Pilot Project - Digital European Platform of Quality Content Providers (2nd Phase)

Interim report – summary

August 2022

Table of Contents

Table of Contents 2
Introduction 3
Legal requirements for the media data space 4
Governance requirements for the media data space 10
Examples of data space applications in the media sector 14
Conclusions 16

Introduction

Data are crucial for economic growth, competitiveness, innovation, job creation and societal progress. In 2020, the European Union announced a European strategy for data, with the aim of creating a data single market for fostering Europe's global competitiveness and data sovereignty. The strategy establishes common European data spaces with the ultimate goal to create a single, interoperable European data space[1]. The development of data spaces in strategic sectors will contribute to accelerating digital transformation objectives defined by the European Commission and included in the "2030 Digital compass: the European way for the digital decade"[2].

The Commission refers to a 'data space' as "*a type of data relationship between trusted partners who adhere to the same high standards and guidelines when it comes to data storage and sharing*"[3]. The key feature of data spaces is that the data are stored by the owner and not centrally and transferred – against compensation or for free – upon agreements between two or more parties participating to the data space[4].

The innovative potential of data spaces is also recognised in the media sector. The Media and Audio-visual Action Plan (MAAP) considers data spaces as a tool able to transform the collaboration between creators, producers, and distributors. According to the MAAP, media data spaces could "*host relevant media data such as content, audience data and content meta-data as well as other types of data on users' behaviours that might be useful to create content better tailored to consumer needs and distribute it more efficiently*".[5] The MAAP also highlights that in a European Media Data Space stakeholders *"could pool together, in a pro-competitive manner and in respect of data protection rules, their content and customer data to produce news targeting their own national audiences*'. Furthermore, it could '*allow for exploitation of synergies with datasets produced by other creative industries'*, such as cultural heritage, film-makers or advertisers.

This document presents a summary of the preliminary findings emerged from the "Pilot project – Digital European Platform of Quality Content Providers (2nd Phase)" carried out by external contractors on behalf of the European Commission. The study analysed and provided preliminary considerations on key legal aspects, as well as on suitable governance models that should be considered for the development of a European Data Space. With regard to the first, an analysis of the rules regulating data exchanges and parties' rights is carried out, whereas the latter focused on the analysis of governance models able to ensure the effective functioning of the data space and parties' trustworthiness. Furthermore, the document displays eleven use cases, developed during the project that could address current or future challenges emerging in the media industry. The information collected is based on a series of consultation activities, including interviews and workshops with relevant stakeholders from the European media ecosystem.

[1] European Commission (2022), Staff working document on data spaces: https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces#:~:text=The%20European%20data%20strategy%20of,meeting%20the%20Green%20Deal%20objectives.
[2] European Commission (2022), Staff working document on data spaces. Available at: https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces
[3] European Commission (2022), Staff working document on data spaces. Available at: https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces
[4] https://gaia-x.eu/what-is-gaia-x/core-elements/data-spaces/
[5] Europe's Media in the Digital Decade: An Action Plan to Support Recovery and Transformation COM (2020) 784 final, available at: *https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0784&from=EN*

Legal requirements for the media data space

The media data space initiative does not operate in a legal and policy vacuum. On the contrary, a media data space must comply with the applicable EU legal framework on both personal data protection and non-personal data, security, fundamental rights, competition law, and other relevant rules. It is beyond the scope of this document to analyse in detail the plethora of legislation applicable to data and data sharing. Ultimately, it is a responsibility of each data holder to ensure legal compliance and to have authorization and/or legal basis for sharing data.

This document provides a preliminary analysis of the s the media data space legal requirements related to the data sharing and of the copyright framework and its related rights.

The Legal Requirements for Data Sharing in a Media Data Space

The analysis focuses on two legislative proposals which, when adopted, will constitute cornerstones of legal framework for data sharing and data spaces: **Data Governance Act (DGA)** proposal and **Data Act** proposal. In what follows, we recall the preliminary legal requirements stemming from these proposals potentially applicable to media data spaces.

These are:

In their contractual negotiations, media data space participants should ensure the **contractual balance** and take into account the potential unfairness of contractual terms, as proposed in the Data Act, when negotiating their data sharing agreement(s). This holds true especially in sub-data spaces and/or use cases with heterogeneity of media outlets in terms of size, position or market dominance (e.g. publicly funded broadcasters vis-à-vis start-ups).

Media data spaces should aim towards a pro-competitive outcome to promote competition and innovation and to offer better products and services to customers and thereby avoid any anti-competitive behaviours (e.g. exclusion of, or discrimination against, certain stakeholders by preventing their effective access to data).[6] Moreover, in accordance with the **FAIR** data principles,[7] common European data spaces should make data "*Findable, Accessible, Interoperable and Re-usable*".

Media data space participants should agree on the role, function and services offered by the **data intermediation service** as defined in the DGA proposal.[8] In the context of media data spaces, data intermediation services could make specific technical infrastructure available for connecting data holders and data recipients as well as offer additional specific tools and services such as temporary storage, curation, conversion, anonymisation, pseudonymisation. According to the DGA proposal, a key element to bring trust for data holder and data users in data sharing services is the neutrality of data intermediaries.[9] The Regulation proposes a number of measures to increase trust in data intermediation services, ranging from a structural separation of the service provider to the prohibition of the use the data for any other purpose. Together with the agreement on a dispute resolution system, the roles and responsibilities of the providers of data intermediation service will constitute an important part of the governance structure.

[6] Europe's Media in the Digital Decade: An Action Plan to Support Recovery and Transformation COM (2020) 784 final, available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0784&from=EN.
[7] Wilkinson MD;Dumontier M;Aalbersberg IJ;Appleton G;Axton M;Baak A;Blomberg N;Boiten JW;da Silva Santos LB;Bourne PE;Bouwman J;Brookes AJ;Clark T;Crosas M;Dillo I;Dumon O;Edmunds S;Evelo CT;Finkers R;Gonzalez-Beltran A;Gray AJ;Groth P;Goble C;Grethe JS;Her, "The Fair Guiding Principles for Scientific Data Management and Stewardship" (Scientific data) <https://pubmed.ncbi.nlm.nih.gov/26978244/> accessed August 10, 2022.
[8] Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) COM/2020/767 final.
[9] Recital 26, DGA Proposal.

Technical protection measures preventing unauthorised use or disclosure of data could include measures such as digital rights management (in the protection of copyright-protected subject matter) or smart contracts (to ensure compliance with the agreed contractual terms for making data available). In particular, the assessment of the technical feasibility of smart contracts and their functionality in the media context will need to take into account the Data Act proposal's essential requirements regarding smart contracts for data sharing, namely: robustness, safe termination and interruption and data archiving and continuity.[10]

Based on the principle of contractual freedom[11], media data space participants remain free to decide whether they make data available for third parties. As an example, a third party to whom data is made available may be a research organisation or a not-for-profit organisation. It is known that (social) media data availability is of enormous benefit to scientific research communities. Access to quality fine-grained impression data plays an immensely positive role for scientific community in measuring the reach of dis- and misinformation. In that regard, it is suggested that the media data space could enhance not only the development of new products and services for media stakeholders and journalists, but it could also benefit scientific research or civil society. It is recommended that the media data space participants specify **conditions for allowing third-party access** to media data. Such conditions should in particular clarify the types of data (personal/non-personal), purposes for which third parties can process the data and foresee a mechanism to allow for a data holder/data user to revoke access or enable discontinuation of data access in case of data misuse. The applicable legal basis for such data sharing needs to be observed by each media data space participant.

The **interoperability** requirements and standards need to be agreed on the sectorial level between the media data space participants. Two recommendations come to fore in that regard. First, minimum common principles and standards should be agreed between the media stakeholders to access, port data and to create value efficiently from data coming from different sources and in different formats. There is a need for a minimal set of commonly agreed cross-sector and cross-border interoperability requirements and solutions.[12] Additionally, common data specifications are to be adopted. Reusable data structures and models (in form of core vocabularies), ontologies, metadata application profile, reference data in the form of core vocabulary, taxonomies, code lists, authority tables, thesauri should also be part of the technical specifications for semantic interoperability. Second, in order to accommodate, to the largest extent, the specificity of the media sector, the standards and specifications should be industry-driven and agreed on in a transparent and, if possible, open-source process. A collaboration with other organizations and initiatives is advisable, such as standards organizations and trade bodies to drive cross-stakeholder consensus and monitoring of specifications and their implementation. Second, according to the Data Act proposal, operators of data spaces will need to comply with the essential requirements regarding interoperability.[13] This should allow to further enhance interoperability with other, related data spaces (e.g. cultural heritage data space, data space for tourism, etc.).

A recurrent subject in data spaces discussions is the notion of **'data control'**. First, the dominant view in legal theory tends to be that data cannot be owned. Among various arguments, many point out to the non-rival, non-excludable, and non-depletable character of data: they can be used more than once without losses in quality. Others exclude data - especially personal data - ownership rights, from a fundamental rights perspective: the

[10] Art. 30 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.
[11] Recital 39, Data Act proposal.
[12] GAIA-X: Technical Architecture, June, 2020 https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-technical-architecture.pdf?__blob=publicationFile&v=5.
[13] Art. 28, Data Act proposal.

fundamental right to the protection of personal data safeguards the personality and the right to informational self-determination of data subjects, not their property. To this end, the potential of data sharing can be achieved if that data is shared amongst businesses, for instance, through the granting of **access rights**, not by exclusive ownership right.[14] At the same time, data sharing is characterized by the opposing demands for access to data and demands for **control of data**.[15] Some media stakeholders are reluctant to provide access to data without adequate incentives and guarantees of **control**. Many fear that data sharing without adequate safeguards can be a significant business risk and/or do not see an added value in sharing data.[16] Hence, a media data space has a potential to address these concerns by enabling access to data while retaining control over data for the participants.

Taking 'data control' as one of the design principles for a media data space requires a series of decisions made by media data space participants. In particular, in line with the Data Governance Act proposal, media data space participants should establish mechanisms of:

- verification of data holder's 'control' over data;
- ensuring that a data holder provides information about data sources (i.e. in case of personal data a legal basis of how the data was obtained), its up-to-datedness, correctness and other relevant details;
- agreeing on the consequences for a data recipient and for a data sharing activity, in the case that a data holder did not have the right to grant access to or to share data as it claimed to have;
- allowing a data holder to require data user identification to verify the data recipient's entitlement to access the data;
- granting a data holder a right to act when data has been provided based on incorrect or false information, has been accessed unlawfully or has been used for unauthorised purposes;
- obliging data recipient to delete and/or abstain from further exploitation of the data which were unlawfully obtained or has been used for unauthorised purposes or has been disclosed to third party without the data holder's authorization;
- preventing further use of such data or derivative data or services developed on the basis of knowledge obtained through such data (this can in particular take place with the use of technical protection measures such as smart contracts);
- clarifying whether and when a data user is required to return the data to the data holder or/and delete the data and terminate data processing activities. Consequences when data has already been combined with other data and/or processed to produce new datasets, services or products should be foreseen.

Finally, the technical means of the implementation of 'data control' should be explored and adapted to the needs of the media data space participants.

[14] Thouvenin F and Tamò-Larrieux A, "Data Ownership and Data Access Rights (Chapter 15) - Big Data and Global Trade Law" (Cambridge Core) <https://www.cambridge.org/core/books/big-data-and-global-trade-law/data-ownership-and-data-access-rights/BC314C63C58A09C4B9C5D55894FE68C6&gt>; accessed June 10, 2022.

[15] "Data Access and Control in the Area of Connected Devices - BEUC" <https://www.beuc.eu/sites/default/files/publications/beuc-x-2018-121_data_access_and_control_in_the_area_of_connected_devices.pdf&gt>; accessed August 10, 2022.

[16] "Business-to-Business Data Sharing: An Economic and Legal Analysis" (EU Science Hub, July 22, 2020) <https://joint-research-centre.ec.europa.eu/publications/business-business-data-sharing-economic-and-legal-analysis_en&gt>; accessed August 10, 2022.

Copyright and Related Rights in Media Data Spaces

The media sector is a significant creator and disseminator of content protected by copyright and related rights. In order to be feasible, a media data space should **comply with the applicable EU rules on copyright and related rights** (the EU copyright *acquis*). The copyright *acquis* applies to works of authorship, i.e., an extensive range of content, including textual, visual and aural creations, original databases, as well as so-called "other subject matter". With the latter expression, the EU legislature refers to non-original productions that are somehow connected or adjacent to the creative process and that are deemed to deserve some form of protection but not full copyright, instead an attenuated form of "related" or "neighbouring" rights. Examples of such related rights are the protection of the investment of phonogram producers, of broadcasting organisations, or of press publishers vis-à-vis information society service providers. Original works and the subject matter of related rights afford their rightsholder a suite of rights of economic exploitation, which however vary in length and intensity depending on whether copyright or related rights are at play. In addition, moral rights may also be present, especially in the case of copyright, rarely for related rights, which however are not part of the copyright *acquis* and are regulated by national law. In order to carry out any activity protected by a right of economic exploitation that involve protected content by anyone other than the right holder, authorisation (either statutory or contractual) is required. In certain cases, the rights pertaining to certain forms of content may be subject to collective management.

A number of different **data types** with different use cases will be exchanged within the media data space and available for media data space participants. Where exchanges involve protected subject matter, these uses should be based on a proper legal basis, commonly either an authorisation by law (e.g. exceptions and limitation) or by the rightsholder (contracts, licenses). The exact taxonomy of data adopted by a media data space is not decisive for the applicability of the *acquis*. "Content data" may be a clearer case of copyright-protected subject matter than others, but this is not guaranteed and depends on the actual content at stake. "Metadata" and "usage data" may be protected by copyright (such as a literary work, albeit unlikely), by related rights (such as the Sui Generis Database right, possibly), or even considered rights management information under the *acquis*. A clarification of the relationship between data types that are relevant for a media data space and copyright notions is advisable.

Additionally, specific aspects of the copyright *acquis* are particularly relevant for the realisability of a media data space. At this time, the report identifies three particular questions that can provide opportunities and challenges for participants of a media data space: (1) the question of whether the exchanged data qualifies as **created or obtained** in the sense of the *sui generis* database right; (2) the extent of **authorisation** (how exceptions and limitation to copyright and the various types of licensing agreements can be optimally combined to obtain reliable, fast and accessible results), and; (3) **technological protection measures**.

Firstly, sets of data that will be exchanged by the media data space may qualify as a "database" - a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means.[17] Whether the data that is collected in such a database has been **created or obtained** entails a difference in treatment within the meaning of the *sui generis* database right. This has important implications for a media data space. On the one hand, data within a database that has been "created" generally does not enjoy protection by the *sui generis* database right.[18]

[17] Art.1(2), Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (Database Directive).
[18] Case C-338/02 *Fixtures Marketing* [2004] ECLI:EU:C:2004:696, paras. 48-52.

This means that other legal regimes, such as that of trade secrets or contracts may more aptly govern such databases. However, in a case in which data is created and if a separate investment in the obtaining, verification or presentation can be identified (i.e. an investment in the creation of the database, not in the creation of data),[19] this could qualify for protection. In this context, participants of a media data space should familiarise themselves with the distinctions between created and obtained data. This would allow participants to determine whether (1) legal alternatives to the *sui generis* database protection are preferrable, e.g. via contractual regulation of their transactions, or (2) the *sui generis* database right protects their database. Furthermore, recent legislative developments via the Data Governance Act and the proposed Data Act seek to prevent holders of data from restricting access via the *sui generis* database right, with implications for the created/obtained distinction.[20] Legal scholars find that such reforms of the *sui generis* database right offer an opportunity for improving data sharing practices,[21] however "risks creating even more fragmentation in the laws of Member States" have been identified.[22] In light of the above and persistent legal uncertainties,[23] it is recommendable for participants of the media data space to explore complementary and/or additional legal vehicles for exchanging data beyond the *sui generis* database right.

Secondly, as stated, data that qualifies for protection under the copyright *acquis*, as well as for protection under the *sui generis* database right, requires a form of **authorisation** for non-rightsholders in order to perform certain acts vis-à-vis the protected subject matter. Statutory authorisations for non-rightsholders under the copyright *acquis* are provided by limitations and exceptions. The *acquis* provides for a range of potential limitations and exceptions for Member States to implement, most notably requiring an exception for temporary acts of reproduction that are (a) transient or incidental (b) an integral and essential part of a technological process (c) has the sole purpose to enable (i) a transmission in a network between third parties by an intermediary, or (ii) a lawful use of a work and (d) which has no independent economic significance.[24] This means that the functions of the media data space organisation relating to temporary acts of reproduction may be covered by this exception in order to intermediate copyright-protected works. Further, voluntary authorisations by rightsholders are generally provided through licenses or contracts. Where "data" is covered by the copyright *acquis*, existing licensing practices can inform the smooth exchange of data within a media data space. Where data is not covered, parties enjoy a wider margin of discretion under the freedom of contract, however, agreed governance rules under the media data space may provide concrete requirements.[25]

[19] Fixtures Marketing, paras. 39-40.
[20] Julie Baloup et al, 'White Paper on the Data Governance Act' (CiTiP Working Paper Series 2021, 23 June 2021), 9; Thomas Margoni et al, 'Chapter X of the Data Act and the Sui Generis Database Right' (*CiTiP Blog*, 14 June 2022) <https://www.law.kuleuven.be/citip/blog/chapter-10-of-the-data-act-and-the-sui-generis-database-right/> accessed 9 August 2022.
[21] Matthias Leistner, 'Protection of and Access to Data under European Law', in Jyh-An Lee et al (eds), *Artificial Intelligence and Intellectual Property* (Oxford University Press 2021).
[22] European Copyright Society, 'Opinion of the European Copyright Society on selected aspects of the proposed Data Act' (*European Copyright Society*, 12 May 2022) <https://europeancopyrightsocietydotorg.files.wordpress.com/2022/05/opinion-of-the-ecs-on-selected-aspects-of-the-data-act-1.pdf> accessed 7 June 2022.
[23] Matthias Leistner, 'Big Data and the EU Database Directive 96/9/EC: Current Law and Potential for Reform' in Lohsse et al (eds), *Trading Data in the Digital Economy: Legal Concepts and Tools* (Nomos 2018) ('Leistner 2018'); Timo Ehmann, 'Big Data auf unsicherer Grundlage - was ist "wesentlich" beim Investitionsschutz für Datenbanken?' (2014) Kommunikation und Recht 394; Mark J Davison and P Bernt Hugenholtz, 'Football fixtures, horseraces and spin offs: the ECJ domesticates the database right' (2005) 27(3) EIPR 113.
[24] Art. 5(1), Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (Information Society Directive); see also regarding the role of Art. 5(1) for text and data analytics: Thomas Margoni and Martin Kretschmer, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' (2022) 71(8) GRUR International 685, 691-694.
[25] A common European media data space should "put in place an appropriate governance structure to ensure fair, transparent, proportionate and non-discriminatory access to, sharing and use of data" (European Commission, 'Commission Staff Working Document on Common European Data Spaces' SWD(2022) 45 final, 3).

In both cases, the use and/or development of standardized, modular and machine-readable licenses is advisable. They could take the form of traditional "proprietary" licenses, but the opportunities of open licenses (e.g. Creative Commons) should be fully explored, especially in the light of the attention paid by the European Commission to open access, open science and open innovation. Overall, these means of authorisation provide participants of the media data space with a range of options for organising exchanges of data.

Thirdly, **technological protection measures** (TPMs) are used to protect the exploitation of different forms of copyright-protected content in a variety of contexts, including in the media sector. In this regard, TPMs provide rightsholders and participants in a media data space a tool to potentially achieve a level of technological control over their content. TPMs are empowered by a legislative framework that protects right holders against the circumvention of effective TPMs, however this protection is not without limits as clarified by the European Court of Justice in the *Nintendo* case.[26] Particularly, this framework regulates some situations where exceptions and limitations under the copyright *acquis* and TPMs clash. However, TPMs have been strongly criticised for inconsistency with international norms, being fragmented and for creating legal imbalances.[27] Looking forward, the market for TPM tools is expected to grow in the near future,[28] and the role of new technologies such as blockchain bear potential and limitations in resolving issues with the current TPM landscape.[29] Generally, the regulation of TPMs may be relevant in the protection of rights under the copyright *acquis* at the technical level of a media data space.

Box 1: Legal requirements take-aways

Data sharing requirements – media data space participants should consider:

- How to establish contractual balance within the media data space;
- How to ensure that data within the media data space is findable, accessible, interoperable and re-usable (FAIR);
- How to ensure that the media data space complies with the requirements for data intermediation services;
- The conditions for providing access to the media data space for third parties;
- Which interoperability requirements and standards to follow, and how to ensure their implementation;
- How to implement 'data control' and how to decide on the design of data sharing within the media data space via agreed mechanisms.

Copyright and related rights requirements – media data space participants should consider:

- Whether content data exchanged between participants will include subject matter covered by copyright and related rights;
- Where relevant, whether the protections of phonogram producers, of broadcasting organisations, or of press publishers vis-à-vis information society service providers apply;
- In the light of the potential heterogeneity of content exchanged, verify whether there are forms of collective management of rights;
- Whether metadata and usage data exchanged between participants will comprise rights management information under the copyright acquis;
- Whether participants consider certain data sets within the media data space as databases covered by copyright and/or the sui generis database right;
- Where relevant, which economic rights under the copyright acquis may be subject to (statutory or contractual) authorisation within the context of the media data space;
- Where relevant, whether and/or how technological protection measures may be implemented within the media data space.

[26] Case C-355/12 *Nintendo* [2014] ECLI:EU:C:2014:25.
[27] P Bernt Hugenholtz et al, 'Study on the implementation and effect in Member States' laws of Directive 2001/29/EC' (Report to the European Commission, DG Internal Market, University of Amsterdam IVIR 2007), 95-101.
[28] IMARC, "PPT- Digital Asset Management Market Growth, Demand and Challenges of the Key Industry Players" <https://zenodo.org/record/4744971> accessed 6 April 2022.
[29] Michèle Finck and Valentina Moscon, 'Copyright Law on Blockchains: Between New Forms of Rights Administration and Digital Rights Management 2.0' (2018) 50 IIC - International Review of Intellectual Property and Competition Law 77.

Governance requirements for the media data space

A data space is a complex structure that involves technical, legal, human and business parameters. In addition to technological enablers such as exchange protocols, common data formats or a reference architecture, a data space requires to develop trust to embolden partners into sharing information that could otherwise be considered too sensitive to be exchanged.

Consequently, a data space requires a governance structure. Data spaces' governance structures involve organisational, operational, business and technological building blocks. These elements are typically translated in a series of documents and structures such as a managing body, a data agreement, a consortium agreement and terms & conditions. In this decoument, a governance structure is defined as the series of institutions, organisations, tools and agreements that have been commonly agreed upon by a number of stakeholders to regulate the data exchange and sharing practices that take place on a specific dataspace architecture.

Achieving such ambitious and complex goal requires to develop a common knowledge among stakeholders. Given the limited literature available in this regard, two main types of architecture were identified:

- **Unstructured data spaces architectures**: built around a single-access point, it can be either altruistic or based on commercial agreements;
- **Structured data spaces architectures**: built around a single access point, it can be based on an **aggregation-anonymisation** architecture or on a **marketplace** architecture.

A sample of eight heterogeneous data space initiatives has been analysed[30], with the aim of defining a set of features for the definition of the data space governance. Moreover, from the analysis of different frameworks[31], the "Knowledge Commons" emerged as the most promising for the analysis of data space governance models. "Knowledge Commons" is both a paradigm (narrative to organise an observed system into an archetype) and a framework (tools and methodology on how to analyse the archetype). As a paradigm, the approach defines any knowledge-sharing initiatives based on common governance mechanisms and structures (either formal or informal) as a "Knowledge Commons". Simple, this definition is technology and industry agnostic — a perspective especially adapted to the current situation where the low number of data spaces initiatives does not allow for intra-industrial comparisons. As a framework, "Knowledge Commons" is an organised set of questions organised under 5 different themes to orient the researcher's efforts. The framework can be used both to analyse written material and as an interview guide.

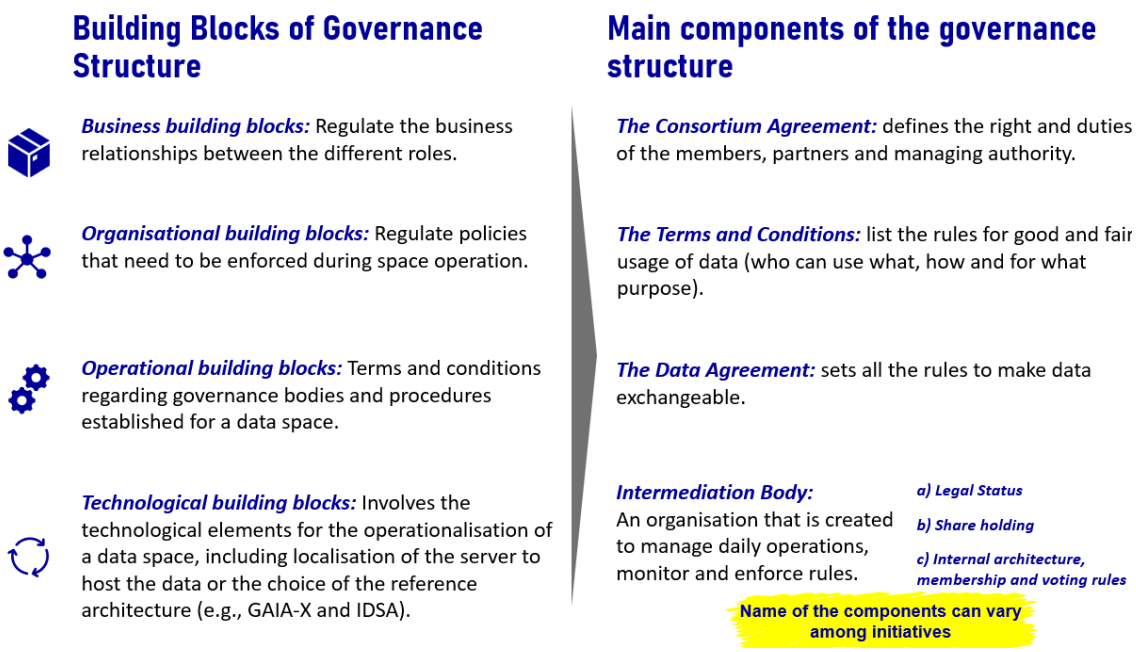
Based on this analysis, a set of key elements was developed for the definition of the media data space governance. These elements – which provide background knowledge that could be applied to any industry data space – are:

1. The most important building blocks to be considered for the construction of a governance model (i.e. "building blocks of governance structure") and the governance structure components that have to be developped. The building blocks include the different elements that are required to establish a functional data space, incuding technological

[30] The data spaces analyzed are the Mobility Data Space, the Health Data Hub, AgDataHub, Drive, The Swiss Media Hub, Skywise, iShare and ECLIC. The sample was selected for its diversity as these 8 initiatives cover 5 different sectors, are based on different architectures, follow different governance models and patterns, use different business models and have different geographies while all being European-led.
[31] The analysed framworks are the Towards European Health Data Space initiative, the Open-DEI framework, and the framework for the analysis of data governance developed by Abraham et all

(e.g. the reference arichtecture and exchange protocols) business (e.g. monetisation schemes), operational (governance bodies and procedures) and organisational (policies to manage the shared data and interactions between members). Some of these blocks will translate in concrete reality, such as the creation of server farm or the production of widgets to track participants' activities. However, any element involved in any of these blocks has important implications and has to be explicitly agreed upon by all partners. The technological infrastructure chosen for the data space, for example, is not as neutral as it might seem. The selection of a type of architecture over another will have important consequences in terms of how data can be shared, protected and activities monitored. Consequently, these building blocks translate in a series of official documents that turn all these elements into legal provisions that have to be consensually chosen and approved. For instance, the restrictions for data usage caused by the technological infrastructure would be specific in the data agreement. Both the components and the building blocks are detailed in Figure 1 below.

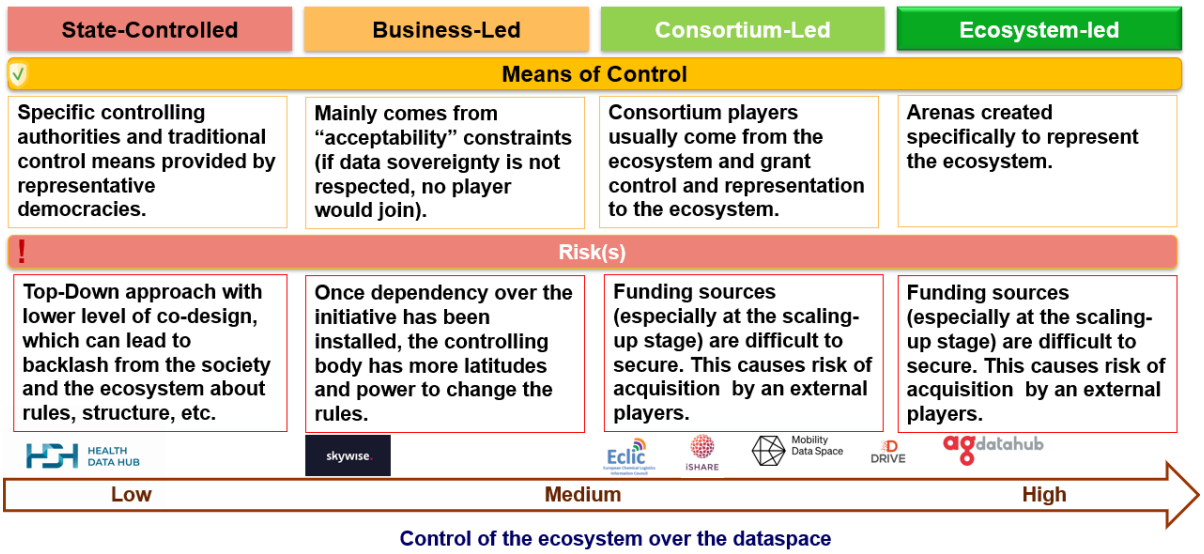
Figure 1: Building blocks and main components for the governance structure of a data space



Source: Developed by the project team based on OPEN-DEI initial list of building blocks

2. The **set of possible governance structures**. Based on the aforementioned components — especially the Consortium Agreement regulating the right and duties of participants and of the data space managing authority, and the caractheristics of this latter — the data space could assume one of the four possibile structures presented in Figure 2. Based on the distribution of rights among the participants and the managing authority, the ecosystem could have an higher or lower level of control. This typology was developed as a result from the case studies to summarise findings and provide some guidance to media stakeholders. Different ways of organising the data space exist, each one carry its own risks and benefits. An inactive ecosystem could leave the initiative for the creation of a data space to a large player, thus facing the risk of loosing control over its data assets. Contrastingly, an ecosystem-led initiative grants more control to the ecosystem, but this type of data space is also more financially fragile.

Figure 2: typology of governance structure for data spaces



Source: Developed by the project team

3. The **main revenues streams of data space business models**. In this regards, the revenue options available to finance a media data space are subscriptions to the data space, commissions to be paid based on what it is exchanged, and funding from public authorities and/or private players. Moreover, it will also be possible to combine these different funding options (See figure 3). As we can see from this typology, data spaces do not generate direct streams of income. Consequently, they depend of a participation fee that has to be taken either directly from participants or on each transation. This involves that the data space has to generate enough gains in terms of operational efficiency to ensure perennity of the initiative. Alterantively, a data space can benefit from the generosity of a sponsor such as the state or a large company (as Airbus did for the development of Skywise).

Figure 3: typology of business models

Business model to make the initiative sustainable (funding the data space).

1. ***Subscription***
    1. ***Simple Subscription (dominating model);***
    2. ***Freemium (ex.: Skywise);***
    3. ***Membership***
2. ***Commission-based:*** taking a fee per exchanged document (was considered for AGDataHub but abandoned).
3. ***Sponsored:*** The initiative is funded by either public authorities or private players (ex.: The Health Data Hub - for now).
4. ***Mix:*** *the norm is to combine different models together. Typically, sponsoring is needed at the launch and scaling-up and additional source of incomes are added afterward.*

Source: Developed by the project team

These elements were discussed with different categories of media stakeholders during co-creation events. These events allowed to develop conclusions and draft key clauses for a licencing agreement covering both the interaction between the data space "organisation" and

the "parties"[32], and bilateral interactions between parties. These conclusions are summarised as key take-aways in the box below.

Box 2 - Governance take-aways based on the interim report and interactive workshops

Regarding the **data space governance structure**:

- **Status of the managing body** – 73% of voting stakeholders favoured the future organisation to be a **"Not for profit";**
- **Mechanism to grant control over the initiative** – 53% of voting stakeholders favoured cross-participation and selection of data space participants as the preferred governance mechanism.
- **Initial coalition** – Large public and private broadcasters/publishers are considered the most important stakeholders to participate in the initial coalition (29% of votes). Moreover, as during the initial phase there is the risk of having participating parties interested in making the initiative fail, it is crucial to start with a smaller number of participants sharing a common interest.
- **Business model** – most stakeholders believe that a **model mixing (at least two of) subscription, commission and sponsorship revenues** is the most appropriate way of funding the media data space.

Regarding **parties' interests on data space architecture:**

- According to stakeholders, the types of content data with the highest potential for bilateral exchange on a marketplace were, in order, content usage data, descriptive metadata and metadata, archives, and collection of fact-checks.
- Regarding the data space architecture, even though the benefits from the marketplace architecture were not denied, the preference leaned toward an **aggregation-anonymization** type of data space focused on increasing participants' operational efficiency.
- A marketplace infrastructure (e.g., for content data and news content) for data exchange could be added to the data space but is not considered the priority by the ecosystem. However, it was also identified that a marketplace could offer a price-setting mechanism that could serve as leverage in bargaining with GAFAM.

Regarding **risks associated with anonymization:**

- No technique for anonymisation is entirely safe;
- The future data space organisation should clarify and inform participants about these risks;
- A data space should remain on top of technological trends and best practice for safety and security.

Regarding **business building blocks:**

- ***Exclusion and fraud rules –*** as the goal of the data space is to integrate all media players, exclusion should not be automatic. A range of sanctions proportionate to the problematic behaviour should be identified and put in place (e.g., exclusion from the use case, etc*).*
- ***Protection against malevolent participants –*** the data space should be structured to enable to withdraw the data from any participant from the aggregated set at any time. Protections against DDoS is important.
- ***Bridging the gap between private and public players —*** no clear consensus was reached on this point as two competing perspectives were identified. On the one hand, a data space built around concrete use cases (e.g., audience metrics) could bring together different players in one common data space. On the other hand, other stakeholders favour a more modest approach to bring together a small sub-set of players to then progressively extend the initiative to the ecosystem.

Regarding **organizational building blocks**:

- ***Auditing rules –*** the auditing body could be a third party, internal committee, or the organization itself. All these modalities have their own risks. However, the preferred option was the definition of an internal committee of expert named by the not-for-profit organisation, because of lower cost, higher transparency and lower risk of including malevolent parties.

Regarding **technological building blocks**:

- ***Functionalities –*** basic functionalities have to be provided by the data space (discovery, etc) and the architecture needs to enable participants to implement innovative solutions on their side (contacting IT companies, etc).
- ***Dynamic dimensions –*** interoperability and the expandability of shared data sets will create new problems and challenges over time. Procedures and mechanisms are needed to address them.

[32] **Organisation** – The body that intermediates the provision of data within the framework of a data space. Within a media data space, this is the institution that manages the exchange of data between participants and is subject to the agreed governance rules. **Party** – individual participant within the framework of a data space that either makes data available or seeks to access data via the data space. Within a media data space, this may be any type of player in the media sector.

Examples of data space applications in the media sector

To highlight the potential of a common European data space in the field of news and media, its application has been explored through the development of 11 media data space use cases. These use cases present possible problems/needs that can be addressed by sharing data with other players to improve their business using a data space. The presented list of used cases have different levels of maturity. Some can be tested in the short term, while others are more innovative and forward-looking, either for the types of problem they aim to address, or for the technologies to be deployed (e.g. applications 10 and 11).

The following table provides an overview these examples, which have been identified through a mix of desk research activities and interviews with relevant stakeholders of the media sector. For each of them, the table displays a brief description of the problem or need it aims to solve, a description of the data space use and the type of data flow.

Table 1 - Media data space use cases

Preliminary list of applications identified and mapped				
	Title	**Problem / Need**	**Description**	**Type of data flow**
1	Adaptive Content Recommendations	A media company has limited availability of usage data that are relevant to its own distribution channels. This limits the potential of algorithms and in general limits the ability to target audience with the right content.	The use case envisages a data space which focuses on sharing usage data specific to consumers, with the aim of improving predictive algorithms that suggest the type of content to be directly shared with the individual consumer.	Cross value chain
2	Shared archiving: enhanced information retrieval	Creating a unified archiving system completely alters the current paradigm whereby each entity has a federated archive locally stored. There is a major issue regarding change management and the shifting sharing paradigm that potentially could limit the applicability of this use case.	The use case approaches the data space with a wide scope of collectively creating a unified media data archive with a federated system that provides the adequate data sovereignty.	All (intra-sector and cross value chain)
3	Optimised Streaming	When streaming over-the-top (OTT) content and video on demand, systems do not typically recognize the specific size, type, and viewing rate of the video being streamed. Video sessions, regardless of the rate of views, are each granted the same amount of bandwidth. This bottlenecking of content results in longer buffering time and poor viewing quality.	The use case refers to a set of technologies used by mobile service providers to improve consumer viewing experience by reducing video start times or re-buffering events.	Cross-value chain
4	Copyright remuneration	Relevant efforts are devoted to ensuring that exclusive copyrights are properly enforced and remuneration goes to the correct right holders across the value chain.	The soft-infrastructure and the rules governing data sharing in a data space could provide an environment where tracking and usage policies over the shared data could be better enforced. Such monitoring of data flows and contents shared would apply to any type of content, from written and audio-visual, including, theatrical, musical and artistic works.	All (intra-sector and cross value chain)

Preliminary list of applications identified and mapped				
	Title	**Problem / Need**	**Description**	**Type of data flow**
5	Anti-piracy system	Copyright protection must be ensured in the whole European media sector. Fighting piracy and guaranteeing a control over your work are crucial determinants in order to avoid financial losses for the entire ecosystem. Moreover, a moral issue raises if right of attribution (writers claim authorship) and right of integrity (publishers are unwilling to accept modification of their work) are not recognized.	The use case envisages a data space to help in streamlining digital content monitoring. Automatic tracking and content recognition would be favored by the dataspace; sharing information on piracy threats and illegal IPTV detected, would help all media stakeholders to timely block illicit activities.	Cross-value chain
6	Audience Analysis	Media entities often conduct audience analysis to improve the quality and relevance of the content to be shared. A media data space based on sharing audience usage data would allow a media entity to increase the basis data which resembles a wider and more diversified audience pool.	The use case approaches the data space by collecting usage data collected by the single data providers and media conveying platforms in order to improve the ability of a single entity to address specific audience interest.	Cross-value chain
7	Translation of content	Media is produced at a national level in the respective language which thus limits the possibility of sharing information with international stakeholders.	The use case approaches the data space by creating a digital space where data is made available in a single language thus making it accessible to the whole European ecosystem.	All (intra-sector and cross value chain)
8	Fake-news detection	With the advent of digital information channels, there has been a growing concern regarding fact-checking and fake-news.	Creation of a single digital space where fact-checking is automated and where fake news are made visible to the whole ecosystem. This would ensure cooperation and coherence in the battle against fake-news to provide verified content.	Intra-sector
9	Addressable Advertising	Advertising represents a major income for media entities. By sharing audience profiles, conversion and attribution data, an entity would be able to improve potential profits by ensuring that ads are customer specific.	The use case envisages a data space which focuses on sharing usage data and advertising-related metrics to address challenges imposed by the removal of 3rd party cookies and mobile identifiers and to foster alliances between media companies for their advertising.	Cross-value chain
10	AI news assistant	In recent years, the amount of information and data published increased rapidly and media / news publishers are facing several challenges. The sector must take action to develop a business model that regenerates how content is created and delivered to the costumers.	The use case refers to an AI-integrated data space – AI News Assistant – that would support publishers in the creation of news stories by providing functionalities for story production and development, story curation and publication, and audience engagement.	Cross-value chain
11	A Media data space in the metaverse	Traditional media seem not to be able to keep pace with disruptive moves made in the Metaverse field by social media platforms. Among media services, journalism – that is suffering a	The use case envisages a dataspace to develop a coherent architecture and tackle simultaneously: integration, standardization, harmonization	All (intra-sector and cross value chain)

Preliminary list of applications identified and mapped				
	Title	Problem / Need	Description	Type of data flow
		profound crisis – could particularly benefit from the Metaverse. However, stakeholders raise concerns around data compliance and standards.	and the facilitation in the data exchange.	

Conclusions

The analysis of the legal framework in which the data space is going to operate as well as its governance model and structure are two key aspects to be considered for the development of a data space in the media sector. With regard to the first, the media data space (MDS) will have to comply with the data sharing legislative framework, and in particular the data sharing requirements will have to be in line with both the Data Governance Act and the Data Act. Moreover, participants will need to understand which data are protected by copyright and related rights, and the extent to which other legal means (e.g. database rights, collective management of rights…) and technological protection measures could be applicable to the exchangeable data. A synthetic overview of the key take aways related to the legal requirements is presented in Box 1.

The document also shares a number of concrete inputs for the governance requirements. These have been defined studying other data space initiatives and then validated through a series of workshops. The key findings mainly concern the data space architecture and the governance structure: most stakeholders expressed their preference for an aggregation-anonymization data space governed by a non-for-profit organization, sustained by a mix of subscription, commission, and sponsorship revenues. Furthermore, the MDS governance should foresee control policies, based on auditing rules and exclusion and fraud rules. A synthetic overview of the governance requirements is presented in Box 2.