

Contents

Executive Summary	2
Introduction	3
Eight Trends in Data Economy	4
From Raw Data to Higher Value Data Products	5
From Licensing of IPRs to Provision of Services	5
From Property Focused Licenses to Context Aware Usage of Data	6
From Clear-cut Layers to Hybrid Technologies	7
From Individual View to Ecosystem Dynamics	8
From Limitless Use to Regulated Use of Data	9
From Open Data to Data Spectrum	10
From Siloed Data to Managed Data Sharing	11
Recommendations for the Stakeholders	12
For SMEs	12
For Supporting Organisations	12
For Policy Makers	13
For Data Spaces	13

Executive Summary

This policy brief highlights legal and business-related learnings from our research on open assets. The learnings include aspects on value creation in the era of data servitisation, mechanisms embedded in hybrid technologies of open assets, dynamics within open data ecosystems and data spectrum, and data management and licensing practices. They are based on our work relating to open assets – open-source software, open data, and open hardware – under the ‘3Os and IP awareness raising for collaborative ecosystems’ (ZOOM) Project.

We offer recommendations aimed at practically all ZOOM stakeholders. Specifically for policy makers we argue that the implementation of the current data regulations should be monitored closely. Agile attitude would be needed in case negative effects on data related value creation take place, especially against the backdrop of contrasting and concurring global phenomena.

Data policy should continue to boost the creation of services for data sharing and intermediation. Healthy competition and level playing fields are key to achieving this as well as investment in upskilling entrepreneurs. The open source software has enabled mechanisms and dynamics to generate heterogenous business models based on community-centred innovation. Data policy should be informed by and consulted with these communities. However, caution must be exercised as these practices cannot simply be copied from open source software to, for example, shared data. Data policy should aim to achieve similar effects by different means. Examples of lessons learnt include how to implement transparency for AI, how to create a nourishing landscape for building communities, and how to enable development of business dynamics.

Introduction

The purpose of this policy brief is to highlight legal and business-related learnings from our research on open assets, i.e. open-source software, open data, and open hardware. What can a concurrent data economy learn from open data and other open assets? The learnings include aspects on value creation in the era of data servitisation, mechanisms embedded in hybrid technologies of open assets, dynamics within open data ecosystems and data spectrum, and data management and licensing practices. They are based on our work relating to open assets – open-source software, open data, and open hardware – under the ‘3Os and IP awareness raising for collaborative ecosystems’ (ZOOM) Project.

The policy brief discusses these legal and business perspectives under the following eight trends of data economy:

1. From raw data to higher value data products
2. From licensing of intellectual property rights (IPRs) to provision of services
3. From property focused licences to context aware usage of data
4. From clear-cut layers to hybrid technologies
5. From individual view to ecosystem dynamics
6. From limitless use to regulated use of data
7. From open data to data spectrum
8. From siloed data to managed data sharing.

These are trends that have been visible for some time now – some longer, some for a shorter period. They encompass changes taking place in several areas: political, legal, economic, social, and technological. These trends are likely to be continued and have implications for different stakeholders, the implementation of the regulations and future policies.

Eight Trends in Data Economy

In discussions on data, we often refer to a wide range of assets: raw data, quantitative data, qualitative data, aggregated data, data records, datasets, and databases. From business and legal perspectives there are clear differences between these categories and in the implications they hold.

In addition to addressing data as an asset, it should also be addressed from the perspective of **intellectual property rights**, i.e. the exclusivity granted to original, human, intellectual input. Intellectual property (IP) contains creations like literature, paintings, images, designs, software, and inventions. And when talking about IPRs granted for such creations we refer for instance to copyrights, patents, trademarks, and database rights.

The above aspects – efforts and intellectual input invested in data as a business asset and as legally protectable intellectual property – form the basis of products and services built on data, all part of the components of concurrent data economy.

In the following, we address our research findings relating to **open assets** from the business and legal perspectives under the headlines of eight trends present in our concurrent data economy. With open assets, we refer to **open-source software, open data, and open hardware**. What should we learn from open assets in terms of value creation in the era of servitisation; mechanisms embedded in hybrid technologies; dynamics within ecosystems and data spectrum; and data management and licensing practices?

ZOOM Deliverables:

D1.1 'Literature review of legal cases in open software, hardware and data'

D1.2 'Licensing framework for choosing the degree of openness of a process, product or service based on constraints of the business model'

D2.1 'Literature review of business cases in 3Os'

D2.2 'Case study report'

D2.3 'Innovation ecosystem report'

From Raw Data to Higher Value Data Products

“Indeed, companies must not rely on the size of their data – it is not useful unless it is applied in an intelligent manner. Therefore, the volume of data is of little importance, since internal data must be combined with external data in order for a company to obtain the most out of its data.”¹ Monino (2021)

Opening or sharing data as one-off datasets does not automatically lead to value creation. **Value needs to be built on top of data.** To give some examples, data also needs to be useful, usable, cleaned, technically and legally accessible, and it must be matched by investments in information, metadata, software, quality management, data analytics capabilities, and social tools that can cultivate the ecosystem around the open data.²

Data as a non-rival resource gets its value in the use in which the data recipient or end-user puts such data. Terms like data product and data-as-a-service emphasize these aspects. For instance, the shift from separate sets of data assets towards data products that may be jointly formulated between data providers and data recipients encompasses more potential to value creation.

This sort of change in the usability and quality of data also has legal implications. Instead of licensing particular data or selling a single data set, business is created by selling good quality of data or real-time access to data. In other words, selling “immediacy and accuracy” by adding value generated by human effort on top of data.³ Mere licensing of protected or unprotected IP is not sufficient anymore. Instead, when the focus shifts towards the value creation mechanisms, the particular field of technology and the dynamics existing within the ecosystem of actors and the community get more emphasis. The future licensing models – value-focused licensing – need to be built on these specific features and dynamics.

From Licensing of IPRs to Provision of Services

The trend towards data-as-a-service gives an additional aspect to the variety of forms the provision of data takes. Through this trend, the provision of data becomes less linked to the rigid licensing of distinct categories of IPRs and puts the aspect of the end-user in the forefront. The higher one looks at the steps of productisation or servitisation, the less weight is given for possible IPRs as such and the more to the value that the end user gets from the data.

¹ Jean-Louis Monino, ‘Data Value, Big Data Analytics, and Decision-Making’ (2021) 12 Journal of the Knowledge Economy 256 <<http://link.springer.com/10.1007/s13132-016-0396-2>> accessed 16 March 2023.

² Serdar Temiz and others, ‘Open Data: Lost Opportunity or Unrealized Potential?’ (2022) 114 Technovation 102535 <<https://linkinghub.elsevier.com/retrieve/pii/S0166497222000827>> accessed 16 March 2023, with its references.

³ Heather Meeker, ‘Beyond Open Data: The Only Good License Is No License’ [2022] PLI Chronicle: Insights and Perspectives for the Legal Community <[https://plus.pli.edu/Details/Details?fq=id:\(352066-ATL2\)](https://plus.pli.edu/Details/Details?fq=id:(352066-ATL2))>.

Initially, the digital economy was shifting the ownership-based economy and physical goods towards licences to IP and fully digital content. For some time now, there has been an increasing trend towards providing digital content as a service. If the content – or data – is provided based on subscription as a service, compared to pure IPR-licences, the usage rights are granted with additional limitations. An example of this is the limitation of the duration of the services. Additionally, the content itself may change dynamically in time. In such cases, the user gets a usage right to the content, however that is not necessarily based on any underlying IPR but is instead a mere contractual commitment. Service providers can build services, irrespective whether there is an underlying IPR. With this trend, also data becomes the object of trade as a service and is no longer commercialised through licensing of databases or similar IP.⁴

Examples:

Examples of this trend of digital business towards subscription of services are cloud services (such as iCloud, Google Drive) and streaming of video and audio content (such as Netflix, Spotify). This trend affects directly how data is processed and consumed in business as well, we have seen for instance services like Azure and AWS, which are widely used.

From Property Focused Licences to Context Aware Usage of Data

The mechanisms embedded in open-source software licensing cannot be copied as such to the licensing of open or shared data. Still, copy-pasting of ill-fitting licence practices often happens. This is since, in many cases, data lacks the element of intellectual input – a lot of data is created automatically. Therefore, the balancing act between revealing the IP to the public (such as a work of art or an invention) and granting of IPR-protection (such as copyright to a work of art or a patent to protect an invention) does not work with regard to data.

One example of problems emerging from the misuse of open-source software licences to data licensing is that mere notice requirements can create obstacles as datasets and records are not used and built in the same way as software packages. Data interacts with software differently than software interacts with other software, this is especially apparent in the case of AI and ML.⁵

During recent years, there have been several initiatives aimed at tackling the problems generated by the use of open-source software licences in open data licensing. These initiatives approach data licensing from the perspective of certain technologies and focus on how the data is actually used within such technologies. **These context-specific licences can make the environment of data licensing clearer and more predictable.** This could lead to more standardised practices affecting the data economy as a whole.

⁴ Hans Graux, 'Data Sharing as a Service: Will Data Services Remove Intellectual Property Rights from the Picture, and at What Cost?' (Publications Office of the European Union 2021) <<https://data.europa.eu/doi/10.2830/815190>> accessed 20 February 2023.

⁵ ibid 3.

Examples:

Montreal Data License Tool addresses the use of data in the field of artificial intelligence (AI) and machine learning (ML).⁶

Cryptographic Autonomy License addresses the use of data in connection with open-source software relating to blockchain technology.⁷

From Clear-cut Layers to Hybrid Technologies

Novel technologies do not always contain as clear-cut layers as we would like to see from the legal or business perspective. **One example of this kind of novel hybrid technology is AI and ML technologies, in which data and software, on one hand, and computer and human, on the other hand, interact in more and more complicated ways.** The hybridity of the technology is also reflected in the hybridity of the intellectual property embedded in the technology.

For example, a dataset used for training an LLM can contain copyrighted material. This is the case for many publicly available datasets⁸. In addition, data or data embedded in datasets or databases, may be subject to other rights, such as rights relating to personal data protection, image rights (e.g., name, photo, and likeness), publicity rights (e.g., misappropriation of name, photo, voice)⁹.

This kind of hybrid technology leads to a system in which different types of data – with different levels of protection and rights – are embedded in datasets and databases, then processed to further data, datasets, databases and even copyrighted work. The system becomes difficult to understand and manage.

The difficulty becomes visible for instance when assessing the definition of open source AI that would aim for a similar type of dynamism as we can see taking place in open-source software communities. In this respect, we should pay attention to the effects the components of AI create and how those components interact with each other. This kind of effects-based, as opposed to rights-based, approach should be considered for defining open source AI. Similar approach can be used in the strive for other types of dynamics in systems encompassing hybrid technologies.

Further reading:

ZOOM Policy Brief #1 'Open Source AI – Building Blocks for a Definition'

ZOOM Policy Brief #3 'Free Software and AI openness: Overcoming challenges in the licensing world'

⁶ Misha Benjamin and others, 'Towards Standardization of Data Licenses: The Montreal Data License' (arXiv, 20 March 2019) <<http://arxiv.org/abs/1903.12262>> accessed 2 November 2022.

⁷ Arthur Brock, 'Understanding the Cryptographic Autonomy License' (Holochain, 22 February 2019) <<https://medium.com/holochain/understanding-the-cryptographic-autonomy-license-172ac920966d>> accessed 26 February 2023.

⁸ Gopi Krishnan Rajbahadur and others, 'Can I Use This Publicly Available Dataset to Build Commercial AI Software? -- A Case Study on Publicly Available Image Datasets' (arXiv, 11 April 2022) <<http://arxiv.org/abs/2111.02374>> accessed 2 November 2022.

⁹ ibid 6.

From Individual View to Ecosystem Dynamics

The trend of servitisation opens new possibilities for building business upon data and highlights the need for novel business design and business models. **It also highlights the need for data sharing between entities and governance mechanisms for such data sharing, for instance through data ecosystems.** Provision of data-as-a-service and the emergence of data ecosystems demonstrate data as a resource with non-rival nature and set the focus to the flow of data between entities.

Traditional “one-way street” open data systems are transitioning towards open data ecosystems. These ecosystems not only prioritise data accessibility but also consider the broader context for the utilisation of data. Within such open data ecosystems, data is not a one-time transaction but is rather continually circulated among the stakeholders. Intermediaries, acting as bridges between data providers and users, play a vital role in open data ecosystems by creating additional value.¹⁰

To navigate the complexity of applying innovation as a business strategy and model, the differences among the types of ecosystems need to be recognised. Clarifying the distinctions among various types of ecosystems is not merely an academic exercise; it is a practical necessity for those seeking to employ innovation as a strategic tool. For open assets, these ecosystems form a layered, multilevel structure, expressed in ZOOM Multilevel Ecosystem Framework.

Further reading:

Van Loenen et al. (2021)¹¹ and Kamarioutou & Kitsios (2022)¹² address open data ecosystems (OD ecosystems) from the perspective of open data. Runeson et al. (2021) broaden the concept of open data ecosystems (ODEs) to also cover other areas of data than mere open data.

¹⁰ Bastiaan Van Loenen and others, ‘Towards Value-Creating and Sustainable Open Data Ecosystems: A Comparative Case Study and a Research Agenda’ (2021) 13 JeDEM - eJournal of eDemocracy and Open Government 1 <<https://www.jedem.org/index.php/jedem/article/view/644>> accessed 16 March 2023.

¹¹ *ibid.*

¹² Maria Kamarioutou and Fotis Kitsios, ‘Bringing Digital Innovation Strategies and Entrepreneurship: The Business Model Canvas in Open Data Ecosystem and Startups’ (2022) 14 Future Internet 127 <<https://www.mdpi.com/1999-5903/14/5/127>> accessed 17 September 2023.

¹³ Per Runeson, Thomas Olsson and Johan Linåker, ‘Open Data Ecosystems — An Empirical Investigation into an Emerging Industry Collaboration Concept’ (2021) 182 Journal of Systems and Software 111088 <<https://linkinghub.elsevier.com/retrieve/pii/S0164121221001850>> accessed 17 September 2023.

From Limitless Use to Regulated Use of Data

From the regulatory perspective, data is currently in the spotlight. The so-called Big Five regulations (Data Markets Act, Data Services Act, Data Governance Act, Data Act and AI Act) are in different phases of implementation. And we should not forget the GDPR, General Data Protection Regulation, setting the European standard for handling personal data. The themes pushed forward with these acts include among other issues data protection, fair competition in the digital markets, obligations for platforms and gatekeepers, obligations for service providers and data intermediaries, re-use of public sector data, data altruism, IoT-data, data spaces and AI transparency and risk assessment. **These regulations will profoundly affect the boundaries on how to conduct data-based business and give incentives for novel types of business in the European data economy.**

The Big Five regulations will also give a boost for open data by affecting the business environment in which data flows. At the same time, the systemic challenges embedded in data sharing and licensing between entities, and on a more general level, within the ecosystems and communities, surface. These challenges contain issues like interoperability, APIs and connectors, access control, data governance and metadata, value generation, regulatory aspects, contractual issues and licensing, and ecosystem building. Due to the multidisciplinary nature of these challenges, they can be tackled only by addressing them from several perspectives: e.g., from technological, business, legal and political.

Learn from open data:

Some of the forerunners in data sharing can be found in public sector data, governmental data and open science. For example, Open Data Directive (2019), formed on the basis laid down by the Public Sector Information Directive (2013), emphasises dynamic data sharing and issues like data access and provision of APIs. For instance, APIs are required for high value data sets like geospatial data, environmental data and mobility data. Typically, these datasets contain data that is not covered by copyrights or database rights. Again, this emphasises the trend of servitisation of data instead of relying on mere licensing of the underlying IPRs. Similar regulations can also be found from other sectors like banking, energy, and automotive industry.¹⁴

¹⁴ Hans Graux, 'Data Sharing as a Service: Will Data Services Remove Intellectual Property Rights from the Picture, and at What Cost?' (Publications Office of the European Union 2021) <<https://data.europa.eu/doi/10.2830/815190>> accessed 20 February 2023.

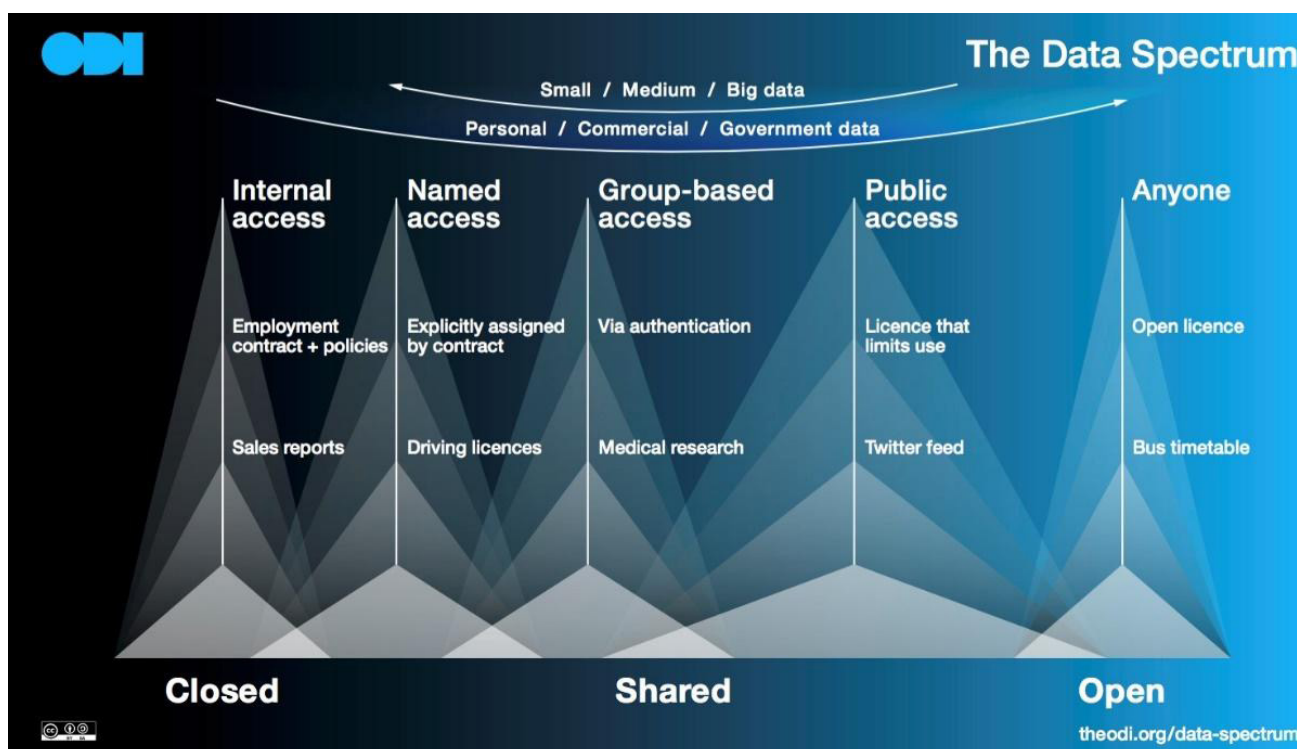
From Open Data to Data Spectrum

The concept of data spectrum gives a good perspective to the business use of data. Business can be created based on different types of data and different types of data can be used simultaneously. To operate in such a business environment, a company needs to understand different spectral areas of data and business opportunities afforded by different types of data.

ODI (Open Data Institute) has visualised this continuum as a data spectrum. In its one end, there is closed data that is accessible internally within a single entity. On its other end, there is the open data licensed under open data licence terms to anyone. In between these extremes there is shared data that can vary from named access, to group-based access and public access. In the data spectrum, the level of openness increases step by step.¹⁵

The concept of data spectrum is especially important for the common European data spaces¹⁶ that also address the whole continuum of data, in addition to the spectrum of sectors and themes they cover. In this work, ODI's sectoral visualisations of the data spectrum can be used to structure these continuums¹⁷.

Data Spectrum:¹⁸



¹⁵ 'The Data Spectrum' <<https://www.theodi.org/about-the-odi/the-data-spectrum/>> accessed 26 February 2023.

¹⁶ Common European Data Spaces | Shaping Europe's Digital Future' <<https://digital-strategy.ec.europa.eu/en/policies/data-spaces>> accessed 5 March 2024.

¹⁷ ibid 15.

¹⁸ ibid.

From Siloed Data to Managed Data Sharing

There is plenty of uncertainty embedded in the sharing of data between entities, and this acts as a hindrance for the use of open or shared data in business. Uncertainty for the use of data stems from several sources:

- There may be a lack of understanding that data has a different nature than copyrighted work or other IP. This leads to unclarity on how data should be passed on to others for reuse. For instance, some of the open licences covering data also cover software, some of them cover databases, and some are designed solely for open-source software licensing.¹⁹
- Combining and mixing different types of data and compliance requirements from regulations and licensing schemes emphasise the need for coherent management of data. Uncertainty creates waste of resources as the actors need to investigate ambiguous licensing terminology, it causes market and incentive imbalances, contractual breaches, and risks to privacy.²⁰
- Additional complexity is introduced in case the data contains personal data protected by regulations like the European GDPR. In these cases, balancing of the risks and harms relating to opening or sharing data and the value that can be generated by the flow of such data becomes essential.²¹

Tackling the uncertainties caused by the systemic challenges of the new data economy requires data management and compliance. In forming data management, licensing processes and compliance for tomorrow's data business, elements and best practices from existing governance and compliance frameworks can be used. These should be developed further and all layers – internal data management, bilateral data sharing and multilateral ecosystem governance – need to be addressed.

Resources:

Only a few of the resources focus specifically on data, most of them focus on open-source software compliance. However, also these frameworks can be used for learnings to be taken into account for open and shared data compliance processes:

- Open Source Compliance in the Enterprise - Handbook
- Open Data Goldbook for Data Managers and Data Holders .

¹⁹ ibid 3.

²⁰ ibid 6.

²¹ Publications Office of the European Union., Capgemini Invent. and European Data Portal., Open Data and Privacy. (Publications Office 2020) <<https://data.europa.eu/doi/10.2830/532195>> accessed 30 November 2022.

Recommendations for the Stakeholders

For SMEs

Remember that regulation does not only set boundaries or limitations for business activities; it also provides possibilities for new and disruptive types of business and higher value data products and services. For instance, the Data Act emphasises the role of vibrant service provider ecosystems. The regulations set the basics of the operating environment, but it remains for an SME to actively build on top of these opportunities.

It is crucial for you as an SME to have a deep understanding of your own technology. However, this is not all, you also need to understand your operating environment, both in business and in legal sense. You should also understand how ecosystems work in your business environment.

Have you already explored all possibilities relating to open data? The utilisation of open data in connection with your own, closed data can be multiplied with shared data. Data sharing environments and data spaces widen the sharing of data to new types of organisations and open new business opportunities based on trust and security.

ZOOM tools you should take a look at:

Training modules for relating to open data and ecosystems

For Supporting Organisations

Hybrid technologies as combinations of software and data will have a profound effect on business and licensing practices. Keeping up with the development of such hybrid assets, for example AI and ML, and related context aware usage practices is essential to be able to support business entities.

There are several tools available that help you to design and create business on top of open-source software (see ZOOM tools). Also consider the possibilities that are embedded in layered ecosystems relating to data and the whole spectrum of data (closed, shared and open). Operating with this kind of versatile set of assets and multidimensional environments introduces complexities, but also vast opportunities for business generation.

Challenge remains on how to find and address all value creation possibilities embedded in new business models relating to the vast spectrum of data. In addition, companies will need support in the participation and creation of communities around open assets. And there is also a demand for generating managed data sharing practices and tools that enable licensing compliance and data management.

ZOOM tools you should take a look at:

Company profiles

Training materials

Supporting organisations materials

For Policy Makers

The implementation of the current data regulations should be monitored closely. Agile attitude would be needed in case negative effects on data related value creation take place. At the same time, it should be recognized that there are several contrasting global phenomena taking place at the same time.

Continue to boost the creation of services for data sharing and intermediation. Healthy competition and level playing fields are keys to achieving this. Also investing in skills and courage for entrepreneurship is needed in Europe.

Open-source software environment has proven mechanisms and dynamics to generate heterogenous business based on sharing and communities built around it. Learnings can be taken from these communities. It should be noted that these practices cannot be copied as such to another field, e.g. from open-source software to shared data. However, the learnings from different fields can be used to aim for similar effects by different means. Examples of these learnings include how to implement transparency for AI, how to create a nourishing landscape for building communities, and how to enable development of business dynamics.

ZOOM tools you should take a look at:

Policy briefs

For Data Spaces

A large number of data regulations has been introduced in Europe. These set out the basis also for data spaces. Regulations impose restrictions, but also open new opportunities for data spaces and business perspectives for data space participants. Several other trends can also boost the operations of the data spaces, e.g. movement towards data services, increase in the role of the ecosystems, and the need for managed data sharing.

Data spaces are core implementers of the European data strategy aiming at the single market for data. This is a demanding role. Even if the target of data spaces is in the future, there are several learnings from the past that should be considered when designing data spaces. Data spaces can learn for instance from open-source software licensing models in creating similar dynamics, the role of communities in open-source development, good practices generated for open data, and recognition of the hybridity involved in novel technologies.

In the data spaces, software and data are in constant interaction. The software components of data spaces will be open-source (Simpl-Open) and these software components will interact with the whole spectrum of data (closed, shared, and open). As software and data will both be shared multilaterally between several entities, due emphasis should be placed on ecosystem and community building, leading to a flourishing landscape of service providers.

ZOOM tools you should take a look at:

Guide to Business Model Innovation for Open Assets