) to facilitate the voluntary sharing of data for objectives of general interest (see Chapter IV of the Data Governance Act).

Data altruism organisations allows data holders and data subjects to select the type of data, the purpose of sharing such data, the duration, and the range of data users who will benefit from it. Engaging in data altruism can help to redistribute the value of data across society and increase public welfare.

Recommendations

→ **Data providers**

- Share data on altruistic grounds
- Collaborate with recognised data altruism organisations
- Establish and register as a recognised data altruism organisation

→ **Data users**

- Credit contributions of data
- Create incentives for data holders to share data on an altruistic basis

→ **Public authorities**

- Assist and support data altruism organisations
- Reward collaboration with recognised data altruism organisations

Selected resources

→ Carballa-Smichowski, B., Duch-Brown, N. and Martens, B., **To pool or to pull back? An economic analysis of health data pooling**. *JRC Digital Economy Working Paper*, 2021, JRC126961.

→ European Union, **Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)**, *Official Journal of the European Union*, L 152, 2022, pp. 1–44, http://data.europa.eu/eli/reg/2022/868/oj

→ Jones, C.I. and Tonetti, C., **Nonrivalry and the Economics of Data.** *American Economic Review*, Vol. 110, no. 9, 2020, pp. 2819–58, doi:10.1257/aer.20191330.

→ Ponti, M. and Craglia, M., **Citizen-generated data for public policy**, Publications Office of the European Union, Luxembourg, 2020, JRC120231.

→ Giovanni, L., Micheli M., Schade S., Kotsev A., Gori M., Codagnone C., **City data ecosystems between theory and practice: A qualitative exploratory study in seven European cities**, *Data & Policy*, Vol. 5, No. e17, 2023, pp. 1-16, doi:10.1017/dap.2023.13.

→ Suman, A. B., Heyen, N. B., & Micheli, M., **Reimagining health services provision for neglected groups: The "personalization from below" phenomenon**. *Frontiers in sociology*, 8, 1052215, 2023.

WHICH LEGAL ASPECTS SHOULD BE CONSIDERED WHEN CREATING, PROVIDING OR USING NOVEL DATA-DRIVEN SOLUTIONS IN DATA SPACES?

ORGANISATIONAL

What is the problem?

The lawful operationalisation of novel data-driven solutions (e.g., web/mobile applications or added value data services) that build on the combination of data shared within data spaces is a multifaceted issue.

On the general level, all entities providing digital services (including data-driven solutions) must comply with legal obligations pertinent to information society services (as defined in Directive (EU) 2015/1535).

Furthermore, the specific obligations for providers of novel data-driven solutions include respecting legal conditions for the use of the data, software products and services throughout the entire data value chain. Data may be subject to certain restrictions and terms of use, and some databases may be protected under copyright and/or *sui generis* database rights. In addition, software providing access to the data and/or complementing their value may also be a subject of legal protection under intellectual protection rights (IPRs).

All the restrictions on the use of data or software must be respected and propagated in the entire digital value chain by means of legal contracts. Different intermediate data providers will make data available through services that are governed by Terms of Service (ToS) and/or Service Level Agreements (SLAs). ToS and SLAs should contain relevant conditions on the underlying data licences, terms of use, relevant restrictions on the use of interfaces, etc. Moreover, these services should be in line with the current European data sharing legal framework,[1] as well as with EU contract law, competition law and applicable sectoral regulation.

Overall, integrating all the components of novel data-driven solutions that are generated as a result of combining data in data spaces, requires the coordination of agreements to correctly manage the legal and organisational aspects and ensure the robustness and stability of these solutions.

1. (EU) 95/46/EC, (EU) 2018/1807, COM/2020/767, COM(2020) 842, COM(2020) 825, COM(2022) 68 final, (EU) 2019/1024, REGULATION (EU) 2019/1150, COM(2020) 823

Scenario

A sustainable-city data application is created within the Smart Cities Data Space to control the level of CO_2 emissions arising from the mobility of individuals. This application offers information on different transport route alternatives and their estimated CO_2 emissions. The relevant city administration keeps a registry of individuals' emissions and offers incentives for people to maintain their emissions below a certain threshold (for instance by reducing their circulation tax).

For this purpose, several data sources need to be integrated: traffic data, public transport network data, data from city mobility services run by private operators/data from private ride-sharing services, data from CO_2 sensors, CO_2 emissions data, mobility patterns, and the location of the individual's mobile telephone device. All these data will have specific licences and terms of use, and throughout the whole data chain none of the steps to integrate additional data should violate these terms and conditions.

The provider of the digital solution must coordinate respect for the legal constraints of all data components cohesively throughout the entire data value chain. For example, if a dataset at the beginning of the chain restricts a specific type of use, then all the intermediate steps and processes further down the chain should also respect and propagate these restrictions downstream. Similarly, the digital solution provider should ensure that each step in the digital value chain complies with the applicable legal framework.

Proposed solution(s)

Each data space should provide support to help disparate data space stakeholders lawfully operate novel data-driven solutions. For example, this support can be provided in the form of tools, templates, guidelines and advice on the legal obligations and applicable laws. Such support will help stakeholders to better coordinate and negotiate with their counterparts and to steer the correct design of rights and responsibility flows within data value chains and the broader digital ecosystem.

Recommendations

→ **Digital solution providers (providers of an overall service covering a whole digital value chain, e.g., application providers)**

- Be aware, and ensure correct propagation of all legal constraints within the digital value chain
- Ensure the application's overall compliance with the legal framework

→ **Intermediate service providers (providers of any service that makes up a given digital solution, e.g., high-value dataset API service provider)**

- Be aware, and ensure correct propagation of legal constraints from upstream components in the digital value chain
- Design and implement lawful constraints and terms of use for services
- Ensure service compliance with applicable legal framework

→ **Intermediate service user (a business or organisation using third-party services to provide their own services or solutions, e.g., an organisation using data components from other stakeholders)**

- Be aware, and ensure compliance with technical and of legal constraints from upstream components in the digital value chain
- Ensure that the uses of the assets accessed are lawful

→ **Data holders**

- Exhaustively describe licensing and conditions for use of data assets

→ **Data space orchestrator**

- Set up legal support mechanisms within the data space
- Ensure stable legal and technical coordination of the overall ecosystem

→ **Public Authorities**

- Effectively communicate the legal framework applicable to data spaces
- Effectively communicate reporting obligations when necessary
- Define oversight mechanisms to facilitate compliance with the legal framework

Selected resources

→ Posada Sanchez, M. and Pogorzelska, K., **API strategy essentials for Public Sector Innovation: Legal and organisational perspective**, EUR 31216 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-56795-0, doi:10.2760/511499, JRC129940.

→ Posada Sanchez, M., Pogorzelska, K. and Vaccari, L., **API strategy essentials for public sector innovation**, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-56796-7, doi:10.2760/203781, JRC129923.

→ Posada Sanchez, M., Pogorzelska, K. and Vespe, M., **The role of Application Programming Interfaces (APIs) in data governance and digital coordination**, European Commission, 2022, JRC128250.

→ Posada Sanchez, M., Vaccari, L. and Pogorzelska, K., **Unfolding opportunities from the use of APIs in Europe**, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-34203-8, doi:10.2760/074141, JRC124758.

4 JRC dashboard on European data spaces

As mentioned in **Section 1.2.3**, the JRC Knowledge Base on data spaces includes an dashboard[28] that allows those interested to identify and retrieve JRC publications that map to the principles, requirements and identified features of common European data spaces.

The aim of this online tool is to showcase the publicly available research by the JRC on topics related to data sharing and data spaces, allowing the dashboard users to retrieve results by different keywords, without having to perform single queries on the JRC Publications Repository. Moreover, this visual tool offers the possibility to restrict or enlarge the search to other terms, and consequently to retrieve additional publications, as well as to download a list of the results and to directly access the full texts of publications. The tool can be accessed through an EU Login account.

The homepage of the dashboard (visualised in **Figure 5**) shows the main functional requirements (see **Section 2.2**), as well as their descriptions, on different buttons that can be selected. Once a requirement has been chosen, three different blocks appear on the page (see **Figure 6**):

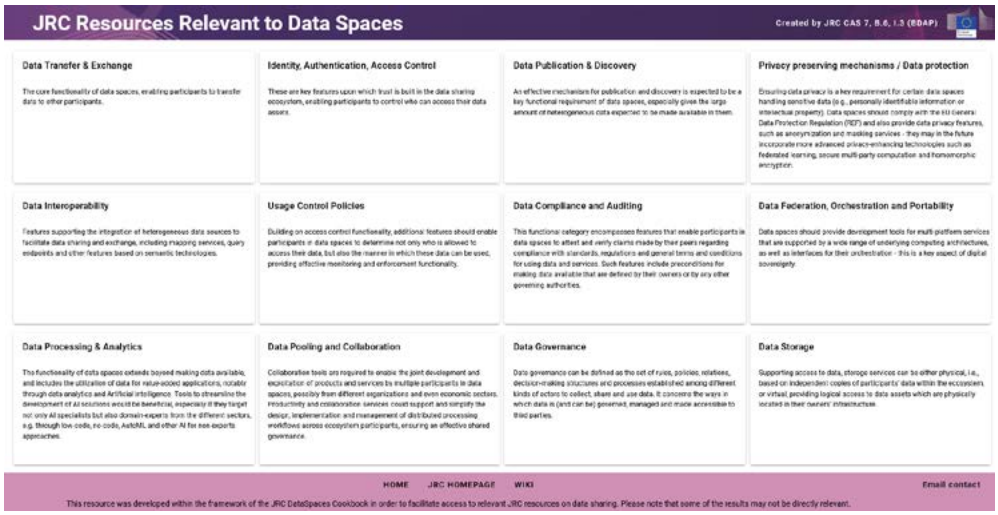
- On the left side is a list of keywords that were used to retrieve relevant JRC publications. Two levels of keywords appear: the first one is composed by an expert-determined combination of words to set the search environment, while the second one constitutes a list of keywords similar to the first level and is automatically retrieved using the SeTA tool (Hradec et al., 2019). All the keywords are by default already selected, and the user can decide to deselect one or more of them, updating the query through the "Retrieve publications" button.
- The central block is composed of the list of publicly available JRC publications that were retrieved through the keywords. The page only shows the title and the year of the publication, but by clicking on each line the user will be redirected to the corresponding page on the JRC Publications Repository, where complete information is available, as well as the publication itself. The list of publications is downloadable in CSV format through the down arrow button.
- The top right panel includes filters that can be applied to the search. By default, the search of the left panel keywords is performed using the title, abstract and keywords of the publications, but through this functionality the user can decide whether to exclude some of these. The slide bar allows the user to expand or reduce the search with respect to year of publication. The bottom right panel includes a bubble chart that gives the user an indication as to the amount of publications retrieved by year.

Regarding the methodology for how the set of keywords was composed, the aim was to create a mix of human-validated terms and machine-identified ones. In particular, starting from an expert curated list of keywords that represents the first level (taking terms from the title of the requirements and fine-tuning the list with some other concepts that could relate well to those topics), we then added to this list a series of similar terms retrieved through a semantic text analysis tool, SeTA (second level). We explicitly decided not to allow users to search open text keywords as this could lead to the retrieval of less relevant publications that do not actually relate to the realm of the requirements that were identified for common European data spaces.

28 https://jeodpp.jrc.ec.europa.eu/eu/dashboard/voila/render/DataSpaces/JRCResourcesData.ipynb

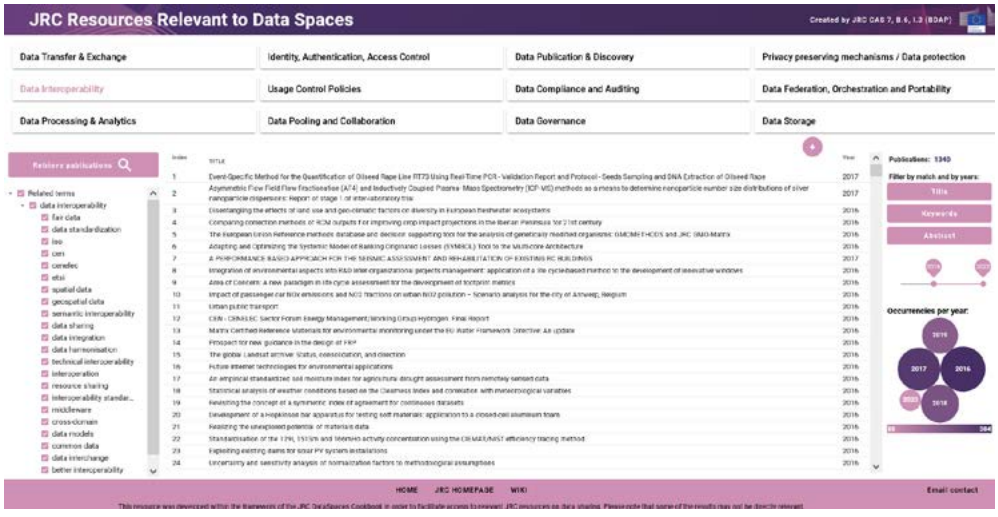
The dashboard is hosted on the JRC Big Data Platform (BDAP) and was created through the Voilà library. The code that performs the search was written in Python, and at the moment the tool searches for publications on a dump of the publicly available JRC publications. In the near future we plan to perform this search on the live database, avoiding the need for periodic updates on the dump and including also the most recent publications on these topics.

→ **Figure 5. Homepage of the dashboard providing access to JRC publications relevant to principles, requirements and features of common European data spaces.**



Source: **JRC dashboard on European data spaces**.

→ **Figure 6. Sample page of the JRC dashboard on data spaces visualised after a requirement is selected.**



Source: **JRC dashboard on European data spaces**.

5 Discussion and conclusions

The ambitious policy agenda defined by the European strategy for data holds great potential for devising approaches that would lead to a fairer, more open and more inclusive data market in Europe. Given the complexity and envisioned impact of the foreseen initiatives, those developments can act as inspiration and trigger similar regulatory developments in other parts of the world according to the so-called "Brussels effect" (Bradford, 2020). Now that the EU policy and legislative initiatives are to a large extent conceptualised, it is important to focus strongly on the implementation aspects of such provisions and "fill" them with concrete substance. Specifically, the projects under the Recovery and Resilience Facility, the Digital Europe Programme, and relevant parts of Horizon Europe, but also similar instruments across the EU Multiannual Financial Framework and others established on the national and sub-national levels, contain the instruments for specific implementation activities.

Within that context, **the present report aims to provide useful scientific insights for the establishment of European data spaces, which can inform the different stakeholders engaged in the establishment of data spaces**. In this report, we (i) distilled evidence from our own activities dedicated to the scientific investigation of technical and organisational enablers for data sharing, and (ii) communicated JRC knowledge that maps to the principles and requirements that we have identified for common European data spaces. In doing so, we did not have the ambition to cover all aspects related to data sharing and we are conscious that there may be many open questions to which we do not (yet) have answers. This is understandable considering that data spaces are a relatively new concept and are still being understood and operationalised in practical terms. A shared understanding and achievement of the high-level objective of the European strategy for data to define a single market for data would be only realised through an inclusive process that engages all key actors. In this regard, we recognise decentralisation as a key aspect in the data spaces endeavour: there is no single organisation that can alone establish and operate data spaces or take top-down decisions to which all the other actors have to adhere, but at the same time the risk that multiple stakeholders set up their own data spaces with related specifications and governance rules shall be avoided.

Also, we are fully aware that further questions, in addition to those discussed in the How-to's (see **Section 3**), may appear and be answered over time, as we explore the needs, issues and unknowns related to the design and operationalisation of common European data spaces. Moreover, while deciding to address general-purpose questions that apply to all common European data spaces, we also acknowledge that the development of each of these data spaces will inevitably demonstrate distinct characteristics based on specific community needs, infrastructures, actors and data sharing practices, and may take different and potentially new directions in the years to come.

For all these reasons, this report is not to be seen as a definitive guide: it should be considered as a living document. As our knowledge and understanding in the domain grows, the JRC Knowledge Base can be further updated and enriched based on real-world implementation evidence and related feedback from actors involved in common European data space initiatives, especially as the development of these becomes more mature. We therefore encourage and welcome any feedback from stakeholders and actors involved in the conceptualisation and implementation of common European data spaces, and will aim to reflect this feedback directly in the online part of the JRC Knowledge Base.

References

- Bodó, B. & Irion, K., Janssen, H. and Giannopoulou, A., 'Personal data ordering in context: the interaction of meso-level data governance regimes with macro frameworks'. *Internet Policy Review*, Vol. 10, No 3, 2021, pp. 1-31, doi:10.14763/2021.3.1581.

- Bradford, A., *The Brussels effect: How the European Union rules the world*, Oxford University Press, USA, 2020.

- Braun, R., Laa, B., and Rohatsch, L., 'Governance challenges of urban dataspace - transdisciplinary perspectives'. In: *Extended abstracts of the Eu-SPRI 2022 conference `Challenging science and innovation policy'*, 2022, pp. 105–107, **http://hdl.handle.net/20.500.12708/63110**

- Data Spaces Support Centre, Starter Kit for Data Space Designers - Interim version | March 2023, 2023a, **https://dssc.eu/wp-content/uploads/2023/03/DSSC-Starterkit-Version-1.0.pdf**

- Data Spaces Support Centre, DSSC Glossary - version 1.0 | March 2023, 2023b, **https://dssc.eu/wp-content/uploads/2023/03/DSSC-Data-Spaces-Glossary-v1.0.pdf**

- Davies, T., 'Data Governance and the Datasphere Literature Review'. *Datasphere Initiative*. 2022, **https://www.thedatasphere.org/wp-content/uploads/2022/11/Data-governance-and-the-Datasphere-Literature-Review-2022.-Tim-Davies.pdf**

- DIGITAL, Call for proposals: Cloud Data and TEF (DIGITAL-2022-CLOUD-AI-02), Version 1.0, 2022, **https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/digital/wp-call/2022/call-fiche_digital-2022-cloud-ai-02_en.pdf**

- European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, *A European strategy for data*, COM(2020) 66 final, 2020a, **https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN**

- European Commission, *Shaping Europe's digital future*, Publications Office of the European Union, Luxembourg, 2020b, **https://data.europa.eu/doi/10.2759/091014**

- European Commission, Regulation of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), *Official Journal of the European Union*, L 152, 2020c, pp. 1–44 **http://data.europa.eu/eli/reg/2022/868/ojhttps://eur-lex.europa.eu/eli/reg_impl/2023/138/oj**

- European Commission, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, Europe's moment: Repair and Prepare for the Next Generation, SWD(2020) 98 final, 2020e, **https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:456:FIN**

- European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 2030 Digital Compass: the European way for the Digital Decade, COM(2021) 118 final, 2021a, **https://eur-lex.europa.eu/resource.html?uri= cellar:12e835e2-81af-11eb-9ac9-01aa75ed71a1.0001.02/DOC_1&format=PDF**

- European Commission, The DIGITAL Europe Programme – Work Programmes, 2021b, **https://digital-strategy.ec.europa.eu/en/activities/work-programmes-digital**

- European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Updating the 2020 New Industrial Strategy: Building a stronger Single Market for Europe's recovery, COM(2021) 350 final, 2021c, **https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021DC0350**

- European Commission, Commission Staff Working Document accompanying the document Proposal for a Decision of the European Parliament and of the Council establishing the 2030 Policy Programme "Path to the Digital Decade", COM(2021) 574 final, 2021d, **https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52021SC0247&from=EN**

- European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Fostering a European approach to Artificial Intelligence, COM(2021) 205 final, 2021e, **https://eur-lex.europa.eu/resource.html?uri=cellar:01ff-45fa-a375-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF**

- European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artifical Intelligence Act) and amending certain Union legislative acts, COM(2021) 206 final, 2021f, **https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52021PC0206&from=EN**

- European Commission, Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act), COM(2022) 68 final, 2022a, **https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022PC0068&from=EN**

- European Commission, Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space, COM/2022/197 final, 2022b, **https://eur-lex.europa.eu/legal-content/EN/TXT/TML/?uri=CELEX:52022PC0197&from=EN**

- European Commission, Commission Staff Working Document on Common European Data Spaces, SWD(2022) 45 final, 2022c, **https://ec.europa.eu/newsroom/dae/redirection/document/83562**

- European Commission, European Open Science Cloud (EOSC), 2022d, **https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en**

- European Commission, Commission Staff Working Document: Impact Assessment Report Accompanying the document Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space, SWD(2022) 131 final, 2022e, **https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022SC0131&from=EN**

- European Commission, Commission Staff Working Document 2023 Annual Single Market Report: Single Market at 30, SWD(2023) 26 final, 2023a, **https://single-market-economy.ec.europa.eu/system/files/2023-01/ASMR%202023.pdf**

- European Commission, Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use, *Official Journal of the European Union*, L 19, 2023b, pp. 43–75, **https://eur-lex.europa.eu/eli/reg_impl/2023/138/oj**

- European Union, Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications), *Official Journal of the European Union*, L 201, 2002, pp. 37–47, **http://data.europa.eu/eli/dir/2002/58/oj**

- European Union, Directive (EU) 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), *Official Journal of the European Union*, L 108, 2007, pp. 1–14, **https://eur-lex.europa.eu/eli/dir/2007/2/2019-06-26**

- European Union, Directive 2011/24/EU of the European Parliament and of the Council of 9 March 2011 on the application of patients' rights in cross-border healthcare, *Official Journal of the European Union*, L 88, 2011, pp. 45–65, **http://data.europa.eu/eli/dir/2011/24/oj**.

- European Union, Consolidated version of the Treaty on the Functioning of the European Union, *Official Journal of the European Union*, C 326, 2012a, pp. 47–390, **http://data.europa.eu/eli/treaty/tfeu_2012/oj**

- European Union, Charter of Fundamentals Rights of the European Union, *Official Journal of the European Union*, C 326, 2012b, pp. 391–407, **http://data.europa.eu/eli/treaty/char_2012/oj**

- European Union, Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC, *Official Journal of the European Union*, L 257, 2014, pp. 73–114, **http://data.europa.eu/eli/reg/2014/910/oj**

- European Union, 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*, L 119, 2016, pp. 1–88, **https://eur-lex.europa.eu/eli/reg/2016/679/oj**

- European Union, Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, *Official Journal of the European Union*, L 117, 2018, pp. 59–68, **http://data.europa.eu/eli/reg/2018/1807/oj**

- European Union, Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, *Official Journal of the European Union*, L 303, 2017, pp. 1–175, **http://data.europa.eu/eli/reg/2017/745/oj**

- European Union, Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, *Official Journal of the European Union*, L 172, 2019a, pp. 56–83, **https://eur-lex.europa.eu/eli/dir/2019/1024/oj**

- European Union, Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services, *Official Journal of the European Union*, L 186, 2019b, pp. 57–79, **http://data.europa.eu/eli/reg/2019/1150/oj**

- European Union, Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act), *Official Journal of the European Union*, L 151, 2019c, pp. 57–79, **http://data.europa.eu/eli/reg/2019/881/oj**

- European Union, Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act), *Official Journal of the European Union*, L 265, 2022, pp. 1–66, **http://data.europa.eu/eli/reg/2022/1925/oj**

- Fritzenkötter, J., Hohoff, L., Pierri, P., Verhulst, S.G., Young, A., and Zacharzewski, A., 'Governing the Environment-Related Data Space'. *TheGovLab*, 2022, **https://files.thegovlab.org/erdgovernance.pdf**
- Granell, C., Mooney, P., Jirka, S., Rieke, M., Ostermann, F., Van Den Broecke, J., Sarretta, A., Verhulst, S., Dencik, L., Oost, H., Micheli, M., Minghini, M., Kotsev, A. and Schade, S., *Emerging approaches for data-driven innovation in Europe: Sandbox experiments on the governance of data and technology*, EUR 30969 EN, Publications Office of the European Union, Luxembourg, 2022, doi:10.2760/511775.
- Hradec, J., Ostlaender, N., Macmillan, C., Acs, S., Listorti, G., Tomas, R. and Arnes Novau, X., *Semantic Text Analysis tool: SeTA*, EUR 29708 EN, Publications Office of the European Union, Luxembourg, 2019, doi:10.2760/57781.
- International Data Spaces Association, *IDSA Rule Book*, White Paper, Version 1.0, 2020, **https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-White-Paper-IDSA-Rule-Book.pdf**
- International Data Spaces Association, *IDSA Rule Book*, White Paper, Version 2.0 (beta), 2023, **https://docs.internationaldataspaces.org/idsa-rulebook-v2/idsa-rulebook**
- International Data Spaces Association, *Governance for Data Space Instances: Aspects and Roles for the IDS Stakeholders*, Position Paper | Version 0.1 | August 2021, 2021, **https://internationaldataspaces.org/wp-content/uploads/IDSA-Position-Paper-Governance-for-Data-Space-Instances-Aspects-and-Roles-for-IDS-Stakeholders.pdf**
- Islam, S., Weber, A. and Tóth-Czifra, E., 'From Green Deal to Cultural Heritage: FAIR Digital Objects and European Common Data Spaces'. *Research Ideas and Outcomes*, Vol. 8, No e93815, 2022, doi: 10.3897/rio.8.e93815.
- Keller, P., Five Things I know about Data Spaces. *Open Future blog*, 2021, **https://openfuture.eu/blog/five-things-i-know-about-data-spaces**
- Kiseleva, A. and De Hert, P., 'Creating a European Health Data Space. Obstacles in Four Key Legal Areas'. *European Pharmaceutical Law Review*, Vol. 5, No 1, 2021, pp. 21–36, doi:10.2139/ssrn.3846781.
- Kooiman, J., *Governing as Governance*, Sage Publications, London, 2003, doi:10.4135/9781446215012.
- Liu, J., 'Social data governance: Towards definition and model'. *Big Data & Society*, Vol. 9, No 2, 2022, pp. 1-14, doi:10.1177/20539517221111352.
- Micheli, M., Ponti, M., Craglia, M. and Berti Suman, A., 'Emerging models of data governance in the age of datafication'. *Big Data & Society*, Vol. 7, No 2, 2020, pp. 1-15, doi:10.1177/2053951720948087.
- Nagel, L. and Lycklama, D. (Eds), 'Design Principles for Data Spaces'. *International Data Spaces Association*, 2021, doi:10.5281/zenodo.5244997.
- Organisation for Economic Co-operation and Development, *Going Digital to Advance Data Governance for Growth and Well-being*, OECD Publishing, Paris, 2022, doi:10.1787/e3d783b0-en.
- SITRA, Rulebook for a Fair Data Economy, 2022, **https://www.sitra.fi/en/publications/rulebook-for-a-fair-data-economy**
- Torre-Bastida, A.I., Gil, G., Miñón, R. and Díaz-de-Arcaya, J., 'Technological Perspective of Data Governance in Data Space Ecosystems'. In: Curry, E., Scerri, S. and Tuikka, T. (Eds) *Data Spaces*. Springer, Cham, 2022, doi:10.1007/978-3-030-98636-0_4.
- World Trade Organization, General Agreement on Trade in Services, 2022, **https://www.wto.org/english/docs_e/legal_e/26-gats_01_e.htm**

List of abbreviations

AI	Artificial Intelligence
API	Application programming interface
BDAP	Big Data Platform
BDVA	Big Data Value Association
B2C	Business to Consumer
CC	Creative Commons
CEF	Connecting Europe Facility
CSA	Coordination and Support Actions
CSV	Comma-Separated ValuesETL - Extract, Transform, Load
DIGITAL	Digital Europe Programme
DG AGRI	Directorate-General for Agriculture and Rural Development
DG CNECT	Directorate-General for Communications Networks, Content and Tecnology
DG ENV	Directorate-General for Environment
DG ESTAT	Directorate-General. Statistical Authority of the European Union
DG DIGIT	Directorate-General for Informatics
DG GROW	Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs
DG JUST	Directorate-General for Justice and Consumers
DGA	Data Governance Act
DSBA	Data Spaces Business Alliance
DG	Directorate General
DISP	Data Intermediation Service Provider
DSSC	Data Spaces Support Centre
DTLF	Digital Transport and Logistics Forum
EDIB	European Data Innovation Board
EHDS	European Health Data Space
EDIH	European Digital Innovation Hubs
EEA	European Economic Area
eIDAS	electronic IDentification, Authentication and Trust Services
EIF	European Interoperability Framework
EMDS	European Mobility Data Space
EOSC	European Open Science Cloud
EU	European Union
EUPL	European Union Public License
FAIR	Findable, Accessible, Interoperable and Reusable
GDPR	General Data Protection Regulation
HPC	High-performance computing
HTML	HyperText Markup Language
HVD	High-Value Datasets
IDSA	International Data Spaces Association
ISO	International Organization for Standardization
JRC	Joint Research Centre
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
LiDAR	Light Detection and Ranging
MIMs	Minimal Interoperability Mechanisms
OGC	Open Geospatial Consortium
RADAR	Radio Assisted Detection and Ranging
RDAO	Data altriusm organisation recognised in the EU
RDF	Resource Description Framework
SDG	Sustainable Development Goal
SDK	Software Development Kit
SDO	Standards Development Organisation
SME	Small and mid-size enterprises
SWD	Staff Working Document
TEF	Testing and Experimentation Facilities
XML	Extensible Markup Language
W3C	World Wide Web Consortium

List of figures

Figure 1: Overview of the legislative framework for the implementation of the European strategy for data. Source: adapted from European Commission 12

Figure 2: Selected characteristics of common European data spaces 14

Figure 3: Overview of the initial breakdown of sector-specific common European data spaces 15

Figure 4: Governance in a common European data space 28

Figure 5: Homepage of the dashboard providing access to JRC publications relevant to principles, requirements and features of common European data spaces 72

Figure 6: Sample page of the JRC dashboard on European data spaces visualised after a requirement is selected 72

List of tables

Table 1: List of selected EU policy provisions translated into functional and non-functional requirements for common European data spaces 22

Table 2: Sample of potentially relevant data sharing initiatives 33

Table 3: List of technical and organisational How to's 35

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct centres.
You can find the address of the centre nearest you online
(**european-union.europa.eu/contact-eu/meet-us_en**).

On the phone or in writing

Europe Direct is a service that answers your questions about the European Union.
You can contact this service:
— by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
— at the following standard number: +32 22999696,
— via the following form: **european-union.europa.eu/contact-eu/write-us_en**.

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website (**european-union.europa.eu**).

EU publications

You can view or order EU publications at op.europa.eu/en/publications. Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre (**european-union.europa.eu/contact-eu/meet-us_en**).

EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex (**eur-lex.europa.eu**).

Open data from the EU

The portal **data.europa.eu** provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.

Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

EU science Hub
Joint-research-centre.ec.europa.eu

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

EU Science

Publications Office
of the European Union