European Commission

Study on technological and economic analysis of industry agreements in current and future digital value chains

FINAL STUDY REPORT

November 2021

A study prepared for the European Commission
DG for Communications Networks, Content & Technology
by:

CARSA ECORYS KU LEUVEN VDI|VDE|IT

Digital Single Market

European Commission

This study was carried out for the European Commission by

Consultores de Automatización y Robótica, S.A.

Ecorys Europe EEIG GEIE

Katholieke Universiteit Leuven

VDI/VDE Innovation + Technik GmbH

This final report was written by Pär Weström, Audren Layeux, Ieva Savickaite and Marco Graziado (CARSA); Andreas Pauer, Luca Turturro, Ann Kristin Gross (Ecorys); Natalie Bertels, Arina Gorbatyuk, Bert Peeters (KU Leuven); Haifa Rifai, Michael Preuß (VDI/VDE IT); Oscar Lazaro (Innovalia Association)

Internal identification

Contract number: LC-01167726

Subject: To carry out a study on technological and economic analysis of industry agreements in current and future digital value chains.

DISCLAIMER

By the European Commission, Directorate-General of Communications Networks, Content & Technology.

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.

ISBN: 978-92-76-43633-1

doi: 10.2759/495071

Digital Single Market

Table of Contents

List of Tables

List of Figures

List of acronyms and abbreviations

ADR	Alternative dispute resolution
AI	Artificial Intelligence
API	Application Programming Interface
CAGR	Compound Annual Growth Rate
CVS	Common Vocabulary Standards
DVC	Digital Value Chain
EBIT	Earnings Before Interests and Taxes
FDI	Federated Digital Infrastructures
FRAND	Fair, reasonable and non-discriminatory
GDPR	General Data Protection Regulation
IA	Industry Agreement
IAA	Identification, Authentication and Authorisation
ICC	International Chamber of Commerce
ICT	Information and Communication Technologies
IDSA	International Data Spaces Association
IIRA	Industrial Internet Reference Architecture
IoT	Internet of Things
IP	Intellectual Property
IPR	Intellectual Property Rights
IT	Information Technology
MRO	Maintenance, Repair and Overhaul
P&P	Plug & Play
PPP	Public–private partnership
RAMI 4.0	Reference Architectural Model for Industrie 4.0
RCA	Relational Contractual Agreements
SEP	Standard Essential Patent
SSOs	Standard-Setting Organisations
T&L	Transport and Logistics

Executive summary

This study was commissioned by the European Commission to support the development of industry agreements as a means to accelerate the digitalisation of European industries and in particular the deployment of industrial data spaces. The overall objective of this study is to provide recommendations and guidance to help industry and policy-makers to further advance industrial digital ecosystems. Thriving digital ecosystems require the development and adoption of shared norms by a critical mass of industry players in order for these rules to become genuine standards. To reach this goal and to create new markets and market opportunities, European industries need to agree on the functions and interfaces for the platforms, reference architectures and interaction protocols that will support the growth of ecosystems. The development of streamlined industrial data ecosystems is necessary to ensure that every business in Europe, whatever the sector, can reap the benefits associated with digital innovation.

"Industry agreements" (IAs) have a key role to play in the development of streamlined and integrated data spaces, as they are designed to establish a common understanding of functionalities, architectures and specifications.

This study builds on the findings of two interim reports that were produced internally. The interim reports identified **eight sectors** in which the implementation of industry agreements have the potential to create high economic impact: mechanical engineering; aerospace; chemicals; (other) manufacturing; transport and logistics; agriculture; health and energy. A list of current and planned IAs in each of these sectors has been compiled.

The findings from value chain analysis work and interviews conducted with key sectoral stakeholders have enabled the research team to identify **eight innovation areas** (one for each sector). These innovation areas can be defined as "areas in which industry agreements have a high potential to create new markets and market opportunities". Based on the results of eight digital workshops, these areas were then clustered into three **areas of innovation that could potentially generate interesting opportunities with a high economic impact**:

1. Data sharing/exchange;
2. Plug & Play/interoperability;
3. Advanced data analytics and Artificial Intelligence.

Together, these three areas cover aspects such as interchangeability, interoperability and processability, all of which are considered to be essential characteristics of a data space and a robust and data-driven European economy.

Each of the three innovation areas provide opportunities for the target sectors, their stakeholders and the European economy to achieve significant economic gains. **Action is needed to support and encourage the development of these three areas in order to drive the transition to digital leadership and strengthen the global competitiveness of European industry.**

However, **the importance of each area and the degree to which they should be prioritised can vary from one sector to another**. In mechanical engineering, for example, the priority is to develop interoperable machine tools to increase operational efficiency, reduce integration costs and generate new business opportunities. By contrast, in the chemical industry, the development of exchangeable formats for chemical molecules and reactions is the mandatory prerequisite for the modernisation of the industry.

In the study, the research consortium also outlined:

(i) an analysis of the **barriers** that prevent the development and implementation of an industry agreement together with possible pathways for their removal;
(ii) tangible evidence of **opportunities**, demonstrating how sectoral agreements can help make European industries more competitive;
(iii) **high-level specifications – "industry agreement templates"** – including technical and legal clauses; and

(iv) a number of **recommendations on how to** remove barriers and **unlock opportunities** to develop sectoral agreements that can produce a high economic impact.

The **barriers** to establishing a robust and successful data ecosystem are typically high. They encompass a wide variety of challenges from technical to legal and economic.

Technical challenges include fundamentals such as assessing data quality and accuracy, achieving semantic and technical interoperability, and ensuring that ecosystem participants have access to data. Creating a data ecosystem also means building a reliable and collaborative network of many diverse actors that are willing to work together and are committed to achieving similar or shared objectives. To facilitate this work, processes, architectures and IT languages require harmonisation. There are also a number of **legal barriers** that need to be addressed. Legal issues, which can for instance be related to intellectual property or the need for control over data, are often considered to have the potential to slow down or even prevent the adoption of industry agreements. This is the case, for instance, with issues such as competition law and liability law. Competition issues can arise where there are problems accessing data (in particular, the failure to provide access) or in relation to evolving standards, whereas liability for damage due to inaccurate data is a recurring issue in all sectors.

These barriers make the development of industry agreements more complex; however, they are also the main justification/incentive for IA development. However, in addition to these barriers for the development of an industrial data space, there also barriers specific to the development of new industry agreements.

From an **economic point of view**, the study identifies a range of barriers that inhibit the development of industry agreements. The structure of an industry or its value chain segments can influence the potential for IA development and implementation. This is most evident where an industry structure is characterised by a prevalence of SMEs (e.g., mechanical engineering, agriculture). The **complexity of the technical aspects** addressed by an IA can also be considered a significant barrier, as well as the **complexity of the regulatory environment** in which an IA operates. The latter holds true particularly when regulations are not harmonised across borders, as a multiplicity of rules at different levels can create considerable uncertainty. The nature of competition and the fear of losing a competitive advantage when disclosing business information are also important factors and potential barriers. IA development can be further complicated by a lack of trust among stakeholders on key aspects, such as how data will be (re)used, the integrity of the systems underpinning IA implementation, or on the definition of the roles, processes, business ethics and general principles that apply to data sharing.

Apart from exploring the barriers to the adoption of IAs, and providing suggestions on how to remove them, the research team also evaluated a range of opportunities in the data economy to develop high-impact IAs. **In each of the three clustered areas, the most significant opportunity has been identified.** For each of these main opportunities, sector-specific opportunities were assessed and IA guidelines were produced. They provide strategic tools and frameworks that are designed to enable and facilitate the development of industry agreements to deliver key economic benefits and unlock industry opportunities in each of the three innovation areas.

Table 1: Innovation areas and main opportunities

Innovation Areas	Main Opportunity	Example of recommendations/specific opportunities
Data sharing/exchange	Development of a data asset value exchange mechanism	Development of “Clearing Houses” to act as intermediaries in the new data economy.
Plug & Play/interoperability	Development of an interoperability framework by mapping standards and closing the gaps	Development of a discussion forum and funding of research programmes for the development of smart farming standards.
Advanced data analytics and Artificial Intelligence	Increase of data quality for advanced analytics and artificial intelligence	Increasing the importance of the “Smart Grid Taskforce” to enable this institution to play a leading role in the definition of data quality standards.

In addition, this report provides **3 IA templates — one for each innovation area**. These templates provide a series of legal and technical specifications to provide a guideline for policy-makers and industrial players that are interested in developing industrial data spaces. Each template features a series of legal and technical clauses that can be discussed and agreed upon during the IA definition process.

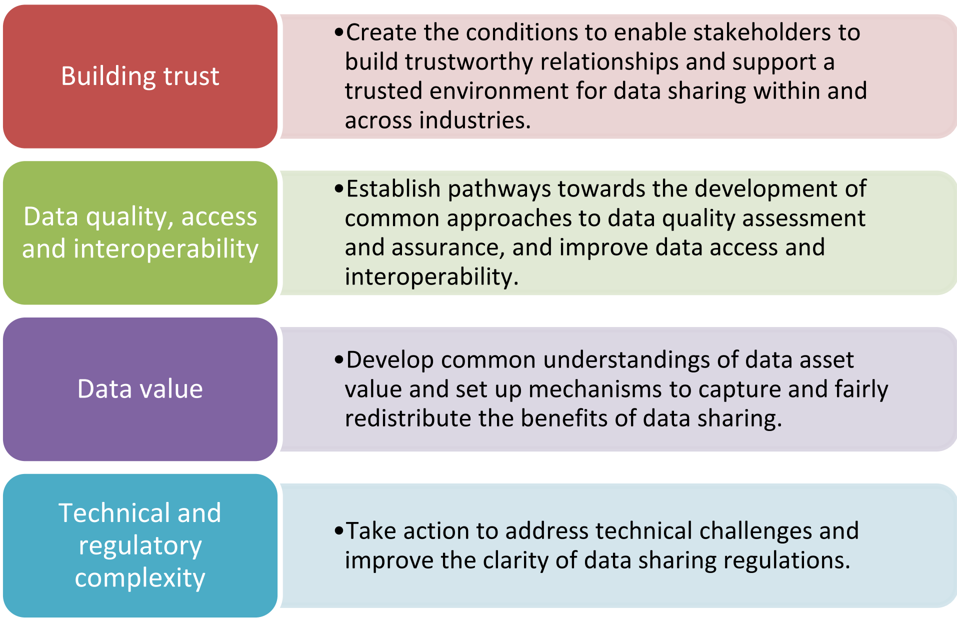
Taken together, these three agreements address the 9 building blocks that are required to develop a common industrial data space: data standards; business models; governance; legal agreements; exchange protocols; identification and authentication; authorisation; metadata; and operational agreements. Policy-makers and industry can use this set of IAs as a strategic roadmap for the digitalisation of industrial processes.

These templates are designed to provide a comprehensive framework. Each template is designed to provide an independent, complete and flexible structure that can be adapted to fit the needs of individual sectors.

Table 2: Summary and description of Industry Agreement templates

Name	Description	Application cases
Data Quality and Data Value Exchange (smart grid and manufacturing applications)	List of legal and technical parameters to achieve greater data quality in industrial value chains.	N/A
Common Ontology-driven Data Documentation (applications in lab robotics and the development of software translation tools)	List of legal and technical parameters necessary for the development of an interoperable digital ecosystem.	Laboratory robotics and translation software.
Shared Data (testing field for common chemical formula and reaction representation)	List of legal and technical requirements for the development of operational data space governance.	Development of standards for the representation of chemical molecules and chemical reaction.

Finally, the research team developed **four key cross-cutting recommendations** to help stimulate greater cooperation in the IA development and implementation process, and to promote the development of common industrial data spaces. For each of these recommendations, specific policy and industry actions are proposed, including an assessment of their potential impact on industrial competitiveness.



1 Introduction

The European Commission (DG CNECT) commissioned this study in response to the growing economic and strategic importance of the Internet of Things (IoT) and the digital transformation challenge. In all industry sector value chains, the effective use of data and digital technologies opens up very significant opportunities for the creation of customer value and is therefore considered a key enabler to increase industry competitiveness.

The **main objectives of the study** are to:

1. Identify 7-10 innovation areas where industry agreements have a high potential to create markets and market opportunities, to validate the need and viability of such agreements for 3-5 areas, and to recommend further actions;
2. Investigate the impact of existing industry agreements in current and future value chains;
3. Provide recommendations to be taken-up by industry, the European Commission, and/or Member States.

In the study's Terms of Reference, the expression "industry agreements" is defined as "agreements on functions and interfaces between industry players that create markets and market opportunities leading to ecosystems and standards". Industry agreements are broadly described as arrangements that facilitate product and service design by industry players. For example, agreements on reference architectures specify how different components can be integrated, what the interfaces are, and how technology will implement the functions specified.

During the course of the study, the technical, economic and legal analyses have gradually become more focused on industry agreements that aim to increase the use, sharing and exchange of **data**, and the creation of **industrial data ecosystems**. In agreement with DG CNECT, the following **study definition** was established:

Industry agreements (IAs)

In the context of this study, **industry agreements (IAs)** are tools designed to address the barriers that hinder the development of industrial data ecosystems to enable industry stakeholders to enjoy the full benefits of the data revolution. They target economic gains for all participants and the creation of new market opportunities, new ecosystems and standards. Industry agreements may encompass different solutions, such as:

- **Federated digital infrastructures (FDI)** of trust which digital value chain service operations will rely upon;[1]
- **Common vocabulary standards (CVS)** and semantics for digital value chain interoperability;[2]
- **Relational contractual agreements (RCA)** that will govern vested interests in the digital value chain.[3]

[1] FDI: Manufacturers can currently enable secure data-sharing collaboration in a number of ways. Although risks will always exist, companies that select the right technology infrastructure and architecture can reduce those risks to an acceptable level. Companies have several options to consider when selecting data repositories, platforms and cybersecurity techniques according to specific requirements for connectivity, computing, performance, cybersecurity, data sovereignty, trust, traceability, availability, transparency and accountability. The development of digital value chains needs to comply with the increasing requirement to leverage a computing continuum across multiple domains and multiple tenants from the assets themselves, through edge and cloud (public & private).

[2] CVS: The operation of digital value chains requires machines, sensors, digital platforms and companies to speak the same language, so that companies can easily aggregate and analyse data. To overcome interoperability issues, data-sharing arrangements in digital value chains require several layers of standardisation.

[3] RCA: Address issues related to data sharing. Examples include rules for accessing and storing data, limitations on aggregation, and the use and further sharing of data, beyond a typical discussion on data ownership. In relational contracts, parties specify mutual goals and establish structures to keep their interests aligned over the long term. This type of contract model is crucial to build trust and to develop long-standing concepts such as digital twins, connected supply chains or digital threads, which are central to the effective and successful operation of digital value chains. Suppliers and OEMs in Japan have started to adopt these contracts.

The final study report is submitted in accordance with the implementation of the service contract No. LC-01167726. The study is carried out by CARSA (Spain) in collaboration with Ecorys (Belgium), KU Leuven (Belgium) and VDI/VDE-IT (Germany) (subcontractor).

Positioning of industry agreements in the current policy context

To boost the future competitiveness of European industry, EU and national policy-makers have launched various research programmes and proposed new legislation to support the creation of a single market for data.

The **European Data Strategy** was released in February 2020. This strategic document provides a framework to stimulate the European data economy, which is expected to grow from €301 billion in 2018 to €829 billion by 2025[4] and develop a competitive single market for data, allowing the free-flow of data within the EU and across sectors for the benefit of businesses, researchers and public administrations. In addition to these objectives, the strategy aims to ensure that digital innovation is aligned with European values. This approach translates into specific policies, such as the release of ethical norms for the development of trusted AI[5] and the European approach to Artificial Intelligence.[6] This initiative includes the development of discussion spaces such as the European AI alliance and the European AI discussion forum.

To complement the strategy, the European Commission also released the **Data Governance Act**[7] in November 2020, which proposes rules to accommodate data sharing. The aim is to foster the availability of data for use by increasing trust in data intermediaries and by strengthening data sharing mechanisms across the European Union.

Research and piloting initiatives were also launched at European level to remove specific technical barriers encountered by industry players. The two most important for the development of a common industrial data space are **GAIA-X and the Industrial Data Space Association** (**IDSA) framework**. Launched in response to the domination of the American GAFA, GAIA-X is developing a European cloud space to protect European data sovereignty and streamline exchanges.[8] The IDSA framework is also a crucial solution component, providing a technical and operational framework for the standardisation and harmonisation of digital infrastructure and systems to enable data exchange.[9]

The European digital ecosystem also features a number of initiatives designed to support **system interoperability**. This objective is pursued through the funding of research programmes and industrial pilots such as Far-EDGE, BOOST 4.0 or the European Interoperability Framework (EIF) for digital health records.[10] However, although these initiatives have produced some positive results, they have also made the digital landscape more complex by developing new and sometimes overlapping initiatives.

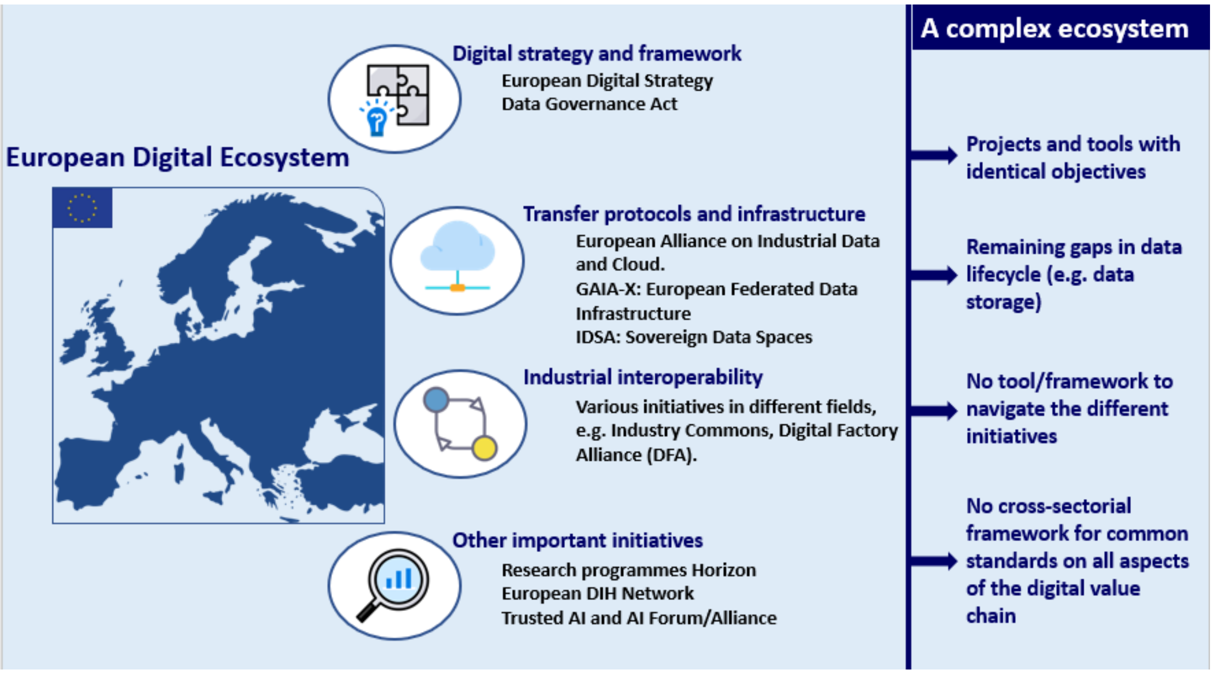
A range of initiatives are also available to support industry players as they undertake their own digital transformation journey. The Digital Innovation Hub (DIH) network provides training and mentoring to support the digitalisation of corporate processes. The Horizon research programmes provide funding for research projects on digitalisation, standardisation and harmonisation.

The different elements of the European Digital Ecosystem are summarised and illustrated in the figure below.

---

[4] EC, 2020, *European Strategy for data*, https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_fr
[5] EC, 2020, *Ethics guidelines for trustworthy AI, https://ec.europa.eu/futurium/en/ai-alliance-consultation*
[6] EC, 2020, *Artificial Intelligence*, https://ec.europa.eu/digital-single-market/en/artificial-intelligence
[7] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767
[8] Lesnumériques, 09/06/2020, *QU'est ce que Gaia-X, le meta-cloud Européen ?,* https://www.lesnumeriques.com/vie-du-net/qu-est-ce-que-gaia-x-le-meta-cloud-europeen-n151195.html
[9] IDSA, 2020, https://www.internationaldataspaces.org/ids-is-not-only-a-technological-standard-but-also-a-universal-legal-framework-to-create-data-driven-business-ecosystems/
[10] EC, 2017, *The new European Interoperability framework, https://ec.europa.eu/isa2/eif_en*

Figure 1: Summary and overview: the European Digital Ecosystem

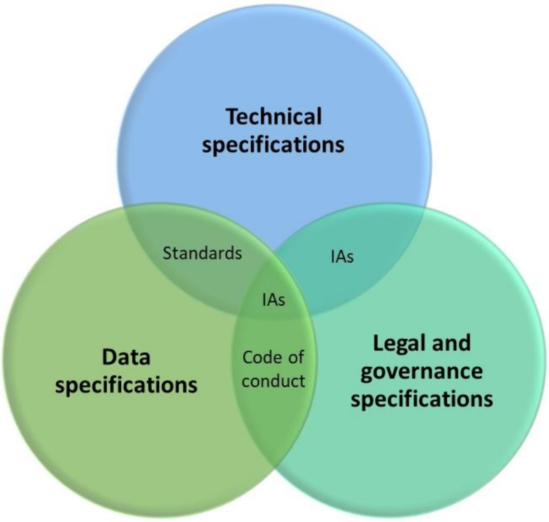


Source: CARSA, 2021. This figure was created using resources from Flaticon.com.

All of these initiatives and pieces of legislation provide a set of measures and framework conditions for the development of trustworthy data-sharing systems and the creation of common European data spaces. These data spaces will offer a secure and trusted environment for companies to share their data based upon voluntary agreements to be developed and implemented by industrial players.

Within this framework, industry agreements (IAs) are intended to be voluntary bi- or multi-lateral agreements that are designed to support the development and functioning of data spaces. Notably, IAs aim to provide a framework for the different building blocks of data spaces, from technical and data specifications to governance and legal dimensions, encompassing all possible stages of the data lifecycle. IAs may take many different forms, and one therefore cannot provide an exhaustive list of contractual arrangements that might be concluded between industry agreement stakeholders, due to their complexity. For example, cooperation between industry agreements stakeholders may take the form of: a rather simple single (or continuous) 'transaction' (e.g., data access agreement, IP licensing agreement, confidentiality agreement, consultancy agreement); a continuous collaborative project (e.g., contract research agreement or collaboration agreement); a joint management and cooperation project involving a large consortium of stakeholders (e.g., consortia agreements); or even entail the establishment of a separate legal entity (e.g., joint venture agreement, agreement establishing a standard setting organisation or a platform).

Figure 2: Veen Diagram on the "USP" of IAs over alternatives



Source: CARSA, 2020

IAs differ from standards and codes of conduct, as these focus on and address specific parts of an industrial data space. Standards can influence new technical and data specifications, but they lack the governance and legal dimensions provided by codes of conduct. Codes of conduct can include some data specification requirements, but they have limited influence on technical specifications and are excluded from the development of new specifications. Codes of conduct and standards can therefore complement each other and serve as instrumental support tools for the development and adoption of industry agreements. For instance, codes of conduct can underpin and facilitate the development of IAs by establishing common principles for data sharing and contractual relations between stakeholders, providing guidance on the use of data and strengthening trust between contractual parties.

Although there is widespread consensus on the importance of these types of agreements, the study finds that industry stakeholders face a number of obstacles that inhibit IA development, with analysis provided in Chapter 2. The study also provides a taxonomy of the key aspects required to support the adoption of common rules by industry actors and the creation of voluntary B2B data sharing schemes. An outline of the core legal and technical elements and contractual clauses required to achieve a sufficient level of data interoperability, exchangeability and quality is provided (see Chapter 3 and Annex I). This exercise can be regarded as a strategic tool to help industry players to develop practical building blocks that can be used for future data spaces.

Outline of the study activities

The study aims to identify key innovation areas where industry agreements – e.g., the standardisation of digital and/or physical processes, devices and systems – have the potential to deliver high economic impact and increase the competitiveness of European industry.

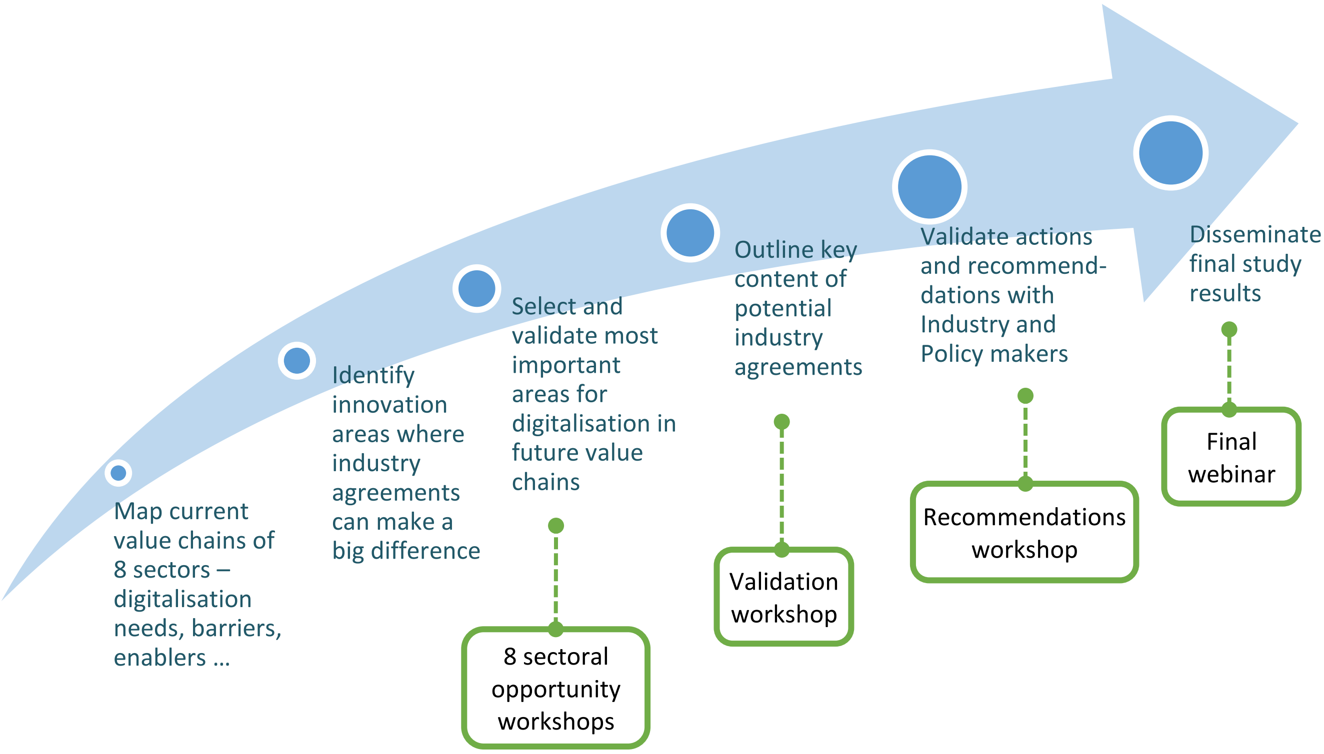
Innovation areas — In this study, **an innovation area refers to** an element of an industry digital value chain for which the development of industry agreements would facilitate digitalisation and create new market opportunities.

The concept is best explained using a practical example. In the shipping industry, the lack of data standardisation and exchange between ports and ships causes a lack of transparency that can result in operational inefficiencies and higher CO_2 emissions.[11] "Ship-port data exchange" is therefore a potentially interesting innovation area for the development of IAs to increase transparency and facilitate data exchange. This could lead to economies in bunker consumption for shipping companies, less CO_2 taxes to be paid by port authorities and could help to free-up shipping capacities.

The detailed methodological approach was described in previous study reports. The following figure provides a quick overview of the main steps involved, including some of the endorsement activities performed during various workshops.

[11] Source: interview with a stakeholder active in the shipping industry.

Figure 3: Outline of project activities



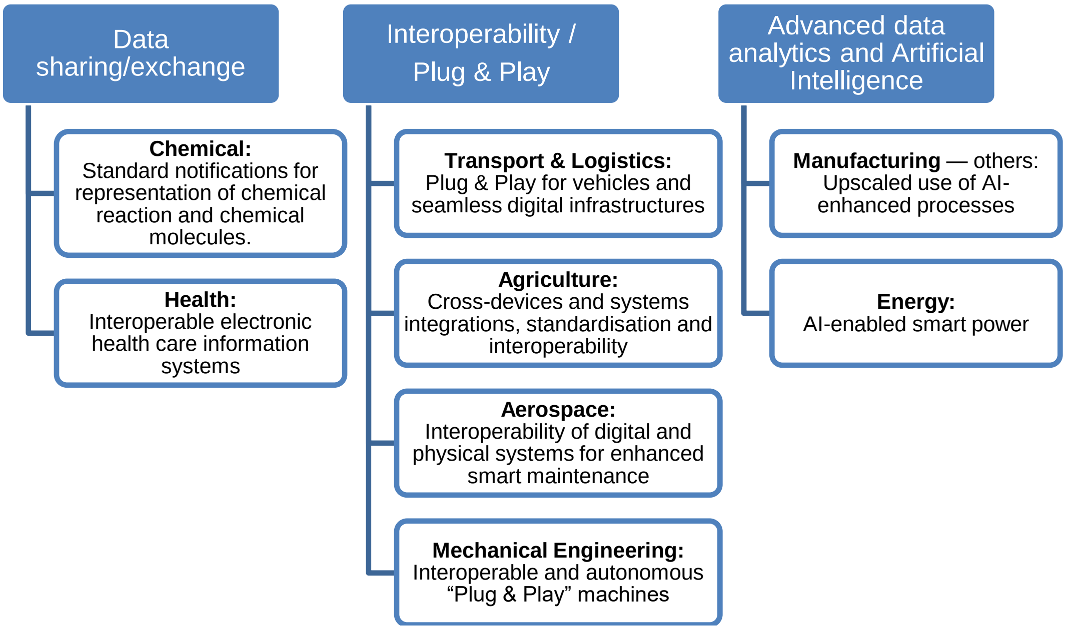
The study examines eight sectors of strategic importance to the sustainable competitiveness of European industry: manufacturing; mechanical engineering; transport & logistics; chemical; aerospace; agriculture; energy; and health. The first interim report analysed European industry at a macro level and assessed each sector's current value chain. It also identified and listed current and emerging industry agreements in each sector.

A series of digital workshops organised by the consortium resulted in the selection of eight innovation areas of opportunity for the development of high-impact IAs (one per sector). Figure 4 presents the selected innovation areas per sector. They are clustered in three groups and all strongly target the increased use, sharing and exchange of data in future digital value chains.

These innovation areas are central to the digital transformation and harmonisation of the European Digital Ecosystem, and can even be viewed as the missing link between current and future digital value chains. In varying degrees, all three areas are crucial for each sector. For example, Plug & Play/interoperability is the key milestone for the development of future mechanical engineering processes. In contrast, smart maintenance in aerospace urgently needs the development of data exchange processes.

These innovation areas cover the three strategic elements that are required to develop an integrated European data space: data must be exchangeable; systems must be able to process data (interoperability); and data must be exploitable for analytics, AI, Machine Learning (ML) and other advanced purposes. In other words, **these three areas are the three keys required to unlock future digital value chains.**

Figure 4: Selected innovation areas by cluster



Source: CARSA, 2020

Whereas the first interim report assessed the use of industry agreements in current value chains, the second interim report provided insight on the IA development in future value chains. Aligned with the selected innovation areas, the second interim report also presented three draft **IA templates**. Together, these templates provide a comprehensive and flexible framework that covers the main elements required to develop a fully integrated industrial data space. Feedback on the draft IA templates was gathered from stakeholders during the validation workshop (see Annex II). In the recommendations workshop, preliminary recommendations were validated by industry and policy stakeholders (see Annex III).

Content and structure of the report

This final report summarises the results of the study, with emphasis placed on the analysis of IAs in future value chains and related recommendations for industry and policy-makers.

Chapter 2 summarises the **barriers** to IA development and uptake. It also highlights **opportunities** for industry to take action and for policy-makers to support European initiatives to unlock the potential that exists in each innovation area.

The main **building blocks** for the **development of high-impact IAs** are summarised in **Chapter 3**. More detailed descriptions of the IA templates, including validation feedback, are presented in Annex I: 9 building blocks for the definition of key components in industry agreements - templates and examples for the development of a streamlined data space.

Final study recommendations and their **potential impact** on European digital competitiveness are presented in **Chapter 4**, including the main findings from the recommendations workshop.

Summarised **outcomes from two workshops** performed in the final phase of the project are presented in **Annex II** and **Annex III**.

2 Analysis of potential industry agreements in innovation areas of high value for Europe

This chapter presents the key findings from the horizontal analysis of three clustered innovation areas that demonstrate significant potential for the development of high-impact industry agreements (IAs):

- Data sharing/exchange;
- Interoperability framework for Plug & Play systems; and
- Advanced data analytics and Artificial Intelligence.

The innovation area analysis evaluates the importance of the business and technological opportunities offered, as well as the economic value that new IAs could help to unlock. Section 2.1 presents a summary of the barriers that prevent industry players in the analysed sectors from developing and implementing IAs. Section 2.2 highlights key opportunities and proposed actions for the development of IAs.

2.1 Summary of barriers affecting the development and uptake of industry agreements

In this section, we outline the key economic and legal barriers identified for the development of industry agreements together with possible pathways for their removal. The suggested pathways cover specific areas where concrete actions could be taken, and which could act as 'pilot' or 'lighthouse' initiatives for the removal of barriers to the development and implementation of an industry agreement.

Key economic barriers

From an economic perspective, a number of barriers were found to impede the establishment of industry agreements across the selected innovation areas and sectors, as illustrated in Table 3 below, according to the frequency of their occurrence and significance[12]. They are presented in more detail in the following sections.

[12] Assessment based on expert judgement, stakeholders' consultation (e.g., workshops, interviews) and market reports

Table 3: Overview of identified barriers to industry agreements across innovation areas

Barriers to IA		*Impact:*			
		Occurrence (3: frequent, 2: common, 1: occasional)			
		Significance (3: high, 2: medium, 1: low)			
Themes	**Sub-theme**	**Impact type**	**Advanced data analytics and Artificial Intelligence**	**Plug & Play/ interoper-ability**	**Data sharing/ exchange**
Industry structure / prevalence of SMEs (Cross-cutting)	Coordination, financial, knowledge constraints	Occurrance	2	2	2
		Significance	2	2	2
Costs of IA development	Technical complexity	Occurrance	3	2	3
		Significance	3	2,5	3
	Regulatory complexity	Occurrance	3	3	3
		Significance	3	3	3
	Coordination complexity	Occurrance	1	3	3
		Significance	2	3	3
Nature of competition and competitive advantage	Competitive advantage	Occurrance	2	3	3
		Significance	2	3	3
	Lock-in	Occurrance	1	3	3
		Significance	1	3	3
High costs of IA implementation	Costs of accompanying investments (capital, processes, knowledge)	Occurrance	3	3	3
		Significance	3	2	3
	Switching costs	Occurrance	1	3	2
		Significance	1	2,5	2,5
Size and distribution of costs and benefits of an IA (non-transparency & asymmetry)	Non-transparency of costs and benefits	Occurrance	2	3	3
		Significance	2	3	3
	Non-transparency of the value of outputs	Occurrance	3	1	3
		Significance	3	1	3
	Misalignment of the distribution of costs and benefits	Occurrance	2	3	3
		Significance	1,5	3	3
Data/system integrity and ownership	Security of commercially sensitive and personal data	Occurrance	3	3	3
		Significance	3	3	3
	Traceability	Occurrance	2	3	2
		Significance	2	3	1,5
	Protection of intellectual property	Occurrance	2	3	3
		Significance	2	3	3

Source: Study/consortium partners – based on own analysis and inputs from industrial experts

Industry structure / prevalence of SMEs characterised by:

- Lack of a 'coordinator'
- Lack of financial resources
- Lack of knowledge

The structure of an industry (or value chain segments) may influence the potential for IA development and implementation. This is perhaps most evident where an industry structure is characterised by a prevalence of SMEs (e.g., mechanical engineering, agriculture). On an individual basis, firms may attribute a low expected value to the (net) benefit of an IA and, therefore, lack sufficient incentive to develop or participate in an IA, even where the collective benefit could be substantially high. Smaller firms may also lack the financial and knowledge resources (capacity) to engage in IA development and implementation. Alternatively, in the absence of a coordinating organisation (e.g., industry association or network, or industry leader), an SME-led industry may lack the institutional capacity to coordinate IA development and implementation.

Note: industry structure is a 'cross cutting' issue that has implications for subsequent barriers described below.

Costs of IA development, associated with:

- Technical complexity
- Regulatory complexity
- Coordination/structural complexity (number of partners)

High (fixed) costs may inhibit IA development. These costs may arise because of the complexity of the technical aspects addressed by an IA or the complexity of the regulatory environment surrounding an IA. Costs associated with coordination, approval processes and/or the diversity of agents involved in IA development may also be a contributory factor.[13] The high cost of IA development may relate, for example, to the specialist expertise/knowledge inputs required to formulate an IA, as well as the volume of effort required (in terms of the quantity of inputs and the time required), including administration and coordination costs. The regulatory context can also add to the complexity of formulating an industry agreement, especially when regulation is not harmonised across borders. For example, an IA in the aerospace sector, involving a large number of operators, reportedly took more than three years to develop[14]. Nonetheless, initial IA development costs can be considered as a one-off investment that reduces the cost of subsequent agreements with new parties, as they do not need to be developed from scratch. Moreover, many stakeholders across the analysed sectors indicate that, compared to development costs, recurring costs of industry agreements are negligible, with benefits significantly outweighing costs.

Insights from existing agreements on reference architectures show that technical complexity is one of the major barriers, especially in large projects that feature many developers and highly interconnected functionality, and in which companies require extra expertise and more tool-orientated solutions.[15]

Overcoming technical and regulatory complexity

Technical complexity and a lack of in-house expertise and skills are some of the main barriers to IA development. Access to tailored **advice and technical expertise through intermediary organisations** and external providers (e.g., excellence centres) may help to limit the impact of those barriers. For example, current structures such as the European Cluster Networks or Digital Innovation Hubs (DIH) could act as **facilitators, helping companies to access technical expertise and upskilling resources.** Support for the creation of IAs could also be included in the scope of current or new initiatives that provide technical expertise to industry, e.g., the Technical Assistance Facility for industrial modernisation, providing business, financial and legal support to interregional projects under the S3P

[13] For example, anecdotal information collected through survey and expert interviews, experiences from different sectors indicate that that the time required to release an industry agreement on the market can vary from 6 months for an agreement involving less than 10 partners to between 1 and 3 years for cases involving more than 50 partners.
[14] according to anecdotal information collected through stakeholder consultation
[15] S. Martínez-Fernández, 2015, *A Survey on the Benefits and Drawbacks of AUTOSAR,* https://ieeexplore.ieee.org/document/7447213

industry platform.

Regulatory complexity, whereby regulatory issues raise concerns and uncertainty about the viability of collaboration opportunities, could be addressed using **regulatory sandboxes**. These allow for collaborative testing grounds for new business models at the edge or outside of the current regulatory framework, under the supervision of regulatory institutions. From the perspective of regulators, it could also help to promote transparency, support regulatory learning and continuous improvement, and enhance innovation capacity.

The development of industry agreements can require extensive synchronisation across operators in different markets, countries and across value chains. Coordination can be particularly challenging in fragmented industries that feature a multitude of players of different sizes, and in which no big players dominate the market (e.g., mechanical engineering, agriculture). In contrast, more highly regulated sectors tend to feature a strong coordination component, especially when regulation is not harmonised across borders.

A lack of coordination can be measured in terms of opportunity costs. In engineering, for example, business intelligence shows that just by retyping the same information in another tool, **an additional labour cost of 20-25% can be incurred**. With the right information interfaces, the effort required to set up new technical systems can be reduced substantially. For automation systems in engineering, labour costs can reach 75% of overall project costs, **20-25% of which could be saved**[16] by improving digital data chain cooperation and information standardisation. **Improved coordination, e.g., to agree on syntax and semantic standards (see next barrier), can unlock huge cost savings**. Industry agreements have the potential to free-up internal engineering and IT resources by allowing information to be easily and securely exchanged between diverse systems to allow seamless integration without the need for costly and time-consuming processes.

Support for standardised Identification, Authentication and Authorisation (IAA)

Agreements on processes for data Identification, Authentication and Authorisation (IAA) are often agreed bilaterally between partners. The process needs to be reset and agreed for each new partner, which is a time consuming and costly exercise that can take months to formulate and requires a high level of resources (e.g., legal expertise) and trust.

iSHARE[17] is one example of a public sector-led initiative that has overcome these barriers. Initiated by a Dutch Government grant programme, iSHARE provides soft infrastructure to support data access. The initiative set up a uniform set of technical and legal agreements or schemes for data access, which enable organisations in the transport and logistics sector to work with common identification, authentication and authorisation methods, eliminating the need to repeatedly make new agreements each time they want to share data. This approach facilitates data sharing with new and previously unknown partners and allows data requests and sharing to be delegated to other parties (e.g., sub-contractors). It also makes it easier for smaller companies to participate in similar schemes, which in turn helps to **break down coordination barriers** and removes or reduces the need for costly and time-consuming data integration, without undermining the security of business sensitive data and risking the loss of competitive advantage.

Implementing a standardised approach to IAA could result in significant savings in terms of time and human resources. For example, in some sectors, a standardised approach could

[16] Insights from stakeholder interviews in the mechanical engineering sector.
[17] https://www.ishareworks.org/en/ishare

free-up 9 to 12 months of engineering and legal resources that would otherwise be needed to develop a new bilateral agreement from scratch[18]. In more fragmented industries, **EU or national public policy initiatives could help to support the creation of soft infrastructure and/or trusted frameworks**, and bring together players across sectoral value chains (e.g., through networking initiatives).

iSHARE as best practice

iSHARE was established in the Netherlands in 2018 to improve processes in the Transport and Logistics sector. Once iSHARE had become an established solution, responsibility for its management and further development was handed over to an industry-led governance group. 15 implementation partners currently provide input, supervise ongoing activities and collaboratively influence the growth and development of the iSHARE scheme. In addition to this "core team", the framework is also supported by more than 50 logistics-related companies, IT experts, research institutes and associations.

iShare aims to enable all logistics industry stakeholders to connect and work together based on mutual trust, irrespective of type, size, modality and jurisdiction. The end goal is to develop standard framework agreements with the essential functional, technical, legal and operational standards needed to support collaboration and data exchange within the iSHARE community. The entry process for new members is straightforward. Membership does not require the use of any additional technologies, nor the disclosure of confidential company data. Instead of storing data on a platform, companies are granted individual access rights which enable them to share data, even with unknown partners. Close cooperation with IDSA ensures compliance with international standards and accelerates the implementation of data sovereignty in Europe.

iShare is identified as a best practice for its exemplary data exchange solution for different kinds of transport and logistics companies (shipping, air cargo, truck transportation), which does not require them to share data on a platform. The trust and security principles at the heart of iShare, its cooperation with IDSA and its focus on data sovereignty in Europe are some of the main factors that make iShare stand out.

Catena-X as best practice

Catena-X is an economic alliance, founded in May 2021 in Berlin and financed by individual membership fees. The association board consists of 15 elected representatives of member companies and organisations like BMW AG, Siemens AG, Volkswagen AG, Daimler AG, BITKOM e.V. and Fraunhofer ISST.[19]

The aim is to create a uniform standard for data exchange along the entire automotive value chain. [20] The Industrial Data Spaces (IDSA) standard for data sovereignty, security and interoperability infrastructure were used as base for the implementation. The main standardisation areas are quality management, logistics, maintenance, supply chain management and sustainability - production and

[18] Insights from stakeholder consultation, from the transport and logistic sector.

[19] https://catena-x.net/en

[20] https://internationaldataspaces.org/catena-x-network-for-cross-company-data-exchange-in-the-automotive-industry-relies-on-ids/

development are planned for the future.[21]

BigchainDB GmbH (Day 1 member of GAIA-X) joined Catena-X on behalf of Ocean Protocol, a decentralized data exchange protocol for data monetization. Ocean will contribute mainly to develop sustainable decentralized business models and incentive mechanisms for the Catena-X network. [22, 23] From the German and European perspective, GAIA-X is a prerequisite and enabler for the simple, efficient, sovereign and secure exchange of data and information in the automotive data economy. Thus, the Alliance will rely on GAIA-X, and depends in part on GAIA-X for its success.

Catena-X was identified as a best practice because of the exemplary solutions concerning data exchange, cooperation, security, interoperability, openness and transparency.

It is necessary to include all relevant players to develop a long-term valid definition of rules and interfaces for an efficient data exchange, so the members and representatives combine the expertise of producers, suppliers, research, associations and politics. The organisational form allows equality and voice for all members.

It is also helpful to build on existing standards and align the target with existing and developing structures und technologies, which are taken into account by cooperating with IDSA and relying on GAIA-X.

Security aspects are not only regarded for the possibility of data-attacks and the necessary measures to avoid them, it is a holistic concept trying to generally hide data from the big data-collectors and -analysts and keeping the opportunity to adapt the protocol if necessary.

Nature of competition and competitive advantage:

- Competitive advantage
- Technology lock-in

Although firms operating in different value chain segments often collaborate with each other (vertical collaboration), direct collaboration on product and production-related data between firms operating in the same value chain segment (horizontal collaboration) is rare and challenging due to issues related to competition and compliance. By necessity, data-sharing arrangements involve companies in the same industry. Self-interest and the need to maintain competitive advantage are strong motivators for product and production data sharing between competitors. The **fear of losing negotiation power or competitive advantage** when disclosing business information is an important barrier to cooperation. Other barriers are the **fear of technology lock-in** when committing to one type of technology and standards, and/or the ‘rapid obsolescence’ of an IA, particularly in rapidly developing technology/innovation areas.

Opportunities for non-competitive cooperation

Opportunities exist for non-competitive cooperation in areas that are not critical to competitive advantage (e.g., cross-sectoral standards on metadata, uniform descriptions of the type of information that metadata must contain and in what form), particularly on technologies that are not core to a

[21] https://www.daimler.com/innovation/digitalisation/industry-4-0/article.html

[22] https://oceanprotocol.com/press/2021-05-20-bigchaindb-bosch-fetchai-join-catenax

[23] https://assets.bosch.com/media/global/research/eot/bosch-eot-catena-x-partner_en.pdf

company's own products and for which standardised solutions could deliver shared benefits.

For instance, insights from the mechanical engineering sector show that companies working as system integrators require detailed descriptions of the devices that they buy. Currently, descriptions are received as PDFs, which are not easy to integrate into engineering value chains, as all information must be re-typed. Providing shared digital device descriptions, as well as the enabling tools to access them independently of the engineering discipline or technology used, would be beneficial.

In the case of more fragmented industries, in particular, EU and national policy-makers could also help to identify cooperation areas, in compliance with competition law.

Industry agreements can be used to **establish minimum quality standards or common performance measures and metrics** to reduce information asymmetries within the supply/value chain and between industry and their customers. Standards can help firms to enhance the quality of their products and the efficiency of their processes. By making technical data more widely available to all firms, they can also support more efficient (less costly) and effective data exchange within the supply chain. However, the development of IAs can also face significant trust-related barriers (e.g., fear of losing competitive advantage).

Relational contracts to unlock data-sharing: need for approaches to increase trust and data access

To enjoy the benefits of data-sharing, all parties need to align their own interests and agree on structures to keep those interests aligned over the long term. **Relational contracts may offer a way to address these challenges**.

Data sharing in manufacturing can improve the quality of an end product (for example, by process optimisation). Suppliers and producers share a common interest to deliver end products of the highest possible quality. Designing a **relational contract that focuses on quality targets for the final customer**, rather than a traditional contract on the quality level that the supplier must provide to the producer, would be much more effective and would help to create more efficient collaborative solutions and win-win outcomes for all parties involved.

In addition, there is a need to investigate **new forms of collaboration and approaches to increase trust and access to data**, and how these would apply to the different sectors. **Data trust**[24] structures could be explored to provide independent data stewardship[25] and help organisations extract value from anonymised data, to allay concerns about how sensitive data is held by third parties.

Stakeholders indicate that these types of solutions have the potential to free-up an enormous number of resources, e.g., bring down costs in the form of **9-12 months for lawyers to negotiate NDAs**.

High costs of IA implementation, associated with:

- Investment costs (capital, processes, knowledge)
- Switching costs

[24] Open Data Institute (ODI), 2018, https://theodi.org/article/what-is-a-data-trust/
[25] Data stewardship is defined as the management and oversight of an organisation's data assets to ensure the provision of high-quality data that is fit for purpose. Data stewards do not own the data but are responsible for tasks such as developing common data definitions, identifying data quality issues and ensuring compliance with specified standards.

High (fixed) costs may inhibit IA take-up and implementation. **High implementation costs may require companies to make significant investments**; for example, companies may have to invest upfront in equipment, information capture and processing, other forms of process reengineering, or knowledge and skills development (i.e., technical, management, business, economics or legal knowledge relevant to all innovation areas, but particularly in less digitalised sectors). These costs are particularly relevant for artificial intelligence, with high initial costs needed to develop and implement AI solutions due to resource heavy research which needs to be conducted beforehand – costs which are disproportionately higher for SMEs. In some sectors, this is also further exacerbated by additional costs such as those for infrastructure upgrade and modernisation (e.g., in the energy sector, and IT infrastructure of hospital information systems in the healthcare sector).

Alternatively, there may be **high switching costs associated with IA adoption**; related, for example, to the cost of forgone investment returns on stranded assets that result from IA adoption and implementation (also highly relevant for interoperability and data sharing). Where these costs are high, it can create business inertia, which limits the take-up of open/interoperable systems. Ultimately, individual firms face a trade-off between the costs of switching and the expected benefits of moving to an 'open' system.

Important barriers to industry agreements arise from the existence of **market failures**. In the AI area, for example, we observe that many companies do not have access to a sufficiently large amount of data to feed their algorithms and face difficulties when **gathering the critical mass of data necessary to feed and train AI solutions**, which is particularly true for SMEs, resulting in **economies of scale** and **virtuous cycles** triggered by greater data flows being lost. For example, to design a robust prediction algorithm that can provide valuable insights, a company needs large amounts of data, with many instances of unexpected machine failures. Prediction algorithms built on inadequate data are ineffective over the long-term and data on unexpected machine failures is rare. As a consequence, most **manufacturers do not have a sufficient amount of data** relating to machine failures, with small manufacturers being affected the most.

These **barriers can be quantified in terms of opportunity costs, by looking at the benefits foregone**, which vary greatly across sectors. An important area in which AI can create value is projection and forecasting, which can help companies to secure important competitive advantage. AI allows businesses to provide better forecasts for their supply chain and design better offerings. For example, **by using sophisticated algorithms, health systems in developed countries can achieve up to 2% in GDP savings through operational efficiencies**, by increasing prediction and prevention capabilities, and a **5–9% reduction in health expenditure** by using machine learning to tailor treatments and keep patients engaged. Similarly**, in some manufacturing use cases, up to 39% of IT staff resources could be freed up** by using AI to fully automate procurement processes and **13% of EBIT improvements could be attained** by using machine learning to predict revenue streams and optimise sales efforts[26]. The benefits of AI-enabled demand forecasting are also impressive in the transport and logistics sector, with **costs related to transportation and warehousing and supply chain administration expected to decrease by 5% to 10% and 25% to 40%**, respectively, with overall **inventory reductions of 20% to 50% percent being feasible**[27].

The size of the market failure and the extent to which these inefficiencies are attributable to the lack of value chain cooperation cannot be clearly established. The ability to unlock all of these benefits relies on different factors, such as access to the technology, the creation of business cases, and the ability to address the scepticism that results from uncertainty about the net impact on industries, labour markets (e.g., job losses) and ethical concerns (e.g., when a machine actively makes decisions on the data).

[26] McKinsey Global Institute, 2017, *Artificial intelligence the next digital frontier?* https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.pdf

[27] McKinsey & Co., 2017, Smartening up with Artificial Intelligence (AI) - What's in it for Germany and its Industrial Sector? https://www.mckinsey.com/~/media/McKinsey/Industries/Semiconductors/Our%20Insights/Smartening%20up%20with%20artificial%20intelligence/Smartening-up-with-artificial-intelligence.pdf

Nevertheless, **the ability and willingness of industry stakeholders to cooperate** and agree on important factors is central **to unlocking the potential of the innovation areas**.

Enabling data access and quality to support the uptake of industry agreements on AI and data-sharing

Data quality is still a huge problem and a barrier to collaborative AI and data-sharing solutions. Companies are often forced to allocate internal resources to burdensome and time-consuming processes to clean data and make it usable, or they may lack the resources or common approaches to do so.

Exchange of good practices and widespread showcasing of successful outcomes would help to build a clearer understanding of the full benefits and added value of collaborative solutions, and would help to encourage businesses to join IAs. Moreover, the definition of clear criteria for assessing the quality of datasets in terms of accuracy, reliability and completeness would be desirable.

Promoting public-private initiatives to ensure access to data and data-related services (e.g., Gaia-X initiative) in the form of trusted and open data infrastructures is key to bringing together European actors to develop collaborative IAs and address existing data market failures. The availability of training data could increase by propagating trust-building concepts for the anonymisation and pseudonymisation of data and by putting in place trust structures for the secure sharing of non-personal interdisciplinary data.[28]

Non-transparency and asymmetry of IA costs and benefits

- Non-transparency of costs and benefits
- Non-transparency of the value of outputs
- Misalignment of the distribution of costs and benefits

IA development and implementation may be inhibited by a lack of information or uncertainty about the size (value) of expected costs and benefits arising from an IA, or on the distribution of costs and benefits across different agents (particularly relevant for interoperability and data sharing). While there may be a consensus of opinion that an IA would deliver a net positive economic benefit, both the overall (aggregate) size and the distribution of benefits (and potential losses) may be difficult to evaluate. Where this is the case, a lack of information (non-transparency) may result in agents under (or over) estimating the potential gains from an IA, which may discourage effort to develop or implement an IA, or they may simply be reluctant to adopt an IA when uncertain about the (net) benefits.

It is not always easy to attribute a concrete economic value to the potential benefits that accrue from an IA (particularly relevant for data-sharing). This occurs, for example, where it is difficult to monetise the value of outputs (e.g., the economic value of data assets). Alternatively, there can be asymmetric information on the economic value of outputs, where some agents are more able than others to evaluate the market (monetary) value of outputs. New challenges revolve around the definition of a price on a dataset which is neither straightforward nor clear, as assigning a price to a data point is not an easy task, making it **hard for parties to assess upfront the value of participation in a potential agreement.**

[28] Bundestag Study Commission on Artificial Intelligence (2020), https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/AIDA/DV/2020/11-09/Executive_Summary_Final_ReportEN.pdf

Data valuation – Methods and Pros&Cons

Data has become one the most valuable assets for modern companies. However, stakeholders still lack the common assessment models they need to estimate the value of data. This leads to difficulties for industrial players to assess the value of participating in the definition of Industry Agreements.

The concept of "fair price for data" relies on the hypothesis that there is an appropriate data valuation and monetisation mechanism. However, the characteristics of data sets makes the creation of such a mechanism a difficult task. Data sets do not have clear economic value deriving from well-established pricing mechanisms.

The definition of data value remains a brand-new topic with no existing data valuation system acting as a common reference across industrial value chains. In fact, different valuation methods will continue to coexist on the data market. Companies engaging in data-trading should chose the most suited approach based on the characteristics of their data-asset, data scarcity and potential use for the traded data.

Currently, three different approaches for data valuation can be identified.

- **Cost value approach** relies on the cost of making data available as a proxy to assess the value of data. Notably, this method looks at the cost to produce and store the data, the cost to replace it and the impact on cash flows if it was lost[29]. This method is relatively easy to carry out compared to other techniques. However, cost-value approach does not accurately assess the real economic value of data as an asset as it leaves aside the data sets' potential to generate value.
- **Market value approach** looks at the exchange price of sold data. In a market-value approach, the value of data is based on the price that the market willingness to pay for a specific data set or comparable assets. The growing practice from industrial players to engage in data exchanges and licensing practices can serve as a starting point to develop price benchmarks. These benchmarks then serve as a reference to assess the value of a data set. As for the cost-value method, market-value leaves aside the intrinsic value of the data set and thus does not assess the data set's potential to generate value. However, with the development of data exchange in the industry, the accuracy of this pricing mechanism will continue to improve.[30]
- **Economic value approach** starts by defining data as an asset that is gaining in value and importance.[31] This approach seeks to determine the business benefits generated by data, looking at the income and value the data set is expected to generate for the acquirer. This approach focuses on identifying and sizing the impact on cash flows for both the data owner and the data acquirer.[32] Theoretically, the economic-value approach is the most solid as it is based on the assessment of the data set's intrinsic value. In this method, data is genuinely defined and traded as an asset. However, this method is very difficult to implement as the value of data varies based on individual use cases. Nonetheless, classes of data are emerging from practice and use cases. Data belonging to an identical class can thus be priced together (e.g. master data used in different processes) — something that greatly simplify the task of defining and assessing a data set intrinsic value.[33].

[29] Anmut (2020), *Different Data Valuation Methodologies: The Pros & Cons,* https://www.anmut.co.uk/different-data-valuation-methodologies/
[30] Deloitte (2020), *Data valuation: Understanding the value of your data assets,* *https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Finance/Valuation-Data-Digital.pdf*
[31]D. Moody, P. Walsh, *Measuring The Value Of Information: An Asset Valuation Approach*
[32] IMDA (2019), *Guide to data valuation for data sharing*, https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Guide-to-Data-Valuation-for-Data-Sharing.pdf
[33] Insights from Recommendations workshop

It can be the case that the distribution of benefits accruing from an IA are not aligned to the distribution of costs. This occurs, for example, where the costs associated with IA development and implementation are disproportionately high for small firms compared to large firms. Alternatively, misalignment can occur where an IA creates benefits ('positive externalities') for agents that do not incur the costs of IA development and implementation; for example, where an IA developed in one value chain segment results in spill-over benefits in other value chain segments. Such aspects can only be discussed and addressed at value chain level. For instance, an agreement on the development of new standards for chemical molecule and reaction would create new opportunities to share information across the entire chemical value chain. Likewise, investments by farmers in new equipment and smart machines also generate important value for the process industry. The process industry does not share in the cost of the farmer's investment, but does enjoy the benefits.

Guidelines and new approaches needed on data valuation to trigger joint investments and solutions

Dissemination of good practices and guidelines to define data valorisation/monetisation models would provide guidance to companies based on best practices and most commonly used methods on how the parties can share the value created, e.g., income-based method, cost-based etc.

Guidelines could include compensation models and valuation methods, e.g., resorting to **shared benefits according to a specific agreement**. For this purpose, new approaches like **data trusts**[34] could be investigated as suitable arrangements to **distribute the benefits arising from data-sharing more equitably**, including monetary benefits – for example, a share in the profits generated by services created from the data. A clear description of how benefits are shared in these types of arrangements would make it easier for industry parties to assess upfront the potential value of participation. Here, the role of intermediary organisations, such as DIHs, could be pivotal by helping to replicate successful applications in other innovation areas and sectors.

Opportunities are also offered by the development of a **Data Asset Value Exchange Mechanism** (see section 2.2).

Data/system integrity and ownership

- Security of commercially sensitive and personal data
- Traceability
- Protection of intellectual property

Economic incentives to IA development and implementation may be inhibited if agents are uncertain about the integrity of the system for collecting, exchanging and storing data and information (i.e., lack of trust in the system) that underpins/supports IA implementation. Issues may arise in relation to the protection of commercially sensitive and personal data, whether this relates to other agents (e.g., competitors) that are permitted to access the system or illegal access (e.g., data breach and data loss).

In the aerospace sector, the fear of data misuse is an important barrier, especially given the highly sensitive nature of safety and MRO-related data. Data-sharing in the maintenance end of the supply chain requires extremely high levels of trust on which operators need to agree, as the information has to be trusted, secured, reliable and error-free. There is highly sensitive safety and vehicle/transport-related data in the transport and logistics sector. In healthcare, the highly regulated and sensitive nature of personal data requires a level of care and discretion that is much higher than in other sectors. In agriculture, different types of data can feed into smart farming solutions and services (e.g., field

[34] Open Data Institute (ODI), 2018, https://theodi.org/article/what-is-a-data-trust/

stations, satellites, sensors, farm logs). Data ownership can be difficult to identify, where the source is not retrievable from specific proprieties.

Specific barriers can be encountered in the different innovation areas. In the interoperability area, there are important security concerns due the increased number of entry points for potential attackers, as well as the lack of traceability in plug & play. In AI, doubts can arise over the ownership of AI produced data. When sharing AI data, how to ensure that a data owner’s consent is passed on to the different purposes and services that are built around the data in the subsequent steps (see also “Key legal barriers to industry agreements” section hereafter) is not an easy question to answer.

Companies value their ownership of data produced and would like to see their Intellectual Property (IP) rights safeguarded. Company concerns about data ownership and rights can limit the appeal of IAs, either discouraging investments or resulting in very lengthy negotiation processes. The lack of trust is a key barrier. Clear rules on data ownership and the management of data outputs would help to facilitate the development of future IAs. The existence of trusted and common frameworks agreed and adopted by industry players would also be beneficial, as it would mean that negotiations (e.g., on IP rights) would not have to start from scratch every time.

Resorting to trusted third parties

Resorting to a **trusted third party (e.g., a service provider or machine supplier) acting as a neutral moderator could help to source the required expertise and facilitate the data sharing relationship** (combine, clean and analyse data). It may even help to resolve conflicts between competitors.

Some good practices are already observed: In the automotive industry, collaborative solutions for end-to-end tracking are already taking place. AutoSphere[35] is a community of automobile OEMs and suppliers, led by large manufacturers such as Honda, Toyota, Nissan and GM. They use a common set of processes and data transactions to more effectively manage their packaging and parts supply chain transactions. They make use of a secure third party that collects, shares and analyses the community’s transactions.

Such initiatives are often driven by large companies. Stakeholders in more fragmented industries (e.g., mechanical engineering, agriculture, healthcare), which are characterised by the predominance of SMEs, find it difficult to initiate these processes and often need public funding to part finance these activities. Current structures and networks such as European clusters or **Digital Innovations Hubs** may offer suitable solutions for SMEs through their “neutral” brokerage role. By connecting relevant stakeholders and providing technical expertise, they offer **fertile ground to lay the foundations for these agreements**. Through **the European Federation of Data Driven Innovation Hubs**[36]**,** communities of practice and collaborations could be stimulated to support arrangements for knowledge/data sharing and to define the baseline for business and sustainability plans, under neutral and non-commercial leadership.

In the manufacturing industry alone, the impacts of improved product tracking through the value chain are **expected to generate up to €35bn**[37], in terms of cost reduction opportunities, part loss mitigation, labour reduction, reduced production interruptions and transportation optimisation.

[35] Surgere’s website : https://surgere.com/autosphere/what-is-autosphere/
[36] https://euhubs4data.eu/#:~:text=The%20European%20federation%20of%20Data,data%20in%20a%20cross%2Dborder
[37] World Economic Forum, 2020, *Share to Gain: Unlocking Data Value in Manufacturing*, https://www.internationaldataspaces.org/wp-content/uploads/2020/02/WEF_Share_to_Gain_Report.pdf

In data sharing, these barriers lead to important inefficiencies, which can be addressed by ensuring business buy-in to an industry agreement strategy that relies on the ability to demonstrate that the long-term added value generated by data sharing and by connecting core offerings is higher than the immediate costs. The **size of these inefficiencies is considerable**. In maintenance, for example, these barriers amount to maintenance costs of 10-40%[38], which could be reduced by improving predictive maintenance. For **aviation** companies, for instance, **delays and cancellations are a huge and expensive problem**, with the worldwide cost of flight delays, often due to maintenance issues, being **estimated at over €22bn**[39]. Up **to 30% of the total delay time is due to unplanned maintenance**,[40] resulting in enormous economic losses for operators in the sector. Lifting barriers to data-sharing and interoperability agreements between MRO operators, manufacturers and airline companies is key to be able to exploit the wealth of safety and performance data that is continuously generated by today's aircrafts. Removing barriers would enable a shift from a corrective approach to a predictive approach and would offer new perspectives in terms of optimising maintenance costs and service delivery.

Manufacturing process optimisation is another area that could benefit from data sharing, providing the opportunity for manufacturers **to create additional value estimated at over €85 billion**, based on best practices[41]. **Over 40% of this value** could be achieved by **tracking products and components in supply chains**. Better visibility of supply chain processes would enable manufacturers to improve production planning, reduce inventory levels and react faster to unexpected events in the supply chain. Although tracking practices already exist, they often fall short of providing full end-to-end visibility. Improved visibility would require all suppliers in a supply chain to combine their data in a single shared system and use a common standard for supply chain transactions.

Key legal barriers

The diverse and fragmented nature of the legal and regulatory landscape presents a number of barriers to the establishment of industry agreements. The legal framework restricts the contractual freedom of the parties involved by stipulating the confines that industry agreements have to respect. The legal barriers to industry agreements do not always stem from the law directly prohibiting or restricting particular agreements at industry level. Many of the identified barriers originate from the legal framework not being fully adapted to the current digital reality, leaving open some of the issues that industry actors are confronted with in their innovation efforts (noteworthy examples are the lack of a clear legal status of data and the issues in applying liability regimes to AI systems).

The objective of this section is to provide a summary of the main legal barriers highlighted by stakeholder input, as well as recommended pathways to overcome these legal barriers. For a more in-depth review, reference can be made to the eight sectoral analyses set out in Annex III to the second interim report.

Table 4: Overview of identified legal barriers to industry agreements across innovation areas

Legal barriers	Main Specifications	Key Action Points
Data ownership/control	• Absence of a clear and overarching legal framework concerning the status of data	• Leverage legal branches (other than property law) such as the regulation of

[38] DIZMO, 2019, The unrealized benefits of predictive maintenance in Industry 4.0, https://www.dizmo.com/the-unrealized-benefits-of-predictive-maintenance-in-industry-4-0/
[39] Kearney, K., sin dato., *7 Ways to Improve Maintenance Costs with the Connected Aircraft,* https://aerospace.honeywell.com/en/learn/about-us/blogs/2019/03/7-ways-to-improve-maintenance-costs-with-the-connected-aircraft
[40] PWC, sin dato., *Predictive maintenance for airlines,* https://www.pwc.com/us/en/industries/transportation-logistics/airlines-airports/predictive-maintenance.html
[41] World Economic Forum, 2020, *Share to Gain: Unlocking Data Value in Manufacturing*, https://www.internationaldataspaces.org/wp-content/uploads/2020/02/WEF_Share_to_Gain_Report.pdf

and ownership fragmentation	• 'Data ownership' primarily regulated on a contractual level • Issues of fair distribution of data benefits	B2B unfair commercial practices • Usage rights for co-generated data
Access to and sharing of quality data	• Need to investigate new forms of collaboration and approaches to increase trust and data access • 'Data quality issue' and a resulting lack of trust • Liability regime for erroneous data remains unclear	• Data sharing and data quality obligations (e.g., 'Droit de suite') • Establishment of standard data sharing agreements • Provide legal support to collaborative data governance mechanisms • Liability for data quality • Engage in (co-)certification of data intermediaries
Data and protected/sensitive information	• Data sets including personal data, commercially sensitive data and IP-protected elements • Stakeholders rely heavily on guidance provided by data protection and other authorities	• Regulatory sandboxes • De minimis rules, fair use exceptions and other measures • Clearing houses, potentially involving multiple competent authorities • Investigate need for regulation of patentability of AI • Investigate need for regulation of IP-rights for AI-generated inventions and creations
Competition law	• Failure to provide access • (Potential) anti-competitive effects of industry agreements, including the development of standards	• Updating EC Guidelines, issuing communications and reviews • Encourage member states' authorities to issue additional guidelines • Data portability • Interoperability duty for dominant platforms
Liability	• Challenging application of liability regimes to AI • Allocation of control and liability/responsibility in a contractual context	• Clarify essential concepts (e.g., 'product' or 'defect') • Horizontal vs sectoral approach • Mapping of AI supply chain • Model agreements/templates • Determine responsible parties (e.g., cheapest cost avoider) • Further develop Artificial Intelligence Impact Assessments
Ethical aspects of AI	• Need for concrete guidelines	• Need for convergence/coherence between applicable regimes • Information duties • Adopt/promote Chief AI Ethics Officer within IAs
Standards, interoperability and general IP rights issues	• Intrinsic conflict/tension between IP and the establishment of standards	• Facilitate access to interoperability information – could be done through, an evolution of patent and trade secrets law • Review SSO financing • Launch open-source standardisation (SSO) initiatives

Data ownership/control and ownership fragmentation

During a series of workshops, 'the absence of data ownership' was identified by stakeholders to be a major obstacle across all innovation areas. The 'data ownership issue' relates to a range of challenges that industry stakeholders are confronted with, such as the distribution of economic benefits of data or ensuring that data is not used to the detriment of the data provider. From a legal perspective, the challenges encountered by industry stakeholders are more generally associated with the **absence of a clear and overarching legal framework to govern the status and value of data.** The issue of 'data ownership' is therefore closely related to the lack of trust and the imbalance of power between stakeholders in (digital) value chains.

The broad 'data ownership' debate has led to a number of policy and scholarly suggestions to help move the discussion forward:

- **Data sovereignty:** this concept involves the technological enforcement of contractual terms to enable data providers to retain some control and self-determination over the reuse of the data they provide. 'Data sovereignty' relies on contractual relationships to provide a clear legal framework. Contract law, particularly in B2B situations is, however, largely regulated at the national level, therefore complicating cross-border agreements.
- **Usage rights for co-generated data:** This approach departs from the creation of 'data ownership rights', and proposes to grant exclusive rights to a predefined data 'owner'. In a context where several entities may have contributed to the creation of data and knowledge (hence the notion of 'co-generation'), clarifying who may have certain entitlements could avoid undue 'enclosure' of data (e.g., by actors well-placed to exert *de facto* ownership in the value chain).
- **Unfair commercial practices:** The fair allocation of data benefits may, however, also be served by leveraging legal branches other than property law. The **regulation of unfair B2B commercial practices could be leveraged to ensure a fair allocation of data benefits across digital value chains**. Unfair B2B commercial practices are mainly regulated at national level. EU harmonisation in this area would benefit the internal market and would provide legal certainty to all digital value chain stakeholders. The main question is whether such regulation should be horizontal (sector-agnostic) or specifically targeted, i.e., at the data economy, at data (transactions), etc. It remains to be analysed whether horizontal provisions regulating unfair B2B commercial practices are sufficient and whether they should be harmonised at EU level in order to cope with all the challenges associated with the (un)fair allocation of data benefits.

Pathways to move forward in the 'data ownership' debate

It is important to recognise and elaborate on current and future strategies in the 'data ownership' debate: to weight the suggestions in an experimental approach and to identify the respective factors for success or failure assigned to data sovereignty, the regulation of data usage rights, collaborative data governance mechanisms and the regulation of unfair commercial practices.

Consideration should be given to leveraging legal branches (other than property law), for example, through the regulation of unfair B2B commercial practices. This type of regulation could be horizontal (e.g., with the French Code of Commerce) or data-specific (e.g., the ALI-ELI Principles for the Data Economy).[42]

Whether such legal protections are sufficient to empower weaker parties in digital value chains and

[42] See the webpage of the project on the ELI's website: https://www.europeanlawinstitute.eu/projects-publications/current-projects-feasibility-studies-and-other-activities/current-projects/data-economy/#:~:text=Principles%20for%20a%20Data%20Economy,which%20data%20is%20an%20asset (last visited 14th June 2021).

to build the necessary trust for them to engage remains a subject for more in-depth analysis. It is clear in any case that the fragmented nature of national legal frameworks and legislation constitutes an obstacle. The adoption of **soft law instruments** could help to overcome this fragmentation. A notable example is the 2018 'Code of Conduct for agricultural data sharing by contractual agreement'[43], which provides non-binding guidance to stakeholders on how to deal with issues of attributing rights to data and protecting data that is subject to the GDPR, intellectual property rights or confidentiality obligations. Industry agreements can also help to provide 'fair' principles for contracting practices by bringing together representatives of stakeholders in the value chain.

Access to and sharing of qualitative and accurate data

The analysis of key barriers affecting the development and uptake of industry agreements has shown that there is a need to investigate new forms of collaboration and approaches to increase trust and access to data. On data quality and data accuracy, for example, it appears that the quality of data is still a major issue and is likely to act as a barrier to collaborative AI and data sharing solutions.

Data sharing obligations - In certain sectors, such as mobility and energy, data sharing is imposed by law.[44] Mandating data sharing may unlock new opportunities in digital value chains. The experience gathered in sectoral data sharing legal regimes, such as in the context of the regulation on EU-wide multimodal travel information services[45], can be used by analogy in other sectors.[46]

Droit de suite

A '**droit de suite**' attached to data obtained thanks to the regulation constitutes an original way to find a ridgeline between two competing objectives: open up data to the benefit of future digital value chains, on the one hand; while on the other hand, preserving the legitimate interests of data providers and/or ensuring that data serves a public interest.

To foster data sharing for a given purpose, policy-makers could consider mandating data sharing, subject to 'droit de suite' obligations attached to the data (i.e., accuracy obligations imposed on data re-users).

Collaborative data governance models

Data sharing may sometimes have to be accompanied by **institutional mechanisms** in order to achieve a given (business and/or regulatory) purpose. This includes collaborative mechanisms, such as (data) pools, (data) commons and certain types of 'data trusts' (mainly in common law countries). This also includes more vertical approaches, such as the creation of data platforms or data hubs, subject to neutrality principles. The analysis of sector-specific case studies shows that data governance mechanisms need to be dynamic and respond to contextual constraints and objectives, related to the economic, technological and legal environment.[47] Whether the sector is characterised by pure market

[43] The latest version of the document (revised in 2020) is available online, https://cema-agri.org/images/publications/brochures/EU_Code_of_conduct_on_agricultural_data_sharing_by_contractual_agreement_2020_ENGLISH.pdf (last visited 14 June 2021).
[44] Ducuing, C. (2019) 'Data as Infrastructure? A Study of Data Sharing Legal Regimes', Competition and Regulation in Network Industries, https://doi.org/10.1177/1783591719895390
[45] Commission Delegated Regulation (EU) 2017/1926 of 31 May 2017 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide multimodal travail information services, OJ L 272/1
[46] Graef, I. and Prufer, J., (2018), 'Mandated Data Sharing Is a Necessity in Specific Sectors', Economisch Statistische Berichten, vol. 103 (4763), p. 298-301
[47] Charlotte Ducuing, 2020, *Beyond the Data Flow Paradigm: Governing Data Requires to Look beyond Data,* Special Issue: Governing Data as a Resource Technology and Regulation

conditions or whether public interest objectives are also present or desirable is also likely to play a crucial role. Although data governance mechanisms should be sector-specific, analysing the empirical factors for success or failure of sector-specific data governance mechanisms can certainly inform other sectors. This type of empiricist and pragmatic approach[48] should be fostered to the benefit of stakeholders and policy-makers alike.[49]

Support collaborative mechanisms

Provide legal support to collaborative data governance mechanisms (i.e., data commons, data pools, data trusts) to ensure sufficient attention is given to the institutional organisation of data rights in the digital value chain.

Liability for inaccurate / erroneous data

The circumstances under which liability may be incurred for damages due to inaccurate data is a recurring question in all sectors. Lack of clarity on this topic has a significant effect on the willingness of industry actors to enter into industry agreements. Again, the absence of a clear legal status for data (e.g., lack of qualification of data as a 'thing' or a 'product') is perceived to be the main challenge. Liability for (inaccurate) data is closely related to the 'data quality' issue, as it decreases trust, the value of data and the relevance of the business case at stake.

In addition to *lex generalis* liability regimes, the sectoral analysis of the study has shed light on interesting sector-specific provisions, which may inspire further (sectoral) legislation. For instance, the Commission Regulation concerning EU-wide real-time traffic information services[50] does not directly deal with liability for data, which is left to national law, though the data quality obligations it lays down are likely to have direct effects on liability. In all likelihood, failure to comply with data quality obligations could indeed trigger fault-based liability.

Liability for data quality

Data quality and liability for erroneous data may be regulated by law, more or less directly, based on a whole range of regulatory tools. The law may simply adopt data quality obligations (such as an obligation to update), which would logically result in liability exposure. Data quality may also arise from safety or cybersecurity obligations that apply to other items (i.e., autonomous vehicles), in cases where data is a necessary input to safety functions.

Data and protected/sensitive information, including: Data protection | Commercially sensitive information | Intellectual property

There are several legal and factual elements that may complicate or hinder the sharing and re-use of data, or at least have the potential to cause a significant chilling effect. This is, for example, the case with data protection, trade secret protection and IP protection, because of the uncertainties that come with these protective regimes:

---

[48] Elinor Ostrom, Governing the Commons: The Evolution of Institutions for Collective Action (Cambridge University Press 1990); Michael J Madison, *Commons at the Intersection of Peer Production, Citizen Science, and Big Data: Galaxy Zoo,* [2014] arXiv:1409.4296 [astro-ph], http://arxiv.org/abs/1409.4296

[49] Charlotte Ducuing, 2020, *Beyond the Data Flow Paradigm: Governing Data Requires to Look beyond Data,* Special Issue: Governing Data as a Resource Technology and Regulation

[50] Commission Delegated Regulation (EU) 2015/962 of 18 December 2014 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide real-time traffic information services, OJ L 157/21.

- The data sets used to train AI could contain personal data that is used in an unlawful manner. At the same time, the technology itself could violate the right to privacy in the course of its function, often in an unforeseeable manner.
- Large and diverse data pools are likely to comprise at least some form of personal or IP-protected data, confronting stakeholders with mixed datasets. With data pools, there is also an additional issue, whereby combined data and analysis reveals unexpected information that could pose a number of problems, from the creation of new personal data, to the discovery of information relating to the core business of participating companies (which could lead to breaches of art. 101 TFEU).

Pathways to dealing with mixed datasets

- Regulatory sandboxes to promote interactions and a better understanding of the mutual effects of law and technology are also foreseen by the Commission in its proposal for AI regulation[51].
- Combination of de minimis rules, fair use exceptions and other measures could alleviate some of the concerns about data that is or may be protected under different legal regimes.[52]
- Industry stakeholders could opt for the use of an entity to serve as a "*screen*" or clearing house.
- A clearing house set-up could also involve multiple competent authorities to enable a holistic approach.[53]

Competition law

Data access and competition

Data access, in particular the failure to provide access to data, may raise competition law concerns: where data is considered essential on the one hand; and on the other, where industry agreements between a limited number of stakeholders may be seen as anti-competitive if other actors are excluded. Alternatively, industry agreements which include data sharing mechanisms may constitute anti-competitive agreements, should business sensitive information be shared (especially between competitors).

Standards and competition law issues

As has been noted in one European Commission requested study, *"there seems to be little doubt that standards set by private associations are caught by the prohibition of Article 81 EC [ex-Article 85]"*.[54] The development of standards by joint undertaking could constitute anti-competitive agreements, which would therefore be prohibited. A distinction seems to be made depending on the source of the standard**. Proprietary standards as well as purely industry-driven standards** could lead to market dominance issues. However, the situation may be different for standards adopted **by standard setting organisations**, which have long been promoted by public policies. Collective standard setting allows for negotiation between participating entities and the adoption of a standard when consensus is reached.

The digital environment is particularly sensitive to competition law issues with respect to (a lack of) standards and interoperability. The fact that some online platforms have grown beyond their initial

[51] European Commission, 2021, COM(2021) 206 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206
[52] See Thomas, J. and Wendehorst, C., (2020) *Response to the public consultation on "A European strategy for data" COM(2020) 66 final*, https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Projects/Data_Economy/ELI_Response_European_Strategy_for_Data.pdf.
[53] https://www.digitalclearinghouse.org
[54] Now Article 101 of the TFEU; See: Falke, J. and Schepel, H. (2000) 'Legal aspects of standardisation in the Member States of the EC and EFTA. Volume 1', European Commission, p. 59.

market to become **vast digital ecosystems** raises, *inter alia,* interoperability issues. Such an ecosystem-based dominance may enable them to impose their own standards and lock users in. The notion of '**data portability** has therefore gained traction, as a means to prevent online platform users from being 'locked in'.

Updating EC Guidelines to account for AI, data and standards

The EC Guidelines will expire in 2022. The Commission opened up consultations in 2020. It was recommended that the updated version should pay due attention to digitalisation, data pooling and algorithms. The updated Guidelines should clarify antitrust liability (e.g., for SSO's and firms employing AI), possibly by revisiting current safe-harbours.

While the consultation and updating process is ongoing, the Commission can clarify these aspects by issuing communications (e.g., the 2017 communication on standard essential patents[55]). Going one step further, the Commission could assess the U.S. Department of Justice example and issue non-binding reviews on whether a specific behaviour is anti-competitive and state its (non-binding) enforcement intentions and the reasoning used in the complex cases that are emerging.

National authorities at Member State level should be further encouraged to issue additional AI and anti-trust guidelines.

The notion of data portability should be further explored to prevent data user lock-in, such as the Commission is already doing, for instance, as part of the proposal for a Digital Markets Act.[56]

An interoperability duty on dominant online platforms should be further assessed.

Liability, including: AI, tort and product liability | AI and contractual liability

Application of liability regimes

The EU White Paper on AI[57] and the accompanying Report on AI Safety and Liability[58], and the New Technologies Formation Expert Report on Liability for Artificial Intelligence and other emerging digital technologies,[59] acknowledge several **challenges** related to tort/product liability and AI. One significant challenge relates to the fact that many of the concepts and terms used still require an **interpretation** by national judges/policy-makers. This can be referred to as the '**AI tort law dilemma'**, which implies that while many initiatives are being adopted at the supranational level, tort law largely remains a national matter. The essential concepts in a potential liability regime will need to be interpreted according to national (case) law. The recent Report of the European Parliament with recommendations to the Commission on a civil liability regime for AI[60] can be taken as an example. Concepts such as due diligence or force majeure, referred to in this document, eventually have to be interpreted by national judges, impeding a harmonised regime.

Against this background, and within the context of industry agreements, one may argue that establishing a potential (EU) regime is already a step ahead of another major challenge, namely the

---

[55] European Commission, 2017, COM(2017) 712 final, https://ec.europa.eu/docsroom/documents/26583
[56] European Commission, 2020, COM(2020) 842 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0842&qid=1624006593011
[57] COM(2020) 65 final, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
[58] COM(2020) 64 final, https://ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020_en_1.pdf
[59] European Commission, 2019, Liability for Artificial Intelligence and other emerging digital technologies, https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608
[60] European Parliament, 2020, Report- with recommendations to the Commission on a civil liability regime for artificial intelligence, https://www.europarl.europa.eu/doceo/document/A-9-2020-0178_EN.pdf

need to **clarify essential concepts** that could actually make a difference when it comes to the allocation of liability.

Clarifying essential concepts

Involve stakeholders to determine the content and meaning of some of the essential concepts, such as '*product*' or '*defect*' within the Product Liability Directive (PLD) or '*due diligence*' (see Report of the European Parliament with recommendations to the Commission on a civil liability regime for AI). Providing clarity on these essential concepts may help to create a clearer and more coherent EU civil (tort) liability regime for AI (cf. within the PLD). This has the potential to reduce risks and increase safety, decrease legal uncertainty and related legal and litigation costs, and enhance consumer rights and trust (cf. European added value assessment civil liability regime for artificial intelligence[61]).

Determine whether horizontal liability legislation may be useful, as opposed to a more sectoral approach. Each sector has its own specific features which could warrant a more granular approach, rather than a one-size-fits-all framework. In this context, industry agreements may provide added value. Action at EU level, such as the adoption of industry agreements, would reduce regulatory fragmentation and costs for AI producers and would help to secure high levels of protection for fundamental and consumer rights in the EU.

Allocation of control and liability/responsibility in a contractual context

Although tort law and product liability are extensively addressed at EU level, the use of **AI in a contractual context** remains an area that needs more attention, especially with regards to the **allocation of control and liability/responsibility in a contractual context:**

- The unique characteristics of AI-systems (e.g., opacity and self-learning abilities) make it challenging to determine which normative frameworks and rules apply when AI is the 'object' of a contract.
- There is some uncertainty regarding the allocation of responsibilities between different economic operators involved in AI system supply chains.
- The cross-border context of the AI supply chain could have an impact on the competent jurisdiction and the applicable rules.

Allocation of control and liability in a contractual context

- Need to map the entire AI supply chain to identify the potential B2B parties involved, their contractual relationships and applicable national legislation (e.g., the Belgian B2B Act contains rules relating to unlawful contractual terms and conditions, the abuse of economic dependence and unfair market practices between undertakings)
- Involve stakeholders in the drafting of model agreements/templates that can be used by other parties in the supply chain (e.g., regarding the allocation of liabilities, exoneration clauses, …) thereby taking into account the national context (e.g., B2B Act)
- Determine which (industry) party is best placed to: (i) identify a risk; (ii) control and minimise the risk; and (iii) manage the risk. Liability could also be allocated to the party best placed to prevent the damage caused by an AI system, thereby incurring the least costs (cf. Calabresi's *cheapest cost avoider*). Several theoretical concepts have been established, but it is now time to

[61] European Parliament, 2020, Civil liability regime for artificial intelligence, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654178/EPRS_STU(2020)654178_EN.pdf

actually determine the responsible party (or parties) via stakeholder involvement (e.g., software developer).

- To determine whether a party (e.g., cheapest cost avoider) can be held liable, the required documents on AI system design, development, operation and decision-making process, etc., need to be stored. All AI system design, development and use information should therefore be maintained, in one form or another. Within industry agreements, the requirement for an Artificial Intelligence Impact Assessment[62] (or the 'Conformity Assessment' in an AI regulation[63]) proposal) could be further developed and operationalised. An alternative is a checklist for organisations, companies and actors involved in the AI supply chain to increase the transparency of AI systems (e.g., identifying actors and determining how they are responsible for transparency or providing explanations).

Ethical aspects of AI

Several documents have already been adopted on ethical aspects of AI. The Ethics Guidelines issued by the High-Level Expert Group on AI and the recent ALTAI assessment list are particularly relevant. However, they are both relatively high level. **More specific and detailed guidelines** for industry parties (e.g., AI developer) would help the implementation of ethical considerations. There is a particular need to determine which ethical guidelines are already part of the legal framework. This analysis would illustrate where additional (regulatory) actions (e.g., within an industry agreement) may be useful.

Further ethical aspects of AI through industry agreements

- Industry agreements should clarify the extent to which AI system designers and users should implement measures or procedures to avoid undesirable consequences that may result from the self-learning/autonomous nature of AI systems.
- Emphasise the need for convergence and coherence between the different (safety) regimes that apply to products and which may incorporate software or be based on AI (e.g., medical devices, tangible products, …).
- There are already several information duties regarding products or services that rely on AI. Nevertheless, AI-specific information duties could also be developed and/or included in industry agreements. Examples could include information obligations for AI systems on: (1) the origin and processing of input data; (2) AI system functionality; (3) how the system was created; and (4) what the benefits and risks are. One example could be to design an 'AI package leaflet/prescription' for end users and customers.
- Adopt/promote the creation of a Chief AI Ethics Officer within IAs. Alternatively, industry agreements could stipulate that each innovation project that receives public funding should contain a Work Package (WP) on ethical aspects of AI.

Standards, interoperability and general IP rights issues

Many authors contend that there is an intrinsic conflict/tension between Intellectual Property Rights (IPRs) and the establishment of standards. For example, the main goal of IPR is to assure property rights, whereas technical standards aim to support specific technologies that are compatible within the larger community. This is even more relevant for interoperability information. In many cases, *"access to interoperability information or to its usage is neither automatic nor free. Access and usage can be*

[62] ECP, 2018, Artificial Intelligence Impact Assessment, https://ecp.nl/wp-content/uploads/2019/01/Artificial-Intelligence-Impact-Assessment-English.pdf
[63] European Commission, 2021, COM(2021) 206 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206

controlled or restricted through three main forms of intellectual property: Patents; Copyright; Trade secrets".[64] As suggested in a 2013 Commission working document[65], facilitating access to interoperability information could be done through an evolution of patent and trade secrets law. This could, for example, involve the granting of automatic licences granting a right to use interfaces, create an interoperability exception, adopt an interoperability directive to deal with the licensing of trade secrets and/or the licensing of patents, or promote non-legislative EU initiatives such as best practices.

European standards setting organisations (SSOs)

- *SSO financing review*

Based on a review of European SSO financing mechanisms and copyright policies, access to interoperability standards could be improved. European standards in CEN/CENELEC are protected by copyright, whereas (most) ETSI standards are available free of charge. Income from the sale of standards is also income for CEN/CENELEC. This model is not sustainable when access to standards is at stake.

- *Open-source standardisation*

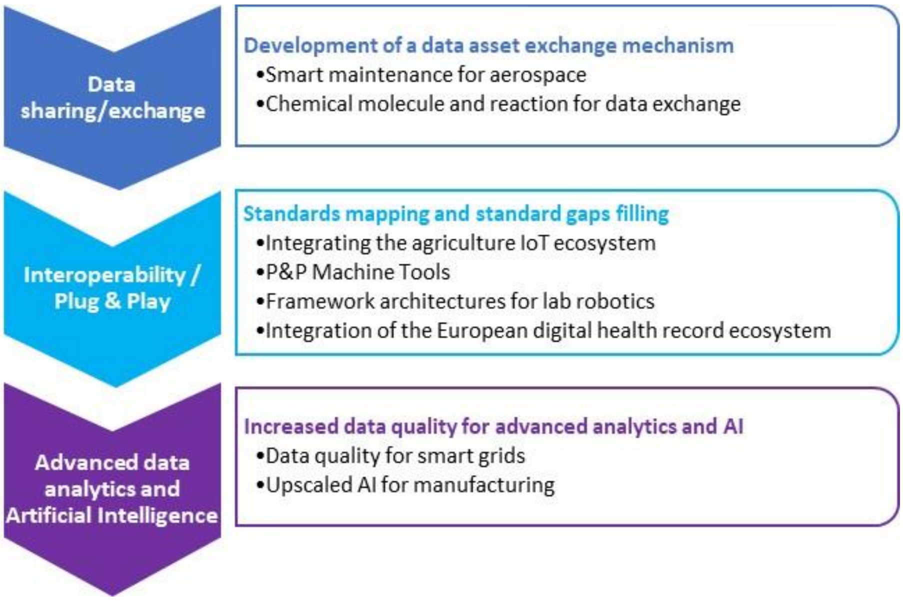
Launch open-source standardisation (OSS) initiatives in European SSOs. OSS may clash with more 'traditional' interoperability standardisation efforts that are based on IP rights. Some open-source initiatives were launched by national standard bodies (e.g., DIN). SSO initiatives to develop common platforms, for example, are valuable. However, Commission-led initiatives have the potential to foster EU-level solutions and prevent fragmentation (see also https://ec.europa.eu/digital-single-market/en/news/standards-and-open-source-bringing-them-together on SSOs and open-source).

[64] Commission Working Document, Analysis of measures that could lead significant market players in the ICT sector to license interoperability information, 2013, SWD(2013) 209 final, p.7.
[65] These suggestions stem from the Commission Working Document, Analysis of measures that could lead significant market players in the ICT sector to license interoperability information, 2013, SWD(2013) 209 final, p. 10-17.

2.2 Key opportunities for the development of new industry agreements

This sub-section presents specific opportunities for each clustered innovation area and broadly scopes their potential economic benefits. They are summarised in Figure 5.

Figure 5: Summary of key opportunities by clustered innovation area



Source: CARSA, 2021

As indicated in the figure above, three main lines of action to orientate corporate and industry efforts have been defined, with one identified for each grouped innovation area. These are best described as strategic tools and frameworks to guide the development of industry agreements, designed to deliver key economic benefits and unlock industrial innovations in each area. In addition, specific examples illustrate how the tools can be made operational.

Data sharing/exchange: developing Data Asset Value Exchange Mechanism

Data sharing and exchange: summary and key opportunities
There is a clear and widespread consensus in virtually all industries about the potential for data sharing and exchange to deliver significant value at every stage in the digital value chain. However, enabling these practices is a complex task, as no clear solutions currently exist. It is more probable, therefore, that the future data exchange market will feature a wide array of competing exchange mechanisms and practices. Examples include: ledger technologies/blockchain; bilateral agreements; altruistic exchanges; private marketplaces; and open data spaces. The structure of the data market can be schematically represented as a series of **three layers: altruistic**; decentralised; and centralised **commercial exchanges and single access point initiatives**. Each layer has its own characteristics and each requires the development of tailored interventions by public authorities? The report strongly supports the claim that the **development of single access points** is the most promising type of initiative currently being implemented and that it offers the **highest potential** for

the development of **public interventions**. This section therefore presents a detailed view on how to **generalise the development of single access points, to upgrade them into clearing houses** with a **business layer**, where possible, and **integrate them into a coherent and networked institution architecture at the European level**.

Following the presentation of the suggested network of **Industrial Data Clearing Houses,** the specific case of **Smart Maintenance for the Aerospace Industry** will be further detailed. Smart maintenance in the field would benefit from the development of a single access point for safety-related information. Leads on how to add a future business layer to this initiative are also presented.

The **development of new standard notations** for the representation of **chemical reaction and molecules** is recommended. The most important system has been reviewed and its shortcomings are identified. This report finds that each notation is currently sub-optimal, resulting in lost opportunities for R&D and the academic world, and in batch chemistry, as well as making data exchange more difficult. As no clear solution is emerging to overcome this barrier, this document recommends the development of a mapping initiative to enable the development of a library of notations that could serve as a basis for the development of new bridging solutions and/or standards.

Introduction: the three layers of Industrial Data Spaces[66]

The complexity of exchanging information excludes the possibility of only one single data exchange mechanism/data space. **The future industrial data space will be composed of at least three layers** (see Figure 6). Their ongoing development will require public intervention.

Table 5: Types of policy actions needed to support the development of Data Asset Exchange Mechanisms by market layer

Layer	Example	Policy intervention
Open Data (altruism)	Data for Good	• Develop a public strategy for the provision and delivery of public information; • Introduce a legal requirement for private players to disclose information produced in the context of a publicly-funded project; • Encourage the introduction of a voluntarily notification procedure for private players and stronger data protection supervision structures.[67]
Decentralised and centralised commercial exchanges	AgriXchange, SnowFlake, consortium Blockchain agreement, undocumented bilateral agreements	• Prevent the development of a monopoly or oligopoly (e.g., Bundeskartellamt vs Facebook) in commercial marketplaces;[68] • Conduct a market study to define market power and disloyal competition practices in the data economy; • Introduce platform regulations with a code of conduct for dominant online platforms and data portability;[69] • Create a unique digital identifier (digital identity) to enable companies in the agricultural sector to interact with public

[66] Please note that this section is focused on the technical analysis. For the pricing mechanisms at play in the data market, refer to the section on “key economic barriers”.
[67] Achim Wambach, 2020, A new competitive framework for the digital economy, https://www.bmwi.de/Redaktion/EN/Publikationen/Wirtschaft/a-new-competition-framework-for-the-digital-economy.pdf?__blob=publicationFile&v=3
[68] Ibidem
[69] Ibidem

		authorities at different levels (e.g., farmers need one identifier for cooperatives, one for European accounts, etc.).[70]
"Single Access Points" as potential future Industrial Data Clearing Houses	MDS, iHub, spatial clearing houses.	• Develop single access point networks where they are lacking in areas with clear potential for economic/societal impact (e.g., single access point for safety data in aeronautics); • Upgrade these access points into "Clearing House" by adding a business layer to the structure to enable the development of new business models and economic transactions; • Develop a coherent institutional architecture to network the different initiatives (e.g., develop an umbrella clearing house for networking, sharing of best practices, developing a repository of repositories.

Public intervention is needed at every level of data exchange/trading. However, due to the characteristics of each level, public policies and interventions for the third level should be prioritised, as they have the highest potential to deliver impact.

In the **altruist layer,** data will be exchanged freely. It involves altruistic initiatives, such as 'Data for Good', which enables benevolent people/organisations to share their information for the common good.

Centralised and decentralised commercial exchanges are by far the most complex layer. They involve a series of potentially competing solutions (competing data marketplaces) or on-the-spot solutions (smart contracts for programmatic exchanges). Concerns over data ownership and technical complexities are creating uncertainty about the market's ability to develop all of the required solutions.

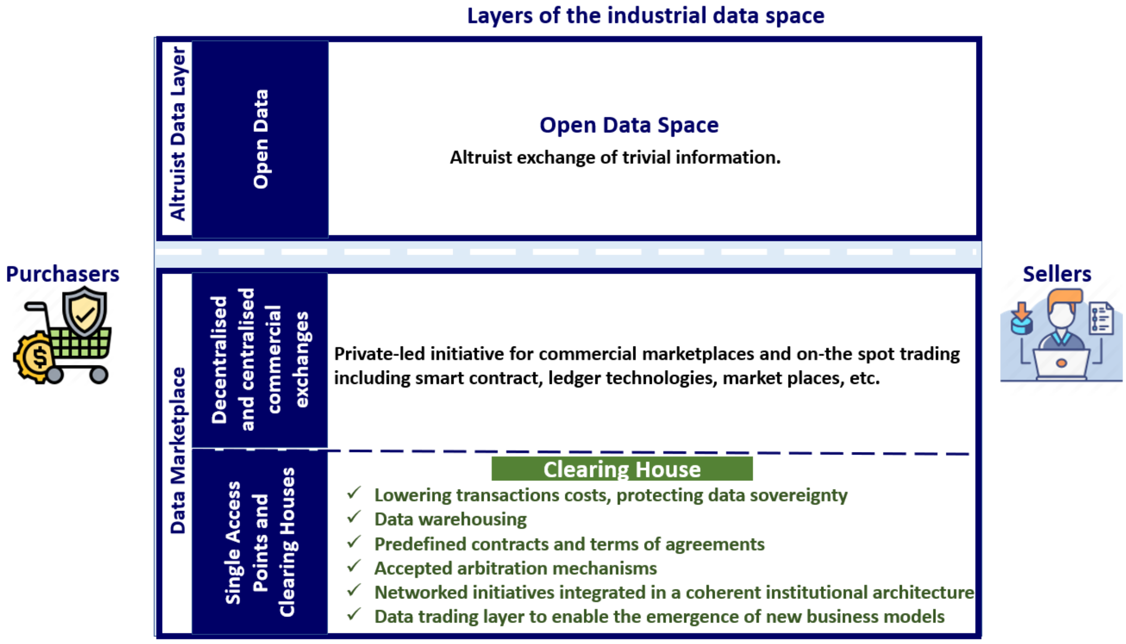
Single access point is not necessarily a separate layer. Indeed, the development of the Skywise initiative in the aerospace sector can be considered an open-data space initiative that could be categorised as a commercial practice. Similarly, a single access point for mobility data could be considered an altruistic exchange mechanism. Consequently, single access points could be seen as technological solutions that provide an interface to access a data space, rather than as a full-fledged market layer.

However, for analytical purposes, this layer is being considered separately. The potential for single access points to offer a repository to access other databases in the field, APIs and data warehousing is very clear. In addition, recent initiatives, such as i-Hub in Australia or the Mobility Data Space in Germany, have started to add business layers to these structures. These improvements are progressively turning these organisations into **Clearing Houses**, enabling them to act as financial and technical facilitators that promote data exchange and the development of new business models.

This report argues that the potential value of this type of initiative is underestimated. More initiatives of this type are recommended and should be developed by market stakeholders and policy-makers.

[70] Comment raised during an interview with a representative of a digital marketplace in the agriculture sector.

Figure 6: Three layers of industrial data spaces and types of Data Asset Value Exchange Mechanisms – the case for the creation of "Data Clearing Houses"



Source: CARSA, 2020. This figure was created using resources from Flaticon.com.

Development of a network of Industrial Data Clearing Houses

Data Spaces are complex ecosystems that bring together a wide variety of players such as certification bodies, manufacturers, software providers, regulating authorities, users, etc. As technology and business models mature, the expectations about what these spaces can bring in terms of innovation and economic impact continues to grow. However, the fragmented nature of these ecosystems is a powerful barrier to streamlined data acquisition, ingestion, sharing and analytics — all of which are crucial for the deployment of a mature industrial data economy.

Many competing initiatives are being launched to find solutions to these challenges. The **Clearing House model** is one of the most promising: "*The **main reason for creating a clearing house** is the desire of users to have a single source for accessing all the available resources. **A spatial data clearing house** is a system to provide this capability serving as a central point for sharing data among data producers and users.*"[71]

The expression "**Clearing House"** has its origin in the financial industry: "*A clearing house is an intermediary between buyers and sellers of financial instruments. It is an agency or separate corporation of a futures exchange responsible for settling trading accounts, clearing trades, collecting and maintaining margin monies, regulating delivery, and reporting trading data*."[72] The expression **Data Clearing House,** on the other hand, originally referred to "the *process of detecting and correcting (or removing) corrupt or inaccurate records from a file exchanged by Mobile Network Operators (MNOs)"*. [73]

[71] Hosep Crompvoets, Wageningen Universtiteit, *National Spatial Data ClearingHouses,* https://edepot.wur.nl/121754.
[72] The two most famous examples of clearing houses are the New York Stock Exchange (NYSE) and the NASDAQ. CFA Institute, Sin dato, *Financial Clearing Houses,* https://www.cfainstitute.org/en/advocacy/issues/central-clearing-houses
[73] RoccoResearch, 03/04/2019, *Results: The leading vendors in Data Clearing 2019,* https://www.roccoresearch.com/2019/04/03/results-the-leading-data-clearing-houses-of-2019/

An **Industrial Data Clearing House** would differ from a clearing house in the financial or telecom industries, but would have to serve a similar purpose. The organisation would act as an intermediary between data providers and users, and would provide data clearing services.[74]

What is an Industrial Data Clearing House? Different perspectives

Different perspectives and definitions exist when defining the role and functions of a clearing house. These visions are not necessarily contradictory; rather, they place emphasis on the specific functions of the clearing house that are most important for the industrial sector they were created for.

- *"A **spatial data clearing house can be** defined as an electronic facility for searching, viewing, transferring, ordering, advertising, and/or disseminating (…) data from numerous sources via the Internet and, as appropriate, providing complementary services. Such a clearinghouse usually consists of a number of servers that contain information (metadata) about available digital data. A clearinghouse is based on a distributed network of people (spatial data suppliers, managers and users) linked electronically (Clinton, 1994; FGDC, 2002). The term 'distributed system' refers to a distributed collection of users, data, software and hardware, whose purpose is to meet some predefined objectives (Bishr and Radwan, 2000). The clearinghouse allows suppliers to make known what (…) data exist, the condition of these data and instructions for accessing these data. Each data supplier describes available data in an electronic form and provides these descriptions (or metadata) to the network using a variety of software tools. Additionally, the data supplier can offer access to his produced data. Users can discover who has what spatial data and their type and quality (Radwan, 2002);*
- *"A **national clearinghouse for spatial data** can be considered as the access network of a [industrial data space], which focuses on the facilitation of (…) data discovery, access and related services. It is not a national repository where datasets are simply stored. It can be seen **as a one-stop shop for all** (…) data, sourced from governmental agencies and/or industrial bodies (Crompvoets et al., 2004)";*[75]
- The **INSPIRE** Architecture and Standards working group (2002) depicts **clearing houses as portals** that feature a suite of commonly used services that provide a starting point and a gateway to the web for a user community;[76]
- The **Mobility Data Space initiative** in Germany defines clearing houses as: "*The clearing house is the system's central logging component and records transactions made within the distributed system in order to make them available to the relevant parties for purposes of billing and quality analysis at a later point in time*."[77]

The commonalities between these definitions enables us to produce a **Definition of an Industrial Data Clearing House**:

- An **Industrial Data Clearing House** is a **single access point** to industrial data for industrial data ecosystems. It can be set up to operate at a national, sectoral, European or international level. These clearing houses act as a portal for remote access and data exchange. They act as a broker for data exchange, they structure the data space to facilitate the retrieval of industrial data, they maintain logs and record information to track activities, and they protect data sovereignty. They also typically add a **data marketplace layer**.

[74] It is not realistic to expect an industrial data clearing houses to provide granular clearing services as in the telecom business as it would require enormous resources and advanced expertise. However, clearing houses can provide high-level clearing services through monitoring participation to the dataspace and acting upon reporting of fraudulent or malicious activities (e.g. through exclusion from the dataspace.

[75] Hosep Crompvoets, Wageningen Universtiteit, *National Spatial Data ClearingHouses*, https://edepot.wur.nl/121754.

[76] INSPIRE, 10/03/2002, *Inspire architecture and standards position paper*, *https://inspire.ec.europa.eu/reports/position_papers/inspire_ast_pp_v4_2_en.pdf*

[77] Mobility Data Space, 15/09/2021, *Architecture and components*, https://www.mobility-data-space.de/en/architecture-and-components.html

A Clearing House would therefore act as a sector-specific **single access point** with a **business layer** to enable the development of new business models for data trading. However, this does not mean that the single access point would warehouse all information in one server farm or in one single data space. Experience in the mobility data space, for example, has shown that these objectives are not feasible — except in the case of niche markets. It is more feasible to use single access points to integrate some information and to provide clear solutions (e.g. metadata repositories, index of databases) to make data retrievable.[78]

The opportunities and potential that clearing houses offer has led to the launch of many initiatives to establish this type of body in different industrial settings.

In the **Spatial Data Space** alone**,** a total of **83 national clearing houses for spatial data** have already been established since April 2005 on the internet. Regional and international houses have also been identified. However, clearing houses tend to be created at national level, because they align better with the characteristics of national data space ecosystems.[79]

Clearing Houses and National Data Spaces [80]

Industrial data spaces tend to develop into **national, local, sectoral or corporate data spaces**. This means that these spaces are deeply defined and limited by the specific features of the digital and physical infrastructure required to develop such a data space: servers, standards, policies, laws, relation between companies, path-dependency phenomenon, etc. An integrated data space is therefore more than a combination of infrastructure, connected users and value-added services. It also involves "*issues regarding interoperability, policies and networks."*

A **clearing house** acts as an **access network that is operational within a specific data space.** To be successful, a clearing house must connect the different components of the national data ecosystem: standards, policies, culture. For example, if we take the example of spatial data: "*A national clearing house for spatial data can be considered as the* ***access network of a national SDI [***Spatial Data Infrastructure***]****, which focuses on the facilitation of spatial data discovery, access and related services. It is not a national repository where datasets are simply stored. It can be seen as a one-stop-shop for all national spatial data, sourced from governmental agencies and/or industrial bodies."*

Another example is the emerging **Innovation Hub for Affordable Heating and Cooling (i-Hub)**, a clearing house for the Australian **Smart Building market**.[81] It is led by a consortium that includes several Australian universities, AIRAH (the Australian Institute of Refrigeration, Air Conditioning and Heating) and the CSIRO (Commonwealth Scientific and Industrial Research Organisation). The objective of the initiative is to increase the availability of data in Smart Buildings, due to the increased deployment of smart devices and the exponential growth in the data those devices produce. "*The i-Hub is curating the* ***Australian Smart Buildings Data Clearing House*** *as a* ***single location for accessing a wide range of energy and building data.*** *It aims to increase the quality and value of data sets, and empower Australian businesses to develop new data analytics services*. " [82]

[78] Interview with a representative of the Mobility Dataspace marketplace.
[79] Hosep Crompvoets, Wageningen Universtiteit, *National Spatial Data ClearingHouses,* https://edepot.wur.nl/121754
[80] Ibidem
[81] HVAC&RNews, 16/11/2020, *Videos explore i-Hub Smart Building Data Clearing House,* https://www.hvacrnews.com.au/news/the-i-hub-project-has-released-two-videos-to-promote-and-explain-its-smart-building-data-clearing-house-initiative/
[82] iHub, 14/09/2021, *Smart Building Data Clearing House,* https://www.ihub.org.au/ihub-initiatives/smart-building-data-clearing-house/

The **Mobility Data Space** initiative led by Acatech in Germany is another revealing example. It involves the development of a **Data Clearing house**[83] in the mobility data space: "*Under the stewardship of the German Academy of Science and Engineering (acatech), a stakeholder dialogue with high-ranking representatives from the automotive, aviation, public transport, freight transport and logistics industries, new mobility service providers and platforms as well as public sector officials, consumer protection and data protection officers is currently taking place in order to jointly create a collaborative trust ecosystem as the basis for the development of this mobility data space (…) the project has won decisive supporters from the automotive industry, which for a long time have been hesitant to join, as well as key stakeholders in the aviation and rail sectors. Part of the success lies in the fact that political supporters ranging up to Chancellor Angela Merkel have often stressed that the prerequisite for the functioning of the mobility data space is the voluntary participation of a variety of participants in the mobility sector.*" [84]

The German federal government granted an €18 million budget to the Academy of Science (Acatech) to set up a new institution to support the development of a mobility data space. Building on the International Data Spaces Association (IDSA) framework, the German initiative will provide the infrastructure required to share data safely and securely. Participants will maintain data ownership and will store the data on their server. The data can be accessed by other certified participants through interfaces. The platform will also include interfaces and billing mechanisms. The strategy behind this initiative is to make the platform attractive enough to reach a "threshold", so that automotive participants will have no choice but to participate if they want to remain competitive. This organisation is developed to: tackle a market need (new business opportunities for big data in transportation); remove technology-related barriers; and offset market risks.[85]

The cost of setting-up an Industrial Data Clearing House

The creation of a clearing house[86] faces two main challenges — the cost of setting one up and access to financial support. Public intervention and funding mechanisms are therefore needed. Information on past initiatives provides insight on the costs involved.

- According to INSPIRE, "*Experiences with clearing house[s] and other applications showed that **€1.5-1.9 million/year were needed.** This included management costs, GIS and internet application development, training, hardware, network server, and the pilot project*."[87]
- In addition, the Wageningen Universiteit notes that: "*Currently, around 500 (non-corporate) SDCs [Spatial Clearing Houses] have been established and it is expected that many more SDCs will be set up in the future. This indicates **that on a global scale hundreds of millions are spent yearly on SDC activities**. Up to now this large investment has rarely been audited or evaluated. A study conducted by the Urban and Regional Information Systems Association cited that while the costs of SDC projects may be relatively easy to assess and highly 'front-loaded', **the benefits are often very difficult to measure** and **may not emerge until well into the life of the SDC** and depend on other factors coming into play.*"[88]

[83] Handelsblatt, October 2020, *Merkel Drängt autokonzerne: BMW, Daimler und VW sollen Datenschatz teilen,* https://www.handelsblatt.com/politik/deutschland/autogipfel-merkel-draengt-autokonzerne-bmw-daimler-und-vw-sollen-datenschatz-teilen/26308418.html?ticket=ST-1376973-6msjtMrmiOfcreSP5PK1-ap2
[84] APCO Worldwide, 7/12/2020, *Has Germany set the European Transport sector on the path to a digital transformation?,* https://apcoworldwide.com/blog/has-germany-set-the-european-transport-sector-on-the-path-to-a-digital-transformation/
[85] Handelsblatt, October 2020, *Merkel Drängt autokonzerne: BMW, Daimler und VW sollen Datenschatz teilen,* https://www.handelsblatt.com/politik/deutschland/autogipfel-merkel-draengt-autokonzerne-bmw-daimler-und-vw-sollen-datenschatz-teilen/26308418.html?ticket=ST-1376973-6msjtMrmiOfcreSP5PK1-ap2
[86] Hosep Crompvoets, Wageningen Universteiteit, *National Spatial Data ClearingHouses, https://edepot.wur.nl/121754*
[87] INSPIRE, 10/03/2002, *INSPIRE Architecture and Standards Position Paper,* https://inspire.ec.europa.eu/reports/position_papers/inspire_ast_pp_v4_2_en.pdf
[88] Hosep Crompvoets, Wageningen Universteiteit, *National Spatial Data ClearingHouses, https://edepot.wur.nl/121754*

- The **Mobility Data Space** clearing house received initial funding of €18 million from the German state.[89] In addition, a further €3 million was allocated to the development of the Mobility Data Marketplace, which will be added to the data space to enable it to operate clearing house functions.[90] It is not clear, however, what costs stakeholders are required to cover.
- The **iHub initiative** has a total budget of 16.72 million Australian Dollars (around €10.36 million). The Australian government, through the Australian Renewable Energy Agency, has provided a subsidy of 6.48 million Australian Dollars (around €4.04 million), which represents around 39% of the total budget.[91]

Clear patterns and frameworks to guide the development of clearing houses are difficult to define: *"**National clearing house implementation can vary enormously**. These factors determine to what extent the clearing house retains control over data (…)* ." [92] In other words, there is not a single model that defines how to build a clearing house. The pattern for the implementation of this type of organisation will depend on the institutional architecture, the adopted meta-data-standards, the digital maturity of the industrial ecosystem, the legal ecosystem, the characteristics of the national data space, public policies, etc.

However, in spite of these disparities, a list of **Best Practices** can be drawn from past experience to support and help stakeholders that are interested in creating a clearing house.

Best Practices for the implementation of an Industrial Data Clearing House

- As the development of a Clearing House is closely linked to the specific features of the connected data space, these institutions are typically developed at a **sectoral level** and tend to start at a **national level**. From a European Union perspective, the aim should be to develop a **transnational network of clearing houses**;
- The development of a clearing house requires the broadest possible involvement of **stakeholders involved in the industrial data ecosystem.** They include software developers, research institutes, certification bodies, regulation authorities and industrial players from the entire value chain;
- Technical solutions to solve concrete problems should rely on **existing standards whenever possible**. The Mobility Data Space uses the Gaia-X and IDSA standard and the existing Mobility Data Marketplace.[93] When confronted with specific difficulties, the i-Hub model made use of the Switch Automation Platform, which is proven infrastructure. "*The complexity was reduced by reprovisioning the Switch Platform in its entirety to reduce the amount of new software development. The DCH is a white-labelled version of Switch launched into its own instance of Microsoft Azure. The amount of new development was limited to the MQTT driver required to consume CSIRO data and the APIs that support third party consumption of DCH data*";[94]
- **Public bodies have a strategic role to play in the removal of key barriers to the development of a clearing house.** Their role involves the provision of financial and policy-making support to

[89] Handelsblatt, October 2020, *Merkel Drängt autokonzerne: BMW, Daimler und VW sollen Datenschatz teilen,* https://www.handelsblatt.com/politik/deutschland/autogipfel-merkel-draengt-autokonzerne-bmw-daimler-und-vw-sollen-datenschatz-teilen/26308418.html?ticket=ST-1376973-6msjtMrmiOfcreSP5PK1-ap2
[90] APCO Worldwide, 7/12/2020, *Has Germany set the European Transport sector on the path to a digital transformation?,* https://apcoworldwide.com/blog/has-germany-set-the-european-transport-sector-on-the-path-to-a-digital-transformation/
[91] ARENA, 15/09/2021, *Affordable Heating and Cooling Innovation Hub (iHub),* https://arena.gov.au/projects/affordable-heating-and-cooling-innovation-hub-ihub/
[92] Hosep Crompvoets, Wageningen Universtiteit, *National Spatial Data ClearingHouses,* https://edepot.wur.nl/121754
[93] Mobilitätsdatenmarketplatz, sin dato,*about the MDM*, https://www.mdm-portal.de/was-uns-aktuell-bewegt/
[94] iHub, Report #001: Final DCH2 Knowledge Sharing Report, *15/12/2020*, https://ihub.org.au/wp-content/uploads/2021/01/DCH2-Final-Sub-project-KS-Report.pdf

help develop a willing coalition: For example, for the MDS: "*Part of the success lies in the fact that political supporters ranging up to Chancellor Angela Merkel have often stressed that the prerequisite for the functioning of the mobility data space is the voluntary participation of a variety of participants in the mobility sector";* [95]

- "***The main critical (clearing house-internal) factors identified for the success of national clearing house** development are: introducing web services to clearing houses (e.g. view services), continuous funding, having a clear vision of the clearing house function, providing good communication channels, building user-friendly interfaces with clear terminology, creating trust in the management environment, motivating data suppliers and web service providers to participate within the clearing house, and ensuring that a motivating environment is created so that clearing house coordinators update the clearing house more regularly*";[96]
- Before launching new clearing house initiatives, it is **suggested to fund research to carry out cross-comparisons analyses** to identify key factors to ensure the success or failure of such an initiative: *"(…) knowledge could be used for the support of future implementation strategies. Factors for consideration could be societal, for instance legal, economic, technological, historical, cultural, demographic, environmental and institutional characteristics of a country, or clearinghouse-internal, such as the network architecture, availability of view services, type of search mechanisms and funding stability*. This complexity also excludes the compilation of simple best practices to replicate across countries;[97]
- A Clearing House is dependent on the development of an embedded **marketplace** for the publication, display, retrieval and trading of data.

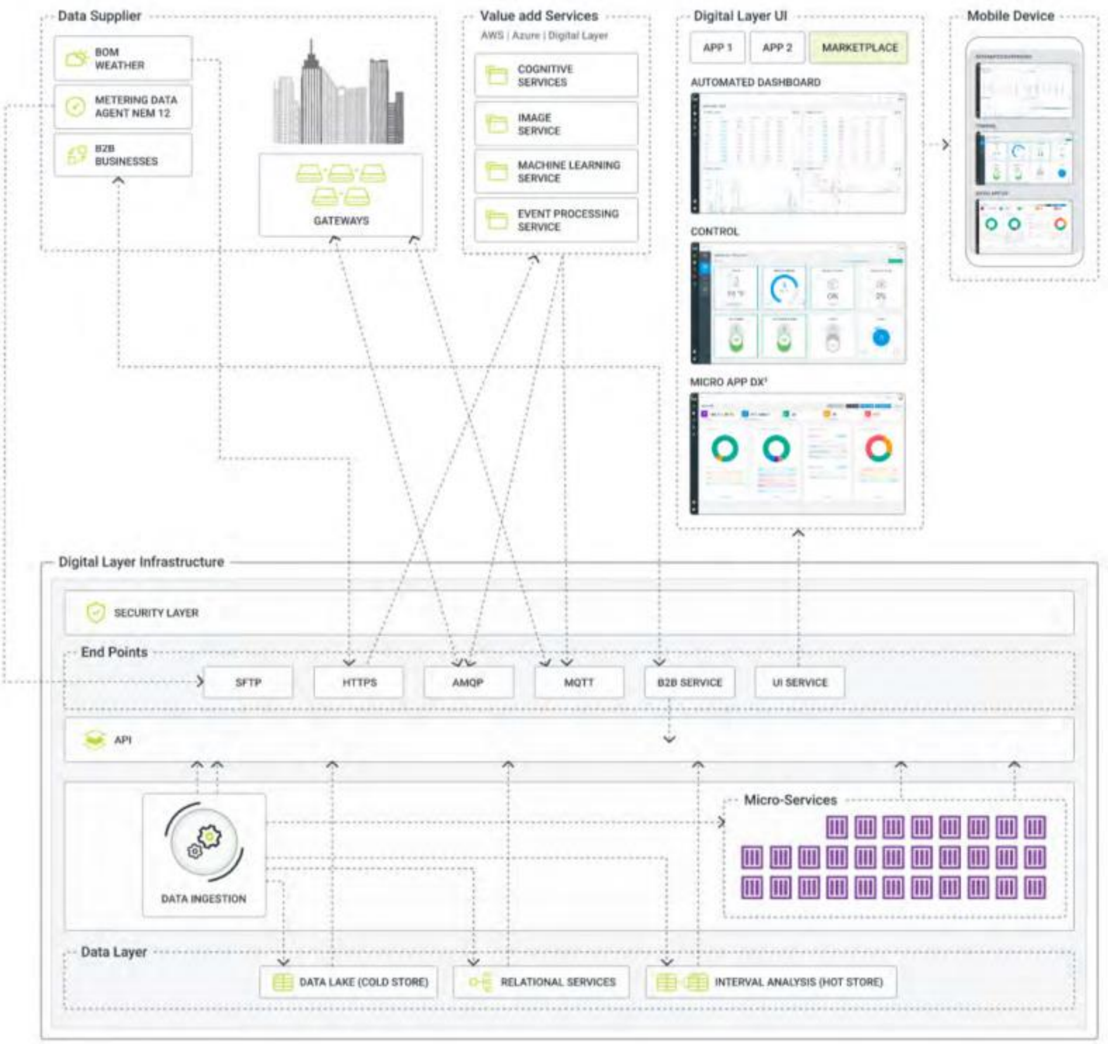
Clearing Houses can be seen as **Open Industrial Data Spaces** that are exclusive to value chain members. Developing this type of institution is complex and faces important barriers, such as: data security; voluntary data exchange; uniform storage formats; competitive neutrality; and missing business models.[98] To date, two models of **clearing house architecture** exist in:

- **Open Data Platform** (i-Hubs, Spatial Clearinghouses);
- **Decentralised network** (Mobility Data Space — MDS).

In **the Open Data Platform** model, the solution is to create a centralised platform with a web portal for common access, data ingestion and acquisition. However, the openness of the model places more constraints on the type of information that can be exchanged. Figure 7 below provides a summary and description of the iHub's architecture.

[95] APCO Worldwide, 7/12/2020, *Has Germany set the European Transport sector on the path to a digital transformation?,* https://apcoworldwide.com/blog/has-germany-set-the-european-transport-sector-on-the-path-to-a-digital-transformation/
[96] Hosep Crompvoets, Wageningen Universtiteit, *National Spatial Data ClearingHouses, https://edepot.wur.nl/121754*
[97] Ibidem
[98] DotMagazine, 01/04/2021, *Data as the key to connected mobility,* https://www.dotmagazine.online/issues/digital-acceleration/digital-trade-routes/connected-mobility

Figure 7:Architecture of the Open Data Platform for clearinghouse (i-Hub)



Source: i-Hub, 2020[99]

The decentralised network, on the other hand, emphasises the importance of data sovereignty, as data is still warehoused on the data providers' servers. "*One-stop-shop window with data packages that individuals have voluntarily made available and that can be purchased by other mobility service providers in search for new business models. In addition to weather and traffic data, this could include information on construction sites in the city, timetable data or even passenger and personal mobility data*."[100]

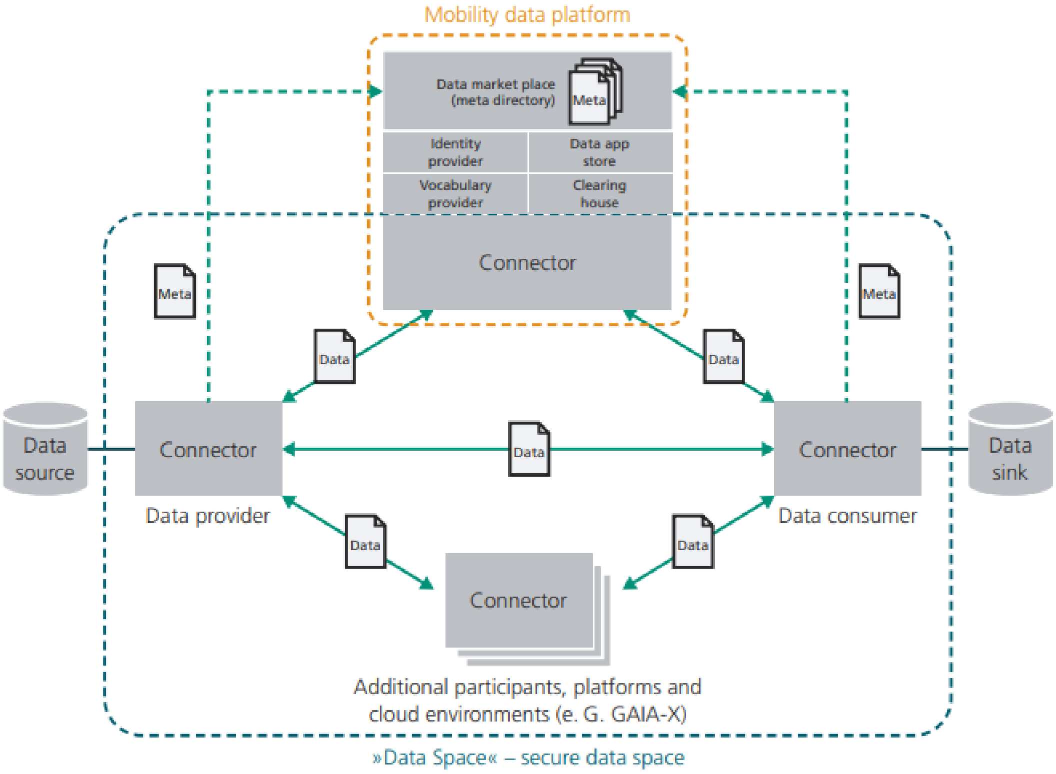
The MDS solutions can be described as "*a mobility data space established across the networked connectors*". It is not a centralised platform, but a network that can be expanded. The data remains under the ownership of the sender and is stored on the sender's network/servers. The system enables secure protection against unrequested access. For data analysis or fusion, access must be done through data apps. In addition, this architecture enables the integration of data from user databases running outside of the connectors. Finally, the architecture has a control layer within the connector to track

[99] iHub, 15/12/2020, *DCH2 reports: learnings from the use of FDD, analytics and demand response applicatiosn in the 3 trial buildings (hosted on the digital layer software platform DCH 1.2)*, https://ihub.org.au/wp-content/uploads/2021/01/DCH2-Technical-Report-Learnings.pdf

[100] APCO Worldwide, 7/12/2020, *Has Germany set the European Transport sector on the path to a digital transformation?*, https://apcoworldwide.com/blog/has-germany-set-the-european-transport-sector-on-the-path-to-a-digital-transformation/

activities, it monitors compliance with governance rules, and ensures that only aggregated results can leave the platform, not the original dataset.[101]

Figure 8: Architecture of the MDS

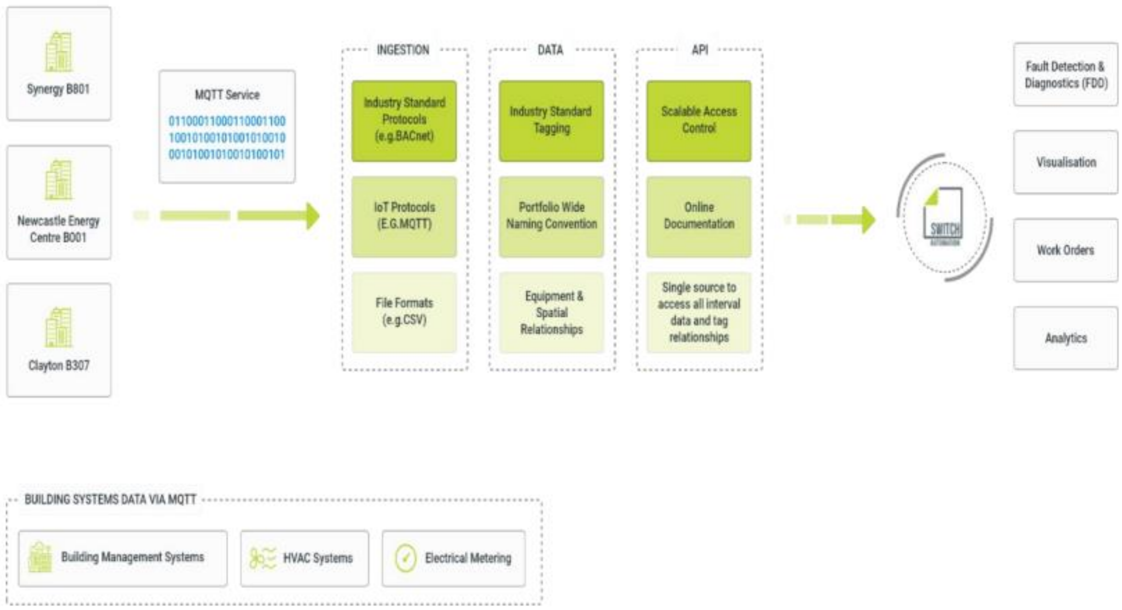


Source: Fraunhofer, 2020[102]

However, both models face the difficult constraints of data ingestion and tagging, as summarised by the i-Hub project.

[101] Mobility Data Space, 15/09/2021, *Architecture and components*, https://www.mobility-data-space.de/en/architecture-and-components.html
[102] Fraunhofer/BAST, 01/03/2020, *Mobility Data Space*, https://www.mobility-data-space.de/content/dam/ivi/mobility-data-space/documents/Mobility_Data_Space_2020_EN_neu.pdf

Figure 9: Clearinghouses and data ingestion, the I-Hub case



Source, i-Hub, 2020[103]

Creating such an organisation is a complex and costly, yet it is also a worthy task. Based on the iHub and MSD model, we can identify two different pathways to create a clearing house.

The Innovation Hub for Affordable Heating and Cooling (iHub) — Australia[104]

Although it is still under development, the initiative provides **a 3-step guide** on **how to develop a clearing house using the Open Data Model**. To this 3-step process, we can add **coalition building** as a preliminary step.

1. **Coalition building**: bringing together interested partners/stakeholders to create the i-Hub consortium;
2. Development of an **Open Data infrastructure** for the Data Clearing House to host data from the ecosystem of stakeholders;
3. *"Development of a **Proof of Concept** Across 3CSIRO campus applying fault detection and diagnostics (FDD) to identify energy and other opportunities." (…) "Objective is to demonstrate Fault Detection and Diagnostics (FDD), visualization, work orders and analytics to provide value from building data to stakeholders";*
4. *"Development of a **marketplace** and creation of an **ecosystem** with demonstration of impact."*

[103] iHub, 15/12/2020, *DCH2 reports: learnings from the use of FDD, analytics and demand response applicatiosn in the 3 trial buildings (hosted on the digital layer software platform DCH 1.2)*, https://ihub.org.au/wp-content/uploads/2021/01/DCH2-Technical-Report-Learnings.pdf

[104] iHub, Report #001: Final DCH2 Knowledge Sharing Report, *15/12/2020*, https://ihub.org.au/wp-content/uploads/2021/01/DCH2-Final-Sub-project-KS-Report.pdf

The Mobility Data Space (MDS) — Germany[105]

The MDS model is still at a very early stage and does not provide a clear list of steps to guide development. However, looking at how the initiative is planned to proceed over the coming years, we can observe similarities and can **identify 5 key steps:**

1. **Launch of the initiative** by the German government with an original funding of €18 million and development of the coalition (consortium). Leading role given to the Acatech academy of science;
2. **Creation of the data room by Acatech** using pre-existing solutions: Gaia-X, IDSA and the Mobility Data Marketplace from the German federated Highway institute (for an embedded marketplace);
3. Creation of the "**DRM Datenraum Mobility GmbH**" as a non-profit organisation to manage the data room. According to the Handelsblatt, one can expect the launch of a tender by Acatech to hand over the management of the DatenRaum to a third party;
4. Launch of **regular operations** in October 2021;
5. Development of a **framework agreement** for participation and data usage agreements. The framework will be presented to industry stakeholders by December 2021.

The i-Hub project provides a list of the key challenges for the creation of a Clearing House, as well as leads on how to remove them. Building on additional desk research, the same analysis can be made for the MDS:

[105] These steps were drawn based on the following sources:

1) IDW, 18/05/2021, *Vernetzer Verkehr: acatech gründet Trägergesellschaft „DRM Datenraum Mobilität GmbH" als Non-Profit-Organisation*, https://idw-online.de/de/news769008;
2) Handelsblatt, 17/11/2020, *Datenschatz Verkehr: VW und Daimler wollen die Pläne der Kanzlerin unterstützen*, *https://www.handelsblatt.com/politik/deutschland/autogipfel-datenschatz-verkehr-vw-und-daimler-wollen-die-plaene-der-kanzlerin-unterstuetzen/26631022.html?ticket=ST-512540-c5LOzyYgi9jBd3MMpVen-ap4*

Table 6: Challenges faced for the development of a clearing house and leads on how to address them

Challenge	Description (from IHub) [106]	Strategy to overcome it: i-Hub model[107]	Strategy to overcome it: Acatech
Complexity	• The complexity of the task requires the building of in-depth domain knowledge that even MNOs often lack. • Complex data ingestion of different data standards and protocols in one single standard to allow comparison, purchase and access. • Unique blend of skills needed to build such a platform (cybersecurity, big data, m2m communication, embedded computing, etc).	• Using proven solutions to build an ecosystem, rather than developing something new from scratch. In this case, the Switch Automation Platform infrastructure was chosen. •" The *complexity was reduced by reprovisioning the Switch Platform in its entirety to reduce the amount of new software development. The DCH is a white-labelled version of Switch launched into its own instance of Microsoft Azure. The amount of new development was limited to the MQTT driver required to consume CSIRO data and the APIs that support third party consumption of DCH data*."	• Using proven solutions and standards: IDSA and Gaia-X for data storage and transfer protocol, Mobility Data Marketplace for exchange and trading, and DATEX II for data model. • However, ongoing experience reveals the emergence of additional complexities linked to the need for additional functionalities, performance requirements and data quality.[108]
Competitive Landscape	*"Noisy and confusing competitive landscape with thousands of companies providing info, data, dashboard and analytics with a handful classified as data platforms with requested functionalities"*	• The problem remains. Organisation of awareness-raising events and past data exchanges helped to bring stakeholders together but is not sufficient to streamline the competitive landscape. • The problem is expected to resolve itself over time as more IT skills enter the market.	• The problem is not fully resolved. However, thanks to the intervention of public authorities, key automotive manufacturers that are vital for the initiative — such as Volkswagen — have accepted to join the initiative.[109]
Customer Understanding	Lack of experience, understanding of complexity of the problem or what the owner can do to help. Lack of understanding of the benefit of open data and exchange and lack of budget for experimentation.	• Awareness-raising activities during industrial fairs and other events. These activities have proven to have positive impact as demonstrated by the number of requests made to the DCH by industry stakeholders to use it for their RFPs. • Nonetheless, the project team identified that these initiatives were insufficient and underestimated the original issue. More effort is needed to educate professionals on the technical complexity of the task. Operational teams should be specifically	• Presentation of the initiative and its potential to different events by official organisations.[111]

[106] iHub, Report #001: Final DCH2 Knowledge Sharing Report, *15/12/2020*, https://ihub.org.au/wp-content/uploads/2021/01/DCH2-Final-Sub-project-KS-Report.pdf
[107] Ibidem
[108] Fraunhofer/BAST, 01/03/2020, *Mobility Data Space*, https://www.mobility-data-space.de/content/dam/ivi/mobility-data-space/documents/Mobility_Data_Space_2020_EN_neu.pdf
[109] TechScurry, 17/03/2021, *Angela Merkel urges: Car companies should join the mobility data room*, https://techscurry.com/general/angela-merkel-urges-car-companies-should-join-the-mobility-data-room/17995/

		targeted, in particular because they have been identified as being resistent to Open Data initiatives. [110]	
Commercial Model	*"Difficulties to establish a correct commercial model. The pricing model success must support market adoption, DCH profitability, partner profitability and return value and benefit to the customer."*	• Several models were developed (but are not yet available for public scrutiny), although no final decision has been made. • Development of a marketplace embedded in the Clearing House. The marketplace is based on three tiers of services: standard offering for access to large datasets via the DCH querying tool; premium services for access to very large datasets and fast retrieval; Application Programming Interface (API) to access both previous services and to provide additional services to build custom applications, dashboards and new services.[112] • Business models are therefore expected to appear over time, thanks to the new opportunities that the clearing house will unlock. • Provision of model contracts and terms of conditions to enable data exchange.	• In terms of the iHub, the clearing house is expected to lead to the development of new opportunities and business models. The marketplace is important to develop market transparency and grant the DCH a broker function. • Private sector stakeholders can develop private marketplaces and connect them to the MDS. Volkswagen, for example, is currently developing one.[113]

[111] See for example: Federal Ministry of transport and digital infrastructure, sin dato, *Open data for smart mobility conference*, https://www.bmvi.de/SharedDocs/EN/Documents/K/open-data-documentation.pdf?__blob=publicationFile
[110] IHUB, 18/10/2020, *i-Hub Switch Data Clearing House, https://ihub.org.au/wp-content/uploads/2021/01/DCH2-Lessons-Learnt-Final.pdf*
[112] Mobility Data Space, 15/09/2021, *Architecture and components*, https://www.mobility-data-space.de/en/architecture-and-components.html
[113] Handelsblatt, 17/11/2020, *Datenschatz Verkehr: VW und Daimler wollen die Pläne der Kanzlerin unterstützen, https://www.handelsblatt.com/politik/deutschland/autogipfel-datenschatz-verkehr-vw-und-daimler-wollen-die-plaene-der-kanzlerin-unterstuetzen/26631022.html?ticket=ST-512540-c5LOzyYgi9jBd3MMpVen-ap4*

The table above lists the main barriers and provide hints on how to address them.[114] However, the question of the key benefits brought by such an organisation are not easy to detail: ***"(…) the benefits are often very difficult to measure*** *and* ***may not emerge until well into the life of the [Clearinghouse]*** *and depend on other factors coming into play."*[115]

Benefits of Data Clearing Houses

The expected benefits can vary depending on the sector in which the clearing house has been created. We can however identify a few examples of the expected benefits, based on the analysed cases:

- **Lower transaction costs** are the main benefit expected from this type of integrated data space. Data marketplaces and data spaces generally help to significantly reduce transaction costs (costs of searching and locating data, cost of initiating and implementing contractual obligations between the data provider and data recipient). Helpful elements include the provision of the platform or the protected trading space itself, uniform data formats, consistent metadata description, model contractual agreements, ... In short, everything that simplifies the exchange relationships between the data provider and x number of data users;[116]
- For **Spatial Clearing Houses**, the expected benefits include: reduced data duplication; improved data sharing and distribution (main benefits); cost savings linked to the previous points; increased consumption (use) of spatial data in the value chain; and improved data market transparency;[117]
- The **iHub** initiative in the Australian Smart Building Sector is expected to generate energy savings, improve efficiency, reduce costs, provide decision-making support and inform policy;[118]
- The **MDS** "(…) *presents a unique opportunity for businesses to connect with leading mobility companies across all modes of transport, engage on the design of a common data ecosystem from the very beginning, enhance innovation through unique access to mobility data for the creation of new business models and find additional means for the monetization of existing company data*."[119]

However, another important challenge remains. The fragmentation of data ecosystems, especially in Europe, creates the risk of additional institutional complexity when creating a **network of Industrial Data Clearing Houses**. As identified in the MDS case: "*The challenge moving forward will be to develop common ground rules in adherence to European data protection regulations and identify opportunities for both businesses and private users to monetise the data sets made available*." [120]

The design of future initiatives should therefore reflect the broader policy context and relevant organisational architectures.

Based on thought experiments and drawing on the key learnings from case studies, we can already identify the main characteristics of the future Clearing House ecosystem, as follows.

[114] A survey ran by the Wageningen Universiteit provides a list of key barriers and the % of voters that identified them as important. These barriers were already covered in the table, under a different name, and are listed here for further clarity: Costs and funding (80%) ; Institutional Problems (33%); Lack of specialised data managers (25%); Data standardisation (23%) ; Lack of harmonised reference systems (3%); Liability problems (12%) ; Inadequate internet Bandwith (16%). Hosep Crompvoets, Wageningen Universtiteit, *National Spatial Data ClearingHouses, https://edepot.wur.nl/121754*

[115] Hosep Crompvoets, Wageningen Universtiteit, *National Spatial Data ClearingHouses, https://edepot.wur.nl/121754*

[116] Interview with a representative of the Mobility Dataspace marketplace.

[117] Hosep Crompvoets, Wageningen Universtiteit, *National Spatial Data ClearingHouses, https://edepot.wur.nl/121754*

[118] iHub, 14/09/2021, *Smart Building Data Clearing House, https://www.ihub.org.au/ihub-initiatives/smart-building-data-clearing-house/*

[119] APCO Worldwide, 7/12/2020, *Has Germany set the European Transport sector on the path to a digital transformation?,* https://apcoworldwide.com/blog/has-germany-set-the-european-transport-sector-on-the-path-to-a-digital-transformation/

[120] Ibidem

Institutional architecture for an Industrial Clearing House Network

Overall, we can identify three layers of governance, with an additional layer of supporting initiatives and organisations.

European Level: EU Main legislative acts and special bodies

The Clearing House institutional architecture will adhere to the characteristics of national data spaces and follow the main lines that can already be observed from the existing initiatives in the field. Clearing houses will bring together large consortiums of private-public stakeholders. A neutral organisation will be created to manage the Clearing House. These institutions will have to coordinate their efforts. This report also stresses the importance of creating an **Umbrella Clearing House** at a European level.

The Umbrella organisation's role would be to:

- Develop a repository of all single access points at sectoral level;
- Provide an arena for the exchange of best practices and key learnings;
- Develop research programmes based on cross-learnings and enable cross-fertilisation.

The main responsibility for managing the Umbrella organisation should be given to the Data Innovation Board. The participation of ENISA and the EU Data Protection Board (especially for the development of research projects) is also recommended. Data Clearing Houses will have to enforce European legislation and can therefore benefit from the participation of these three bodies. The same observation can be made the other way around, as these institutions should also benefit from the lessons learned from Clearing Houses initiatives.

Network of Sector-specific Industrial Data Clearing Houses

These clearing houses would fulfil two functions: they would act as single access points (though a mix of API access, metadata repositories and data warehousing); and would enable the development of business activities and business models. Clearing Houses can provide model contracts and terms of agreements for data exchange and arbitration mechanisms for litigation claims (exclusion from data spaces, in case of fraudulent activities and flagging fraudulent activities to other members of the data space). These sectoral and national organisations can be brought together in sub-organisations (such as NapCore[121] in the mobility data space, which provides a forum for data space members, the organisation managing the data space and other stakeholders) that would then report to and exchange best practices through the Umbrella organisation.

National And Sectoral Ecosystems

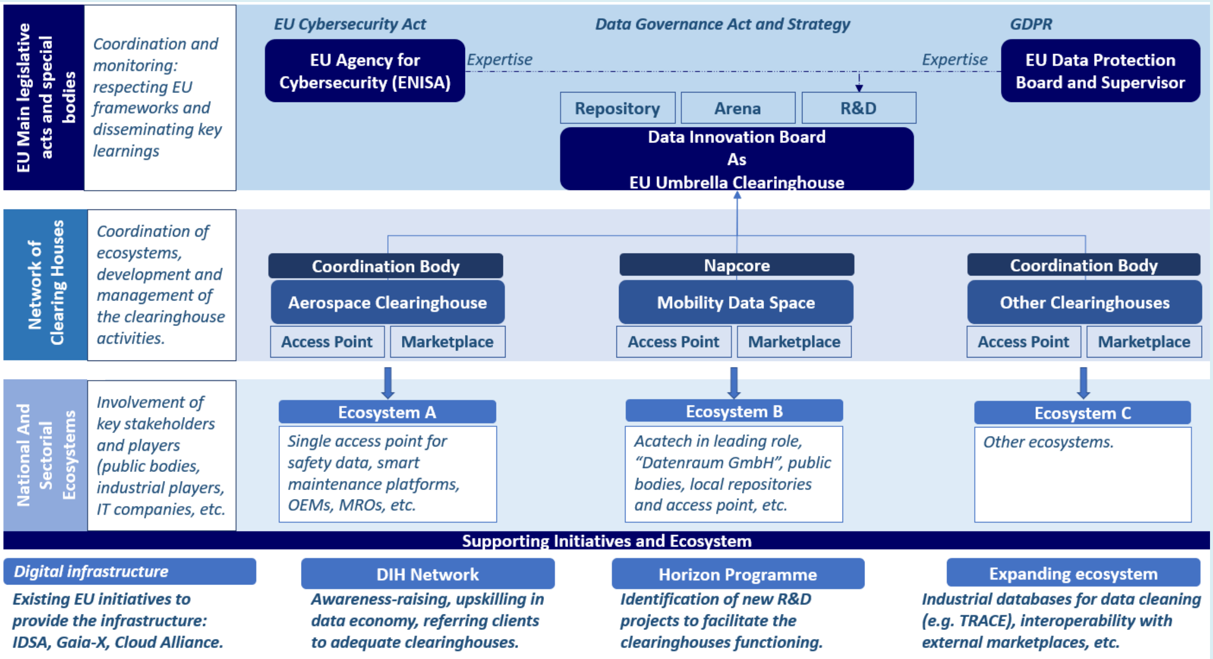
Each Clearing House represents a consortium in charge of a neutral organisation that manages the clearing house and that brings together digital value chain stakeholders. These ecosystems are dynamic, as stakeholders can come and go from the clearing house.

Supporting Initiatives and Ecosystem

The EU can provide further support through its other related initiatives, such as the DIH network, to raise awareness about the existence and role of clearing houses and refer potential customers to them. The Horizon Programme can be leveraged to fill the technical gaps. Existing initiatives such as IDSA and Gaia-X are vital to develop a common digital infrastructure for the clearing houses. Finally, the ecosystem should be perceived as **dynamic**, as more initiatives and organisations can be invited to join over the long-term (e.g., industrial databases to serve as a data repository for data cleaning).

[121] Napcore, 16/09/2021, National Access Point Organisation for Europe, https://www.gov.si/en/registries/projects/napcore-koordinacija-nacionalnih-tock-dostopa-eu/

Figure 10: Overall view on the European Clearing Houses ecosystem



Source : CARSA, 2021

Smart maintenance for Aerospace

Relatively under-digitalised, **smart maintenance for the aerospace industry** remains the clearest opportunity for the development of high-impact data sharing.[122] The opportunity offered by digitalisation presents an economic growth potential of around **10%** over the coming years.[123] In addition, smart maintenance is an important societal challenge. According to the UK Civil Aviation Authority, based on feedback from field workers, mistakes in safety reporting remain a common practice in aeronautics. This has led to the emergence of the unofficial practice of managing a "Black Book" in the industry — a book that tracks all of the different parts of the reporting system that need to be carefully checked for potential mistakes. Although it is difficult to quantify the number of cases of inaccurate data that have led to accidents, for the UK alone, a screening of the Air Accident Investigation Reports and Recommendations has allowed the authority to identify 102 incidents related to maintenance. 74 of those have highlighted shortcomings in maintenance data as the root cause, 12 of which led to serious consequences, resulting in a total of 143 fatalities and 92 serious injuries.[124]

Currently, OEMs and MROs (Maintenance Repair and Overhaul) companies are increasing the use of "**Electronic Maintenance Records" (EARMs**) to digitise and track information. However, these digital records tend to be siloed and kept in-house, and there is a reluctance to share information. This leads to stakeholders — even public authorities — facing difficulties when trying to access this information, including for accident investigations.[125]

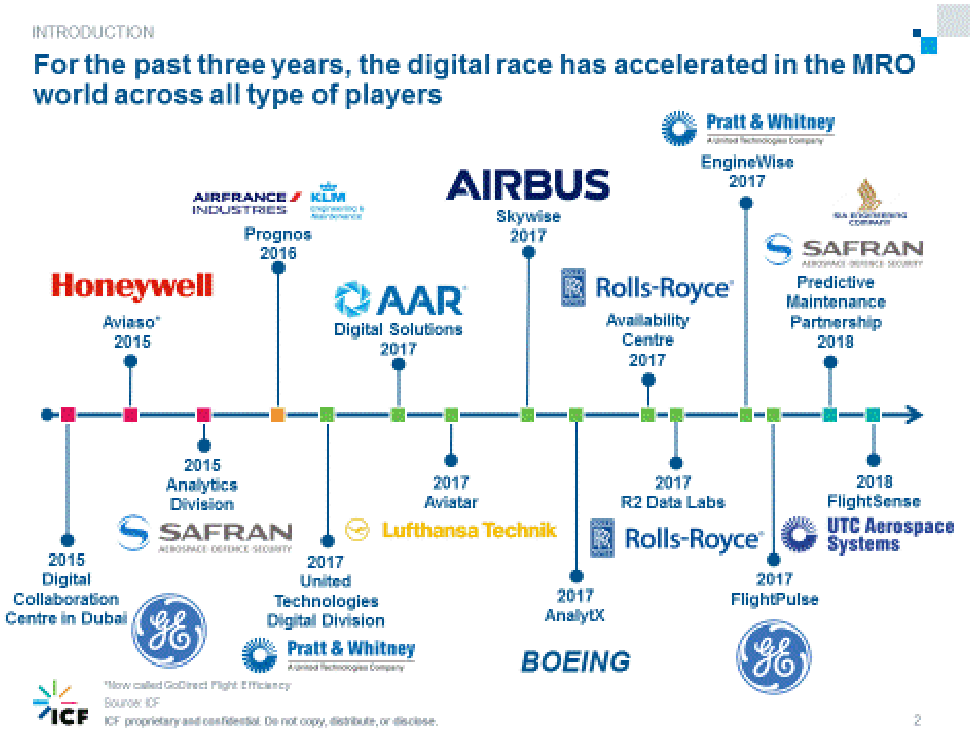
However, several initiatives have already emerged for the development of **single access points** in the aeronautics industry. We can illustrate this with two important initiatives in Europe:

[122] From workshop on aerospace.
[123] Lucht- en Ruimtevaart nederland, 01/03/2018, *Roadmap Aeronautics 2018-2025*, https://www.nlr.nl/wp-content/uploads/2018/03/Roadmap-Aeronautics-2018-2025.pdf
[124] David Hall — UK civil aviation authority, 4/04/2002, *Aviation Maintenance Data in the United-Kingdom*, https://www.faa.gov/about/initiatives/maintenance_hf/library/documents/media/human_factors_maintenance/aviation_maintenance_data_in_the_united_kingdom.pdf
[125] Fernadnez Antonio, 16/02/2019, *How Blockchain could enhance aircraft maintenance*, https://datascience.aero/blockchain-enhance-aircraft-maintenance/

- **Skywise**. Launched by Airbus and Palantir (large start-up providing digital service in data analytics), the initiative creates an Open Data Space for data sharing. Data is anonymised to protect data sovereignty and ensure anonymity. Data is then stored on Skywise servers. Data analytics services are provided on the platform. They enable, for example, access to exploitable information on certain types of aircraft operated by other companies and data for predictive maintenance, etc. The initiative already brings together 90+ airlines, 10+ suppliers and more than 8,000 types of aircraft. According to Airbus, the platform has had an important impact on operational efficiency, for example, by accelerating the production time for the A350 by 33%;[126]
- **The European ATM Information Management service** and its related **European Aeronautical Information Services (AIS)** (which for example includes legislation texts, maps, list of airports, etc).[127] This initiative aims to provide a single access point to information held by national civil aviation authorities, air navigation services providers and military administrations. Providers maintain ownership of shared information that can be accessed by another organisation and the general public.

In spite of these initiatives, the objective of bringing together a single access point is not yet achieved, most notably, because several of these initiatives are privately-led

Table 7: Fragmentation of the Smart Maintenance Platform landscape



Source: ICF[128]

[126] Frederic Sutter (Airbus digital transformation leader), 02/10/2019, *Skywise*, https://www.aaco.org/Library/Assets/Skywise%20by%20Fredric%20Sutter%20-%20Airbus.pdf
[127] Eurocontrol, sin dato, *European AIS database*, https://www.eurocontrol.int/service/european-ais-database
[128] Aircraft IT maintenance opérations, 01/01/2018, *Big Data: Racing to platform maturity*, https://www.aircraftit.com/articles/big-data-racing-to-platform-maturity/

The missing single access point for the Aerospace data space

- Platforms created by OEMs are unlikely to develop into single access points because of the reluctance of competitors (e.g. Boeing and Airbus) to participate in each other's initiative:
  - Initiatives for **cross-platform interoperability** are missing, which makes it difficult to develop an integrated and seamless smart maintenance data space;[129]
  - These initiatives are focused on aeronautics but do not yet include space activities;
  - A lack of communication between large stakeholders and SMEs remains a challenge. Even when they do collaborate, SMEs struggle to obtain data from large stakeholders and IPR agreements do not necessarily allow SMEs to access and use data, as access to data is often restricted (even the Skywise initiative has several levels of benefits based on price). Industrial agreements could be a solution to help overcome this barrier and create more open data spaces;[130]
- Safety-related information is usually not registered on this platform. For example, it is not available through Skywise.[131] A single access point for safety information is still missing, but should be developed for the common good:[132]
  - The FAA in the US and the ASA in Europe define requirements on how to perform maintenance, but not on how it is documented;
  - Standards for data formats, data content and the platform architecture for information exchange would speed-up the development of smart maintenance and enhance maintenance efficiency.

Legal Action Points

- ✓ Considering the safety critical nature of aerospace, institutional and regulatory intervention may be conceived in parallel with industry work to develop standards and norms for data sharing. It may be advisable, in combination with industry agreements, to develop regulatory frameworks and standards for hardware and data interoperability (either at national or EU level);
- ✓ Foster interoperability through an evolution of patent and trade secrets law by means of automatic licensing, by providing an interoperability exception or by promoting non-legislative EU initiatives such as best practices;
- ✓ Data sharing obligations, already mandated by law in certain sectors, could also be extended to the sharing of maintenance data, subject to further research;
- ✓ Provide additional clarity vis-à-vis the legal status of data;
- ✓ Clearly delineate responsibilities so that stakeholders can assess their exposure to liabilities and associated risks.

As already detailed in the previous section on clearing houses, the quest for a single access point is unlikely to materialise. Instead, it must be admitted that industries will be organised around a multiplicity of access points. Single access points tend to focus on specific parts of the digital value chain and the information that they bring together. Depending on the type of information, the level of openness and the role of public authorities will vary. The key objective is therefore to create repositories and develop platform interoperability.

However, one more access point is missing and should be developed in aeronautics: a single access point for safety data. A **single access point for safety could serve as a starting point for the development of an Aerospace Industrial Data Clearing House.**

[129] From the workshop on aerospace.
[130] Workshop on the aerospace industry.
[131] JDA Journal, sin dato, *Airbus' and Delta's SKYWISE is missing a key focus*, http://jdasolutions.aero/blog/airbus-deltas-skywise-missing-key-focus/
[132] The following two points are taken from an interview with the manager of a smart maintenance platform in the aeronautics industry.

A single access point as a first step towards an Aerospace Industrial Data Clearing House

- **The French Aerospace Valley** hosts a series of top-class R&D centres, manufacturing activities and industry stakeholders and could serve as an interesting case for the development of a pilot project (with opportunities for extension and replication). The valley goes through two French regions: Nouvelle-Aquitaine and Occitanie. Together, these two regions represented a cumulative GDP of €323.013 million in 2015.[133] The region features 1,900 companies with 2 of the 3 French "Grande école pour l'aéronautique" — centre for teaching and R&D in aerospace;[134]
- The key stakeholders in the region (MROs, OEMS, airports) should work together to develop a **Single Point of Access for Aerospace safety data:**
  - The **Aviation Safety Information Analysis and Sharing (ASIAS) programme**[135] aims at a similar objective and could serve as a source of inspiration. According to David Ryan, VP at MP aviation Holdings, the development of the initiative was facilitated by the organisation of regular meetings in the regional ecosystem. It created a culture of sharing and enabled the development of informal data sharing practices: *"(…) I would tell you that these face-to-face meetings provide the most value," he says. "People are sitting around the table, everyone has signed an NDA-type agreement and there's a high level of trust that lets us focus on our goal: to have everyone learn a particular lesson once—collectively." (…)"This could be as simple as flight departments at your local airport getting together quarterly to share their experiences about the facility. Or, virtually every regional association around the country has a safety committee and/or a Safety Day, and lots of data sharing happens at committee meetings*."[136]
  - Furthermore: *"(…) company participates in data sharing on a regional level through the Southern California Aviation Association and the SHARE (Safety Hazard Awareness Reporting and Empowerment) program it started years ago. Using the program's online portal, a flight department can enter safety reports and related data, and other participants can access that information*." [137]
  - Such an initiative would require — and would also justify — the involvement of **public authorities.** The access point could also act as a **repository** for local, national and international initiatives. The responsibility for managing this repository should be given to Eurocontrol;
  - The developed platform should build on existing technology standards and solutions (especially IDSA, GAIA-X) and should ensure interoperability with other platforms (especially with Eurocontrol and national AIS platforms);
  - Billing and marketplace mechanisms are unlikely to take place at this stage. However, **this single access point could be a starting point for a more open culture in data sharing**. Developing these platforms would create new standards and unlock new opportunities. The single access point could then be extended to incorporate other altruistic information. An agile IT architecture could be added to make it a dynamic platform for data providers to add new sources and connect their own digital marketplaces.

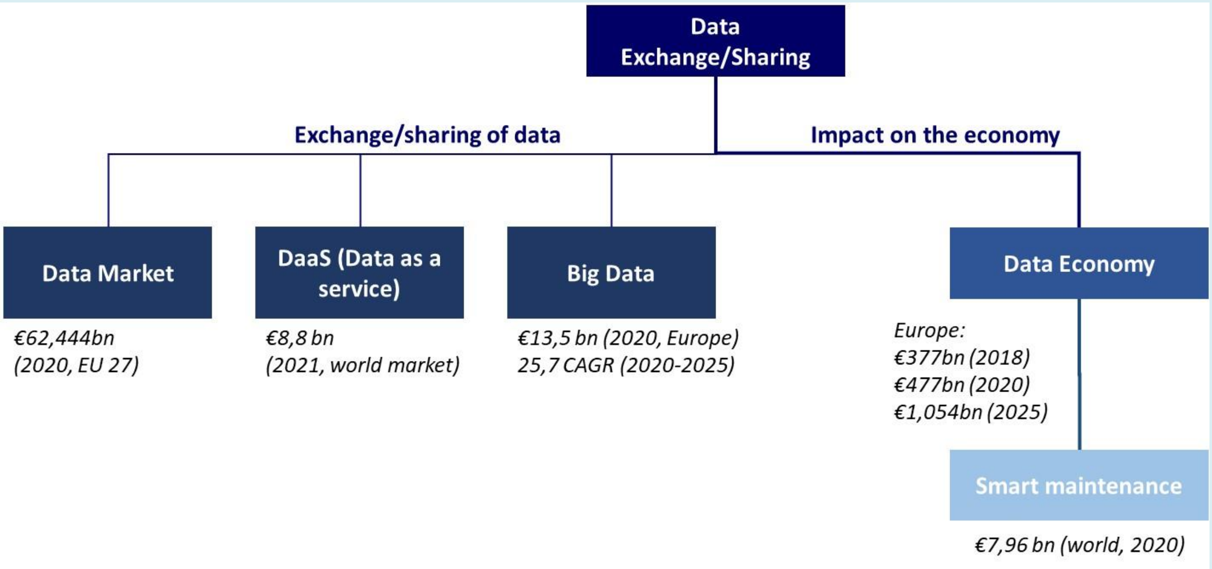
---

[133] Insee, 2018, *Produits intérieurs bruts régionaux et valeurs ajoutées régionales de 1990 à 2015*, https://www.insee.fr/fr/statistiques/1893220
[134] Aerospace Valley, *Sin Dato*, *Le territoire*, https://www.aerospace-valley.com/page/le-territoire
[135] *"The U.S. Federal Aviation Administration (FAA) often describes Aviation Safety Information Analysis and Sharing (ASIAS) as an evolving program enabling "users to perform integrated queries across multiple databases, search an extensive warehouse of safety data, and display pertinent elements in an array of useful formats."* Skybrary, sin dato, *Aviation Safety Information Analysis and Sharing*, https://skybrary.aero/index.php/Aviation_Safety_Information_Analysis_and_Sharing_(ASIAS)
[136] Aviation Safety, 18/03/2020, *The benefits of Aviation Data Sharing and how to get started*, https://www.global-aero.com/the-benefits-of-aviation-data-sharing-and-how-to-get-started/
[137] Ibidem

Assessing the economic impact of smart maintenance

- Maintenance costs reduced by 10-40%;[138]
- Worldwide, maintenance represents approximately 10% of an airline's operational costs;[139]
- For example, Air France – KLM spent 11% of its €24 billion operational expenses in 2019 on maintenance;[140]
- Average cost of maintenance for civilian companies is ~€1017 per flight hour (IATA, 2016);[141]
- Fragmented market as 80% of MROs are SMEs;[142]
- Lower risk of fatal accidents.

Figure 11: Size of data sharing/exchange markets



Note: values converted from USD to EUR using the current exchange rate last value of the calendar year (for forecast, current exchange rate as of 12/11/2020) .

Source: CARSA based on different sources (mergermarkets, BigDataCoe, Accenture, EC, etc.) [143]

[138] DIZMO, 2019, *The unrealized benefits of predictive maintenance in Industry 4.0,* https://www.dizmo.com/the-unrealized-benefits-of-predictive-maintenance-in-industry-4-0/

[139] IAC Partners, *sin dato, Point de vue: les entreprises de maintenance aéronautiques ont le vent en poupe,* https://www.iacpartners.com/blog/quel-avenir-pour-la-maintenance-aeronautique

[140] Aeronewstv.com, 2015, *Compagnies aérienne : les plus gros postes de dépense,* https://www.aeronewstv.com/fr/transport/compagnies-aeriennes/4674-compagnie-aerienne-les-plus-gros-postes-de-depense.html#:~:text=Le%20troisi%C3%A8me%20poste%20de%20d%C3%A9pense,au%20total%20l'an%20pass%C3%A9.&text=La%20maintenance%20p%C3%A8se%20pour%2011%25.

[141] IAC Partners, *sin dato, Point de vue: les entreprises de maintenance aéronautiques ont le vent en poupe,* https://www.iacpartners.com/blog/quel-avenir-pour-la-maintenance-aeronautique

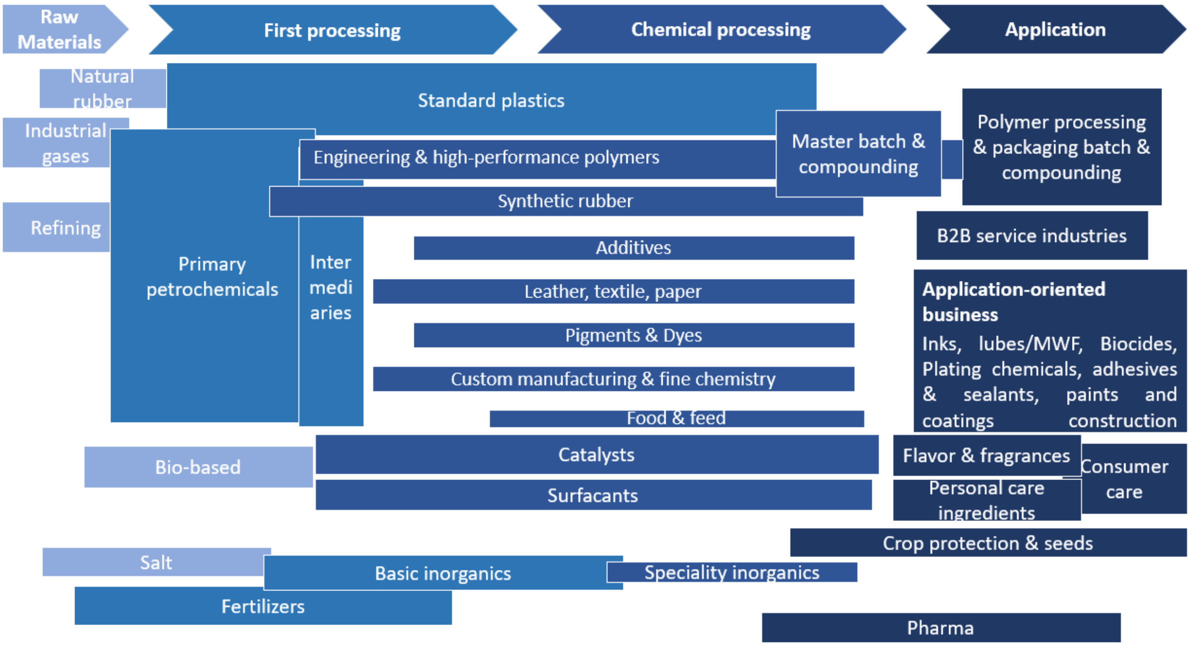
[142] PIPAME, 2010, *Maintenance et réparation aéronautique,* https://www.entreprises.gouv.fr/files/files/directions_services/etudes-et-statistiques/etudes/aeronautique_maintenance/aeronautique_maintenance.pdf

[143] Compilation of data from Marketandmarkets, 2020, *Big Data Market*, https://www.marketsandmarkets.com/Market-Reports/big-data-market-1068.html, Accenture and WDT, Sin dato, *Value of data, the dawn of the data marketplace*, https://datamakespossible.westerndigital.com/wp-content/uploads/2018/05/Western_Digital_VALUE_OF_DATA_Dawn_of_the_Data_Marketplace.pdf, EDM Final Sutdy report, 2020, BigDataCoE, Sin dato, *Big* data in Europe: new environment, new opportunities

Standards for chemical molecule and chemical reaction representation[144]

The chemical industry is a complex and fragmented industry. The challenges encountered by petrochemicals stakeholders differ considerably from the difficulties met by pharmaceutical companies. For example, pharmaceutical companies are faced with much higher legal constraints for molecule development and testing.

Figure 12: Chemical value chain structure



Source: KPMG Deal Capsule[145]

Going into the details and technicalities of all notation systems would go far beyond the scope of this report and section. Furthermore, this is a fast-moving field with quick updates and a high level of technicalities. The objective of this document is therefore to provide a broad layout that defines the problem at stake and that sketches out possible ways to overcome the current hurdles faced by researchers and industry stakeholders working on the development of standardised notation systems for **molecule representation** and **chemical reaction representation.**[146]

While this focus will make this section particularly interesting to academics and those engaged in R&D activities, application of new **notation standards** is expected to generate positive effects for the entire chemical industry. Indeed, new standards in the field represent important steps on the path towards improved **knowledge extraction** for chemical applications.[147] Furthermore, these new standards have the potential to unlock important new business opportunities for cheminformatics and lab robotics. New, standardised and more efficient ways of representing and storing chemical information will open the door to new applications

[144] **Special thanks to Goodman Jonathan and Alexei Lapkin as well as their research team**. In addition to their inputs, they kindly shared a research document that was still in the process of peer-reviewed when this section was written. This document allowed the research team to gain clarity and further clarify the technical aspect of the notification system in chemistry. The upcoming document **also includes a series of clear recommendations** and suggestion for next research step and researchers/people interested in the topic are encouraged to contact Alexei Lapkin from Cambridge university on the matter. The upcoming paper should be published under the following reference: **Wigh Daniel, Goodman Jonathan, Lapkin Alexei, *An updated review of Molecular representations in the Age of Machine Learning.***

[145] CHEManager, 2018, *Life sciences- und chemiebranchen sehen kooperationen als attraktive alternativen su M&A*, https://www.chemanager-online.com/themen/management/life-sciences-und-chemiebranchen-sehen-kooperationen-als-attraktive-alternative-zu

[146] Consequently, a series of elements are being left aside: material information, analytical/testing laboratory information and presentation of data. For example, designing a polymer involves more elements and process chemistry is interested in more dimensions than molecules. For example, designing a polymer involves more elements and process chemistry is interested in more dimensions than molecules. Similarly, dimensions about catalysis and synthesis and combination of materials is left aside from this section.

[147] Interview from a representative of a standardisation body in the chemical industry.

and opportunities in both fields. Finally, standardisation will enable a greater exchange of information, leading to replicability of chemical experiences and chemical synthesis while generating new industrial applications.[148]

Standard notations for the representation of chemical molecules are necessary to enable human to human communication, and they also make information machine-readable. While the first aspect (human readability) has been achieved by existing standards, new debates and challenges have emerged with the rise of **cheminformatics** and related techniques (deep learning, machine learning, etc). Each representation has its strengths and weaknesses. But the main limits arise when these notations are fed into computers to develop new applications. AI and ML processes for advanced cheminformatics use and need to be fed a large amount of information that is not necessarily available. In addition, some notation systems will have a "blind spot" that will influence the final result produced by the machine.

For example, the Simplified Molecular Input Line Entry System (SMILES) is a line notation that describes chemical structures using ASCII strings. Already well-used in the industry, it has enabled the development of extensions for speciality chemicals and other applications (e.g. SMIRKS and SMARTS). The line format makes it more compact, and that makes it an interesting option for warehousing and the retrieval of database information. However, when used in cheminformatics, some "limitations" of the notation have appeared, in terms of what machine learning techniques can learn from it (e.g. built-in limitations of what the notation can represent and information it can "store" translate into limitations in the results the ML techniques can produce and the 3D molecule visualisations they enable).[149] The International Chemical Identifier (inChI) was developed over the course of the 2000's as an open standard based on the SMILES system. Although it has gained wide popularity in the industry, the notation has also faced challenges, such as the need for long notation which makes data storage in chemical databases more difficult.[150]

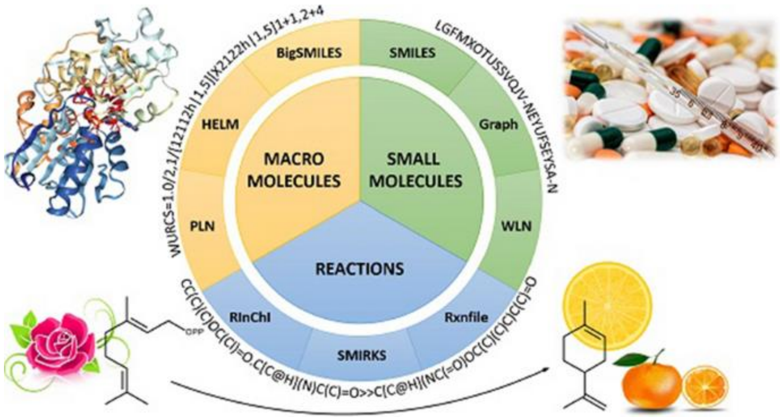
Finally, other systems for the representation of chemical compounds that feed machine learning and cheminformatics systems exist. Examples include connection table, featured-based and computer-learned representations (computer-learned representations are fed with data models that enable the computer to develop the representation by itself).[151]

Other standards for chemical notations exist, although they are not all as well-used by industry. For example, the Representation of Organic Structures Description Arranged Linearly (ROSDAL) notation, the Wiswesser Line Notations (WLN) and the SYBILL standards are not currently well-used.[152] Some other initiatives have also been explored with varying degrees of success. The case of **chemical ontologies** is a good example. Ontologies are a technique used to transform human knowledge in a format that is very easy for machines and algorithms to read and process.[153] However, ontologies for chemicals involve a lot of different types of information that have to be identified, synthetised and then stored. The amount of information required indicates the complexity of the task and explains why many initiatives in the field have been abandoned.[154] In fact, some experts doubt whether the development of a high-impact chemical ontology is even feasible.[155]

[148] Opportunities are taken from the Workshop on the chemical industry.
[149] Based on the Workshop on the chemical industry, on Chemistry and on LibreTexts, 12/08/2020, *Line Notation Smiles and InChI, https://chem.libretexts.org/Courses/Fordham_University/Chem1102%3A_Drug_Discovery_-_From_the_Laboratory_to_the_Clinic/05%3A_Organic_Molecules/5.08%3A_Line_Notation_(SMILES_and_InChI) and on* Institute of Science and technology, sin dato, *Chemistry of Eletronic Materials,* https://www.sathyabama.ac.in/sites/default/files/course-material/2020-10/UNIT-5_2.pdf
[150] Based on the Workshop on chemical industry and on Chemistry LibreTexts, 12/08/2020, *Line Notation Smiles and InChI, https://chem.libretexts.org/Courses/Fordham_University/Chem1102%3A_Drug_Discovery_-_From_the_Laboratory_to_the_Clinic/05%3A_Organic_Molecules/5.08%3A_Line_Notation_(SMILES_and_InChI) and* on Institute of Science and technology, sin dato, *Chemistry of Eletronic Materials,* https://www.sathyabama.ac.in/sites/default/files/course-material/2020-10/UNIT-5_2.pdf
[151] These four classes of representations were detailed and described in the document previously mentioned shared by Alexei Lapkin. Interested parties should refer to this document once available.
[152] Confirmed by two experts in the chemical engineering industry.
[153] Hastings Jana et all, 19/08/2011, *Chemical Ontologies: what they are, what are they for and what are the challenges*, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3083557/
[154] Pachl Christian et all, sin dato, *Overview of chemical ontologies*, https://arxiv.org/abs/2002.03842
[155] Based on the Workshop on chemical industry

An ideal solution not yet been found and the industry does not believe that **a one-size-fits-all solution will ever be developed.** Instead, a combination of solutions based on common needs and features will remain the norm in the industry. In fact, the chemical industry tends to combine different standards for specific tasks, as some of the methods are complementary (see Figure 13).

Figure 13: Example of mapping of different standards for Molecule representations in AI-driven drug discovery



Source: Journal of Cheminformatics[156]

However, challenges remain in the current notation landscape. As already mentioned, notations systems are the key to unlocking new opportunities in cheminformatics. The new applications they will enable will create new business opportunities for the industry.

Standards for chemical molecules are an essential building block for the chemical value chain. They enable the constitution of databases, data retrieval and the creation of a common language for data exchange. Without these standards, data would not be retrievable on the internet or from databases. However, the existence of different standards makes the development of databases more complex. Currently, two alternative solutions exist to “bridge” between different formats in different databases — none of which is satisfactory. The **lookup** works like a thesaurus, enabling the location of an identifier from one database to another (example: CACTUS, UniChem, PubChem Identifier Exchange services). **Translation software** (also referred to as “**Data Model Transformation Tools”** in the industry and in this report) sets a list of rules to convert one type of representation into another format. However, they typically lose too much information in the process — some types of information stored in one format can be impossible to translate into another format. Examples of translation toolkits include: “*RDKit, OpenBabel, Chemistry Development Kit, among others (see Blue Obelisk)*”.[157]

An integrated ecosystem for chemical molecule and reaction representation

No clear solution is emerging in the industry to tackle the problem. The complexity of the situation excludes the adoption of a “top-down” approach and a “one-size-fits-all” solution is not deemed feasible.

- A **pathway towards a solution** can be sketched out as follows:
  - There is no clear view on how many standards there are and what their specific features would be (standards for AI, SMIRKS for reaction, InChI for stereochemical, etc). A **mapping initiative at this level has yet to be implemented and is a precondition for an effective solution to the problem**;

[156] David Laurianne et all, 17/09/2020, *Molecular representations in AI-driven drug discovery: a review and practical guide*, https://jcheminf.biomedcentral.com/articles/10.1186/s13321-020-00460-5
[157] Libretext Chemistry – Belford Robert, 23/06/2019, *Cheminformatics*, https://chem.libretexts.org/Courses/University_of_Arkansas_Little_Rock/ChemInformatics_(2017)%3A_Chem_4399_5399

- Mapping should include a comparative analysis with a clear research framework to identify the strengths and weaknesses of existing solutions when using cheminformatics;
- The initiative should include the broadest possible coalition of stakeholders (academics, manufacturers, etc) to ensure efficient formats for all types of information;[158]
- Development of a new notification system is an advanced and technical task that requires the involvement of a broad coalition of value chain stakeholders, from R&D and academic specialists in chemicals, to **IT specialists** (data scientists, experts in Machine Learning, etc). An Open Data culture in the domain should be encouraged (for example using the GitHub platform to share advancements in the domain);
- Molecule file formats should be standardised (see 2.3 section on lab robotics);
- An **industry agreement is drafted in Annex I** to guide the development of a new standard or library of standards in the field;

- Two additional comments include:
  - The **Smart Specialisation Platform for Industrial Modernisation (S3P industry)** is the ideal platform for the development of networking initiatives and the funding of R&D projects in Europe in this area;[159]
  - Different sectors (biology, chemistry, etc.) and stakeholders operating in parallel value chains act in silos and are often unaware that they face similar problems. Connecting them could lead to cross-fertilisation.

Although tackling the topic of cheminformatics is beyond the scope of this section, a quick summary on how cheminformatics can help leveraging incomplete data is also interesting. **Advanced machine learning techniques**, for example, are needed to minimise limitations of notations techniques and should be **seen as a solution to some of these challenges**.

Cheminformatics, machine learning and artificial intelligence as tools to work with current limitations

New techniques should be explored that take cheminformatics applications into consideration (currently missing in the industry), with a view to developing innovative solutions to minimise current limitations that are unlikely to be fully resolved in the short or medium-term:

- Machine Learning and AI techniques need to be fed a significant amount of data to perform well. However, difficulties associated with chemical notations and data encoding are a significant problem. **Frugal AI and frugal Machine Learning solutions** should therefore be explored further by industry. For example, a research team has demonstrated that their ElemNet model based on frugal ML techniques is able to perform well by leveraging only 2% of a OQMD (open quantum materials database) dataset for training, compared to more classic ML techniques;[160]
- Deep Learning techniques offer important opportunities for the development of aided drug-design, quantum chemistry, etc, and typically outperform traditional machine learning algorithms;[161]
- The improvement of AI techniques will help to solve current challenges, and the more advanced those techniques become, the more they will be able to exploit the full potential of existing datasets. For example: "*A recent deep learning framework has been successfully applied to the task of molecular generation using attention mechanisms. We have benchmarked the resulting Transmol method utilizing the MOSES benchmark. The results demonstrate a number of the advantages when*

[158] Comment raised in the Workshop on the chemical industry.
[159] Smart Specialisation Platform, Chemical, *sin dato,* https://s3platform.jrc.ec.europa.eu/chemicals
[160] Jha Dipendra et all, 06/11/2018, *ElemNet : Deep learning the chemistry of materials from only elemental composition*, https://d-nb.info/1175382124/34
[161] Goh Gareett et all, sin dato, *Deep Learning for Computational chemistry*, https://arxiv.org/ftp/arxiv/papers/1701/1701.04503.pdf

using this attention-based methodology in comparison with earlier approaches. In addition, such model architecture allows the generation of new focused molecular libraries using two seeds."[162]

Best Practices examples of the use of Machine Learning in chemistry are summarised below (from Artrith et all[163])

Table 8: Best practice examples of the use of Machine learning in chemistry

Key point	Recommendation
Data Source	*"Listing all data sources, documenting the strategy for data selection, and including access dates or version numbers. If data is protected or proprietary, a minimally reproducible example using a public dataset can be an alternative."*
Model Choice	*"Justifying your model choice by including baseline comparisons to simpler — even trivial — models, as well as the current state-of-the-art. A software implementation should be provided so that the model can be trained and tested with new data."*
Model Training and validation	*"Stating how the training, validation, and test sets were obtained, as well as the sensitivity of model performance with respect to the parameters of the training method, for example, when training is repeated with different random seeds or ordering of the dataset. Validation should be performed on data related to the intended application."*
Code and reproducibility	*"The full code or workflow is made available in a public repository that guarantees long-term archiving (for example, an online repository archived with a permanent DOI). Providing the code not only allows the study to be exactly replicated by others, but to be challenged, critiqued and further improved. At the minimum, a script or electronic notebook should be provided that contains all parameters to reproduce the results reported."*

Source: Taken from Artrith et al[164]

Impact of molecule representation and chemical reaction standards

- Standards will help to develop Artificial Intelligence solutions for the chemical industry and promote their uptake. Artificial intelligence is expected to generate an income growth of around +1/2.03% for the chemical sector;[165]
- A UK study found that molecular modelling has the potential to achieve a 1% increase in GDP. E-Infrastructure and new and advanced computing could potentially account for 6% of that increase;[166]
- Faster time to market and new business opportunities for cheminformatics and lab-robotics.[167]

[162] [Thesis] from Zhumagambetov Rustam, April 2021, *Towards automated molecular search in drug space*, https://nur.nu.edu.kz/bitstream/handle/123456789/5495/Thesis%20-%20Rustam%20Zhumagambetov.pdf?sequence=1&isAllowed=y
[163] Artrith et all, 31/05/2021, *Best Practices in machine learning for chemistry*, https://www.nature.com/articles/s41557-021-00716-z?proof=t%29.
[164] Ibidem
[165] ITU, 2018, *Assessing the economic impact of artificial intelligence*, https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-ISSUEPAPER-2018-1-PDF-E.pdf
[166] Goldbeck Gerhard, 2012, *The economic impact of molecular modelling of chemicals and materials*, https://www.researchgate.net/publication/262232197_The_economic_impact_of_molecular_modelling_of_chemicals_and_materials
[167] Comment raised in the Workshop on the chemical industry.

DAWEX as best practice

DAWEX[168] is one of the most advanced **data exchange and data marketplaces**. Companies and organisations can manage data traffic by sourcing, monetising and exchanging data directly with each other, completely securely and in compliance with applicable regulations. DAWEX offers...

... trust, by acting as a **trusted third-party** and offering a platform with carefully vetted participants;

... individual solutions with **flexible operating models**, from private data exchange to a global data marketplace that includes multiple use cases (e.g., data sharing, data sourcing, data monetising);

... regulatory and **contractual compliance** by direct licensing between the provider and acquirer, and local data protection regulations, and privacy-by-design.

Because of its third-party role, and the privacy and flexibility of its operational model, DAWEX is perceived to be a trustworthy and competent service provider. They received the "Pioneer award" from the World Economic Forum and act as consultants in this field at the highest level. More than 8,000 companies worldwide use this solution and more than 50 % are European countries.

DAWEX was founded in France in 2015 and has offices in Lyon, Paris, Montreal, and (in the near future) in Tokyo. The business model focusses on data exchange options and services in all variations - from a simple data marketplace to an individually designed use case. All personal and activity-related data is encrypted, clients can control the visibility and availability of their data, and DAWEX does not explicitly collect, share or sell any user profile-information.[169] The company acts as a trusted third-party that provides a platform with carefully vetted participants. There is no focus on special sectors, but they do describe use cases for automotive, banking and finance, agriculture, aerospace, health, government and the environment on their website.

DAWEX was identified as a best practice because of the very advanced services they provide and the wide variety of data exchange options they support. Their third-party role, and the privacy and the flexibility of their operational models, are key best practice features. The "Pioneer award" they have received from the World Economic Forum, their high-level consultancy activities in the field, and their growing number of users (>8,000) are all signs of international acceptance and they validate DAWEX as a best practice.

[168] Website of the initiative: https://www.dawex.com/en/data-exchange-platform/
[169] Website of the initiative: https://www.dawex.com/en/data-exchange-platform/

Developing an interoperability framework for autonomous Plug & Play Systems: gaps and an overabundance of standards

Interoperability: summary and key opportunities

The quest for interoperability and autonomous Plug & Play systems has led to the development of a wide variety of standards in diverse industries over the past decades. Although these initiatives were necessary, the constant addition of new projects and standards are starting to pile-up, adding new layers of complexities to an already difficult task. Consequently, any new attempt to streamline the digital landscape and develop system interoperability, regardless of the area or industry, should **start with mapping initiatives, standards and gaps**.

Four innovation areas with especially interesting economic potential for the development of interoperability are analysed in this section.

1. IoT ecosystem in agriculture;
2. Plug & Play machines in mechanical engineering;
3. Framework architecture for laboratory robotics in chemistry;
4. Digital health record systems in healthcare.

In **agriculture**, effort should focus on the complex ecosystem offered by fixed farming equipment. In spite of the existence of genuine technical complexity, this report supports that the main barriers to interoperability are caused by a lack of consideration for economic incentives in existing technologically-driven initiatives for farming equipment interoperability.

The development of **plug & play machine tools,** requires the development of new neutral toolboxes (translation software and Asset Administration Shell are two powerful examples of this new approach). Intended for manufacturers, these solutions will contribute to silo integration of machineries at the shopfloor level (vertical integration), while serving as a neutral interface for standardisation and exchange along the value chain (horizontal integration).

The development of a **framework architecture for laboratory robotics** is an ambitious target for the chemical industry. Niche robots for R&D and academic labs currently require research teams to develop homemade software and language to operate the machines. In addition to this difficulty, laboratory 4.0 are difficult to connect to integrate to other functions in cognitive plants. Building on existing initiatives (SiLA, Allotrope and Obelisk Group), there is a need for the creation of an open-library of standards (ontologies, semantics, etc) that could serve as the basis for the development of operating systems for niche robots. Furthermore, common reference architecture could be developed in partnership with lab teams in other parallel industries for easier integration in cognitive plants.

Finally, the **digital health record systems** need to be integrated to enable rapid information exchange between healthcare providers. This requires the development of an inclusive approach for the standardisation of systems to enable the quick approval from patients.

Developing Plug & Play systems goes beyond the development of standards series. Indeed, autonomous systems require agreements between organisations to connect systems and devices. To reach the distant objective of autonomous Plug & Play systems, the four layers of interoperability need to be jointly addressed in a common framework:

Layer 1: *technical* (communication and infrastructure protocols);
Layer 2: *syntactic* (data formats, encodings and compressing);
Layer 3: *semantic* (shared understanding of data exchange meanings);

Layer 4: *organisational* (ability of organisations to exchange information across sectors, departments and borders).[170]

Achieving system interoperability requires the development of standards where they are lacking. However, rather than focusing on developing new standards, it is essential to first map and disseminate existing solutions (e.g. UA OPC in mechanical engineering or RAMI reference architecture).

New standards need to be made compatible with pre-existing solutions and cannot be developed in silos. Industry players tend to work in silos, developing in-house or collective solutions to resolve the specific technical difficulties they are facing.[171] A more integrated approach is necessary, although this can only be achieved with the support of policy-makers through the development of R&D and pilot projects (especially for the development of innovative emerging approaches, such as translation software and Asset Administration Shell) and the funding of existing initiatives for the development of common model and semantic/ontology mapping across sectors and industries (the DEMETER project in smart farming is highlighted as a positive model to replicate in other fields).

Ontology alignment and harmonisation are solutions for interoperability development and emerging solutions for interoperability.

However, there is a growing consensus that the development of taxonomy (hierarchical classifications of terms) is an imperative first-step which will enable the ontology to be developed.[172] Taxonomy bridging and mapping is a required first step to achieve vertical, syntactic and semantic interoperability. For example, in mechanical engineering, a number of standard ontologies and semantics already exist. To make them interoperable implies greater complexity. The importance of legacy systems in this industry also adds further complexity. An initiative to develop new standards in machinery should therefore start by mapping what already exists to ensure that existing standards (e.g., translation software) are interoperable.

Standardisation gaps should also be identified and addressed through the development of new standards — giving consideration to meta-formats and taxonomy to avoid incompatibility between new formats and pre-existing standards. This approach should complement the development of new standards and an interoperability framework through taxonomy mapping. The development of large-scale mapping initiatives, under the supervision of standardisation bodies and/or European organisations and European research programmes, represents a great milestone on the journey towards interoperability.

Integrating the Agriculture IoT ecosystem

The development of smart farming is one of the most promising development in agriculture. Smart farming creates new opportunities to feed the need of a growing world population and higher productivity with lower environmental impact. However, two barriers stand in the way of a fully modernised agricultural sector: the cost of new technologies and the "*lack of SF* [Smart Farming] *system's capabilities to interoperate with each other*."[173]

Enabling the uptake of Smart Farming

- Public intervention is needed to deploy new 5G infrastructure in rural areas to enable the uptake of new smart farming solutions. Greater capital-intensity will help to speed-up the deployment of 5G.

[170] ETSI, (2018), *Cross-fertilisation through alignment, synchronisation and exchanges for IoT*, https://european-iot-pilots.eu/wp-content/uploads/2018/11/D06_02_WP06_H2020_CREATE-IoT_Final.pdf
[171] Based on the results from the workshop on manufacturing and the workshop on mechanical engineering.
[172] Costin Aaron, Issa Raja, *(2017), The need for taxonomies in the ontological approach for interoperability of heterogeneous information models*, https://ascelibrary.org/doi/abs/10.1061/9780784480830.002
[173] ResearchGate, 01/06/2019, *IoT and data interoperability in agriculture*, https://www.researchgate.net/publication/333853181_IoT_and_data_interoperability_in_agriculture_A_case_study_on_the_gaiasense_TM_smart_farming_solution

However, costs will start to increase exponentially for the most difficult 20% to cover.
- Farming is a fragmented market especially on the production side. Subsidies will be required to modernise equipment and to provide training to upskill the farming workforce.

The cost of new systems has been briefly addressed in the economic analysis (section 2.1 on economic barriers). Consequently, this section will focus on the difficulties related to interoperability and how they could be addressed. Before going further, it is however necessary to understand that challenges in interoperability of agriculture machines greatly vary between **fixed and mobile equipment**.

Mobile farming equipment involves less devices and connectivity. For example, a tractor cannot be connected to more than a few add-ons. Furthermore, the ISO11783 (commonly referred to as the "ISOBUS") standard is already well implemented and very powerful in addressing most of technical difficulties in that field. Through its different features and sub-versions (mainly the TECU, UT, AUX-N, TC-BAS and TC-GEO), this standard already makes BUS interoperability for connectivity in mobile equipment much more streamlined. (The BUS is the part of the hardware that ensures connectivity between a machine and a connected device). Furthermore, solutions exist to retrofit legacy systems and make them compatible with new developments of the ISOBUS standard.[174]

The **fixed farming equipment** is a much more complex landscape. It involves more machines (milking robots, sensors, stationary systems) and software (platforms, accounting, field management). Consequently, efforts to develop interoperability in this field are more complex: "(…) *the field of **interoperability in fixed agricultural equipment is extremely broad,** with a lot of organizations, initiatives, standards and formats. In the field of animal production, we're going to touch the breeder, the veterinarian, the council, genetics, mass distribution and so on. In crop production, we will get closer to the farmer, advisors, the collection and storage organizations, agricultural chains, the consumer (…). And the more people there are, the more different needs and interests there are, and the more difficult it is to agree on standards and exchange formats*" [175]

However, these difficulties do not mean that no initiatives exist. On the contrary, many very promising and efficient initiatives that target the development of interoperability exist, as illustrated in Table 9.

Table 9: List of main interoperability initiatives in Smart Farming

Initiatives	Structures	Maturity
ADAPT + référentiels	AgGateway	Operational
AgDatahub	AgDatahub	Operational
Agripilot	Open Geospatial Consortium (OGC)	Under development
AgriRouteur	DKE-Data	Operational
BDCA	Brazilian Association of Machinery and Equipment Industry	Under development
DAPLOS, eDAPLOS + référentiels	AgroEDI Europe	Operational
DataConnect	John Deere, CNH, Claas, 365FarmNet	Operational
DataLinker	Red Meat Profit Partnership (RMPP)	Operational
ISOXML	AEF	Operational
JoinData	JoinData	Operational
MyEasyFarm	MyEasyFarm	Operational

174 Aspexit, 17/02/2021, *Standards and data exchange in agriculture*, https://www.aspexit.com/standards-and-data-exchange-in-agriculture/
175 AEF, ISOBUS Library, https://www.aef-isobus-database.org/isobusdb/login.jsf

Nevonex	Bosch	Operational
Numagri	FNSEA, Coopération agricole, Groupe Avril, APCA	Under development
DEMETER	European initiative for the development of common data models and semantic interoperability mechanisms.	Under development
ATLAS	European project for data interoperability network.	Under development
Internet of Food and Farm (IFF)	European initiative with a series of project that cover interoperability in smart farming.	Under Development

Source: Original list from ASPEXIT[176] with addition from CARSA.

On closer examination, it appears that **most of the existing challenges for the development of technical interoperability can already be tackled within existing initiatives**. For example, an important barrier for streamlined interoperability is the lack of cross-device interoperability. Typically, standards are developed specifically for one type of device (smart irrigation, sensors in pork farms, etc) and tend not to be compatible (e.g. the irrigation sensor would not be able to communicate with a smart meter). However, Demeter and IFF are already producing interesting outcomes that demonstrate that the development of new cross-device standards is already on the way. Similarly, the overabundance of standards causes new forms of complexity to emerge, making the development of common data models and semantic interoperability through taxonomy-driven initiatives a priority — this is the current focus of the DEMETER initiative.[177]

Furthermore, other standards exist that tackle the additional needs of the digital value chain. We can, for example, mention the ISO standard for blockchain in agriculture[178] or the research projects funded by the European Innovation Partnership for Agricultural Productivity and Sustainability (EIP-AGRI). In addition, research programmes for smart/precision farming have been conducted by the AOITI.[179]

This situation reveals that the **key barrier to the technical integration of IoT systems depends more on economic incentives than on technological factors.**

Interoperability generates clear economic gains for IT specialists, but the advantages brought by innovative solutions are not necessarily obvious to farmers. Whereas the cost of acquiring new devices and integrating IoT ecosystems is heavy on the upstream (farming) part of the value chain, the benefits are mainly concentrated on the downstream (IT provider, consumers, distribution, etc) part of the chain.[180] Moreover, according to Honey Bee Canada (manufacturer of smart farming devices), this discrepancy is the main reason behind the lack of interoperability in smart farming. According to this company, many organisations have weaponised the lack of interoperability to create lock-ins and dependencies. Honey Bee even argues that although the definition of technical standards should be left to industry, enforced interoperability is the only way to escape this situation.[181]

[176] Aspexit, 17/02/2021, *Standards and data exchange in agriculture, https://www.aspexit.com/standards-and-data-exchange-in-agriculture/*
[177]Collaboration and involvement of trusted organisation are essential for the development of common data models and interoperability-based mapping of existing platform, APIs and ontologies. These elements were raised by stakeholders during the mini-workshop on agriculture.
[178] ISO, *ISO and agriculture,* https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100412.pdf *and* ISO, 2020, *Getting big on Blockchain*, https://www.iso.org/news/ref2540.html
[179] Euractiv, 2016, *Smart Farming trying to find its feet in EU agriculture, https://www.euractiv.com/section/agriculture-food/news/smart-farming-trying-to-find-its-feet-in-eu-agriculture/*
[180] Insight gathered from the mini-workshop on agriculture organised in the context of this study.
[181] Honey Bee Manufacturing, Q4 2020, *Interoperability in Farming operations, https://www.pc.gov.au/__data/assets/pdf_file/0008/270575/sub002-repair-attachment2.pdf*

It is difficult to fully support the radical view advocated by Honey Bee. Nonetheless, it highlights the fact that the key to IoT interoperability in agriculture is no longer to be found in new technological research. **What is currently missing is to embed technical initiatives in an approach that makes them economically viable solutions for the entire ecosystem.**

Integration of the smart farming ecosystem

- Support the development of **Atlas, Demeter and IFF** and the development of taxonomy-driven initiatives and cross-device standardisation/interoperability. Industry players are encouraged to engage in/with these initiatives. The current standards landscape makes the development of new standards a counterproductive approach;
- Most effort should be focused on **fixed farming equipment**, where most technical challenges are concentrated;
- **The role of policy-makers** will be crucial for the removal of barriers to smart farming interoperability, developing 5G infrastructure in rural areas and organising events to raise awareness about the benefits of smart farming to farmers;
- Tackling the imbalance of benefits across the value-chain is a difficult task. Subsidies to support the modernisation of farming equipment is one possibility. Another would be to encourage greater involvement of the **farming sector** in technical initiatives to ensure that the needs of the sector are built into the digital value chain.

Impact of greater and faster market uptake of smart farming

- The main impact of this type of initiative would be to create a more integrated ecosystem and accelerate the uptake of smart/precision farming solutions;
- Devices, sensors and systems would be (more) interchangeable, increasing the bargaining power of final consumers and potentially decreasing prices;
- Improve the industry's environmental footprint and reduce the use of scarce resources (water, seeds, etc);
- The market for smart farming production is expected to reach €6.6 billion by 2023;[182]
- Important impact on price competitiveness. For example, the impact of precision farming is said to have "saved" olive oil production in Greece. Countries with a lower level of technology uptake face much higher prices. Italy ranks first (€469.8 per 100kg, Spain is lowest with €221.1 per 100kg and Greece reached a level of €260.1 per 100kg;[183]
- However, an investment of around €200 billion is needed to provide rural areas with broadband access of at least 100mb per second (European Investment Bank).[184]

[182] EuroScientists, 2020, *Will smart farming gain a foothold in Europe*, https://www.euroscientist.com/will-smart-farming-gain-a-foothold-in-europe/

[183] Euractiv, 2020, *Precision Farming Saved extra virgin olive oil, Greek producers say, https://www.euractiv.com/section/agriculture-food/news/precision-farming-saved-extra-virgin-olive-oil-greek-producers-say/*

[184] Michalpoulos Sarantis, 2020, *EIB official: €200 billion needed to build broadband infrastructure in EU rural areas,* https://www.euractiv.com/section/agriculture-food/news/eib-official-e200-billion-needed-to-build-broadband-infrastructure-in-eu-rural-areas/

OPEN DEI as best practice

OPEN DEI[185] supports the implementation of next generation digital platforms in manufacturing, agriculture, energy and healthcare. The project aims to detect gaps, encourage synergies, support regional and national cooperation, and enhance communication between the innovation actions implementing the EU Digital Transformation strategy. The project is an H2020 coordination and support action (CSA) that involves the harmonisation of data space design principles and data ecosystem development by cross-sector task forces, sector-specific coordination with the support of working groups, community building and standardisation. To increase the adoption of digital platforms, industry ambassadors and project partners provide knowledge transfer support to businesses, especially in the target sectors, with knowledge transfer activities, including common data space principles and data ecosystem development.

OPEN DEI is identified as a best practice driver of digital platform innovation, standardisation processes and data ecosystem development, with the ability to increase awareness and facilitate the acceptance, adoption and customisation of cross-sector industry agreements.

Plug & Play machine tools

The digitalisation of the shop floor and the growing integration of IT processes in manufacturing activities make system integration both more complex and necessary for machine manufacturers. Machinery has been continuously increasing its complexity since the 80s. Pushed by market demand, ICT technologies have become key enablers to support manufacturing demands, which have been extended with the use of digitalisation technologies. As a result, complexity has increased, making system integration more challenging.

Figure 14: Evolution of production paradigm in mechanical engineering

Paradigm	Craft Production	Mass Production	Flexible Production	Mass Customization and Personalization	Open Complex and Adaptive Production Systems
Paradigm started	~1850	1913	~1980	2000	2020
Society needs	Customized Products	Low cost products	Variety of products	Customized products	Customized on-demand products
Market	Very samll volume per product	Steady demand	Smaller volume per product	Global manufacturing and fluctuating demand	Global manufacturing and fluctuating demand
Business Model	Pull Sell-design-make-assemble	Push design-make-assemble-sell	Push-Pull design-make-sell-assemble	Pull design-sell-make-assemble	Pull design-sell-make-assemble
Technology Enabler	Electricity	Interchangeable parts	Computers	Information technology	Information and communication technology
Process Enabler	Machine tools	Moving assembly line	Flexible Manufacturing Systems, robots	Reconfigurable Manufacturing System	Self-organizing agents

Source: Manufuture, 2003[186]

The complexity of the task leads manufacturers to outsource system integration. However, the procedure is both costly and risky: assessing the benefits and costs associated with system integration is not always

[185] Website of the initiative: https://www.opendei.eu/

[186] Lanz Minna et all, 06/06/2012, *Towards adaptive manufacturing systems — knowledge and knowledge management systems,* *https://www.intechopen.com/chapters/36408*

feasible until a project is completed. According to Connecting Software, trillions are spent every year on industrial IT systems, the majority of which is spent on system integration.[187]

The use of Plug & Play machines is a worthy objective, not least because of the savings that could be achieved. However, achieving interoperability is a complex task. To better understand the challenge, a difference must be made between the four types of interoperability that currently exist in the context of mechanical engineering:[188]

1. **Syntactic interoperability** focuses on data formats rather than data content. Example: Hypertext Markup Language (HTML), Extensible Markup Language (XML) or Abstract Syntax Notation One (ASN.1)");
2. **Semantic interoperability** focuses on the content of data and harmonises human/machine interpretations of data points. Example: Resources Definition Framework and XML;
3. **Factory interoperability** (vertical integration) targets the development of machine-to-machine interoperability within the factory (including shopfloor and other activities). However, current paradigms for vertical integration seek to define industry-wide standards or full machine-to-machine interoperability, which is both costly and unrealistic;[189]
4. **Cloud-Manufacturing Interoperability (Horizontal Integration)** is a cross-cutting domain that involves transport protocols, policy interoperability and harmonisation of system behavioural interoperability. Example: ISO/IEC 199941 (a standard that provides guidelines for interoperability and portability in cloud computing).

Big steps have already been made in the area of mechanical engineering interoperability with the development and the adoption of multiple standards. Nonetheless, there is still a long way to go to achieve full interoperability.

Machine vendors have traditionally developed **their own proprietary standards** to secure their IPR and ensure differentiation from their competitors. At the same time, the fragmented proprietary standards landscape has helped those vendors to maintain their market share, locking-in their customers — especially in "Niche" markets. By developing and adopting **open standards**, system providers are able to open up new market opportunities, reduce development costs, gain customer recognition and attract new customers. Machine vendors have been able to protect their know-how and IPR, while integrating open standards to ensure interoperability with other vendors. **Standards are not static but dynamic elements.** Standardisation organisations at national and international level develop and maintain standards, in parallel with technology developments by experts in the domain, to ensure a consensus is maintained.[190] Standards are periodically updated. Depending on the standardisation organisation, this usually occurs every 5 to 10 years.

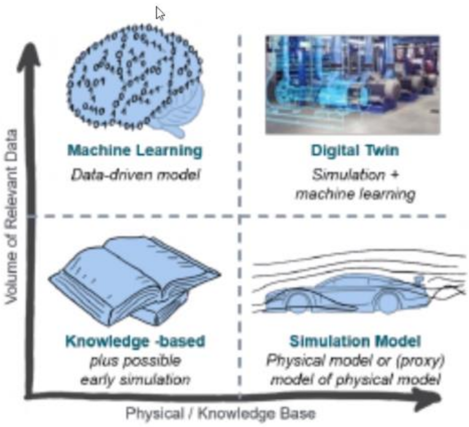
Several standardisation initiatives exist. An example is the Universal Machine Tool Interface (UMATI[191]). UMATI brings together several standards from the VDMA and the OPC foundation. However, it is difficult for manufacturers to know what standard to adopt: standard competition makes the longevity and success of a standard uncertain, while adoption costs remain an expense for the adopter.[192]

Covering all challenges to interoperability for mechanical engineering in this section would not be feasible. Instead, this report **advocates for the development and adoption of neutral toolboxes** ("Vendor-Neutral" solutions that enable enhanced interoperability) across the value chain. These tools are necessary to enable **fully streamlined data continuity** across industrial processes.

---

[187] Connecting Software, 09/07/2021, *Infographics: Why are IT systems integration costs so high,* https://www.connecting-software.com/blog/infographics-why-are-it-system-integration-costs-so-high/
[188] Zeid Abe et all, 11/02/2019, *Interoperability in Smart Manufacturing: research challenges,* https://www.mdpi.com/2075-1702/7/2/21/pdf
[189] Ibidem
[190] Elements based on the workshop on mechanical engineering and an interview with a researcher in mechanical engineering.
[191] UMATI brings together several standard from VDMA and OPC to enable machine to machine communication. See UMATI Website, https://umati.org/about/#vision
[192] Results from the mini-Workshop on Mechanical Engineering.

Figure 15: Key aspects for data continuity



	Physical Models	Data Driven Models
Data Reliance	**Small data**	Big data
Domain Reliance	Deep domain expertise	**Little domain knowledge**
Fidelity & Robustness	**Highly non-linear and complex relations**	Limited to data and model complexity
Adaptability & Deployability	Complex and time consuming setup	**Rapidly adapt to a specific problem**
Interpretability	**Consistent physical meaningful**	Physics agnostics surrogate

Source: Hartman Dirk and van der Auweraer Herman[193]

Data continuity is a process that flows throughout the entire value and data chain. It starts with the physical object that is then turned into a "digital artifact" that is then uploaded into the platform (data integration). Data sets and digital artifacts can then circulate across the value chain and through data streams in the factory. The more streamlined these data flows, the more we can speak about "data continuity". To achieve this objective, four different interoperability challenges must be jointly addressed. However, the main barrier to continuity is that organisations have historically been organised in "*(…)* ***different teams that work in insolation*** *from each other. Once a task is completed, they will hand over documents and 3D models to the next team. This approach does not address data silo issues and information that is often lost or lacks traceability*".[194]

To simplify, this section will provide suggestions on how to achieve greater data continuity in three steps:

1. Data ingestion using data transformation model;
2. Data threading;
3. Asset Administration Shell (AAS).

Technical complexity and economic interest make the development of fully integrated Plug & Play systems a distant dream, rather than a realistic target. Instead, a "Neutral Toolbox" solution developed by a manufacturer-neutral organisation[195] would act as an intermediary for machine integration and communication between machines parks running on different standards. Machine parks are progressively being integrated in "silos" that separate machines, based on the standard that they operate on, and are connected to other silos through these new solutions.[196]

To integrate physical elements in an industrial dataspace (vertical or horizontal), industrial data must first be **integrated** in a **data hub**. It follows that heterogeneous sources are connected to the platform for data integration and ingestion. This situation raises challenges related to the protection of **data integrity** (making sure that "data is complete, consistent and accurate". The process requires the definition of parsing/cleaning rules (identifying records, fields and other components of the input data and check if data is well-formed), **data transformation** (translation of input data into the form expected by the primary computer) and **data**

[193] Hartman Dirk and van der Auweraer Herman, 28/01/2020, *Digital Twins*, *https://arxiv.org/abs/2001.09747*
[194] Yeh Pang Toh et all, 01/01/2021, *Developping a Digital Twin and Digital Thread Framework for an "Industry 4.0" Shipyard*, https://www.researchgate.net/publication/348771602_Developing_a_Digital_Twin_and_Digital_Thread_Framework_for_an_'Industry_40'_Shipyard
[195] In other words, a company that would not be linked to machine manufacturers.
[196] . The OPC UA follow a somehow similar pattern. The standard is independent from manufacturers and works by mapping machines, devices and capability description to enable a "*system and supplier neutral exchange of data by means of uniform interfaces*".

aggregation (pre-analytics computations that result in aggregated information about the input). Together, these steps are commonly referred to as "Execute-transform-load" (ETL). [197]

To achieve the difficult task of data transformation, the development of **data transformation model tools** (translation software) has been identified as a best practice by industry. As fairly recent technologies, which have achieved a maturity level of TRL1 or TRL2, these solutions enable **vertical integration**.[198] Machine silos running on different standards are connected to servers/processors equipped with translation software. These connectors translate and act as link between the different silos, enabling vertical integration.[199] In addition, these models enable **streamlined data integration** that foster **data continuity.**

The most common use type of data transformation is **syntactic models:** "(…) *often used to perform transformations such as changing date format from MMDDYY to MM-DDYYYY or YYYY-MM-DD, changing city names to their respective countries, or applying a formula to convert liters to gallons"*. However, the **limitations of this approach have led to the development of semantic transformation:** *"Syntactic transformation does not work for transformations that require searching for reference data that contain mappings between the input and output values, such as company name to stock ticker symbol or event to date. In order to address the limitations of syntactic transformation, semantic transformation was developed*." The limitations of classic data transformation models are that they: do not support semantic-aware data; are schema dependent; do not focus on a meaningful semantic relationship for data integration from different sources; and do not support the sourcing of new information using data inference and reasoning.[200]

Semantic transformation is the emerging and most **promising** way of **mapping and translating data**, as it is the only version that also tackles the question of data meaning. However, it requires that "*data elements in the source and destination systems have "semantic mappings" to a central registry or registries of data elements.*"[201] [202]

After the mapping stage has been completed, the semantic transformation requires an **ontology alignment**/**matching** step. This step is the most difficult in the process, as it requires an alternative approach, especially when data models do not have direct one-to-one mapping/matching of elements. The alignment is achieved through a combination of techniques to calculate the similarity of measures using key parameters (weights, thresholds, etc) and external resources (thesaurus, dictionary). Four methodologies exist to establish ontology alignment:

---

[197] Balijepalli Saketh, 01/01/2019, *Best Practices in Using Semantic Transformation in Data Integration to Address Data integrity issues, https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/24786/Balijepalli_2019.pdf?sequence=1&isAllowed=y*

[198] Results from the mini-Workshop on Mechanical Engineering.

[199] The "Ontology Alignment Evaluation Initiative" launched in in 2004, is an interesting initiative developing and assessing the quality of ontology alignment initiatives in different industries. Website of the Ontology Alignment Evaluation Initiative: https://oaei.ontologymatching.org/ One can also find an example of ontology mapping http://www.ontobee.org/tutorial/ontobeep

[200] Balijepalli Saketh, 01/01/2019, *Best Practices in Using Semantic Transformation in Data Integration to Address Data integrity issues, https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/24786/Balijepalli_2019.pdf?sequence=1&isAllowed=y*

[201] *"There are **three types of Semantic equivalence for mapping activities**: class-equivalence (indicating that class or "concepts" are equivalent. For example: "Person" is the same as "Individual"), property equivalence (indicating that two properties are equivalent. For example: "PersonGivenName" is the same as "FirstName"); instance equivalence (indicating that two individual instances of objects are equivalent. For example: "Dan Smith" is the same person as "Daniel Smith")."* ITEA3, 01/12/2020, *Machinaide, Deliverable D1.2, State-of-the-art description* (…) ***A semantic mapper** is a tool or service that aids in the transformation of data elements from one namespace into another namespace. A semantic mapper is an essential component of a semantic broker. Essentially the problems arising in semantic mapping are the same as in data mapping for data integration purposes, with the difference that the semantic relationships are made explicit through the use of semantic nets or ontologies which play the role of data dictionaries in data mapping*." ITEA3, 01/12/2020, *State of the art description*, https://itea4.org/index.php/project/workpackage/document/download/7292/D1.2%20State-of-the-Art%20Description_V1.pdf

[202] **"***"Semantic transformation makes use of over 100 million reference tables and online resources to look up meaningful information and transform the data accordingly, leading to more current and less obsolete data, ultimately resulting in improved data integrity"* (see Balijepalli Saketh, 01/01/2019, *Best Practices in Using Semantic Transformation in Data Integration to Address Data integrity issues, https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/24786/Balijepalli_2019.pdf?sequence=1&isAllowed=y)*

- **Terminological methods**: based on comparisons of terms, strings and texts;
- **Structural method**: calculate similarity between entities using structural information forming hierarchy of characteristics;
- **Extensional method:** infer similarity from the analysis of extensions;
- **Semantic methods:** this methodology uses an external ontology as common reference to proceed to ontology alignment (such models are defined by OPC-UA standards or STEP 242, Industry Commons, NIST through the ontology foundry).

The main motivation driving the development of semantic transformation was to **address the shortcomings** of syntactic transformation, especially in terms of **data integrity (**e.g. data duplicates, lack of data currency, inaccuracy, etc). [203] Transforming data means going beyond the simple question of the format, it also involves semantic alignments which pose the delicate question of **ontology alignment and semantic transformation**: *"The information needed for system design, reconfiguration planning and reactive adaptation, is based on extensive knowledge from different distributed sources in various fields of expertise. A major problem is the poor interoperability between the information and design support systems used to create, save, manage, and utilize this information. The majority of these systems use their own proprietary data structures and vaguely described semantics (…) Also, the different actors involved in the process—humans or organizations—may have different understanding of the used terms, leading to interoperability issues (…) Retrieving and utilizing information from multiple diverse sources puts high demands on semantic integration solutions*."[204]

Developing Semantic transformation

- **Semantic transformation can be defined** as "*the process of using semantic information to aid in the translation of data in one representation or data model to another representation or data model. Semantic transformation takes advantage of semantics that associate meaning with individual data elements in one dictionary to create an equivalent meaning in a second system. An example of semantic transformation is the conversion of XML data from one data model to a second data model using formal ontologies for each system such as the Web Ontology Language (OWL). This is frequently required by intelligent agents that wish to perform searches on remote computer systems that use different data models to store their data elements. The process of allowing a single user to search multiple systems with a single search request is also known as federated search*";[205]
- As this technology is recent and still at an early TRL stage, it is difficult to make advanced recommendations. It is however **clear that semantic transformation is the most promising field** for the development of **data transformation models**. Collaborative efforts that bring together value chain stakeholders (IT specialists, factory engineers, academics) are needed to develop this methodology further:
  - **A template industry agreement** specific to the development of **translation software** in **semantic transformation** is provided in **Annex I**;
- More effort is needed to develop **cross-value chain ontologies**, as the **ontology alignment step** could be greatly facilitated by the development of **vendor-neutral third-party ontologies**, using clear **formal engineering ontologies** with frameworks that should be publicly disclosed to data engineers:
  - ***"Formal engineering ontologies*** [a field that studies the methods for developing formalised ontologies] *are emerging as popular solutions for addressing the semantic interoperability issue in heterogeneous distributed environments and for bridging the gap between the legacy systems and organizational boundaries"* [206]
  - The **use of reference tables is also recommended as best practice for the development of**

[203] Balijepalli Saketh, 01/01/2019, *Best Practices in Using Semantic Transformation in Data Integration to Address Data integrity issues*, https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/24786/Balijepalli_2019.pdf?sequence=1&isAllowed=y
[204] Järvenpäa Eeva et all, 14/11/2017, *The development of an ontology for describing the capabilities of manufacturing resources*, https://d-nb.info/1165373912/34
[205] ITEA3, 01/12/2020, *Machinaide, Deliverable D1.2, State-of-the-art description*, https://itea4.org/index.php/project/workpackage/document/download/7292/D1.2%20State-of-the-Art%20Description_V1.pdf
[206] Järvenpäa Eeva et all, 14/11/2017, *The development of an ontology for describing the capabilities of manufacturing resources*, https://d-nb.info/1165373912/34

semantic transformation: *"Using online reference tables to obtain data dynamically is one of the most efficient ways to avoid obsolete data and helps with keeping the data current (Abedjan et al., 2015). This approach offers promise to organizations that embrace and implement the concept of semantic transformation for maintaining high data integrity, increasing confidence levels when making important decisions based on the data*." [207]

In addition to the **data ingestion stage,** the development of **threading techniques** is vital to achieve **data continuity** in industrial processes. These techniques also require the development of **vendor-neutral solutions**.

Developing Digital Threading for Data Continuity: best practices from Lockheed Martin[208]

- **Definition of digital thread:** "*Digital thread creates a closed loop between the digital and physical worlds, transforming how products are engineered, manufactured and serviced. Digital threads seek to create simple universal access to data. They follow a single set of related data as it weaves in and out of business processes and functions to enable continuity and accessibility*";[209]
- The current **lack of connection "back to model-based engineering requirements"** could be addressed through greater use of Digital Threading. Currently, the following elements are lacking:
  - Part numbers are sometimes missing;
  - Serialisation (serial numbers) can be inconsistent (too often ad-hoc solutions without centralised overseeing);
  - Lack of feature level tracing;
  - No culture of "delivering data artifacts" with manufactured items;
  - Lack of manufacturing semantics and ontologies (different OEMs will provide the same item with different identifiers).
- Going ahead would require the development of:
  - **Vendor-neutral "Record of Authority"** (common naming and identifying of parts and items valid across all manufacturers;
  - Manufacturers must practice **systematic serialisation** of all part numbers with the development of clear instructions for the workforce and a clear link with the digital artifact;
  - **Standardisation of digital artifacts** for streamlined integration of the entire digital value chain (from supplier to final customer, the delivery of standardised digital artifacts will become a normal expectation).

Ensuring data continuity in industrial processes also requires the development of **Asset Administration Shells**. That requires both the **Application-Lifecycle Management (ALM)** and the **Product-Lifecycle Management (PLM)** to be streamlined.

PLM is a holistic term that refers to all of the steps that surround the management of a product. However, traditional PLM systems are not designed to manage the creation of software and digitised products and systems, which has led to the development of ALM by software engineers: "*The goal of ALM is to provide a comprehensive technical solution for monitoring, controlling, and managing software development throughout the application lifecycle*."[210] The integration of digital and physical processes requires the

[207] Balijepalli Saketh, 01/01/2019, *Best Practices in Using Semantic Transformation in Data Integration to Address Data integrity issues,* https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/24786/Balijepalli_2019.pdf?sequence=1&isAllowed=y

[208] Lockheed Martin, 14/05/2021, *Usability of manufacturing data for analytics,* https://www.nist.gov/system/files/documents/2021/04/26/Wednesday%20-%2001%20-%20Usability%20of%20Manufacturing%20Data%20for%20Analytics.pdf

[209] IPC, 21/09/2021, *What is digital thread?,* https://www.ptc.com/en/blogs/corporate/what-is-a-digital-thread

[210] Deuter Andreas, Imort Sebastian, 23/06/2021, *Product Lifecycle management with the asset administration shell,* https://www.mdpi.com/2073-431X/10/7/84/pdf

integration of PLM/ALM processes. However, there remains a lack of standards to achieve this. The **Open Services for Lifecycle Collaboration (OSCL) standard** already partially aim to achieve this. However, as many digital solutions remain software-specific, integration still requires that solutions from different vendors to brought together. Currently, this step requires the **development of custom-interfaces**[211]

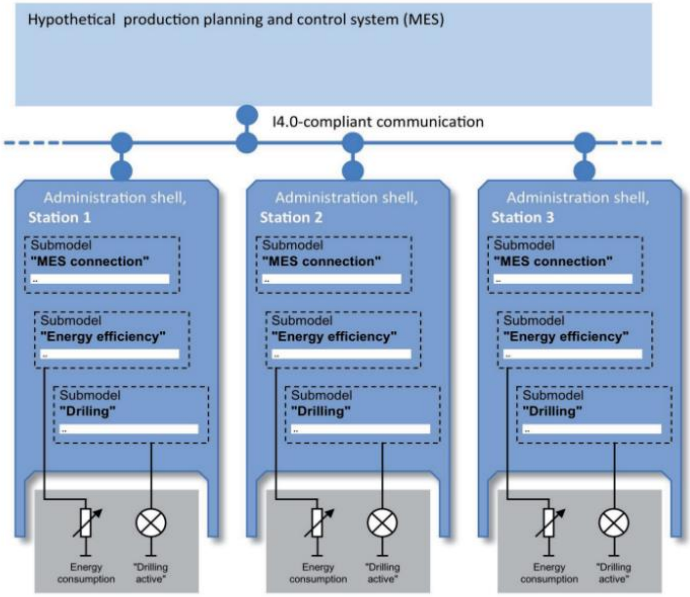
Similarly, developing integrated PLM processes is complex "*Numerous tools and data models are used throughout this process. In recent years, industry and academia have developed integration concepts to realize efficient PLM across all domains and phases. However, the solutions available in practice need specific interfaces and tend to be vendor dependent*."[212] There is simply no all-encompassing PLM data model. However, there are phase-specific and domain-specific standards. For example, Siemens developed the PLM XML schema *"(…) it supports application integration through workflows. PLM XML is a very comprehensive model. However, it is a proprietary format, and there is no widespread acceptance in practice*." [213]

A solution to tackle these difficulties and enable data continuity is the **development of a vendor-neutral AAS solution**. "*The Asset Administration Shell (AAS) aims to be a standardized digital representation of an asset (e.g., a product). In accordance with its objective, it has the potential to integrate all data generated during the PLM process into one data model and to provide a universally valid interface for all PLM (Product LiveCycle Management) phases*."[214]

AAS acts as a neutral interface for machine data exchange within the factory (vertical integration) but also along the value chain (horizontal integration). [215] **The AAS can be defined as follow**: "*The basic idea of I4.0 components is to surround every Industrie 4.0 asset with an asset administration shell that can provide a minimal but sufficient description according to the Industrie 4.0 use cases. At the same time, it is important that we are able to map existing standards in accordance with the definition of the asset administration shell in question*."[216] It is also important to note that these AAS have models but also sub-models by type of machines (drilling equipment, etc). The AAS can act as a neutral toolbox to enable and streamline exchange and proceed by silo integration.

[211] Deuter Andreas, Imort Sebastian, 23/06/2021, *Product Lifecycle management with the asset administration shell,* https://www.mdpi.com/2073-431X/10/7/84/pdf
[212] Ibidem
[213] Ibidem
[214] Ibidem
[215] Industrial internet consortium & platform industrie 4.0, *Digital Twin and asset administration shell concepts and application in the industrial internet and industrie 4.0,* https://www.iiconsortium.org/pdf/Digital-Twin-and-Asset-Administration-Shell-Concepts-and-Application-Joint-Whitepaper.pdf
[216] ZVEI, 01/04/2017, *Example of the Asset Administration Shell for Industrie 4.0 components,* https://www.zvei.org/fileadmin/user_upload/Presse_und_Medien/Publikationen/2017/April/Asset_Administration_Shell/ZVEI_WP_Verwaltungschale_Englisch_Download_03.04.17.pdf

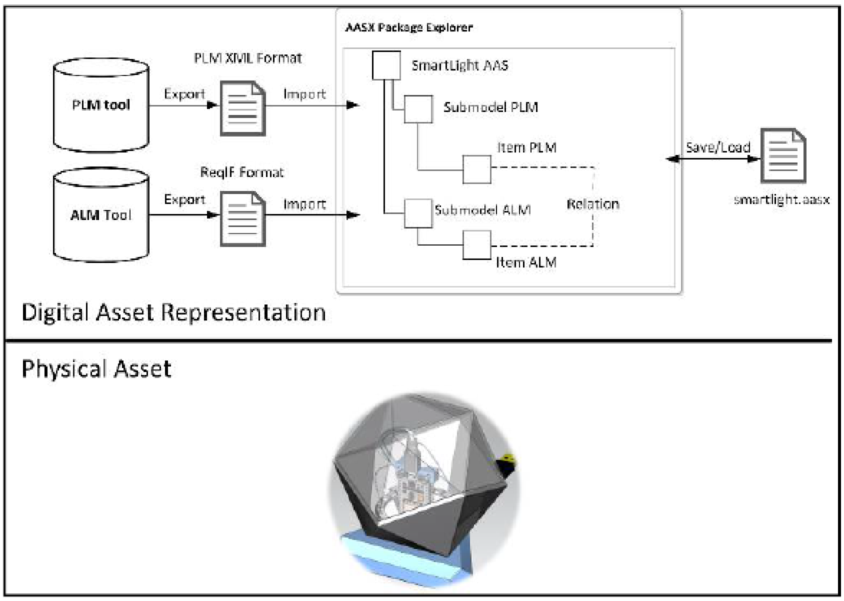
Figure 16: Hypothetical production planning and control system from the ZVEI to demonstrate the silo integration using the AAS



Source: ZVEI[217]

AAS can work as vendor-neutral digital artifact enriched with PLM/ALM data and create links between systems.

Figure 17: AAS for Data continuity in PLM and ALM processes



Source: Deuter Andreas, Imort Sebastian, 2021 [218]

[217] ZVEI, 01/04/2017, *Example of the Asset Administration Shell for Industrie 4.0 components*, https://www.zvei.org/fileadmin/user_upload/Presse_und_Medien/Publikationen/2017/April/Asset_Administration_Shell/ZVEI_WP_Verwaltungschale_Englisch_Download_03.04.17.pdf

[218] Deuter Andreas, Imort Sebastian, 23/06/2021, *Product Lifecycle management with the asset administration shell*, https://www.mdpi.com/2073-431X/10/7/84/pdf

The ability of the AAS to enable neutral exchange across the value chain between components and functions makes it a fundamental element of the future interoperability landscape. Furthermore, the concept of AAS *"enables interoperability across value stream of an asset and enables substitution of one asset by another without changing the information model and API of its digital twin. This series of specifications help developers who want to implement an interoperable digital twin for Industry 4.0 or industrial applications*." [219]

The AAS acts as a "*technology neutral information model*" Furthermore, its standardised way of providing information allows it to overcome the problem of "information silos" within a factory. In a **brownfield scenario**, it "*enables interoperability across multiple proprietary digital twins (…) In **greenfield scenarios**, the AAS can directly be adopted to implement digital twins. By following a single standard for digital twins, all industries, especially end users, benefit from an open and standardized metamodel, standardized data models with homogenized semantics (e.g., technical and operational data) and standardized APIs and infrastructure services*." [220]

Vendor-Neutral AAS solution for data continuity: benefits and key requirements

- Several barriers block the development of streamlined data continuity in industrial processes, looking at both the PLM and ALM systems:[221]
  1. "*Due to **individual item naming** in the systems, different interpretations of an artefact occur within companies;*
  2. *Different **data formats** are used for the same processes.*
  3. **Lack of data accessibility:** *"Data should be accessible in such a way that it can also be used in areas for which it was not intended when it was created. This also includes independence from the location as well as from the company.*
  4. ***The completeness of data** cannot be guaranteed because it is often stored in different data repositories or even exists as documents in digital or paper form.*
  5. *Data access is hindered **by data security** requirements.*
  6. *Effort is needed to make the data of **one PLM system accessible to other systems***."

As already mentioned, a vendor-neutral AAS would be a key solution for the integration of PLM and ALM system. However, the **AAS concept and specifications do not detail how to develop a vendor-neutral AAS.**

- **The AASX Package Explorer** is an open-source software tool enabling "*a user to create, edit, and view an AAS. Furthermore, it provides access to the AAS via an Open Platform Communications Unified Architecture (OPC UA) or Message Queuing Telemetry Transport (MQTT) interface*." This tool is an interesting support for the **development and implementation of AAS**;
- Developing the requirements for these solutions will always be case-specific. However, based on the Siemens case study, Deuter Andreas and Sebastian Import identified the **9 key requirements for the development of a vendor-neutral AAS.** These requirements could be used as a starting point for the development of an **Industry Agreement for a vendor-neutral AAS:**

---

[219] ZVEI, 01/04/2017, *Example of the Asset Administration Shell for Industrie 4.0 components*, https://www.zvei.org/fileadmin/user_upload/Presse_und_Medien/Publikationen/2017/April/Asset_Administration_Shell/ZVEI_WP_Verwaltungschale_Englisch_Download_03.04.17.pdf
[220] Ibidem
[221] Deuter Andreas, Imort Sebastian, 23/06/2021, *Product Lifecycle management with the asset administration shell,* https://www.mdpi.com/2073-431X/10/7/84/pdf

Table 10: Key requirements for the development of a vendor-neutral AAS

General Transversal requirements	
R1	*The concept must be based on a standard in order to enable vendor-independent integration.*
R2	*The underlying data models must provide comprehensive access to all data created for or by an asset in order to include data that than PLM/ALM data.*
Specific Requirements	
R3	It must be possible to link a PLM work item with a ALM item and vice versa
R4	it must be possible to link a ALM item with a PLM document.
R5	It must be possible to access the data of any attribute of a PLM work item from the linked ALM item and vice-versa.
R6	It must be possible to link more than one PLM work item with one ALM item in a single action.
R7	When creating a traceability report in the PLM system, all linked ALM items must be included and vice versa.
R8	The status of a PLM item can only be changed if the status of the linked ALM item has a dedicate status.
R9[222]	PLM and ALM must be integrated in the AAS to cover IPR continuity in a dynamic data environment.
R10	Agile architecture for the addition of case-specific requirements.

Source: Taken from Deuter Andreas and Imort Sebastian[223] (with modifications from CARSA)

Even though **benefits are hard to** monetise, we can foresee that generated benefits could reach at least -50% reduction in time and personnel requirements. Commissions would be reduced by a -50% as well, and digital twin capabilities would be faster to deploy. Furthermore, importing errors related to system integration would be removed.[224] We can already have a first overview on the impact such solutions could bring in "an ideal world scenario". System integration is a costly and cumbersome procedure that is typically outsourced because of the difficulty to carry it out. Standardisation would enable these tasks to be progressively carried-out internally and would help to reduce costs. Complexity, and therefore cost, would gradually decline, thereby freeing up resources that could be reallocated to other activities. Plug & Play machines would also be interchangeable, increasing competition intensity between suppliers and driving down prices. These impacts are, however, unlikely to materialise in niche markets, where natural monopolies will remain.[225]

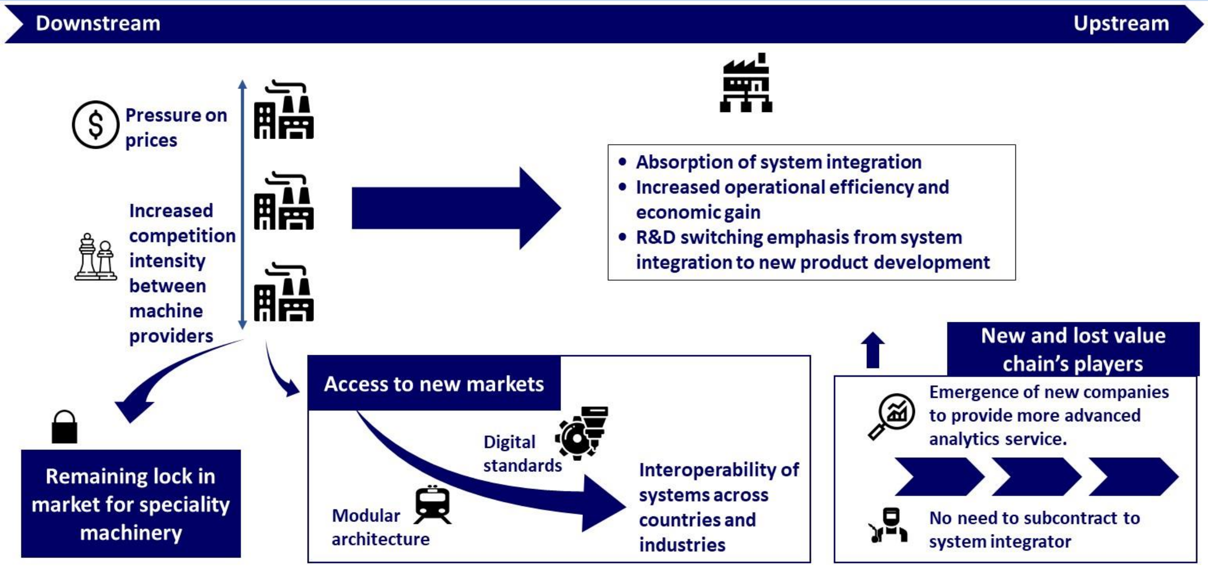
Impact of Plug & Play Machine-tools

- The **achievement of better data integration and continuity aim to produce** *"(…)* ***actionable business intelligence (BI)*** *(…) Business intelligence plays a key role in decision-making in the business world (…) BI initiatives are enabling businesses to gain "insights from the growing volumes of transaction, product, inventory, customer, competitor, and industry data generated by enterprise-wide applications such as enterprise resource planning (ERP), customer relationship management (CRM), supply chain management (SCM), and knowledge management"*[226]

[222] This requirement was added by CARSA.
[223] Deuter Andreas, Imort Sebastian, 23/06/2021, *Product Lifecycle management with the asset administration shell,* https://www.mdpi.com/2073-431X/10/7/84/pdf
[224] These benefits were detailed by research and development manager in mechanical engineering with expertise in virtual twinning.
[225] Kerber Wolfgang, Scweitzer Heike, 2017, *Interoperability in the digital economy*, https://www.jipitec.eu/issues/jipitec-8-1-2017/4531
[226] Balijepalli Saketh, 01/01/2019, *Best Practices in Using Semantic Transformation in Data Integration to Address Data integrity issues,* *https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/24786/Balijepalli_2019.pdf?sequence=1&isAllowed=y*

- According to a study, mechanical engineering companies involved in standardisation committees have achieved a higher growth in turnover and international sales than their counterparts, up by 23% and 20% respectively;[227]
- Up to 60% in cost savings from improved operational efficiency according to an interview with a stakeholder involved in the industry;
- Other gains. E.g., energy consumption cut by approximately 15% by automated machine tools during machining, if they use automated energy monitoring systems that require a Plug & Pay machine;[228]
- Digital transformation of mechanical engineering is already expected to generate a cut in annual operational costs by 3.6%;[229]
- Progressive integration of the Product Development Process through the integration of the data ecosystem leading to less errors and misinterpretation of data. This is expected to translate into lower costs, improved quality and a faster development lead time for new products;[230]
- Competitiveness gains are especially important for the European Mechanical Engineering industry as the industry is confronted by low structural growth and is dependent on its ability to remain competitive in foreign markets.[231]

Figure 18: Impact of P&P machine tools on mechanical engineering: "Ideal World Scenario"



Source: CARSA, 2020. This figure was created using resources from Flaticon.com.

[227] CETIM, 2017, *It pays to take part in the standardisation process*, https://www.cetim.fr/en/News/headlines/In-brief/It-pays-to-take-part-in-the-standardisation-process
[228] Vijayaraghavan A, Dornfeld D, 2010, *Automated energy monitoring of machine tools*, https://core.ac.uk/download/pdf/191543802.pdf
[229] European Commission, 2017, *Digitising mechanical engineering: leveraging the potential of the cloud and data*, https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/DTM_%20Digitising%20mech%20eng%20v2.pdf
[230] Aubry Alexis, 2017, *Semantic interoperability for an integrated product development process*, https://hal.archives-ouvertes.fr/hal-01537847/document
[231] VDMA, 2017, *Mechanical engineering demands a decade of innovation*, https://bayern.vdma.org/en/viewer/-/v2article/render/16646489

Manufacturing Global Response Initiative (MGRI) as best practice

Manufacturing Global Response Initiative[232] is a joint effort set by the World Economic Forum (WEF) and the Digital Factory Alliance (DFA) that calls for public authorities, digital and manufacturing industries of all sizes to join forces for coordination, collaboration, better preparedness, agility, trusted information sharing mechanisms, anticipation, availability, efficiency and efficacy to address future global outbreaks. MGRI will follow the resiliency compass, presented in a recent WEF white paper.[233]

The initiative should articulate the trusted international digital factory community for manufacturing repurposing knowledge sharing and manufacturing resilience leadership.

The initiative is setting up a federation of rapid response manufacturing networks and harmonising a set of manufacturing repurposing protocols, digital infrastructures and investments in the modernisation of European factories at large. Such Plug & Respond mechanisms enables the flexible set and ramp-up of manufacturing networks as a service and on-demand taking advantage of digital technologies such as data spaces, blockchain and advanced manufacturing technologies such as 3D printing.

Development of a framework architecture for the standardisation of lab robotics

As always for chemicals, the question of vertical integration and interoperability of systems depends on the segment of the value chain that is investigated. Pharmaceuticals and petrochemicals will face different challenges and technical difficulties. However, the general framework for the development for interoperability of machines and systems in manufacturing settings has already been clarified in the section on “Plug & Play machine tools”. Consequently, this section will focus specifically on the question of lab robotics in the chemical industry.

Two situations should be distinguished:[234]

- **Niche markets**. For advanced research done in academic settings or niche chemicals, specific robots have to be tailor-made to meet specific technical needs. These machines are often delivered without any ontologies/semantics or even an operating system;[235]
- In the context of **cognitive and smart factories**, the real challenge is the **vertical integration of R&D activities and robots.** Large manufacturers have the means to develop on-the-spot solutions and systems, but they lack the resources to develop interoperable systems. The key challenge is how to connect laboratory 4.0 solutions to the rest of the production process and activities.[236]

To achieve this, three layers of interoperability in smart manufacturing must be considered: **data transfer protocols**; **data representation and presentation** (see section on molecule standards); and **semantics/understanding of data**.[237]

[232] https://digitalfactoryalliance.eu/manufacturing-resiliency-manufacturing-global-response-initiative/
[233] World Economic Forum, July 2021, White Paper, *The Resiliency Compass: Navigating Global Value Chain Disruption in an Age of Uncertainty*, https://www.weforum.org/whitepapers/the-resiliency-compass-navigating-global-value-chain-disruption-in-an-age-of-uncertainty
[234] Distinction based on the results from the Workshop on chemicals.
[235] Interview with an academic active in the chemical industry.
[236] Results from the Workshop on chemicals.
[237] Lundholm et all, 01/12/2016, *Integration of Digital Factory with smart factory based on internet of things*, *https://www.researchgate.net/publication/306023791_Integration_of_Digital_Factory_with_Smart_Factory_Based_on_Internet_of_Things*

Integrating Laboratory 4.0 in chemical cognitive plants

Barriers to interoperability in chemical plants are multiple and must be taken into consideration simultaneously for vertical integration of Laboratory 4.0 in cognitive plants. Bottlenecks and difficulties occur when modern labs are connected to production processes.[238]

- Standards are needed to integrate the entire Laboratory 4.0 environment. For example, the **Interoperability of Electronic Lab Notebooks** should be taken into consideration as part of the integration of the Laboratory 4.0 digital ecosystem. These devices typically have their own metadata standards and often also feature semantic heterogeneity. Research data infrastructure in the lab (e.g. NFDI4CAT or NFDI4Ing) ingests data processes by ELNs, hence the need for agreed protocols, APIs and programmatic interoperability (i.e., minimum standards in practically dealing with data);[239]
- **File formats harmonisation**. Various formats are used for different types of chemical applications and are not necessarily compatible or readable by other software/machines.[240] The most important formats are the MODL, Molfile, SDfile, RDfile, SMILES, PDB file, CIF, JCAMP, CML. [241]

Table 11: Main data formats used in the chemical industry[242]

Format type	Codename of the format
Document formats	mrv, Marvin Documents (MRV) cdx, cdxml, ISIS/Draw sketch file (SKC) skc, ChemDraw sketch file (CDX, CDXML)
Molecule file formats	mol, rgf, sdf, rdf, csmol, csrgf, cssdf, csrdf, cml, smiles, cxsmiles, abbrevgroup, peptide, sybyl, mol2, pdb, xyz, inchi, name
Graphics formats	jpeg, msbmp, png, pov, svg, emf, tiff, eps
Compression and encoding	gzip, base64

- **In addition**, technical barriers specific to chemical engineering exist and must be addressed in R&D processes. [243] For example,
  1. **Terminologies** used by different industry sub-fields should be harmonised. The same words can designate different realities. For example, in "**chemical crystallography" vs "protein crystallography"** "*In coordination chemistry as well as CX a `ligand' is an ion or molecule (a functional group) that binds to a central metal atom to form a coordination complex. In biochemistry it is a substance that forms a complex with a biomolecule to serve a biological purpose. In protein-ligand binding, the ligand is usually a molecule that produces a signal by binding to a site on a target protein. This is in effect the same definition for two very different aspects. (…) this definition, and the field in which it is applied, is rarely clearly listed and confusion is experienced when inter-disciplinary research is conducted*."
  2. **Electron-density maps versus peak list**. MX and CX can make different uses of electron-density maps can lead to a different resolution and software architecture.

Several important initiatives already exist in the industry to tackle these barriers. For example, **SILA** (standardisation in lab robotics), is an open-source initiative that is developing interfaces for systems and data

238 Workshop on the chemical industry.
239 Task force Charter, sin dato, *Semantic Interoperability*, https://www.eosc.eu/sites/default/files/tfcharters/eosca_tfsemanticinteroperability_draftcharter_20210614.pdf
240 Brin Alice & Helliwell John, 06/08/2019, *Why is interoperability between the two fiels of chemical crystallography and protein crystallography so hard*, https://journals.iucr.org/m/issues/2019/05/00/lt5022/
241 Ibidem
242 Institute of Science and technology, sin dato, *Chemistry of Eletronic Materials*, https://www.sathyabama.ac.in/sites/default/files/course-material/2020-10/UNIT-5_2.pdf
243 Brin Alice & Helliwell John, 06/08/2019, *Why is interoperability between the two fiels of chemical crystallography and protein crystallography so hard*, https://journals.iucr.org/m/issues/2019/05/00/lt5022/

management systems.[244] This is also the case for NCCLS in the US, which develops modular architecture standards (for firmware and hardware) for clinical labs, with a view to achieving Plug & Play interoperability.[245] Finally, the Allotrope foundation (interoperability initiative) and the Obelisk Group (standardisation for cheminformatics) are also important initiatives.

What is lacking, however, is an inclusive effort, involving all digital value chain stakeholders, to build a coherent and consistent framework to develop cross-domain standards and architectures.

Standardisation for lab automation/robotics

To integrate and standardise Lab Robotics in chemical plants, new standards are needed. Simultaneously, work is needed to integrate Laboratory 4.0 with other production functions.

The following **pathway** is suggested to address these barriers:

- Develop a common **framework architecture** to integrate the different standards and IAs for lab robotics.[246] This work should involve the participation of the entire digital value chain and should build on the results from the Obelisk Group, SiLA and the Allotrope foundation. The main step would be to replicate/extend these initiatives to cover the dimensions and barriers that they do not yet cover (e.g. SiLA is focuses on pharmaceuticals);
- Develop an **Open Library** of standards, ontologies and semantics that can be reused to develop digital systems for niche robots;
- **A draft industry agreement** to serve as support for the inclusive development of new interoperable standards is provided in Annex I;
- Develop **Large-Scale Pilot projects** in Europe to test and operationalise these frameworks in different industrial settings (pharmaceutical, chemical, clinical, etc.);
- **Research initiatives** funded by public authorities should develop niche standards for lab robotics (start-ups and academics are the ideal candidates to participate in these research programmes). These initiatives should be connected with the **framework architecture initiatives,** starting with the mapping and development of a library of solutions that can be extended to niche markets and serve as a library of solutions for small businesses;
- Develop a common platform to bring together users and developers from different industries that are active in Laboratory 4.0 (biotechnology, clinical lab, chemical lab, etc.);
- Cross-fertilise, reuse and upscale existing solutions across domains and industries.

Impact of an industry agreement for standardised lab robotics

- The total world laboratory automation market is estimated to have reached €3.44 billion in 2017, €3.55 billion in 2019, and is expected to reach €4.4 billion by 2022 and €5.8 billion by 2027;[247]
- The lab robotics market is estimated to have reached €157.3 million in 2019 and is expected to reach €232.4 million by 2025;[248]

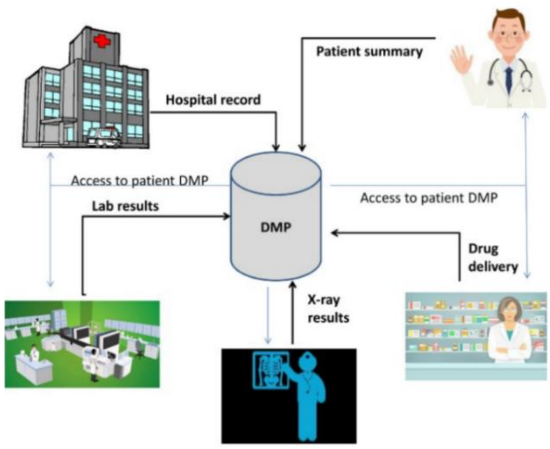
[244] SILA, Sin dato, *SILA 2 is now officially released*, https://sila-standard.com/
[245] Hawker C.D. Schlank M. R., 2000, *Development of standards for laboratory automation*, https://pubmed.ncbi.nlm.nih.gov/10794772/
[246] For a sketch of what a framework architecture could look like, kindly refer to Roche, 2018, *Intelligent data lab automation*, http://www.globalengage.co.uk/prdi/Haffner.pdf
[247] Results originally provided in USD were converted using the USD to EUR exchange rate of 12/11/2020. Data come from Marketsandmarkets, 2017, *Lab Automation Market*, https://www.marketsandmarkets.com/Market-Reports/lab-automation-market-1158.html and Verifiedmarketresearch, 2019, *Lab automation market size by application*, https://www.verifiedmarketresearch.com/product/lab-automation-market/.
[248] MordorIntelligence, 2020, *Lab robotics market*, https://www.mordorintelligence.com/industry-reports/laboratory-robotics-market

- Operational efficiency gains are difficult to measure as they depend on the activity being carried out in the lab. For example, a study in total laboratory automation in a clinical chemistry laboratory showed that automation has resulted in the price of the produced chemical test falling from 0.79 dollars to 0.15 (-81%). In addition, the turnaround time for results availability was substantially reduced;[249]
- "(…) *the total market for automation systems and equipment would be significantly greater with standards than without standards especially if customers were not forced to purchase everything from one vendor, and that there might be competitive pricing and new technology fostered via the standards."* [250]

Integration of digital/electronic health record systems

An electronic health record (EHR) can be defined as "*a digital version of a patient's paper chart. EHRs are real-time, patient-centred records that make information available instantly and securely to authorized users. While an EHR does contain the medical and treatment histories of patients, an EHR system is built to go beyond standard clinical data collected in a provider's office and can be inclusive of a broader view of a patient's care*."[251] ERH offer several key advantages as they contain a patient's entire medical history (diagnoses, medication, allergies, etc), they grant access to evidence-based tools that support medical decision-making and they automate provider workflows. Furthermore, the key feature of ERH is that they enable data sharing between providers and healthcare organisations (laboratories, specialists, medical imaging facilities, pharmacies, emergency facilities, and school and workplace clinics), and therefore consolidate all information gathered from all clinicians involved.[252]

Table 12: Schematic representation of an Electronic Health Record



Note: DMP is the French expression for EHR.
Source: Burnel Philippe[253]

The EHR is therefore the foundation on which the digitisation and digitalisation of the healthcare industry will be built. However, in spite of all the promises this novelty offers, stakeholders are finding it difficult to exploit them. The complex task of making eHealth systems interoperable can be divided into three steps:

- **Digitisation and nationwide standardisation of EHRs**;
- **Developing cross-platform interoperability.** Once digitised, the information in the EHR must be uploaded onto digital platforms. However, many different platforms are built on proprietary systems, which can lead to a lack of compatibility and insufficient interoperability;
- **Cross-border interoperability** for a truly European eHealth data space.

[249] Clin Chim Acta, 2003, *The effects of total laboratory automation on the management of a clinical chemistry laboratory*, https://pubmed.ncbi.nlm.nih.gov/12589970/
[250] Hawker C.D. Schlank M. R., 2000, *Development of standards for laboratory automation*, https://pubmed.ncbi.nlm.nih.gov/10794772/
[251] KMENDIS, sin dato, *Electronic Health Records, https://kmendis.net/electronic-health-records/*
[252] Ibidem
[253] Burnel Philippe, 09/09/2018, *The introduction of electronic medical records in France: More progress during the second attempt*, https://www.sciencedirect.com/science/article/pii/S0168851018303245

Each of these three points will be scrutinised and guidelines/recommendations will be proposed to show how higher interoperability at each level can be achieved, starting with the question of developing common nationwide **eHealth Records.**

Lessons from France: developing a nationwide standard for EHR[254]

- **The French initiative was launched in 2008** with the aim of developing a common nationwide Electronic Health Record. It was expected to generate a total ROI of around €3.5 billion. The project cost a total of €210 million and only generated 160,000 file openings. The project was almost abandoned because of how little success it was achieving;
- The project was brought back to life in 2013. New strategic guidelines and orientation helped to make the new version of the project a success. It can now be highlighted as a **Best Practice:**
  - **Co-definition.** The new initiative was launched by the French Health Ministry, following recommendations made by an internal working group that brought together a wide range of different stakeholders. Several modifications were envisioned from the start:" The *DMP* [French expression to designate an EHR] *has been clearly positioned as* ***a professional tool*** *that benefits and remains* ***under the control of the patient*** *(in accordance with general patient rights).* ***A number of functional changes*** *were planned: creation of DMPs by the patients themselves, automatic documentation of DMPs using the data resulting from claim procedures, and accessibility to DMPs via mobile apps*"; [255]
  - Stronger policy-involvement was key and was achieved by putting a **single organisation in charge of the project:** the National Health Insurance System. The **choice of the organisation** was correct, as it is an independent organisation that is closely linked with the professionals and patients familiar with their interests and needs;
  - The **first initiative failed mainly because of resistance from practitioners and doctors**. The initiative was managed in a "top-down" manner and did not give enough consideration to the requirements/needs of practitioners and patients. The renewed project was designed to include all key stakeholders. For example, directly contacting patients (by email), providing incentives for doctors, and developing integrated solutions that are easy for software companies to use;
  - Finally, **other good practices can be learned from:** "*The deployment of the policy linked political aspects (communication about public health goals), strategic aspects (goals are integrated in the fees' negotiation of the different healthcare professionals (physicians, nurses, midwives and pharmacists) and operational aspects (supporting software companies and professionals to guarantee easy use and reliability of software).*" [256]
- **<u>The key success factor</u>** for the implementation of a nationwide EHR is the **commitment and involvement of all stakeholders**. Conversely, the **most <u>critical failure factor</u> stems from the "***negative reaction to any change from the medical, nursing and administrative community is considered as the most critical failure factor. (…) An issue that should never be overlooked or underestimated is the* ***alignment between the functionality of the new EHR system and user requirements***." [257]

The French case provides clear lessons and key guidelines for the development of an integrated nationwide EHR. However, this is only the first step of the process: uploading digitised information onto digital platforms, which then poses another challenge – to develop an interoperable health data space: "***EHR interoperability refers to the digital health record systems' ability to openly communicate with each other****, allowing providers*

[254] Burnel Philippe, 09/09/2018, *The introduction of electronic medical records in France: More progress during the second attempt*, https://www.sciencedirect.com/science/article/pii/S0168851018303245
[255] Ibidem
[256] Ibidem
[257] Fragidis Leonidas, Chatzoglou Prodomos, 02/03/2018, *Implementation of a nationwide electronic health record (EHR),* https://pubmed.ncbi.nlm.nih.gov/29504871/

to access and exchange patient healthcare information. The systems must process the data and present the information in an easily accessible manner to qualify as truly interoperable."[258]

However, important initiatives for technical interoperability already exist. The three main ones are:

1. ***"Fast Healthcare Interoperability Resources (FHIR***): *Protocols for exchanging care data at the clinic and administrative levels;*
2. ***Health Level 7 (HL7):*** *Standard controlling the exchange of data through the v2 messaging system;*
3. ***Health Relationship Trust (HEART):*** *Profiles that allow patients to control access to their healthcare data."* [259]

In spite of these initiatives: "(…) [In the US] *the vast majority of* ***EHR systems in use today do not allow practices to share patient records with ease****. Instead, they must exchange information through fax and other archaic methods — or cycle through more than a dozen standalone systems. At the average hospital, for example, providers have to tap into 16 distinct EHR systems to retrieve information and update their patients' records*." [260]

The same comment can be made in Europe. **For example, in Belgium,** we can identify 6 EHR platforms that are currently in use. Xperthis is used by 95% of Walloon hospitals, Primuz is used in Brussels, Cerner in Anvers, Epic in one hospital based in Brussels, and Nexuzhealth and Xperthis that are both mainly used in Flanders. Legacy systems that are expected to be replaced soon were not considered in this list.[261]

The difficult task of developing cross-platform interoperability

- Even though room remains for the development of new technical initiatives for technical interoperability, the main reason behind the lack of technical initiatives is the **lack of uptake of existing interoperability initiatives by industry**. The main reasons for this situation are:[262]
    - **Many platform providers lack the incentives** to develop interoperability, as proprietary standards create market lock-in which they benefit from. The belief that greater market share can be achieved through this type of lock-in remains strong in the value chain;
    - This leads to a **lack of market-demand for these standards** (for example in tenders and published call for procurements);
    - The lack of **awareness and education** among healthcare providers and patients leads to a low level of activism. Interoperability is a complex topic that many people do not fully understanding. Greater awareness would help to create a demand for the development of interoperability which would then increase market pressure on solution providers to meet that demand;
- Following the key lessons established in France for the development of a common nationwide EHR model, the healthcare sector would benefit from the launch of **public events promoting a cultural switch**;
- New interoperable solutions should also place more emphasis on **data retrievability**, a dimension that is not really addressed by the HLF 7 FHIR.[263]

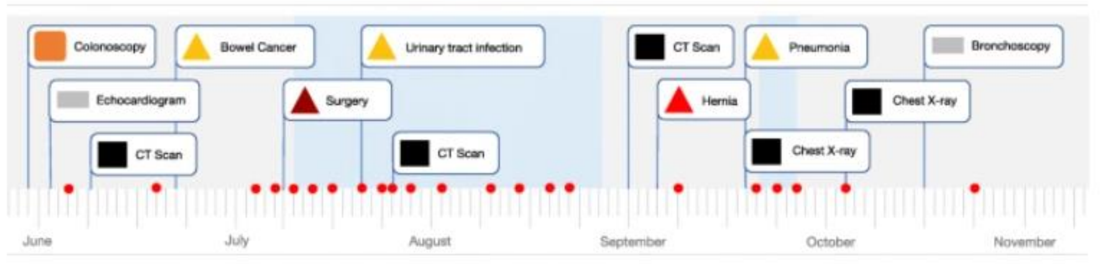
As already mentioned, the **co-design of the platform and apps** is crucial. **Patients must be in control** to protect their rights and alleviate their genuine concerns over a potential loss of data sovereignty.

[258] PROGNOCIS, 14/06/2019, *HER Interoperability — Why is it important*, https://prognocis.com/ehr-interoperability-why-is-it-important/
[259] Ibidem
[260] Ibidem
[261] UZA, 2020, *Integrated HER's in Belgian hospitals, what about Antwerp*, *http://www.karva.be/wp-content/uploads/2019/09/Reinhart-Maertens-Integrated-EHR%E2%80%99s-in-Belgian-hospitals_20190830.pdf*
[262] This point and the following bullet points were raised during the workshop on the healthcare industry.
[263] PROGNOCIS, 29/09/2016, *Is HER interoperability possible with a national patent identifier*, https://prognocis.com/is-ehr-interoperability-possible-with-a-national-patient-identifier/

Co-definition of EHR systems and best practices for user-friendliness

- **The importance of co-defining a platform's interfaces**. Existing solutions are frequently under-used by their users because of a lack of user-friendliness. **For example, a study carried out with Belgian physiotherapists** found that most of the functionalities were not used simply due to a lack of knowledge of their existence. When faced with new digital tools, users tend to stick to what they are familiar with and know how to use. More effort is needed to co-define solutions to develop more user-friendly devices that will enable practitioners to use advanced services and make their use common practice;[264]
- **All platforms must be built on the idea of ensuring patient control.** EHR systems that do not meet this requirement are very likely to be rejected by a majority of the public, leading to eventual abandonment of the initiative. Australia has established a best practice as it is the only country that allows patients to **control who sees the information and to request corrections to their own health data.[265]**
- Even though co-definition should be left to stakeholders, a study developed by Warren Leigh[266] et all identified concrete guidelines for the design of **Electronic Health Record interfaces that enable users to easily learn to use advanced functions**;
- Design principles should be based on simplicity, beauty, clarity and user-friendliness. Design should be intuitive to ensure access to accurate information that can be rapidly assessed and understood:
  - Principles of information hierarchy (e.g. highlighting in bold letters) for more important information;
  - Interactive features such as infographics are very useful to help patients understand medical data (including colour coding for the gravity of conditions);
  - Timeline design, anatomical-based design, calendar design and other abstract design can support data understanding;
  - Colour and shape coding for enhanced user-friendliness.

Table 13: Example of a prototype of user-friendly EHR interface.



Prototype EHR interface design based on input from workshop participants

Source: Weight et all[267]

Finally, on the topic of cross-platform interoperability, even though **public authorities have a vital role to play**, local stakeholders will not remain passive. The full integration of the landscape is unlikely to take place directly at national level. It will more likely be done through regional/local initiatives that will then have to be bridged by developing cross-platform interoperability.

[264] Ghent University — Faculty of Medicine and health sciences, 2017 [Thesis], *The use of Electronic Health Records in Belgian Physiotherapy*, https://libstore.ugent.be/fulltxt/RUG01/002/350/192/RUG01-002350192_2017_0001_AC.pdf
[265] Osma Hafizah, 12/12/2018, *Australia leads the world in personal control of electronic health records*, https://www.healthcareit.com.au/article/australia-leads-world-personal-control-electronic-health-records
[266] Warren et all, 03/12/2019, *Working with patients and the public to design an electronic health record interface: a qualitative mixed-methods study*, https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-019-0993-7
[267] Ibidem

Example of pathway for the definition of a common platform: University Hospital from Anvers[268]

The university hospital in Anvers (commonly referred to as "UZA") has been working on digital health record system integration. Launched in May 2020, the hospital has developed a multi-hospital EHR platform using the following process:

- Selection of a high-end **integrated platform** (in this case, the Cerner Millenium platform was selected after a market study and a tender process);
- Establishment of a **consortium of partners** (hospitals and clinics) to join the project (in this case, AZ Klina, AZ Monica and Helix Network joined the initiative). Convincing partners to join required the development of intermediate solutions as part of a transparent long-term shared objective;
- Joint (all partner) development of the **integrated multi-hospital HER**, based on one data model, one record and respect for operational individuality.

Added to the EHR and the cross-platform interoperability challenge, the subject of **cross-border interoperability is crucial, especially in a European context.** The development of a **European Health Data Space** has become one of the priorities of the European Commission for the 2019-2025 period. The future health data space will be built on three pillars: governance, data quality and interoperability of infrastructures.[269] One can highlight that very positive features are covered in the plan (data reusability, mapping of RGPD practices, work for syntactic and semantic interoperability, policy actions, development of FAIR principles for health data).

Work to develop an integrated and interoperable eHealth data space began in 2008, when the EU undertook important efforts to "*build a European framework for interoperable exchange of eHealth information within EU Member States, starting with Patient Summary and the ePrescription services, through the epSOS (Smart Open Services for European Patients) initiative.*" The epSOS (launched in 2008, including the participation of 25 countries and 50 beneficiaries) and the OpenNCP (software implementation of epSOS) are the two main initiatives that have been tested in European projects such as eStandards, Trillium II and International Patient Summary project (IPS). "*On this basis, the European Commission commissioned the European Committee for standardization (CEN) to create the European Standard 17269 titled 'The Patient Summary for Unplanned, Cross-border Care."*[270]

The main goal of **epSOS** is to "*develop a practical eHealth framework and an Information & Communication Technology infrastructure for enabling interoperable access to patient health information, with respect to basic documents namely Patient Summary (PS) and ePrescription (eP)*". The **epSOS vision** dictates "(…) *that each country must deploy a National Contact Point for eHealth (NCPeH), acting as a bidirectional interface between the existing national IT infrastructures and those provided by the common European infrastructures supporting eHealth functions."*[271]

[268] UZA, 2020, *Integrated HER's in Belgian hospitals, what about Antwerp*, http://www.karva.be/wp-content/uploads/2019/09/Reinhart-Maertens-Integrated-EHR%E2%80%99s-in-Belgian-hospitals_20190830.pdf
[269] European Commission, sin dato, *Espace européen des données de santé*, https://ec.europa.eu/health/ehealth/dataspace_fr
[270] Nalin Marco et all, 01/06/2019, *The European cross-border health data exchange roadmap: Case study in the Italian setting*, https://www.sciencedirect.com/science/article/pii/S1532046419301017
[271] "***The PS is a standardized set of patient data including:***
1. General information about the patient (name, birth date, gender, etc.);
2. A medical summary consisting of the most important patient data (e.g., allergies, current medical problems, etc.)
3. A list of current medication including all prescribed medication that the patient is currently taking
4. Information about the PS itself (e.g., when and by whom was the PS generated or updated).
The eP includes two main processes:
1. the electronic prescription of drugs, transmitting the information to the pharmacy where it is being retrieved;
2. eDispensing (eD), i.e. the retrieval of an eP, and dispensing of the drug to the patient, and the submission of a report for the medicine dispensed."
Nalin Marco et all, 01/06/2019, *The European cross-border health data exchange roadmap: Case study in the Italian setting,* https://www.sciencedirect.com/science/article/pii/S1532046419301017

Other key European initiatives for cross-border eHealth interoperability [272]

- The **"European Committee for standardisation group of Health Informatics (CEN/TC 251)" which** released a standard for Patient Summary through the IPS project, along with an implementation guide (developed in partnership with HL7) using the "European Guidelines on Cross-Border care" as a starting point. A draft IPS standard, "prEN 17269 – The Patient Summary for Unplanned, Cross-border Care", was approved in January 2019;
- ***"prTS 17288-****The International Patient Summary: Guidance for European Implementation Technical Specification", specifies how the standard PS can be deployed within a European context. These aspects will be tested in the frame of the Trillium II project (…) Trillium II is advancing the IPS standard to enable people to access and share their health information for emergency or unplanned care anywhere and as needed (…)."*
- Legal actions have been undertaken to ensure secure the identification of patients and healthcare providers. Furthermore, "*new guidelines specify how patients, upon explicit request, can have a summary of their EHR available when visiting another country in the EU. Similar guidelines have been adopted by the eHN in 2014 for the eP minimal dataset".* (…)
- "*To stimulate the development of generic cross-border eHealth services, 16 Member States received financial support under the Connecting Europe Facility (CEF). CEF is a key EU funding instrument of the Innovation and Networks Executive Agency (INEA).*"

Finally, the **NCPeH** and the KONFIDO initiatives must be further detailed as there are crucial elements in the European interoperability framework

"The eHN (eHealth Network) released the "Guideline on an Organisational Framework for eHealth National Contact Point". The main architectural element of this framework is the ***NCPeH****, which constitutes the* ***country's communication gateway*** *providing the interface (not only technical) between the National Infrastructure (NI) and the EU network of other Member States' NCPeH, as well as with the central EU services. The NCPeH must be recognizable both in the EU domain (with the NCPeH of other countries) and in the national domain, acting as the main interface between the two. Every NCPeH can work in two different scenarios, when a patient is travelling abroad for any reason (holiday, study, work relocation, etc.): Country-A: It is the Country of Affiliation (…) Country-B: It is the Country of Treatment"*[273]

The **KONFIDO** ("Secure and Trusted Paradigm for Interoperable eHealth Services). This EU-funded project explores new cutting-edge technologies (homomorphic encryption, physical unclonable functions, event management systems, blockchain-based auditing, etc.) to provide new, secure and ethical options for the protection of eHealth cross-border information sharing. Pilot projects were conducted to test the approach and key lessons and recommendations for the integration of eHealth can already be made, based on the results of these pilots (launched in Spain, Denmark and Italy in 2019).[274]

Based on the KONFIDO project, the following key barriers to the development and deployment of an interoperable European eHealth data space have been identified:[275]

1. *"Not all EU Member States are aligned with the JASeHN agreement;*
2. *Different consent mechanisms exist among Member States;*
3. *Lack of standard EHR system in Member States;*
4. *Different implementation of EU regulations among Member States;*
5. *Different information workflows among National Infrastructure and healthcare organisations;*

[272] Nalin Marco et all, 06/2019, *The European cross-border health data exchange roadmap: Case study in the Italian setting,* https://www.sciencedirect.com/science/article/pii/S1532046419301017
[273] Ibidem
[274] Ibidem
[275] Natsiavas Pantelis et all, 18/06/2021, *Developing an infrastructure for secure patient summary exchange in the EU context: lessons learned from the KONFIDO project,* https://journals.sagepub.com/doi/full/10.1177/14604582211021459

6. *Lack of harmonisation in rules, processes, and safeguards;*
7. *NCPeH deployments in Member States are still in early stages;*
8. *Lack of the budget to address security aspects by healthcare organisations*."

Developing European eHealth Interoperability

The complexity of the task and the importance of stakeholder involvement means that the effort of developing interoperability should be seen as **a shared agenda** that requires a cultural shift in healthcare organisations: *"(…) takes a collaborative effort by government, the private sector, health care providers and others stakeholders, focusing on several critical building blocks to build towards a nationwide interoperable health information infrastructure*."[276]

The main learning is that the principal barriers were not technical; rather, they were **cultural, administrative and organisational**. Building on the lessons from KONFIDO and national initiatives to develop interoperability and nationwide EHR, we provide a list of barriers to the development of interoperable transnational EHR systems and propose technical countermeasures in the table below.

Table 14: Key technical barriers and recommendations for their removal

Barrier	Description of the problem and suggested countermeasures
Lack/absence/immaturity of existing infrastructure	• Policy-makers should prioritise the development of eHealth infrastructure at national level and the adoption of adapted initiatives;[277]
Semantic interoperability[278]	• The epSOS framework provides a template for the exchange of patient summary information; • Interlinking with widely-accepted knowledge structures could enhance data value: *"To this end, using Semantic Web technologies and the Linked Data paradigm could facilitate the data handling based on FAIR principles and provide a gateway to relevant data models, also enabling "intelligent" reasoning"*;[279]
Difficulties related to the integration of reference implementation frameworks or libraries[280]	• *"OpenNCP framework architecture patterns in the KONFIDO pilot could perhaps be revisited to facilitate its integration in other IT systems or its extension in order to fit in a wide range of setups"*;

Cultural resistance and lack of IT competencies	• The main barriers and enabling factor for technological uptake;[281]
	• Addressing this barrier requires:
	  • Improved retention of skilled IT professionals in healthcare;
	  • Basic IT skills training for the entire healthcare workforce and the recruitment of professional information and IT specialists to identify and implement local solutions;
	  • Support for clinicians to help them develop specialist IT skills and undertake leadership roles, working closely with CIO/CMIO/CCIO/CNIO partners;
	• Development of a new operational culture in healthcare institutions that promotes data sharing within and between departments;[282]

[276] Office of the National Coordinator for Health Information technology, 2020, *Connecting health and care for the nation, a shared nationwide interoperability roadmap*, https://www.healthit.gov/sites/default/files/nationwide-interoperability-roadmap-draft-version-1.0.pdf
[277] Natsiavas Pantelis et all, 18/06/2021, *Developing an infrastructure for secure patient summary exchange in the EU context: lessons learned from the KONFIDO project*, *https://journals.sagepub.com/doi/full/10.1177/14604582211021459*
[278] Ibidem
[279] Ibidem
[280] Ibidem
[281] Workshop on the healthcare industry.
[282] From workshop on healthcare

Proprietary vendor or healthcare system interests[283]	• **Core technical standards and functions:** Includes consistent data formats and semantics, standard secure services, consistent secure transport techniques, accurate identity matching, and reliable resource location; • **Certification to support adoption** and optimisation of digital health products and services: Includes assurance to stakeholders that digital health is interoperable; • Development and adoption of digital data standards and functions that enable information sharing as part of a coherent interoperability strategy. Developed solutions should be based on standards and widely-accepted practices;[284] • See suggestions for enhanced cross-platform interoperability and the development of a nationwide her;
Fragmentation of the healthcare system[285]	• Development of an approach that supports localities and enables them to plan their own route to a paperless health and care system; • Building effective leadership and capabilities to drive digital across the health and care system;
Data quality[286]	• Development of clear data cleaning rules for incomplete record data entry; • Alignment of system structures address different pathways and scenarios, instead of developing a "one data structure fits all" approach.

Impact of the digital health record system

Interoperable EHR is the priority for the healthcare sector. Based on the digitisation of basic information, EHR is key to enabling new advanced analytics (a large data pool is needed to develop new functions), data sharing and exchange (interoperability enables exchangeability):

- **Life-saving impact** such as immediate access to urgent information for unconscious patients. Digitalised solutions provide more opportunities for data screening and error identification. As is the case in the USA, the Health Information Management Systems Society's (HIMSS) recommendation to Congress states that: "*Patient data mismatches remain a significant and growing problem. According to industry estimates, between 8 and 14 percent of medical records include erroneous information tied to an incorrect patient identity*. ***The result is increased costs estimated at hundreds of millions of dollars per year to correct information. These errors can result in serious risks to patient safety***";[287]
- Health spending amounts to 10% of GDP across the EU as a whole.[288] According to a PwC report (2015), the digital health records system could save around €88 billion in health costs;[289]
- Digital health record systems could enable 11.2 million people with chronic conditions and 6.9 million people at risk of developing chronic conditions to extend their professional lives and improve productivity. That would add a further €93 billion to EU GDP;[290]

[283] Based on NHS England, 01/02/2016, *Joint report on international success factors for adoption and use of digital health in the US and NHS England*, https://www.healthit.gov/sites/default/files/adoptionreport_-branded_final4.pdf
[284] Natsiavas Pantelis et all, 18/06/2021, *Developing an infrastructure for secure patient summary exchange in the EU context: lessons learned from the KONFIDO project*, https://journals.sagepub.com/doi/full/10.1177/14604582211021459
[285] Based on NHS England, 01/02/2016, *Joint report on international success factors for adoption and use of digital health in the US and NHS England*, https://www.healthit.gov/sites/default/files/adoptionreport_-branded_final4.pdf
[286] Based on NHS England, 01/02/2016, *Joint report on international success factors for adoption and use of digital health in the US and NHS England*, https://www.healthit.gov/sites/default/files/adoptionreport_-branded_final4.pdf
[287] Prognocis, 26/08/2016, *Is HER interoperability possible with a National Patient Identifier¸https://prognocis.com/is-ehr-interoperability-possible-with-a-national-patient-identifier/*
[288] Eurostat, Healthcare expenditure across the EU: 10% of GDP, https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20190904-1
[289] PwC, 2015, Healthcare: A digital divide? Insights from PwC's 2015 Global Digital IQ Survey. Available online: https://www.pwc.com/gx/en/healthcare/pdf/healtcare-digital-divide.pdf
[290] Ibidem

- Connected and compatible Digital Health Record systems would accelerate and enhance European response capabilities to future health crises and pandemics, such as COVID-19;
- Inoperable EHR systems offer a series of key benefits such as: greater accessibility; new possibilities for data aggregation; access to a greater knowledge base; integration to Clinical-Decision support systems (CDSS); better communication with patients; and improved communication between providers.[291]

Increase data quality for advanced data analytics and artificial intelligence

Data quality and data value exchange: summary and key opportunities

A sufficient level of data quality and Information Quality (IQ) is the mandatory prerequisite for the development and uptake of advanced analytics and Artificial Intelligence. Currently, many companies fail to leverage big data due to a lack of good quality data. In addition, data analytics projects tend to dedicate most of their effort on developing and improving datasets, rather than data analysis.

Achieving higher data quality output is not a simple technical task. It requires the adoption of a quality assessment framework and the implementation of adjusted corporate data governance and data management processes. The development of common guidelines for Data Quality Assessment Frameworks and certification could help to increase industrial data quality across all sectors.

Two industrial sectors should be prioritised for the development and implementation of new data governance/management processes:

- **Data quality for smart grid, advanced energy analytics and AI;**
- **Upscale of AI applications for “manufacturing — other”.**

Although **smart grids** remain at an early stage of development worldwide, the question of data quality already appears to be a priority for utility companies. The need for “smarter” power grids that are able to manage bi-directional energy flows and higher connectivity is an indication that data is expected to grow exponentially in the near-term. Tackling data quality challenges at an early stage would help to provide solutions to challenges that more digitalised sectors are already struggling to overcome. Examples include challenges related to the management of increasing volumes of data and the existence of conflicting standard in the value chain. The energy sector should already start to develop an adapted Data Governance structure and Data Quality Assessment Framework. These tools would enable the industry to develop and implement new IoT solutions more efficiently and would help to prevent the emergence of issues that are more difficult to address at a later stage.

Most of the industrial sectors analysed in this report are manufacturing sectors. A number of segments are also grouped together under the “**manufacturing — other”** label. Many of the recommendations proposed in this document (e.g., interoperability for mechanical engineering or data quality framework) also hold for these sectors. Due to the diversity of sectors grouped under this label (wood machineries, pulp and paper, etc), it was not possible to develop a specific set of recommendations. Transversal guidelines on how to foster the development of Artificial Intelligence across all manufacturing sectors are therefore provided. The development of a data governance structure and data quality framework are a necessary first step for the uptake of AI solutions. However, the specific characteristics of AI-driven solutions require two additional adjustments: the development of specific metrics for dataset bias and the creation of a data office position in corporate data governance.

[291] World Health Organization, 01/06/2017, *Handbook for electronic Health Records Implementation*, https://www3.paho.org/ict4health/index.php?option=com_docman&view=download&slug=drafthandbookehr&Itemid=320&lang=es

Data is gradually becoming an integral part of many business operations and some of today's most successful companies use data analysis as a key component in practically all of their processes. Substantial benefits can be achieved through a greater uptake of advanced data analytics and innovative applications, such as artificial intelligence for smart grid and machine learning for the identification of micro-flaws in production outputs. More generally, advanced analytics can generate 4 main types of benefits:

1. More informed decision-making and improved risk mitigation;
2. Easier and more accurate audience targeting;
3. Increased profitability through improved operational efficiency and sales strategy;
4. Improved customer satisfaction through greater product and service customisation.[292]

The creation of data models and certification systems would help to develop industrial data quality. According to IBM, greater data quality could generate additional gains of approximately €2.6 trillion per year for the US economy.[293]

Example of opportunities linked to greater data quality: the market potential of AI

The global AI market is expected to grow at a CAGR (Compound Annual Growth Rate) of 30% between 2019 and 2026, generating revenues from € 24.7 million in 2018 to €202.4 million in 2026.[294]

However, the potential of the technology is still underexploited, as demonstrated by a study from Accenture, which showed that greater market uptake of AI could double the annual economic growth rates of countries such as the US, Japan, Germany, Italy, Spain and the UK.[295]

To achieve data quality in different industries, data models, certification and training are necessary. The EC institutional framework could be leveraged to support and enhance industrial efforts.

Building on existing European bodies to improve and ensure data quality

- The development of certifications schemes, data models and guidelines to improve data quality should build on the work performed by existing initiatives, such as the European AI Alliance and the Task Force for smart grid;
- Guidelines should include:
  - Data Integrity and Data Entry Standards to add data into a business system;
  - Anomaly detection solutions to flag "bad" data and identify any anomalies that deviate from a dataset's normal behaviour, which can affect data quality;
  - Data Quality (DQ) Assessment models to assess the quality of datasets;
  - Time series datasets to track business metrics over time;
- Training should be provided to equip data experts with the skills they need to implement data models and assess data quality.

Note 1: These guidelines should apply to anyone involved in data collection and entry, and to those responsible for managing these types of information assets. Guidelines should also consider data collected

[292] LOTAME, 2019, Why Is Data Quality Important?,, https://www.lotame.com/why-is-data-quality-important/#:~:text=Improved%20data%20quality%20leads%20to,in%20consistent%20improvements%20in%20results
[293] IBM, 2019, Extracting business value from the 4 V's of big data, https://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data
[294] GMI Research, 2018, Artificial Intelligence (AI) Market, https://www.gmiresearch.com/report/global-artificial-intelligence-ai-market/
[295] European Parliament, 2020, Opportunities of Artificial Intelligence, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652713/IPOL_STU(2020)652713_EN.pdf

from partners and external data sources.

Note 2: The purpose of these guidelines is to establish a set of requirements and standards to ensure that the data collected, stored and processed is of adequate quality for its intended present and future use. Good data quality will be achieved when data requirements, in terms of accuracy, validity, reliability, timeliness, relevance and completeness, are met.

Data quality for smart grid, advanced energy analytics and AI

The **smart grid** has become the most promising field to streamline energy flows and meet new challenges: decreasing energy loss for environmental purposes; creating "two-way" grids to connect renewable energies to the grid; etc.[296] Although smart grid is not a new concept, a truly "smart grid" has yet to become operational anywhere in the world.[297]

In spite of this challenge, Data Quality for smart grid is already an important question. The development of common data models at an early stage could help to avoid the future data standards landscape becoming fragmented. Although it may seem to be counterintuitive, the **growth of data quality requirements for energy grids should be a current priority.** The use of smart devices, IoT and connected systems will result in a continuous exponential increase in the volume of data in energy grids. Data volumes could be multiplied by x1000, requiring utility companies to invest in costly new data management solutions. The challenge is that, to be effective and useful, these newly created data sets need to be timeless and 100% accurate.[298]

The structure of the energy market indicates that the surge in data production will be accompanied by growing data quality challenges that may **lead to the emergence of new challenges rather than solutions to current challenges**. This is at least partly caused by the fact that utilities want to maintain a single source of reliable information, which often leads to data records being siloed in different departments:[299] "*Electric utilities are large organisations with many discrete sub-organisations, with each managing various programmes, processes and systems. Typically, these organisations work separately, in silos, often duplicating and not sharing data.* ***As a result, harmonising data systems and processes with increasing volume of data aggravates problems associated with a lack of data governance****. Today's mandate is, therefore, to evolve a governance model that ensures process, system and data alignment to meet modern grid demands*. (…) *Data governance enables the utility to aggregate data across multiple processes and systems, and requires blending accountability, agreed service levels and measurement. Adopting a strong governance model will improve the approach to data life cycle*."[300]

In other words, data quality issues cannot be addressed by looking solely at technical requirements. A better approach for the improvement of **data quality in energy would be to address the challenge from a Data Governance perspective**.

In addition to their siloed approach, utility companies tend to resort **to large data recoding** (even when unnecessary) and often prefer **"pen & paper".** For example, information about facilities is either mapped electronically or on paper. To verify that information, utilities are then required to organise a "field survey", in which employees physically check all facilities and ensure the accuracy of the recorded information. The need

[296] Frontiers in Big Data, 13/05/2021, *Data Consistency for Data-Driven Smart Energy Assessment*, https://www.frontiersin.org/articles/10.3389/fdata.2021.683682/full

[297] *"No fully functional smart grid of any meaningful scale exists yet. At this point, the easiest quantifiable metric for measuring SG development progress is the number of smart meters installed, because these devices are typically seen as the prerequisite for making many smart grid promises come true. Using that basis, it seems that Italy and Sweden are essentially tied."* Power, 01/01/2020, *Which country's smart grid is the smartest,* https://www.powermag.com/which-countrys-grid-is-the-smartest/

[298] Power Grid International, 01/01/2010, *Maintaining Intelligent Networks — Smartgrid & Data Quality*, https://www.power-grid.com/news/maintaining-intelligent-networks-smartgrid-data-quality/#gref

[299] Kotagiri Sunil, 06/03/2019, *Data Quality and Governance Critical for Utilities, https://www.tdworld.com/smart-utility/data-analytics/article/20972300/data-quality-and-governance-critical-for-utilities*

[300] Ibidem

for a trained power linesman — in short supply in all utilities — to manually inspect all facilities makes it a prohibitively expensive procedure. Switching to a data-model using digital imagery technologies could dramatically reduce the cost of these types of procedures, as new inspections could be done from the office. Furthermore, paper-based recording is not always reliable. For example, a “tracklog” is necessary to check the difference between the “as planned” and “as constructed” characteristics of energy facilities. In addition, difficulties in recording information clearly and consistently can lead to a loss of information across departments, causing different departments to complain about data unreliability.[301]

Although they offer a clear solution for industry**, data management techniques are currently fragmented with no common approach to their use**: *“In smart grids, stakeholders currently work with data management techniques that are unique and customised to their own goals, thereby posing challenges for grid-wide integration and deployment*.”[302]

To achieve higher quality, a Data Quality Assessment Framework and a new data-driven approach need to be defined. Achieving this require to develop a new **Data Governance architecture for data exchange between utilities (macro-level)**[303] as well as the development of internal **data management procedures (micro-level).** To extract value from data, utility companies must adopt genuine data-driven procedures and new data-oriented paradigms. For example, utilities still prefer to use a model-based and an as-designed-model rather than an as-operated model (data and models based on how machines and systems are operated rather than how they were designed).[304]

Data quality must also be improved to ensure consistency between data and operations. That requires an Intelligent Data Management Solution to align utility processes and system data: “***Finding the right model and system to align this data is the first step to obtaining high quality, actionable data and improving modern grid service quality.”*** [305]

Key principles for Data Governance and Data Management in utilities

- Data quality cannot be achieved solely through the use of new technologies. It also requires change at an organisational and **Data Management/Governance** level: *“So, in addition to energy management, smart grids require great data management to be able to deal with high velocity, important storage capacity and advanced data analytics requirements. (…) So, the main goal of utilities now is the ability to manage high volume data and to use advanced analytics to transform data collected [in]to information, knowledge and actionable plans.”*
- Key guidelines to achieve better data governance include:
    - Utilities should not focus on information flows within single departments but rather look at the **entire Work Flow Process.** For example, the sharing of design information from planners in a common ecosystem would remove pen & paper processes, and would help to ensure tracked information is more reliable and timeless;[306]

[301] Power Grid International, 01/01/2010, *Maintaining Intelligent Networks — Smartgrid & Data Quality*, https://www.power-grid.com/news/maintaining-intelligent-networks-smartgrid-data-quality/#gref
[302] Sundarajan Aditya et all, 24/06/2020, *Adapting big data standards, maturity models to smart grid distributed generation: critical review*, https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/iet-stg.2019.0298
[303] Kotagiri Sunil, 06/03/2019, *Data Quality and Governance Critical for Utilities, https://www.tdworld.com/smart-utility/data-analytics/article/20972300/data-quality-and-governance-critical-for-utilities*
[304] Model-based computational approach “are developed by using physical properties and parameters of the modelled system. (…) in recent years many advances occurred in techniques for signal processing and data analytic (…) re making the data-driven approach more viable and effective, especially because of the independence of the possible approximation of the system model and uncertainty of the parameters used”. Chicco Gianfranco, 13/05/2021, *Data Consistency for data-driven Smart Energy Assessment, https://www.frontiersin.org/articles/10.3389/fdata.2021.683682/full*
[305] Kotagiri Sunil, 06/03/2019, *Data Quality and Governance Critical for Utilities, https://www.tdworld.com/smart-utility/data-analytics/article/20972300/data-quality-and-governance-critical-for-utilities*
[306] Power Grid International, 01/01/2010, *Maintaining Intelligent Networks — Smartgrid & Data Quality*, https://www.power-grid.com/news/maintaining-intelligent-networks-smartgrid-data-quality/#gref

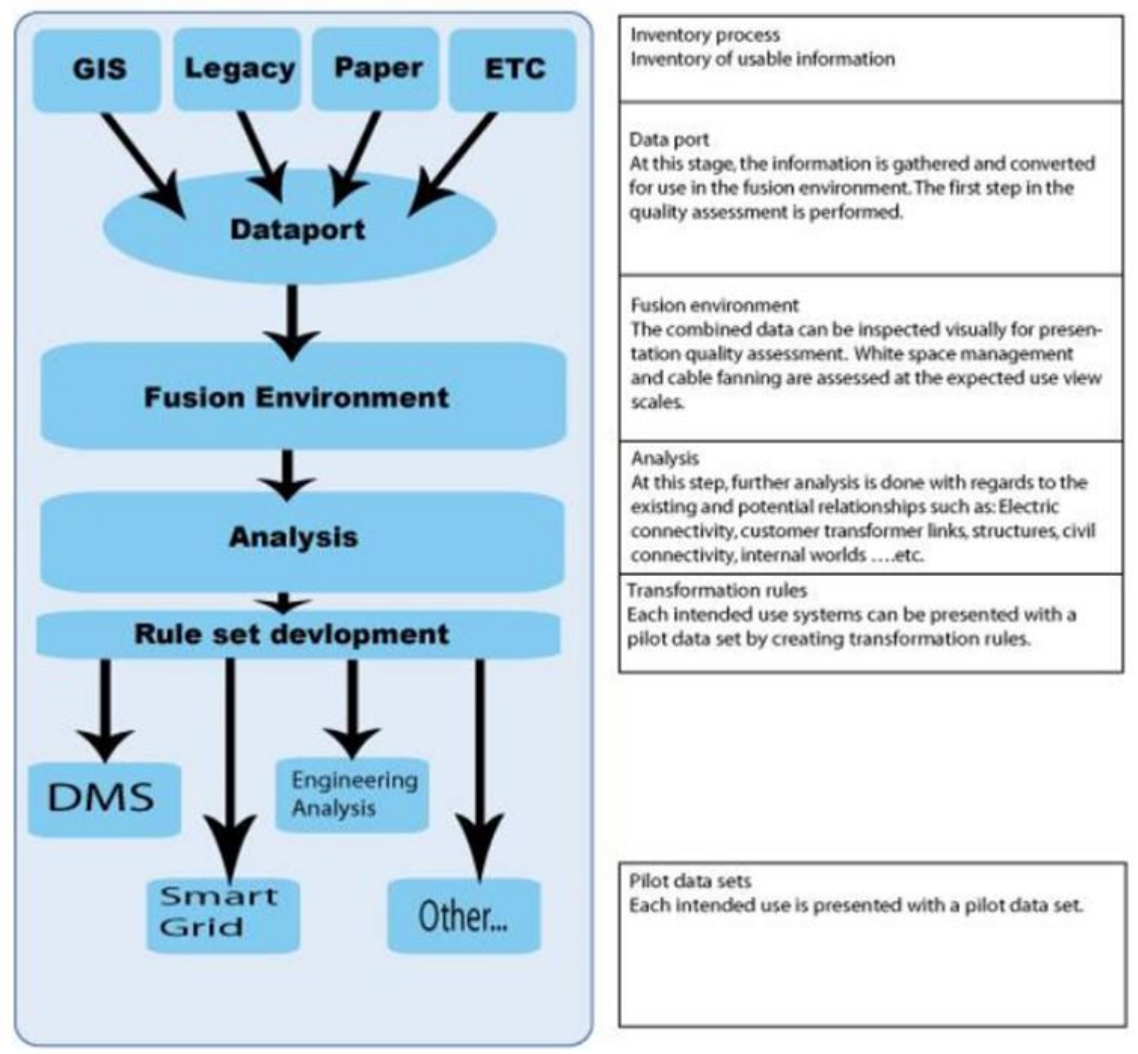
- Development of a **Single Point of Truth** and Data Stewardship. A single point of access and a single point of contact (data steward) for data types would reduce data silos and improve data accessibility across the entire organisation;[307]
- The **field workforce** should be provided with modern technologies to improve connectivity to central offices. Field workers are the only ones that are really familiar with physical infrastructure. By "*enabling field-workers to participate in maintaining records data, data quality and timeliness is improved and field-workers are more productive";*[308]
- **Power plants databases** should be **standardised** and **merged** (interconnected);[309]

- **Data governance should be more inclusive** and should include different types of energy suppliers/utilities. Gas, electricity and other providers should agree on a model for data exchange and information sharing supported by the development of new platforms:[310]
  - **Future platforms** should involve a range of stakeholders (e.g. electrical, gas and heating sources) to provide information that goes beyond consumption data;
  - **Open data platforms** can help to collect, clean and organise useful data. [311] Similar to other industries, there are three broad types of open data platforms that can co-exist with each other: online Open data platforms for altruistic data sharing; corporate Open data platforms for internal data sharing; and consortium Open platforms for data sharing between selected partners. The use of completely open data platforms with full transparency is an unrealistic objective;
  - **A more open data ecosystem** can also lead to greater market stakeholder involvement and an improved cost-benefit to ecosystem participation;
- The principle of **parsimony for data recording** should be observed to ensure that data quality is prioritised over quantity/volume: "*There have been cases where network operators amassed volumes of data that were never actually used — such attempts only result in wastage of time and resources. Therefore, data must always be gathered for a specific purpose and not just for the sake of record keeping* ";[312]
- Utilities tend to rely on their own platforms, architectures and single data types (smart meters and Phasor Measurement Units (PMU)). These elements therefore tend not to be integrated, leading to data redundancy, noise and silos. Utilities should consider the **creation of data hubs** and the improvement of data management processes. These hubs would enable to integrate their data in data fusion ecosystem and improve resiliency of critical infrastructure.[313]

---

[307] Ibidem
[308] Ibidem
[309] Gotzens Fabian, 2019, *Performing energy modelling exercises in a transparent way – The issue of data quality in power plant databases*, https://www.sciencedirect.com/science/article/pii/S2211467X18301056
[310] Workshop on Energy.
[311] An example of this is the *Open Energy Data Portal* is an important initiative that centralises a lot of important information on data usage, production and energy assets.
[312] Kotagiri Sunil, 06/03/2019, *Data Quality and Governance Critical for Utilities,* https://www.tdworld.com/smart-utility/data-analytics/article/20972300/data-quality-and-governance-critical-for-utilities
[313] Department of Energy Pages, 07/04/2019, *Big Data Analytics in Smart Grids*, https://www.osti.gov/pages/biblio/1639296

Figure 19: Example of a data management process with development of a "Data Hub" for Smart grids



Source: JCMB Technology position paper[314]

A new **Data Governance model** is a necessary first step to enable the implementation of truly "Smart" grids and to enable energy organisations to address the challenges associated with the Data economy. The definition of a **common Data Quality Assessment Framework** should be tackled simultaneously as these frameworks are an integral part of data governance and data management processes.

[314] JCMB, sin dato, *Data Quality Assessment Process*, http://www.jcmb.com/jcmb2/images/PDF/smart_data_smart_grid_whitepaper.pdf

Key principles for a Data Quality Assessment Framework in Energy

- The development of **data quality models** should be done in close cooperation with ongoing efforts to develop **interoperable**[315] devices and systems for Smart Grid. The development of **Common Data Models** and interoperability is a requirement for smart grid uptake and should be connected to data quality concerns, as these new models need to translate key principles for data quality into concrete action points:[316]
    - To define data models, a **common data classification** must first be developed to clarify how the Smart Grid should operate;[317]
    - CIM (Common Information Model) and SAREF are promising examples of interoperability models in the energy sector.[318]Currently gaining traction, CIM provides a common vocabulary and basic ontology to tackle problems in **energy systems that are still lacking a consistent model to represent electricity networks and underlying assets, the fragmentation of data formats, and the lack of data quality and uniformity**;[319]
- Three recommendations for data quality and the development of data quality models:
    - The **Smart Grid Taskforce**[320] should be upgraded to **extend its focus to the question of data quality**.
    - The **INTERRFACE project** is a €20.9 million EU project[321] for the alignment of national and vertical data models for utilities. The next step for this project would be to disseminate its results throughout the utilities sector;[322]
    - The **Smart Data Models** initiative (driven by partners such as FIWARE) provides higher quality data models, including for utilities;[323]
- Increasing data quality in energy typically involves screening datasets for three types of problems: incomplete data, outlier data and noisy data — although outlier data is the main type of challenge that utilities face.[324] [325] To assess data quality, datasets need to be measured to ascertain whether their requirements/formats are "fit for purpose". Metrics therefore need to be defined based on a number of key parameters;[326]

[315] Interoperability is an important part of the data ecosystem, space however does not allow to dig-further into this. Standard information models for smart grid interoperability already exist (e.g. IEC 61850, IEC 61850-90-7, IEC 61970/61968, IEEE 1815, and IEEE 2030.5). However, "(…) *there is no standard information models to describe interoperability among various big data analytics platforms, architecture, and their operational integrations with utility decision frameworks. Furthermore, storage, usage, dissemination, and sharing of data with utility operational frameworks are not unified. Interoperability between various cloud computing service vendors is necessary. Therefore,* ***extensive R&D is needed to develop interoperability*** *among different devices, network operations, data analytics platforms, big data architecture, data repository, and information models."* Department of Energy Pages, 07/04/2019, *Big Data Analytics in Smart Grids*, https://www.osti.gov/pages/biblio/1639296 . Furthermore, there is no unique reference architectures for the energy sector for digital platforms. However, projects like FIWARE and SEF model exist that could connect electrical facilities with the use of AI.
[316] Results from workshop on Energy.
[317] Results from workshop on Energy.
[318] Workshop on energy.
[319] Catapults, 2019, *Energy Data Taskforce Appendix 6*, *https://es.catapult.org.uk/wp-content/uploads/2019/06/EDTF-Report-Appendix-6-Standards.pdf*
[320] The "**Smart Grid Taskforce"** was created in 2009 to develop recommendations on interoperability and enable the development of smart grids in the EC: The cornerstone of the taskforce is to develop interoperability and system harmonisation through the development of recommendations and documentation. Smart Grid Task Force, https://ec.europa.eu/energy/topics/markets-and-consumers/smart-grids-and-meters/smart-grids-task-force_en?redir=1 ; European Smart Grids Task Force, 2019, *Towards interoperability within the EU for electricity and gas data access & Exchange*, *https://ec.europa.eu/energy/sites/ener/files/documents/eg1_main_report_interop_data_access.pdf*
[321] INTERFFACE, sin dato, *The Project*, http://www.interrface.eu/
[322] Workshop on energy.
[323] SmartDataModels, Website, https://smartdatamodels.org/
[324] Das Sarasij, Radhakrishan Asha, 01/12/2018, *Quality Assessment of smart grid data*, *https://www.iitk.ac.in/npsc/Papers/NPSC2018/1570474605.pdf*
[325] Chen Wen (et al), 2017, *Data quality of electricity consumption data in a smart grid environment*, https://www.sciencedirect.com/science/article/abs/pii/S1364032116307109
[326] Teaoman Duran Timocin, 15/08/2020, *Data quality in the interface of industrial manufacturing and machine learning*, https://www.diva-portal.org/smash/get/diva2:1469008/FULLTEXT01.pdf

- The European Open Science Cloud launched a taskforce to define and implement FAIR metrics for data quality and a common quality assessment process for multiple disciplines.[327] The methodology is focused on identifying key data quality standards and requirements and the development of a common framework. A **similar project for the energy industry would be useful**.

The development of metrics and a framework for Data Quality in utilities would need to be defined by the industry. However, Mouzhi et all[328] has already identified the key technical elements that should be taken into consideration for the definition of **an Industry Agreement for Data Quality Assessment in the Energy sector.**

Table 15: Main data quality problems for Smart Grid and suggestions for their removal: key elements to consider for a common Data Quality Assessment Framework

Data Quality Problem	Description	Potential Countermeasure
DQP1: Duplicate Data	Duplicate data from different smart meters.	Cross-linking of data from multiple devices and examination of repeating sequences.
DQP2: Missing or Incomplete Data	Lack of data (e.g. because of technical failures).	Linear interpolation and creation of daily load profiles for historical patterns recreation).
DQP3: Zero Records Semantics	Difficulties in interpreting zero-value range.	Reasonability test and comparison of data with other devices.
DQP4: Data Outliers (out-of-range)	Data busts or overly low/high values.	Reasonability test, anomaly detection algorithm.
DQP5: Measurement Errors	Datapoints caused by measurement errors.	Reasonability test and cross-device comparison of data, signal analysis for outlier detection.
DQP6: Non-trustworthy data-points	Intentional manipulation of data points.	Historical data for statistical-based detection.
DQP7: Data Anonymisation	Aggregation of attributes and features for the protection of privacy can make data analysis more difficult.	Hashing and signature for protection of data integrity.
DQP8: Timing Issues	Inaccurate data recording that complicates data integration and analysis.	Cross-device comparison of recorded data values.
DQP9: Biased data sets for artificial Intelligence[329]	Datasets can present bias (e.g. gender biased datasets).	Define metrics and assign a trusted AI data officer for data monitoring (see next section).
DQP10: Data quality requirements for Cybersecurity[330]	Quality of information shared about threat intelligence for higher cybersecurity.	• Definition of quality requirements for shared threat intelligence; • Improved dashboards to navigate existing information (including functions to customise information) on existing threats; • Traceability and provenance of the threat intelligence must be visible in the curated database; • Common data entry rules; • Automated integration of external sources;

[327] EOSC taskforce FAIR metrics and Data Quality Charter, https://www.eosc.eu/sites/default/files/tfcharters/eosca_tffairmetricsanddataquality_draftcharter_20210614.pdf
[328] Mouzhi Ge et all, 22/03/2019, *Data Quality Management Framework for Smart Grid Systems*, https://www.researchgate.net/publication/331951606_Data_Quality_Management_Framework_for_Smart_Grid_Systems
[329] DQP9 was added to the list by CARSA
[330] Added to the list by CARSA. Content of this row is mainly based on: Mussman Andrea et all, 24/10/2016, *Data quality challenges and future research directions in threat intelligence sharing practices,* http://library.usc.edu.ph/ACM/SIGSAC%202017/wiscs/p65.pdf

		• Metamodel for threat information that fits with stakeholder models. Intelligence can be grouped by type of vulnerabilities for easier data retrieval. Data correction should happen at the main source.

Source: With the exception of DQP9 and DQP10, the list is taken from from Mouzhi Ge et al.[331]

It is not possible to address all Smart Grid energy quality topics within the space allocated to this section. Effort was therefore focused on data governance and a data quality framework as they have been identified as the main enablers of greater quality and new analytics in utilities. Table 16 provides a short overview developed by "Bhattarai et all" [332] of the barriers obstructing the uptake of analytics in energy, as well as insights on how to remove them.

Table 16: Barriers and potential solutions for the uptake of Big Data Analytics in Smart Grid

Barriers	Solution
No standardised architectures or platforms for the deployment of big data analytics in smart grid. Most of the current big data platforms in utility industries rely on cloud computing.	• Deployment of efficient distributed platforms, such as Hadoop, Cassandra, and Hive is appropriate for big data analytics; • Development of holistic and modular energy big data analytics architectures, as well as corresponding computational platforms.
Advanced computational analytics. Owing to the huge volume of smart grid data, distributed and parallel intelligence is normally needed to effectively address data computation and handling challenges. Distributed computing and parallel intelligence are effective for addressing local grid issues and challenges, but they need to be coordinated to preserve global visibility.	• Development of effective distributed intelligence and coordination algorithms; • R&D on advanced approaches, such as metamodeling, dimensionality reduction, fog computing and edge computing, should be conducted to reduce computational and communication burdens.
Integration with advanced visualisation. Existing smart energy big data analytics schemes do not incorporate visualisation solutions.	• Integration of advanced visualisation with automated operation to provide directive information to operators and avoid the need for intuitive decision-making; • Co-design of smart grid big data analytics and advanced visualisation mechanisms.
Advances in algorithms. Smart grid data will continue to grow in volume, variety, and veracity. This reality will make it more difficult to adopt new ML and advanced AI techniques to exploit buried information and correlations between data sources and develop human-free decision-making processes.	• Scalability of ML models is critical; • Future R&D efforts on ML should focus on scalability, computational efficiency and accuracy to enable the uptake of new innovative techniques.
Value proposition to different stakeholders. A structured business model is necessary to fulfil financial goals and stakeholder requirements. The success of big data analytics in the utilities industry is contingent upon the active participation of electric utilities, customers, and system operators. Identifying revenue streams and the development of proper business models are critical to the success of big data deployment in smart grid. More importantly, the costs associated with the adoption of big data by all stakeholders should be justified. Those costs should also be broadly accepted by stakeholders, including policy makers,	• Future research should focus on techno-economic studies to quantify the technical and economic values of big data to electric utilities, system operators, and customers; • Workforce training will be required for data analysis interpretation and to better understand the capability and limitations of these tools; • Access to data will only provide a return on investment for utilities when professionals fully understand the capabilities and tools available,

[331] Mouzhi Ge et all, 22/03/2019, *Data Quality Management Framework for Smart Grid Systems*, https://www.researchgate.net/publication/331951606_Data_Quality_Management_Framework_for_Smart_Grid_Systems
[332] Bhattarai et all, 07/04/2019, *Big Data Analytics in Smart Grids*, https://www.osti.gov/pages/biblio/1639296

regulators, utilities and the consumer.	requiring changes in undergraduate and graduate curricula to include data science topics for future power engineers.

Source: Taken from Bhattarai et al.[333]

Impact of improvement data quality in a smart grid environment

- Early adoption of data quality models would provide solutions to challenges that would otherwise arise at a later stage (conflicting data standards, noisy data, data redundancy, etc);
- Adoption of artificial intelligence applications can help the energy sector and utilities to achieve GDP gains of roughly 12% by 2030, including a 5% increase in GDP through productivity gains, and a 7% increase through product enhancement gains;[334]
- Smart meters and machine learning can enable customers to reduce their monthly bills by €8-25;[335]
- Smart sensors and machine learning can optimise asset yields, increasing energy production by 20%;[336]
- 10-20% EBIT (Earnings Before Interests and Taxes) improvement by using machine learning to enhance predictive maintenance, automate fault prediction, and increase capital productivity.[337]

Upscale of AI application for manufacturing — others

Notwithstanding a few adjustments, many of the recommendations and key findings presented in the previous section (e.g. advice on the development of interoperability in lab robotics) can be applied to the “manufacturing – other” sub-sector grouping. The data quality development framework presented in the previous section was also tailored to fit requirements related to Artificial Intelligence (see DQP9).[338] However, Artificial Intelligence (AI) raises specific concerns frequently referred to under the “Trusted AI” label. Datasets that feed artificial intelligence can be biased, which could result in AI systems taking discriminatory decisions or even causing physical harm.

Removing Bias in datasets for Artificial Intelligence

- Companies that are willing to implement AI solutions in their industrial systems should nominate a **Data Officer for Trusted AIs:** [339]
    - The function can be merged with other roles in the data governance (e.g. sustainability) model, especially in SMEs;
    - The role of the office is to ensure and monitor data quality requirements, especially requirements that aim to avoid bias, ensuring non-discrimination and risk-assessment;
- Definition of **“Fairness Metrics”** to monitor decision-making algorithms (e.g. neutral demographic

[333] Ibidem
[334] PwC, 2018, *An introduction to implementing AI in manufacturing*, https://www.pwc.com/gx/en/industrial-manufacturing/pdf/intro-implementing-ai-manufacturing.pdf
[335] European Commission ,2017, *Harnessing the economic benefits of Artificial Intelligence*, https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/DTM_Harnessing%20the%20economic%20benefits%20v3.pdf
[336] Ibidem
[337] Ibidem
[338] Data quality is crucial for the development of AI in manufacturing. However, nearly 80% of generated data is erroneous or incomplete, which is especially problematic as AIs will only be as good as the dataset used. The problem of data quality is by far the most important initial barrier to the development of advanced AI and new applications for analytics, including machine learning. It is, for example, estimated that between 80% and 90% of data analysis projects are spent making the data reliable enough so that the results can be trusted. Decide, 2019, *The data boom and the possibilities it offers, https://decidesoluciones.es/en/data-quality-or-data-quantity-which-is-most-important-for-ai/* and K Q Wang, 2014, *Analysis of Data Quality and Information Quality problems in digital manufacturing*, https://www.researchgate.net/publication/251860789_Analysis_of_Data_Quality_and_Information_Quality_Problems_in_Digital_Manufacturing
[339] Interview with an expert in the provision of training for AI services

metrics):
- Increase awareness about **the dangers associated with this type of method. For example,** efforts to develop "blinding algorithm" techniques tend to backfire and create more biased data;[340]
- Metrics should be developed using a human-centric approach, with monitoring performed by the Data Officer for Trusted AIs. The final decision-making should not be automated and given to algorithms but to the Data Officer.

It should not be forgotten that AI, and especially Machine Learning (ML), techniques can also be used to **improve datasets**. Advanced AI techniques such as frugal AI and ML can reduce the amount of data used and they can also work on unstructured datasets to label, curate and exploit data. AI can extract information from unstructured data (optimisation, forecasts, plans). Datasets produced by AI systems must also be shared with digital value chain stakeholders (e.g. suppliers) to justify decision-making, such as reduced production planning. How AI datasets will be produced is not yet fully clarified, as this is a brand-new domain; however, it is important for the uptake of AI technologies.[341]

Legal Testbed as best practice

Legal Testbed[342] is a German project that was launched in 2019 and is funded by the Federal Ministry of Economics. It aims to develop a publicly accessible digital testbed for autonomous and automated business processes, e.g., for AI-trained software agents. Smart contracts are technical already possible, but they also raise complex legal and IT security issues that may act as a barrier to businesses, and particularly SMEs, limiting uptake of IoT technologies. The aim of the Legal Testbed project is to overcome these issues by ...

... offering **tools to test** the legal compliance of individually designed smart contracts;

... offering **concept and implementation support** to enable SMEs to create and use (automated) hybrid services;

... creating and testing a new **cooperation model for dynamic IoT value chains**;

... providing **actionable recommendations** on new legal standards for policy-makers and companies.

Legal Testbed was identified as a best practice because it is the first initiative to address the need for legal support to assist (small and medium-sized) enterprises with automated business processes. It helps companies to find legal solutions to the challenges that IoT technologies pose. Legal Testbed results and their market relevance are continuously checked by reputable partners, such as Platform Industry 4.0 and the International Data Spaces Association.

Standards for AI data quality are also enabling digital manufacturing and the improvement of data quality in manufacturing.[343] The European Commission has already begun to set the scene by defining ethical standards for AI.[344] The creation of the European AI alliance forum and the High-Level Expert Group on Artificial Intelligence (AI HLEG) are very positive and necessary first-steps for the creation of standards. However, these two initiatives should be extended to help these ethical standards to translate into operational requirements.

[340] Kleinberg Jon et all, 01/01/2018, *Advances in Big Data Research in Economics — Algorithmic fairness*, https://www.cs.cornell.edu/home/kleinber/aer18-fairness.pdf
[341] From Workshop organised on "Manufacturing — Others".
[342] Website of the initiative: https://legaltestbed.org/en/start/
[343] Diab Wael, 2019, *AI Standards help accelerate digitalization of smart manufacturing*, https://etech.iec.ch/issue/2019-06/ai-standards-help-accelerate-digitalization-of-smart-manufacturing
[344] European Commission, 2018, *Ethics guidelines for trustworthy AI,* https://ec.europa.eu/futurium/en/ai-alliance-consultation

These requirements should be linked to the question of data quality. Indeed, a good data set is also an ethical data set.

Upgrading the AI High-Level Expert Group and the European AI Alliance?

- Creation of a working group in the AI-HLEG to define/discuss fairness metrics that could be added to a Data Quality Assessment Model for manufacturing processes;
- Development of a reference framework for data-based AI quality that should define:
  - A common Data Quality Assessment Model for data fairness;
  - A common set of Data Quality Rules that companies can use to analyse a specific dataset and assess whether it meets the defined requirements;
  - A common standard for Quality Assurance Methods that addresses trusted-AI problems and "Bias identification and cleaning".

Note 1: In the development of Artificial Intelligence, "reliability and validity" are the two main conditions to consider in a data quality assessment. "Measurement" and "representation" error are part of the data source to be considered.

Note 2: The assessment model must consider information related to data collection and the methodology used for data capture, as well as data and metadata level descriptions.

Impact of AI on manufacturing

- Around 70% of companies would adopt at least one type of AI technology by 2030, whereas less than half of large companies would deploy the full range;
- AI in the manufacturing market is expected to be valued at €0.93 billion in 2020 and to reach €14 billion by 2026 with a CAGR growth of 57.2% during the forecasted period;
- The adoption of artificial intelligence applications can increase the contribution of the manufacturing sector, achieving GDP gains of about 13% by 2030: including a 5% increase in GDP through productivity gains and an 8% increase through product enhancement gains;[345]
- In the long-term, AI may lead to scientific breakthroughs that could even create entirely new and unforeseen industries;[346]
- In the supply chain, AI is expected to increase material lead times by up to 30% and ensure a 3-5% improvement in production yields;[347]
- 10% yield improvement for integrated-circuit products using AI to improve the R&D process;[348]
- 39% IT staff reduction by using AI to fully automate procurement processes;[349]
- Up to a 13% improvement in EBIT by using machine learning to predict service revenue sources and optimise sales efforts.[350]

[345] PwC, 2018, *An introduction to implementing AI in manufacturing*, https://www.pwc.com/gx/en/industrial-manufacturing/pdf/intro-implementing-ai-manufacturing.pdf
[346] OECD, 2018, *Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*, https://doi.org/10.1787/sti_in_outlook-2018-en
[347] European Commission ,2017, Harnessing the economic benefits of Artificial Intelligence, https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/DTM_Harnessing%20the%20economic%20benefits%20v3.pdf
[348] Ibidem
[349] Ibidem
[350] Ibidem

3 Building blocks for the development of a European industrial data space and new high-impact industry agreements

An industrial data space is a complex ecosystem that brings together various stakeholders. In spite of different motives and objectives, these players share a common interest in sharing information and exchanging data about related industrial activities.

However, industry actors can be reluctant or discouraged from entering the data economy due to the lack of infrastructure that is required to serve their needs and protect their interests. To exchange sensitive information, industry players require a series of guarantees to help develop trust, protect data sovereignty/ownership, ensure technical feasibility and maintain a level playing field for all in future data spaces.

The infrastructure necessary for the development of an industrial data space can be divided into two broad categories: soft and hard infrastructure.[351] Hard infrastructure entails all types of physical equipment and devices, including servers, cables, etc. In contrast, soft infrastructure includes a broad range of legal and operational agreements.

The complexity of the task at hand can, however, be reduced if summarised as a series of building-blocks that must be tackled simultaneously. These 9 blocks can be organised into three categories. Technical specifications cover 4 blocks: data standards; exchange protocols; identification and authentication; and authorisation. Legal and organisational includes 4 blocks: business models; governance; legal agreements; and operational agreements. Data specifications contains the final block: metadata.

All efforts to develop a streamlined and efficient data space have to respect 4 key principles:

1. Creating trust among partners;
2. Data sovereignty/ownership/reuse rules;
3. Ensuring that the data space offers a level playing field (to avoid disruption to competition and the emergence of monopolies/oligopolies);
4. Enabling a decentralised soft infrastructure (to ensure that the data space cannot be controlled by some providers).

The objective of the building blocks has been clarified. However, there is a need to explain how this framework can be translated into a methodology that can resolve technical challenges, incentivise business activities in areas with high potential for technology R&D, and unlock important market opportunities.

The complexity of modern industries and the differences between them require diversified policies to tackle similar problems in different manners across varied industries. Policy-making should therefore be driven by two key concerns: avoid recreating a level of complexity that is greater than exists in the current industrial landscape; make sure not to impose technical requirements on industrial processes.

Gaia-X and IDSA as best practices

The development and uptake of industry agreements by industry players in the digital value chain can be considered to be a relatively new field. Nonetheless, Gaia-X and IDSA already shine as best practices that can serve as an inspiration and model for subsequent efforts.

[351] OpenDEI
, 2020, *Design principles for dataspaces*, https://design-principles-for-data-spaces.org/

The **International Data Space Association (IDSA)** was founded as a non-profit organisation in Germany in 2016. It aims to develop a reference architecture for international data spaces, including a governance model and an adoption strategy. IDSA members represent many different industry sectors that are based in most European countries. IDSA results are publicly available and can be used by any company or organisation.

The IDSA reference architecture solves specific problems related to data exchange and places strong emphasis on data sovereignty — the terms and conditions are always defined by the data owner. A decentralised infrastructure enables secure and distributed network operations that are designed to achieve economies of scale and prevent individual network partners from exerting disproportionate influence on the network. The architecture is flexible and enables different source of data to be aggregated in a common data space. Data can for example sourced from industrial processes, platforms, connected devices, etc.[352]

IDSA is identified as a best practice because it has become one of the most important and internationally recognised associations. Other organisations or enterprises build on their results or directly cooperate with IDSA.

Launched as a Franco-German initiative in 2019, **Gaia-X** is quickly gaining momentum. While IDSA covers the "exchange protocol" part of the transfer of information, Gaia-X also covers the storage of information. Aiming to enhance and protect European data sovereignty, GAIA-X works by linking different initiatives through open interfaces and standards to make data available to broad audiences while enabling the development of new innovation platforms.

The key elements that make Gaia-X and IDSA best practices to follow are their ability to address very specific technical problems in a clear and efficient manner, while respecting general principles that are indispensable for any IA to become successful:

- Openness and transparency;
- Placing data sovereignty and ownership rights at the core of the initiative;
- Development of interoperability and connectivity, drawing on existing solutions rather than replacing them;
- Boosting trust and ensuring the emergence of a level playing field.

Table 17. Key requirements for a data space covered by Gaia-X and IDSA (non-exhaustive list)

Requirements for operation data space 1	Matching with existing software/hardware	
Key specifications for an industry data space	**IDSA**	**GAIA-X**
Interoperability	X	X
Data portability	X	
Security	X	X
Usage control	X	X
Authentication	X	
Privacy and confidentiality protection	X	X
Storage		X

[352] Fraunhofer, 2016, *Industrial Data space*, https://www.fraunhofer.de/content/dam/zv/en/fields-of-research/industrial-data-space/whitepaper-industrial-data-space-eng.pdf

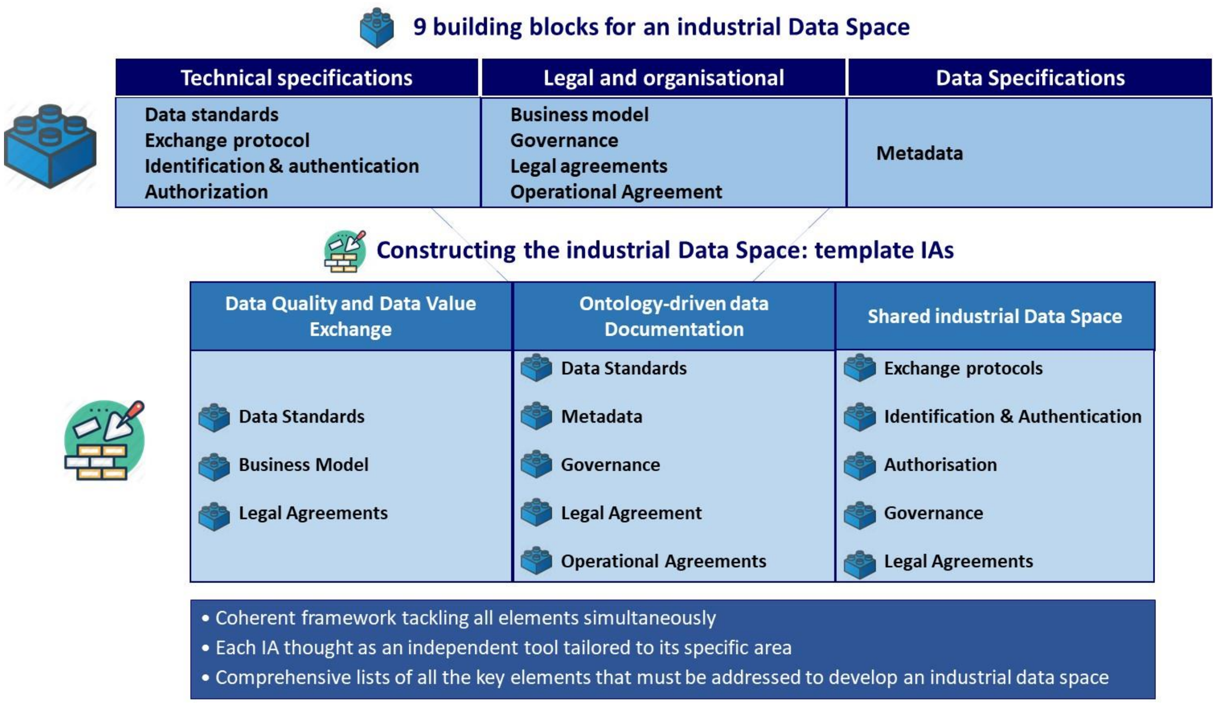
The framework developed in this report is based on a series of 3 industry agreements. Taken together, the 3 IAs listed below make it possible to simultaneously tackle the 9 building blocks and build a robust and streamlined industry data space:[353]

- Data Quality and Data Value Exchange;
- Common ontology-driven Data Documentation;
- Shared Data Spaces.

Figure 20 summarises how the framework is built, as well as its key strengths. The specifications and components for each IA are detailed in Annex I.

This framework enables industry players to identify and develop their own initiatives, thereby avoiding the need to impose concrete technical specifications on industrial processes. Furthermore, these IAs do not aim to add a new layer of complexity. Instead, they are industry-driven frameworks that draw on existing initiatives and previous standardisation efforts.

Figure 20: Industry agreements as a consistent framework for an industry data space



Source: CARSA, 2021. This figure was created using resources from Flaticon.com.

[353] The 9 building blocks were originally developed by Innopay, see for example, Innopay, 2020, *Data sovereignty and soft-infrastructure: key enablers of the European data economy*, https://www.innopay.com/en/media/616/download

4 Recommendations for industry and policy makers

The fragmentation of the current European data landscape affects both technical and data specifications, and legal and governance dimensions, and has important economic consequences. It results in business operation inefficiencies, limited economies of scale and missed business opportunities. There is a need, therefore, for greater cooperation among industry players in distributed value chains to establish common frameworks and rules for data sharing.

Industry agreements (IAs) that, for example, define core legal and technical elements and contractual clauses, are powerful tools. They provide industry actors with the key building blocks with which to establish rules for common data spaces, such as voluntary B2B data sharing schemes[354], and to stimulate innovation. However, this study reveals that many significant barriers are inhibiting IA development and implementation, the increased deployment of data sharing technologies and solutions, and the potential that enhanced industrial ecosystems have to unlock new market opportunities.

This concluding section presents four key cross-cutting recommendations to help stimulate greater cooperation in the IA development and implementation process and promote the development of common industry data spaces. For each of these recommendations, specific policy and industry actions are proposed, including an assessment of their potential impacts on industrial competitiveness.

- **Recommendation 1**: Create the conditions to enable stakeholders to build trustworthy relationships and support a trusted environment for data sharing within and across industries;
- **Recommendation 2**: Establish pathways towards the development of common approaches to data quality assessment and assurance, and improve data access and interoperability;
- **Recommendation 3**: Develop common understandings of data asset value and set up mechanisms to capture and fairly redistribute the benefits of data sharing;
- **Recommendation 4**: Take action to address technical challenges and improve the clarity of data sharing regulations.

Recommendation 1: Create the conditions to enable stakeholders to build trustworthy relationships and support a trusted environment for data sharing within and across industries

In addition to technical challenges, the study finds that a lack of trust between industry stakeholders is one of the main obstacles to industry agreement development and implementation. This lack of trust can take different forms:

- Doubts on how data will be used or reused when they are further aggregated (i.e., uncertainty about the scope of secondary usage);
- Uncertainty about the definition of roles and processes, business ethics, and the general principles that apply to data sharing agreements;
- The absence of a legal framework for data sharing and uncertainties about the practical implications of existing regulations for data sharing agreements, which often means that businesses are unsure if they are allowed to share data and under what conditions;
- Lack of trust in the integrity of systems used to collect, exchange and store data and information (i.e., distrust in the system), and which are required to underpin or support IA implementation;
- Issues relating to the protection of commercially sensitive and personal data, for example, where competitors are permitted to access the system or the possibility of illegal access (e.g., unintentional exposure or malicious reverse engineering, data breaches and data loss). For example, the probability

[354] During the course of the study, as explained in the introduction, the analysis has shed light on the prominence of industry agreements, particularly those that target the increased use, sharing and exchange of data, and the creation of industrial data ecosystems.

that large and diverse data pools will contain at least some form of personal or IP-protected data means that there is a possibility that combined data analytics could result in an unanticipated exposure of information, such as personal or commercially sensitive data belonging to participating companies (potentially in breach of art. 101 TFEU).

To build trust in data sharing systems, solutions to technical challenges (e.g., secure technical infrastructure to support secure communications between parties), legal conventions and operational guidelines are needed. To build trust between data owners and users, appropriate consent tools are required. Other importance aspects include access and rights for data (re)use, and the modalities (e.g., where data is stored, governance models) and time limits on use. This implies the need for an efficient data governance mechanism (e.g., a set of rules of a legislative, administrative and/or contractual nature that determine the rights to access, process, use and share data) that is capable of creating and maintaining trust between all participants in a data space.

Other factors, such as data quality, clarity of the regulatory framework, a demonstrated value proposition and transparent financial dynamics, will also influence trust among participants in a data space. Some of these aspects are addressed in more detail in subsequent recommendation areas. Furthermore, building trust is not an action that one single stakeholder can achieve in isolation; rather, it is the result of active trust building actions by various players that interplay over time.

Specific actions for industry to implement to build trust[355]

1. **Develop technical specifications/clauses** focusing on:
   a. Operational agreements (governance models, shared operational processes, certification etc.) to establish trust for data re-use and to clarify the scope of secondary usage. This includes common criteria for admission and certification, withdrawal, warnings, suspension, exclusion, incident management and change management;
   b. Technical agreements (e.g., common standards platforms, building reference architectures[356], interoperability frameworks, assessments and criteria) to assure security and interoperability;
   c. Conditions for granting third party access to interoperable solutions, permissible types of data re-use.

   Examples *of industry initiatives that could be considered best practices, are: i) iShare*[357]*, a framework of agreements that has been developed to overcome most of the barriers to data sharing; ii) Dawex Data Exchange Platform*[358]*, which supports technical solution for data access, data licensing formats to fit different needs, traceability by the data provider, trust between the data provider and the data user, multiple business models for data exchange, and compliance with rules and regulations; iii) LEGAL TESTBED*[359]*, which is a German publicly-funded project addressing trust and automated contracts.*

2. **Develop technical specifications/clauses to address the dynamic aspects of data** e.g., governance models or suitable data rights management frameworks to enforce data usage rights and manage dynamic IPs. There are constant changes to the data in the pool, e.g., when data is aggregated or a new dataset is created. As a consequence, data parameters, for example about ownership, can change and may not be anticipated by legal agreements upfront. It is therefore an absolute priority to establish common rules to track and trace how evolving IP rights are deployed in the common data space. Use case scenarios can be deployed to shed light on potential technical issues, highlighting them in advance and enabling solutions to be developed before issues arise.

[355] Ordered based on prioritisation emerging from the recommendations workshop.
[356] Which includes structures, building blocks and requirements.
[357] See Chapter 2.1 for further details
[358] See Chapter 2.2 for further details
[359] See Chapter 2.2. for further details

3. **Endorse available reference architectures as a general guide/roadmap** to organise elements, roles, functionalities and responsibilities (liabilities) to enforce data-related contractual clauses in IAs, e.g., through data security, data sharing rights, description of the relationships between IA parties, and time limits on the use of a particular dataset. Convergence among different initiatives is needed to highlight and raise awareness of the current possibilities.
4. **Develop sectoral and cross-sectoral codes of conduct** laying down general principles (e.g., data access, control, portability, privacy and security, liability and IP rights) for data sharing and to provide greater insight on contractual relations and guidance on the use of data. Setting out default principles and clarifying stakeholder roles can help to increase trust. Codes of conduct should be complemented by clear implementation rules and conditions to ensure maximum stakeholder endorsement. An existing industry example is the EU code of conduct on agricultural data sharing[360].
5. **Develop standardised trustworthiness profiles** that describe the capabilities and expectations of partners that must achieve a required level of trustworthiness. Initially, different certifications could be provided by the companies themselves and checked manually. In the longer-term, the process could be performed automatically, with additional rules in place that combine technical details with legal aspects and a standardisation of consent.
   An existing industry example *is the collaboration work between RRI, Japan and PI4.0, Germany*[361] *to create ad-hoc trustworthy relationships.*

Specific actions for policy-makers to support the creation of trust[362]

1. **Support the development/dissemination of guidelines and model agreements or standardised templates and clauses for B2B data sharing** e.g., through publicly accessible inventories of templates/model agreements, checklist and guidance documents, sharing of best practices, networking initiatives, etc.
2. **Provide legal clarification on the concept of 'data intermediary'** (as intended by the Data Governance Act) to enable these entities to act credibly as a trusted and neutral service provider for contractual commitments. Given the diversity of settings (e.g., data providers and users, the nature of the data, the purpose of its (re-)use, the technical environment, the sector, etc.), a one-size-fits-all business model should not be prescribed. Areas where clarification is needed include: the types of actors to be included in the EU framework; their functioning in practice and compatibility with existing legislation, notably the GDPR; and their compliance with ethical standards, as well as possible accountability and effective enforcement mechanisms.
3. **Support the development of innovative data governance structures and sector-specific collaboration models** such as data trusts[363], data cooperatives, and data commons, which can facilitate public and private data sharing in ways that preserve privacy (independent data stewardship and anonymisation of data) and contractual control over data.
4. **Develop and disseminate EU-wide Data Governance practices/guidelines** to clearly define rules of conduct for the fair use of exchanged data. This should be done after testing different approaches to assess the impact of data sovereignty and to explore means for Conformity Assessment (e.g., voluntary or licensed certification for companies, data sharing assurance agreements).
5. **Promote the further use and evolution of high-impact technology though research and**

[360] https://cema-agri.org/images/publications/brochures/EU_Code_of_conduct_on_agricultural_data_sharing_by_contractual_agreement_2020_ENGLISH.pdf
[361] Plattform Industrie 4.0 (PI4.0), Germany and Robot Revolution & Industrial IoT Initiative (RRI), Japan, 2020, *"IIoT Value Chain Security – The Role of Trustworthiness"*, https://www.plattform-i40.de/PI40/Redaktion/EN/Downloads/Publikation/IIoT_Value_Chain_Security.html
[362] Ordered based on prioritisation emerging from the recommendations workshop.
[363] https://theodi.org/article/what-is-a-data-trust/

innovation actions, e.g., privacy enhancing technologies – such as differential privacy, homomorphic encryption, and federated machine learning – to create opportunities for data sharing while preserving individual privacy.

6. **Promote data sharing facilities and environments for the collaborative testing of new business and innovation models to enforce data usage rights.** Common or shared testing spaces should enable safe experimentation with models under different dynamic scenarios, e.g., the network of European Digital Innovation Hubs (Big Data Innovation Hubs network) and the BDVA I-Spaces[364].
7. **Promote opportunities/options for cooperation in areas that are not critical to competitive advantage to showcase the benefits of data sharing**, particularly on technologies that are not core to a company's own products, on which companies do not need to compete, and for which standardised solutions could deliver shared benefits and competitive advantage. Promotion actions could include:
    a. In the case of more fragmented industries, EU and national policy-makers could help to identify cooperation areas, in compliance with competition law;
    b. Promotion of cross-sector standards on metadata by EU policy-makers, i.e., uniform descriptions of the types of information the metadata must contain and in what form.
8. **Provide clarification/guidance on how to implement data anonymisation** in ways that are legally accepted. To the extent that data still qualifies as personal (which includes pseudonymous data), many obligations exist that can sometimes be hard to meet and which can hamper innovation.

Actions that help to build trust are expected to significantly improve industrial competitiveness, as illustrated in the figure below. Our analysis reveals an expectation that standardised/model agreements will help to orientate data sharing by companies and provide greater confidence in negotiated contracts, thereby reducing the negotiation process and associated transaction costs. More broadly, an improved innovation ecosystem based on trustworthy relationships should lead to a wider adoption of digital technologies, thereby enabling the appearance of novel - both radical and disruptive - business models that are completely integrated into the system. This should help to generate new market value and ensure the financial security of future market players.

Clarifying essential notions of data intermediaries (e.g., type of actors, functioning, ethical standards and accountability) is expected to lead to regulatory harmonisation, enabling intermediaries to act credibly as a trusted service provider for contractual commitments. This is expected to reduce pre-contractual costs (e.g., search costs for appropriate data and partners and the cost of negotiating), post-contractual risks, and monitoring costs for participants, to help overcome data market failures. It should also enable data sharing transactions that would otherwise not materialise because of the perceived risks.

[364] https://www.bdva.eu/node/1172

Figure 21 Main impacts of proposed actions on industrial competitiveness

Digital competitiveness components	Company level				Value chain level			
Actions	Reduced transaction/ compliance costs	Increased productivity	Reduced labour costs	Improved innovation capability/ digital skills	Wider adoption/ investment in digital technologies	Improved innovation ecosystem	Regulatory framework harmonisation	Wider technology access/ awareness
Industry	High	Medium	Low	Medium	High	High	Low	Low
Policy	High	Medium	Low	Low	Low	Medium	High	Low

Source: Study/consortium partners – own analysis validated by industry experts; *Note: Assessment based on the top ranked actions according to stakeholders consulted: i) Industry action - Develop guidelines, model agreements and standardised clauses (e.g., technical, operational); ii) Policy action - Legal clarification on the concept of data intermediaries. Their impacts on a range of digital competitiveness components are assessed, with a distinction made between company and value chain levels. For more information, please see Annex III on the outcomes of the recommendations workshop.*

Recommendation 2: Establish pathways towards the development of common approaches to data quality assessment and assurance, and improve data access and interoperability

Data quality requirements can pose a significant challenge to the adoption of collaborative agreements across all of the sectors analysed. Shared data should adhere to established quality standards to ensure that its use will provide accurate results, meaningful insights and a reliable basis for decision-making. This is applicable to all data sharing arrangements and is especially true for artificial intelligence solutions, which rely on unbiased and standardised datasets to feed the algorithms. Due to the wide variety of purposes and types of data, actors and diverse scenarios, stakeholders often adhere to different principles and practices and lack common frameworks to measure and assess data quality, which is a barrier to collaborative agreements.

Accessibility to high-quality datasets is an ongoing challenge for most organisations in the selected sectors, as is access to interoperable solutions, due to the lack of commonly agreed standards. Standards are crucial, but it is also important to build capacities to develop platforms and functionalities that are easy to use and adapt to all scenarios.

Moreover, the provision of frameworks to support claims concerning data quality and the circumstances under which liability may be incurred for damages, due to inaccurate data, is a common challenge in all sectors. Lack of clarity on this topic has a significant effect on the willingness of industry actors to enter into industry agreements. The absence of a clear legal status for data (e.g., lack of qualification of data as a 'product') is perceived to be the main challenge. Liability is mainly regulated at national level, which constitutes another layer of uncertainty for industry actors entering into cross-border agreements.

Specific actions for industry to implement to improve data quality, access and interoperability[365]

1. **Develop contractual clauses to increase the quality and value of data, such as:**
   a. **Standards & methods for data quality assurance** (common mechanisms to collectively ensure the high quality and accuracy of data and its continuous consistency and maintenance). These clauses will help to build a common data space where companies can rigorously test and control the data quality of each dataset;
   b. **Mutual agreements on metadata models and common industry ontologies/semantics**, with common vocabularies to be respected for data sharing within and across DVCs, as a necessary step towards syntactic and semantic interoperability;

   ***Existing examples** of industry-driven international initiatives in this field are:*
   I. *The Industrial Ontologies Foundry*[366] *is working with OntoCommons*[367] *to develop common industry ontologies and semantics as an important prerequisite for data quality;*
   II. *"FAIRification"*[368] *concept developed by GO FAIR, a stakeholder-driven and self-governed initiative that aims to implement the FAIR data principles;*
2. **Include procurement requirements and the obligatory use of standards** (e.g., on ethical principles and how they should be safeguarded) to increase accountability for misuse/inaccurate data and vendor lock-in in contracts on interoperability;
3. **Identify a list of standards (e.g., ISO, OPC etc.) that need to be supported in data spaces** for each sector and level of interoperability (e.g., machine level, line level, factory level, supply chain level, data, etc.) to help the convergence of various initiatives and ongoing work in different fields.

Specific actions for policy-makers to support improvements in data quality, access and interoperability[369]

1. **Promote the development and use of standardised data licensing models/agreements** to facilitate new collaborative approaches, based on existing examples and best practices;
2. **Promote the definition of harmonised and clear criteria/guidelines to assess the quality of datasets**, in terms of accuracy, reliability and completeness**.** Guidelines on resolving data quality problems could include data integrity and data entry standards**;**
3. **Support the development of a data quality certification framework or data quality seals** to provide industry with a systematic approach and framework to assess data quality seals (e.g., according to FAIR principles). This framework should support the certification of organisations that meet data quality and dataset requirements, while preventing data quality assessments from being performed in silos;
4. **Support the development of European cloud capacities and data pooling across different sectors** (i.e., an interoperable EU industry data space focused on EU data sovereignty) and invest in the development of new standards, tools and infrastructure to process and store data;
5. Since data is generally considered of high quality if it is fit for the use case required/meant, **promote common or shared spaces for data testing, in which various types of data quality can be tested in diverse scenarios** (e.g., BDVA i-Spaces).
6. **Coordinate industry-driven efforts (e.g., through dedicated coordination bodies such as the**

[365] Ordered based on prioritisation emerging from the recommendations workshop.
[366] Industrial Ontologies Foundry (IOF), https://www.industrialontologies.org/
[367] Ontology-driven data documentation for Industry Commonshttps://ontocommons.eu/
[368] GO FAIR, *"FAIRification process"*, https://www.go-fair.org/fair-principles/fairification-process/
[369] Ordered based on prioritisation emerging from the recommendations workshop.

envisaged European Data Innovation Board) to identify a list of standards that need to be supported in the data spaces, and then support efforts to bring the relevant players together to define standards and drive convergence work towards the achievement of common standards.

7. **Set up independent subjects/bodies to serve as data quality test engines** that are tasked with independently verifying the quality and appropriate use of data, in line with principles and standards. They could also act as coordination bodies that bring relevant stakeholders together to test and implement solutions in different industry settings and disseminate results.

Improving data quality, access and interoperability is expected to positively affect industrial competitiveness, as illustrated in the figure below. Developing common standards and methods to increase the quality of data is expected to substantially reduce transaction costs. In particular, pre-contractual transaction costs could be lowered by proposing standardised clauses in data sharing contracts to facilitate negotiations. Time-consuming processes to clean data and make it usable can be reduced by the provision of commonly agreed methods/standards to measure and assess the quality of data. Adhering to established and shared quality standards is expected to ensure that the use of data will provide accurate results, meaningful insights and a reliable basis for decision-making, improving productivity for all parties involved.

Similarly, standardised data licensing models/agreements are expected to facilitate trustworthy access to data and reduce transaction costs for companies. Common/standardised contractual provisions (e.g., redistribution obligations) in data licensing agreements can be used to encourage sharing by limiting the liability of the data provider and ensuring that those downstream can identify the data source. This would encourage new collaborative approaches and improve the innovation ecosystem, in a similar way to open-source software. The identification of a common list of standards to be supported in data spaces should help to accelerate the adoption of digital technologies and encourage greater levels of investment. This would give greater confidence to companies and reduce the risk of potential lock-in and/or of 'rapid obsolescence' when committing to one type of technology and/or standard.

Figure 22 Main impacts of proposed actions on industrial competitiveness

	Digital competitiveness components							
	Company level				Value chain level			
Actions	Reduced transaction/ compliance costs	Increased productivity	Reduced labour costs	Improved innovation capability/ digital skills	Wider adoption/ investment in digital technologies	Improved innovation ecosystem	Regulatory framework harmonisation	Wider technology access/ awareness
Industry	High	Medium	Low	Medium	High	High	Medium	Low
Policy	High	High	Low	Low	Medium	High	Medium	Low

Source: Study/consortium partners – own analysis validated by industry experts; *Note: Assessment based on the top ranked actions according to the stakeholders consulted: i) Industry action - Develop contractual clauses to increase the quality and value of data (common standards and methods); ii) Policy action - Promote the development and use of standardised data licensing models/agreements. Their impacts on a range of digital competitiveness components are assessed, with a distinction made between company and value chain levels. For more information, please see Annex III on the outcomes of the recommendations workshop.*

Recommendation 3: Develop common understandings of data asset value and set up mechanisms to capture and fairly redistribute the benefits of data sharing

To give data providers a degree of control over 'their' data, the concept of "data sovereignty" has emerged. Data sovereignty involves the technological enforcement of contractual terms to enable data providers to retain some control over the reuse of data. Data sovereignty relies on the premise that a market for data can produce a fair "price for data" and strongly relies on the hypothesis of an appropriate data valorisation and monetisation mechanism. However, when it comes to the value of data, establishing a "fair price for data" may be complicated by the very nature of data, as its value increases the more it is reused and repurposed.

There are different streams of work that relate to data value and valuation[370], and which touch upon the notion of data as an asset (i.e., value for the user, in terms of future business benefits), as opposed to a resource (i.e., cost incurred to make data available, as a proxy for value). However, stakeholders still lack the common assessment models they need to assess the value IA participation upfront.

Data holders and re-users may not be able to reach an agreement without an appropriate data monetisation mechanism and where they fail to capture (and monetise) the gains from data sharing. The study finds that the development and implementation of IAs may be hampered by a lack of information or uncertainty about the magnitude (value) of expected costs, the benefits that can be derived from an IA, and how they are distributed. This could happen, for instance, where the IA-related costs are disproportionately high for smaller companies compared to larger companies, or where an IA creates benefits for stakeholders that have not shared in the cost of IA development or implementation. A lack of information (non-transparency) can also result in stakeholders under- or overestimating the potential gains from an IA, which could discourage their development and/or implementation.

There is therefore a degree of uncertainty about who should reap the economic benefits of IA-based data sharing, how data providers can ensure that they receive some of the profits that are generated from the data they share and how to ensure that data is ultimately not used in a way that would go against the interests of the data providers. From a legal perspective, the challenges encountered by industry stakeholders are more generally associated with the absence of a clear and overarching legal framework to govern the status and value of data.

These types of uncertainties make it difficult for data providers, to assess their data's perceived value against a range of potential risks (e.g., perceived loss of control over data, loss of trade secrets and data policy breaches). This highlights the importance of establishing business models and mechanisms to capture and monetise the benefits of data sharing and redistribute them to those parties that feel unfairly treated or are at risk of losing out.

Specific actions for industry to implement to improve data valuation[371]

1. **Develop clause(s) to guarantee fair treatment.** The agreement must provide a level playing field and guarantee that there will be no special treatment of specific actors to ensure that the value of shared data is preserved. This type of guarantee would provide assurances to data sharing partners that the value of their data (a business asset) is protected. Performance obligations and data sharing rules are two possible clauses that could be included in industry agreements.
2. **Develop/agree on common valuation models** based on existing examples and good practices to

[370] For more details, see "*Non-transparency and asymmetry of IA costs and benefits" in* Section 2.1 "*Key economic barriers*"
[371] Ordered based on prioritisation emerging from the recommendations workshop.

enable the parties involved to assess the potential value and benefits upfront.

3. **Adopt new collaborative arrangements and legal constructs for the fair allocation of IA costs and benefits**, based on proven examples, e.g., data trust, data cooperatives.
4. **Review specific use cases** and identify what value a company gets in return when sharing a certain type of data for a certain period of time, with a view to developing the taxonomy of data needed to identify certain value classes. The value of data depends on how it is used, although when one assesses different types of data, there is a class of data for which a general asset value might be assignable.

Specific actions for policy-makers to support improvements in data quality, access and interoperability[372]

1. **Promote the dissemination of good practices and guidelines to define data valorisation/monetisation models** to provide best practice guidance and recommend the most common methods used to enable parties to share the value created from data sharing, e.g., income-based method, cost-based.
2. **Define specific "co-generated data" rights** (e.g., for IoT data in industrial settings) and distribute them to co-contributors – i.e., horizontal rights granted to (co-)generators of data, as well as different rights and obligations that the classic notion of "property" entails and which are to some extent separated and distributed over the value chain, with several entities contributing to the creation of data and knowledge. Clarify who may have certain entitlements to avoid any undue data 'enclosures' (e.g., by actors that are well-placed to exert de facto ownership in the value chain).
3. **Leverage legal branches other than property law to fairly allocate the benefits of data**. Regulation of unfair B2B commercial practices could serve the purpose. EU harmonisation of unfair B2B commercial practices fit for the data economy, either horizontally or specifically targeted at data (e.g., ELI-ALI Principles), would help to provide legal certainty to all digital value chain participants.
4. **Support the development of a Data Asset Value Exchange Mechanism, through a network of Clearing Houses**, to enable organisations to exchange data while providing protection against counterpart risks (e.g., fear of the loss of data sovereignty). These houses should be created at sectoral level, with a regional/national focus to begin with. They should bring key public players (e.g., academy of science) together to develop a data exchange architecture using the IDSA framework and define model contracts for data exchange using distributed ledger (smart contract trading as in derivatives markets).[373] An umbrella organisation should coordinate the network, provide API access, certifications and drive the harmonisation of European laws and practices. This umbrella function is proposed to be given to the "European Data Innovation Board".

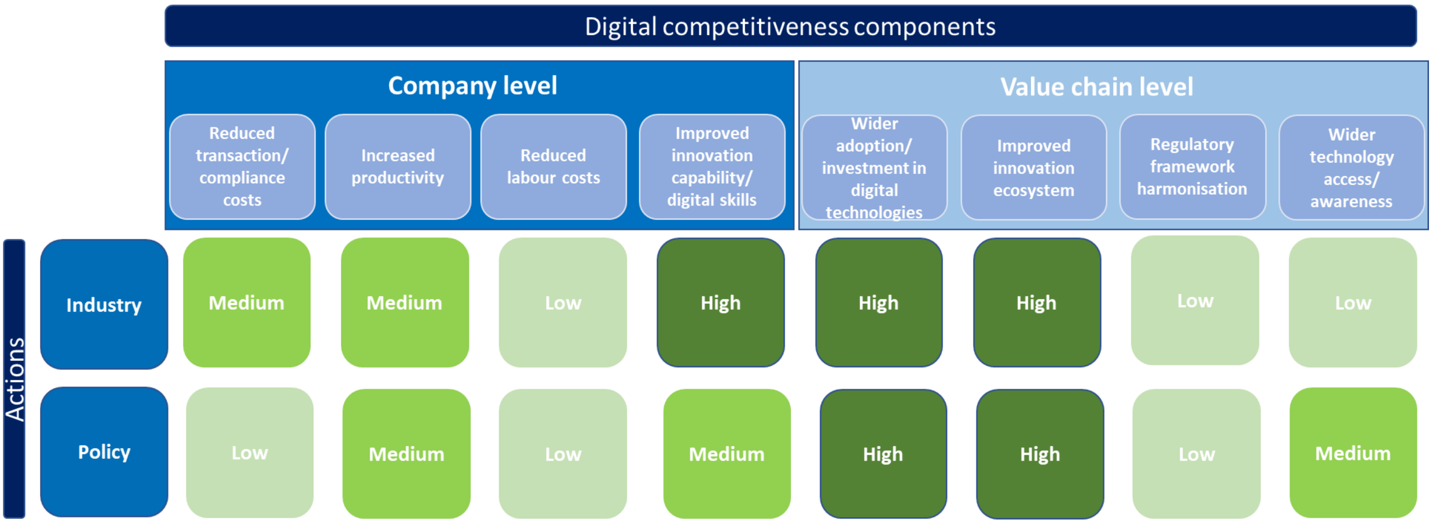
Actions to define common rules for data valuation and to protect the value of data shared are expected to deliver a number of benefits that will improve industrial competitiveness, as shown in the figure below. Due to the wide range of potential data and use cases, and the relative infancy of data valuation, there are no simple or common methods. It is therefore difficult to assess the potential value and benefits of data sharing upfront. Guidance and common practices on data valuation models should help to create business models to capture and monetise the benefits of data sharing, by setting up a mechanism for the redistribution of gains to those

[372] Ordered based on prioritisation emerging from the recommendations workshop.
[373] See Chapter 2.2 for further details.

parties that feel unfairly treated or are at risk of losing out by sharing their data. This would help companies to assess the value of data sharing upfront and balance their data's perceived value against potential risks of exposure. It would also help to accelerate the adoption of digital technologies and encourage greater levels of investment.

Similarly, innovative collaboration arrangements (e.g., data trust, data cooperatives) should include clear mechanisms for the redistribution of benefits. This type of provision would make it easier for industry parties to assess the potential value of data sharing upfront and improve the overall innovation ecosystem. Greater clarity and transparency, with regards to the magnitude of expected benefits that an IA can deliver and how they can be fairly distributed, should also help to accelerate digital technology adoption and investments, and increase productivity.

Figure 23 Main impacts of proposed actions on industrial competitiveness



Source: Study/consortium partners – own analysis validated by industry experts; *Note: Assessment based on the top ranked actions according to the stakeholders consulted: i) Industry action - Promotion of innovative collaborative arrangements for the fair allocation of costs and benefits; ii) Policy action - Dissemination of good practices and guidelines to define common data valuation/monetisation models. Their impacts on a range of digital competitiveness components are assessed, with a distinction made between company and value chain levels. For more information, please see Annex III on the outcomes of the recommendations workshop.*

Recommendation 4: Take action to address technical challenges and improve the clarity of data sharing regulations

The study finds that high fixed costs may be caused by the complexity of the technical aspects addressed by an IA or by the complexity of the regulatory environment in which an IA operates. This complexity may actually inhibit IA development. The high costs of IA development may be associated with, for example, the need to involve specialist expertise to help formulate an IA, the volume of effort required (in terms of quantity of inputs and the time required), including administration and coordination costs (particularly challenging in fragmented industries). Other important barriers could include a lack of relevant knowledge and skills (i.e., technical, management, business, economics or legal), or the need to make upfront investments in equipment, information capture and processing, or other forms of process reengineering.

In addition, the development of industry agreements can require extensive synchronisation work, bringing together many organisations that operate in different markets, countries and value chains. Coordination is also particularly challenging in fragmented industries that feature a multitude of players of different sizes.

Regulatory environments can also add to the complexity of IA development work, especially where regulations are not harmonised across borders. A myriad of rules at different levels (i.e., regional, national, EU, international) can create a degree of confusion and legal uncertainty. Industry agreements are confronted with a very complex legal framework, featuring a variety of complementary legal instruments that each impose requirements and limitations on the contractual freedom of the parties involved. The grey areas are large and the legal uncertainty is perceived by companies to be a considerable barrier, especially in relation to new digital issues for which there is currently little case law and good practice. For instance, some data pooling agreements between big companies could lead to anticompetitive agreements, according to competition law, and may therefore be prohibited. Another major challenge is the lack of clarity on how data is classified (e.g., non-personal or anonymised data) and what obligations are applicable.

This regulatory background may prove to be very restrictive and could deter stakeholders from engaging in innovative cooperation (e.g., due to high compliance costs or concerns about anti-competitive effects).

Specific actions for industry to implement to address technical and regulatory complexity[374]

1. **Leverage existing trusted entities** such as the network of European Digital Innovation Hubs (DIHs) – e.g., Big Data Innovation Hubs network – European cluster networks and the BDVA i-Spaces **for safe experimentation and validation under recognised labels**. These entities can act as intermediary organisations and support B2B data sharing with a range of services, such as match-making and the provision of enabling infrastructure, tools, technologies and technical/legal advice, thereby providing fertile ground on which to lay the foundation for industry agreements.
2. **Promote IA use cases in operations to identify technical issues in advance of future demand,** including how to address and resolve them. This would provide guidance and would help to make technical complexities more foreseeable and easier to address.
3. **Develop and include further information duties in industry agreements** to help reduce grey areas and provide greater clarity on technical aspects, e.g., information obligations for AI systems on: (1) the origin and processing of input data; (2) AI system functionality; (3) how the system was created; and (4) what the benefits and risks are.

Specific actions for policy makers to address technical and regulatory complexity[375]

1. **Clarify essential concepts of liability**, e.g., the content and meaning of essential concepts such as 'product' or 'defect' within the Product Liability Directive (PLD) or 'due diligence', to support the development of a clear and coherent EU civil (tort) liability regime for AI;
2. **Provide/promote unambiguous legal interpretations, guidelines or model contracts/ contractual clauses** that companies can use to provide greater clarity and legal certainty on liability, data protection and competition law. Practical guidelines on how competition law applies would be particularly helpful, as they would enable companies to determine on a case-by-case basis whether their cooperation agreements are compatible with competition rules, e.g., for data pooling and data sharing between competing companies, specifically with regards to the Horizontal Guidelines;
3. **Support the establishment of European coordination bodies for data sharing** to enable and support the exchange of best practices and promote guidelines for common practices, e.g. promote initiatives that involve stakeholders in the drafting of model agreements or templates that could be

[374] Ordered based on prioritisation emerging from the recommendations workshop.
[375] Ordered based on prioritisation emerging from the recommendations workshop.

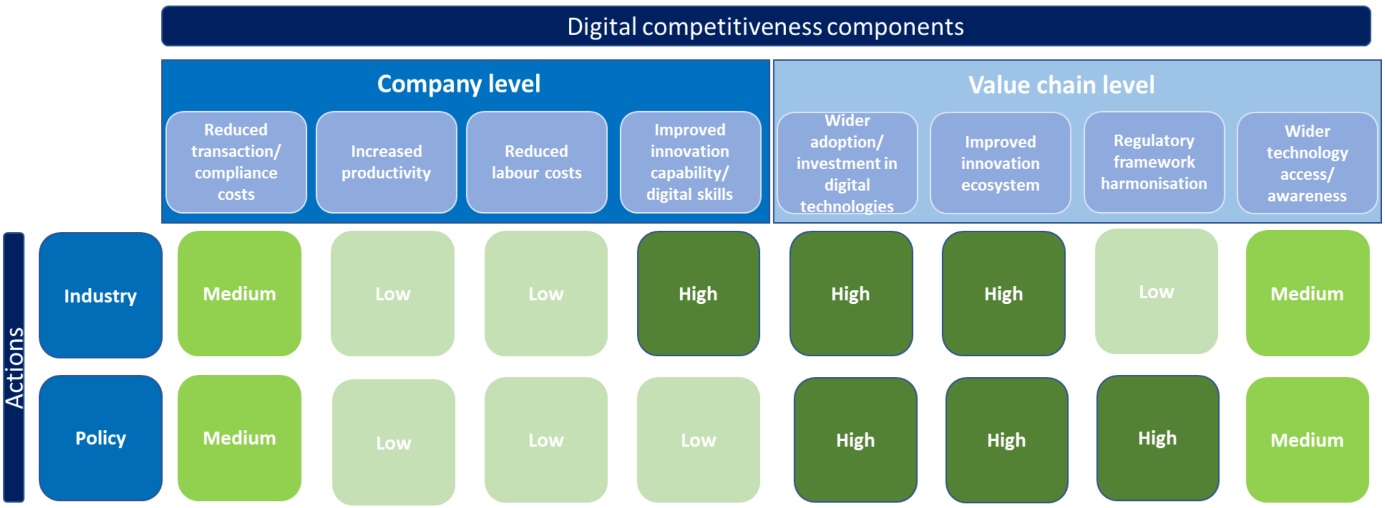
used by other supply chain parties (e.g., regarding the allocation of liabilities, exoneration clauses, etc). This role could be taken up by the envisaged European Data Innovation Board;

4. **Support the creation of (EU-wide) soft infrastructure and/or trusted frameworks for cross-sectoral data governance**, ensuring the convergence of existing initiatives and bringing players together across sectoral value chains to break down coordination barriers;
5. **Promote regulatory sandboxes to address regulatory issues** that raise concerns or uncertainty about the viability of collaboration opportunities. This would enable the set-up of collaborative spaces in which to test new business models at the edge or outside of the current regulatory framework. Testing would be under the supervision of regulatory institutions (e.g., data protection authority, market authority etc.) that could also explore and test innovative approaches to regulation.

Overcoming the challenge of technical and regulatory complexity should help industry to unlock significant competitiveness gains. The use of trusted entities (e.g., Big Data Innovation Hubs network) should facilitate stakeholder engagement (reducing pre-contractual costs) and make it easier for them to access the technical/legal expertise they require. This should lead to better stakeholder engagement throughout the value chain and help to improve their digital skills and innovation capabilities. Wider access to enabling technologies is also expected to improve stakeholder awareness and encourage greater investments in innovative solutions.

The overall innovation ecosystem is also expected to benefit from regulatory harmonisation and actionable guidelines on how relevant laws apply to the data economy. They should help to build trust between stakeholders when negotiating contracts and sharing their data. In addition, greater legal certainty will lower compliance costs for companies, increase confidence and encourage the wider adoption of innovative solutions.

Figure 24 Main impacts of proposed actions on industrial competitiveness



Source: Study/consortium partners – own analysis validated by industry experts; ***Note:*** *Assessment based on the top ranked actions according to the stakeholders consulted: i) Industry action - Leverage existing entities for safe experimentation and validation, seek tailored advice and technical expertise; ii) Policy action - Provide additional guidance/practical guidelines on how relevant laws apply (e.g., competition law). Their impacts on a range of digital competitiveness components are assessed, with a distinction made between company and value chain levels. For more information, please see Annex III on the outcomes of the recommendations workshop.*

Bibliography

Accenture and WDT, Sin dato, Value of data, the dawn of the data marketplace, https://datamakespossible.westerndigital.com/wp-content/uploads/2018/05/Western_Digital_VALUE_OF_DATA_Dawn_of_the_Data_Marketplace.pdf ,

Achim Wambach, 2020, A new competitive framework for the digital economy, https://www.bmwi.de/Redaktion/EN/Publikationen/Wirtschaft/a-new-competition-framework-for-the-digital-economy.pdf?__blob=publicationFile&v=3

AEF, ISOBUS Library, https://www.aef-isobus-database.org/isobusdb/login.jsf

Aeronewstv.com, 2015, Compagnies aérienne : les plus gros postes de dépense, https://www.aeronewstv.com/fr/transport/compagnies-aeriennes/4674-compagnie-aerienne-les-plus-gros-postes-de-depense.html

Aerospace Valley, Sin Dato, Le territoire, https://www.aerospace-valley.com/page/le-territoire

Agarwal Rachit, Gomez David, 2016, Unified IoT ontology to enable interoperability and federation of testbeds, https://hal.inria.fr/hal-01386917/document

Aircraft IT maintenance opérations, 01/01/2018, *Big Data: Racing to platform maturity*, https://www.aircraftit.com/articles/big-data-racing-to-platform-maturity/

Anmut (2020), *Different Data Valuation Methodologies: The Pros & Cons,* https://www.anmut.co.uk/different-data-valuation-methodologies/

Anticipatory Regulation: 10 ways governments can better keep up with fast changing industries, https://www.nesta.org.uk/blog/anticipatory-regulation-10-waysgovernments-can-better-keep-up-with-fast-changing-industries/

APCO Worldwide, 7/12/2020, *Has Germany set the European Transport sector on the path to a digital transformation?,* https://apcoworldwide.com/blog/has-germany-set-the-european-transport-sector-on-the-path-to-a-digital-transformation/

ARENA, 15/09/2021, *Affordable Heating and Cooling Innovation Hub (iHub), https://arena.gov.au/projects/affordable-heating-and-cooling-innovation-hub-ihub/*

Artrith et all, 31/05/2021, *Best Practices in machine learning for chemistry*, https://www.nature.com/articles/s41557-021-00716-z?proof=t%29.

Aspexit, 17/02/2021, *Standards and data exchange in agriculture, https://www.aspexit.com/standards-and-data-exchange-in-agriculture/*

Aubry Alexis, 2017, Semantic interoperability for an integrated product development process, https://hal.archives-ouvertes.fr/hal-01537847/document

Aviation Safety, 18/03/2020, *The benefits of Aviation Data Sharing and how to get started,* https://www.global-aero.com/the-benefits-of-aviation-data-sharing-and-how-to-get-started/

Balijepalli Saketh, 01/01/2019, *Best Practices in Using Semantic Transformation in Data Integration to Address Data integrity issues, https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/24786/Balijepalli_2019.pdf?sequence=1&isAllowed=y*

Bhattarai et all, 07/04/2019, *Big Data Analytics in Smart Grids*, https://www.osti.gov/pages/biblio/1639296

BigDataCoE, Sin dato, Big data in Europe: new environment, new opportunities, http://cdn.bdigital.org/PDF/BigDataCongress15/OpenEvidence_FranciscoLupianezVillanueva.pdf

Brin Alice & Helliwell John, 06/08/2019, *Why is interoperability between the two fields of chemical crystallography and protein crystallography so hard, https://journals.iucr.org/m/issues/2019/05/00/lt5022/*

Bundestag Study Commission on Artificial Intelligence ,2020, https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/AIDA/DV/2020/11-09/Executive_Summary_Final_ReportEN.pdf

Burnel Philippe, 09/09/2018, *The introduction of electronic medical records in France: More progress during the second attempt*, https://www.sciencedirect.com/science/article/pii/S0168851018303245

Catapults, 2019, Energy Data Taskforce Appendix 6, https://es.catapult.org.uk/wp-content/uploads/2019/06/EDTF-Report-Appendix-6-Standards.pdf

CEMA-AGRI, 2020, EU Code of conduct on agricultural data sharing by contractual agreement, 2020, https://cema-agri.org/images/publications/brochures/EU_Code_of_conduct_on_agricultural_data_sharing_by_contractual_agreement_2020_ENGLISH.pdf

CETIM, 2017, It pays to take part in the standardisation process, https://www.cetim.fr/en/News/headlines/In-brief/It-pays-to-take-part-in-the-standardisation-process

CFA Institute, Sin dato, Financial Clearing Houses, https://www.cfainstitute.org/en/advocacy/issues/central-clearing-houses

CHEManager, 2018*, Life sciences- und chemiebranchen sehen kooperationen als attraktive alternativen su M&A*, https://www.chemanager-online.com/themen/management/life-sciences-und-chemiebranchen-sehen-kooperationen-als-attraktive-alternative-zu

Chen Wen (et al), 2017, Data quality of electricity consumption data in a smart grid environment, https://www.sciencedirect.com/science/article/abs/pii/S1364032116307109

Chicco Gianfranco, 13/05/2021, *Data Consistency for data-driven Smart Energy Assessment, https://www.frontiersin.org/articles/10.3389/fdata.2021.683682/full*

Chondrogiannis Efthymios, Sin dato, An intelligent ontology alignment tool dealing with complicated mismatches, http://ceur-ws.org/Vol-1320/paper_16.pdf

Chu, Hung-Ju, Y Randy, Chen Su-Shing, 2005, Semantic Association of taxonomy-based standards using ontology, http://ceur-ws.org/Vol-156/paper4.pdf

Claudia Imhoff, 2018, Advanced Data Lineage: The #1 Key to Removing the Chaos in Modern Analytical Environments, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE36VUW

Clin Chim Acta, 2003, The effects of total laboratory automation on the management of a clinical chemistry laboratory, https://pubmed.ncbi.nlm.nih.gov/12589970/

Connecting Software, 09/07/2021, *Infographics: Why are IT systems integration costs so high,* https://www.connecting-software.com/blog/infographics-why-are-it-system-integration-costs-so-high/

Costin Aaron, Issa Raja, 2017, The need for taxonomies in the ontological approach for interoperability of heterogeneous information models, https://ascelibrary.org/doi/abs/10.1061/9780784480830.002

D. Moody, P. Walsh, *Measuring The Value Of Information: An Asset Valuation Approach*

Das Sarasij, Radhakrishan Asha, 01/12/2018, *Quality Assessment of smart grid data,* *https://www.iitk.ac.in/npsc/Papers/NPSC2018/1570474605.pdf*

David Hall — UK civil aviation authority, 4/04/2002, *Aviation Maintenance Data in the United-Kingdom*, https://www.faa.gov/about/initiatives/maintenance_hf/library/documents/media/human_factors_maintenance/aviation_maintenance_data_in_the_united_kingdom.pdf

David Laurianne et all, 17/09/2020, *Molecular representations in AI-driven drug discovery: a review and practical guide*, https://jcheminf.biomedcentral.com/articles/10.1186/s13321-020-00460-5

Decide, 2019, The data boom and the possibilities it offers, https://decidesoluciones.es/en/data-quality-or-data-quantity-which-is-most-important-for-ai/

Deloitte (2020), *Data valuation: Understanding the value of your data assets,* *https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Finance/Valuation-Data-Digital.pdf*

Department of Energy Pages, 07/04/2019, *Big Data Analytics in Smart Grids*, https://www.osti.gov/pages/biblio/1639296

Deuter Andreas, Imort Sebastian, 23/06/2021, *Product Lifecycle management with the asset administration shell,* https://www.mdpi.com/2073-431X/10/7/84/pdf

Diab Wael, 2019, AI Standards help accelerate digitalization of smart manufacturing, https://etech.iec.ch/issue/2019-06/ai-standards-help-accelerate-digitalization-of-smart-manufacturing

Digital Clearinghouse website, https://www.digitalclearinghouse.org/

DIZMO, 2019, The unrealized benefits of predictive maintenance in Industry 4.0, https://www.dizmo.com/the-unrealized-benefits-of-predictive-maintenance-in-industry-4-0/

DotMagazine, 01/04/2021, *Data as the key to connected mobility,* https://www.dotmagazine.online/issues/digital-acceleration/digital-trade-routes/connected-mobility

Ducuing Charlotte, 2020, Beyond the Data Flow Paradigm: Governing Data Requires to Look beyond Data, https://techreg.org/index.php/techreg/article/view/49

Ducuing Charlotte, Dutkiewicz Lidia and Miadzvetskaya Yuliya, 2020, Legal and Ethical Requirements (TRUSTS Trusted Secure Data Sharing Space)

Ducuing, Charlotte., 2019, 'Data as Infrastructure? A Study of Data Sharing Legal Regimes', https://doi.org/10.1177/1783591719895390

ECP, 2018, Artificial Intelligence Impact Assessment, https://ecp.nl/wp-content/uploads/2019/01/Artificial-Intelligence-Impact-Assessment-English.pdf

Energy Systems Catapult, 2019, Energy data Taskforce: a strategy for a modern digitalised energy system, https://es.catapult.org.uk/reports/energy-data-taskforce-report/

EOSC taskforce FAIR metrics and Data Quality Charter, https://www.eosc.eu/sites/default/files/tfcharters/eosca_tffairmetricsanddataquality_draftcharter_20210614.pdf

ESMA, EBA and EIOPA, 2020, FinTech: Regulatory Sandboxes and Innovation, https://eba.europa.eu/sites/default/documents/files/documents/10180/2545547/154a7ccb-06de-4514-a1e3-

0d063b5edb46/JC%202018%2074%20Joint%20Report%20on%20Regulatory%20Sandboxes%20and%20Innovation%20Hubs.pdf

ETSI, 2018, Cross-fertilisation through alignment, synchronisation and exchanges for IoT, https://european-iot-pilots.eu/wp-content/uploads/2018/11/D06_02_WP06_H2020_CREATE-IoT_Final.pdf

EUHubs4Data website, https://euhubs4data.eu/#:~:text=The%20European%20federation%20of%20Data,data%20in%20a%20cross%2Dborder

Euractiv, 2016, Smart Farming trying to find its feet in EU agriculture, https://www.euractiv.com/section/agriculture-food/news/smart-farming-trying-to-find-its-feet-in-eu-agriculture/

Euractiv, 2020, Precision Farming Saved extra virgin olive oil, Greek producers say, https://www.euractiv.com/section/agriculture-food/news/precision-farming-saved-extra-virgin-olive-oil-greek-producers-say/

Eurocontrol, sin dato, *European AIS database*, https://www.eurocontrol.int/service/european-ais-database

European Commission, 2013, Commission Working Document, Analysis of measures that could lead significant market players in the ICT sector to license interoperability information, SWD(2013) 209 final, https://ec.europa.eu/digital-single-market/en/news/analysis-measures-could-lead-significant-market-players-ict-sector-license-interoperability

European Commission, 2014, Commission Delegated Regulation (EU) 2015/962 of 18 December 2014 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide real-time traffic information services, OJ L 157/21, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32015R0962&from=en

European Commission, 2016, DEI Working Group 2, Strengthening Leadership in Digital Technologies and in Digital Industrial Platforms across Value Chains in all Sectors of the Economy, https://ec.europa.eu/futurium/en/system/files/ged/dei_working_group_2_platforms.pdf

European Commission, 2017, Commission Delegated Regulation (EU) 2017/1926 of 31 May 2017 supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide multimodal travel information services, OJ L 272/1, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R1926&from=EN

European Commission, 2017, Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee, COM(2017) 712 final, https://ec.europa.eu/docsroom/documents/26583

European Commission, 2017, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Building a European Data Economy, SWD(2017) 2 final, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017DC0009&from=EN

European Commission, 2017, Digitising mechanical engineering: leveraging the potential of the cloud and data, https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/DTM_%20Digitising%20mech%20eng%20v2.pdf

European Commission, 2017, Harnessing the economic benefits of Artificial Intelligence, https://ec.europa.eu/growth/tools-databases/dem/monitor/sites/default/files/DTM_Harnessing%20the%20economic%20benefits%20v3.pdf

European Commission, 2018, Ethics guidelines for trustworthy AI, https://ec.europa.eu/futurium/en/ai-alliance-consultation

European Commission, 2019, Liability for Artificial Intelligence and other emerging digital technologies, https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=36608

European Commission, 2020, COM(2020) 64 final, https://ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020_en_1.pdf

European Commission, 2020, COM(2020) 65 final, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

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

European Commission, 2020, ERAvsCorona Action Plan, https://ec.europa.eu/info/sites/info/files/research_and_innovation/research_by_area/documents/ec_rtd_era-vs-corona_0.pdf

European Commission, 2020, European Strategy for data, https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_fr

European Commission, 2020, Online consultation – a European Strategy for data, https://ec.europa.eu/digital-single-market/en/news/online-consultation-european-strategy-data .

European Commission, 2020, Proposal for a regulation on contestable and fair markets in the digital sector (Digital Markets Act), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0842&qid=1624006593011

European Commission, 2020, Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, https://ec.europa.eu/info/sites/info/files/report-safety-liability-artificial-intelligence-feb2020_en_1.pdf

European Commission, 2021, Proposal for a regulation laying down harmonised rules on artificial intelligence, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206

European Commission, sin dato, *Espace européen des données de santé, https://ec.europa.eu/health/ehealth/dataspace_fr*

European IP Helpdesk, 2019, Your Guide to IP and Contracts, https://www.iprhelpdesk.eu/sites/default/files/2018-12/european-ipr-helpdesk-your-guide-to-ip-and-contracts.pdf.

European Parliament, 2016, Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, OJ L 157/1, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016L0943&from=EN

European Parliament, 2020, Civil liability regime for artificial intelligence, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/654178/EPRS_STU(2020)654178_EN.pdf

European Parliament, 2020, Opportunities of Artificial Intelligence, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/652713/IPOL_STU(2020)652713_EN.pdf

European Parliament, 2020, Report- with recommendations to the Commission on a civil liability regime for artificial intelligence, https://www.europarl.europa.eu/doceo/document/A-9-2020-0178_EN.pdf

European Research Area Committee, 2012, COM, European Research Area Guidelines on Intellectual Property (IP) Management in International Research Collaboration Agreements between European and Non-European Partners, https://ec.europa.eu/research/innovation-union/pdf/international_cooperation_guidelines_erac_kt_group.pdf

European Smart Grids Task Force, 2019, *Towards interoperability within the EU for electricity and gas data access & Exchange,* *https://ec.europa.eu/energy/sites/ener/files/documents/eg1_main_report_interop_data_access.pdf*

European Smart Grids Task Force, 2019, Towards interoperability within the EU for electricity and gas data access & Exchange, https://ec.europa.eu/energy/sites/ener/files/documents/eg1_main_report_interop_data_access.pdf

EuroScientists, 2020, Will smart farming gain a foothold in Europe, https://www.euroscientist.com/will-smart-farming-gain-a-foothold-in-europe/

Eurostat, Healthcare expenditure across the EU: 10% of GDP, https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20190904-1

Falke, J. and Schepel, H.,2000, Legal aspects of standardisation in the Member States of the EC and EFTA, https://op.europa.eu/en/publication-detail/-/publication/9acd4259-13c6-436f-8c12-2740ccf7bb71

Federal Ministry of Economic Affairs and Energy in Germany, 2020, https://www.bmwi.de/Redaktion/EN/Artikel/Digital-World/GAIA-X-Use-Cases/covid-19-dashboard-und-hub.html

Federal Ministry of transport and digital infrastructure, sin dato, *Open data for smart mobility conference*, https://www.bmvi.de/SharedDocs/EN/Documents/K/open-data-documentation.pdf?__blob=publicationFile

Fernadnez Antonio, 16/02/2019, *How Blockchain could enhance aircraft maintenance*, https://datascience.aero/blockchain-enhance-aircraft-maintenance/

Fragidis Leonidas, Chatzoglou Prodomos, 02/03/2018, *Implementation of a nationwide electronic health record (EHR),* https://pubmed.ncbi.nlm.nih.gov/29504871/

Fraunhofer/BAST, 01/03/2020, *Mobility Data Space*, https://www.mobility-data-space.de/content/dam/ivi/mobility-data-space/documents/Mobility_Data_Space_2020_EN_neu.pdf

Frederic Sutter (Airbus digital transformation leader), 02/10/2019, *Skywise*, https://www.aaco.org/Library/Assets/Skywise%20by%20Fredric%20Sutter%20-%20Airbus.pdf

Frontiers in Big Data, 13/05/2021, *Data Consistency for Data-Driven Smart Energy Assessment*, https://www.frontiersin.org/articles/10.3389/fdata.2021.683682/full

Ghent University — Faculty of Medicine and health sciences, 2017 [Thesis], *The use of Electronic Health Records in Belgian Physiotherapy, https://libstore.ugent.be/fulltxt/RUG01/002/350/192/RUG01-002350192_2017_0001_AC.pdf*

GMI Research, 2018, Artificial Intelligence (AI) Market, https://www.gmiresearch.com/report/global-artificial-intelligence-ai-market/

Goh Gareett et all, sin dato, *Deep Learning for Computational chemistry*, https://arxiv.org/ftp/arxiv/papers/1701/1701.04503.pdf

Goldbeck Gerhard, 2012, The economic impact of molecular modelling of chemicals and materials, https://www.researchgate.net/publication/262232197_The_economic_impact_of_molecular_modelling_of_chemicals_and_materials

Gorbatyuk, Arina, 2019, 'Rethinking Intellectual Property Ownership in the Context of Open Innovation', KU Leuven Ph.D. Thesis https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS2815296&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US&fromSitemap=1

Gorbatyuk, Arina, 2020, 'The Allocation of Patent Ownership in R&D Partnerships: Default Rules v. Contractual Practices', https://script-ed.org/article/the-allocation-of-patent-ownership-in-rd-partnerships-default-rules-v-contractual-practices/

Gotzens, Fabian, 2019, Performing energy modelling exercises in a transparent way – The issue of data quality in power plant databases, https://www.sciencedirect.com/science/article/pii/S2211467X18301056

Gov.uk, 2018, Energy Data Taskforce, https://www.gov.uk/government/groups/energy-data-taskforce

Graef, I. and Prufer, J., 2018, Mandated Data Sharing Is a Necessity in Specific Sectors, https://ssrn.com/abstract=3206685

Hagedoorn, 2002, Inter-firm R&D Partnerships: An Overview of Major Trends and Patterns since 1960', https://doi.org/10.1016/S0048-7333(01)00120-2

Halt, G., 2014, Intellectual Property in Consumer Electronics, Software and Technology Startups

Handelsblatt, 17/11/2020, *Datenschatz Verkehr: VW und Daimler wollen die Pläne der Kanzlerin unterstützen, https://www.handelsblatt.com/politik/deutschland/autogipfel-datenschatz-verkehr-vw-und-daimler-wollen-die-plaene-der-kanzlerin-unterstuetzen/26631022.html?ticket=ST-512540-c5LOzyYgi9jBd3MMpVen-ap4*

Handelsblatt, October 2020, Merkel Drängt autokonzerne: BMW, Daimler und VW sollen Datenschatz teilen, https://www.handelsblatt.com/politik/deutschland/autogipfel-merkel-draengt-autokonzerne-bmw-daimler-und-vw-sollen-datenschatz-teilen/26308418.html?ticket=ST-1376973-6msjtMrmiOfcreSP5PK1-ap2

Hartman Dirk and van der Auweraer Herman, 28/01/2020, *Digital Twins, https://arxiv.org/abs/2001.09747*

Hastings Jana et all, 19/08/2011, *Chemical Ontologies: what they are, what are they for and what are the challenges*, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3083557/

Hawker C.D. Schlank M. R., 2000, Development of standards for laboratory automation, https://pubmed.ncbi.nlm.nih.gov/10794772/

Holzapfel, H. and Königs, M., 2018, New German Law on the Protection of Trade Secrets, https://www.mwe.com/insights/new-german-law-protection-trade-secrets/ .

Honey Bee Manufacturing, Q4 2020, *Interoperability in Farming operations, https://www.pc.gov.au/__data/assets/pdf_file/0008/270575/sub002-repair-attachment2.pdf*

Hosep Crompvoets, Wageningen Universtiteit, National Spatial Data ClearingHouses, https://edepot.wur.nl/12175 .

HVAC&RNews, 16/11/2020, *Videos explore i-Hub Smart Building Data Clearing House*, https://www.hvacrnews.com.au/news/the-i-hub-project-has-released-two-videos-to-promote-and-explain-its-smart-building-data-clearing-house-initiative/

IAC Partners, sin dato, Point de vue: les entreprises de maintenance aéronautiques ont le vent en poupe, https://www.iacpartners.com/blog/quel-avenir-pour-la-maintenance-aeronautique

IBM, 2019, Extracting business value from the 4 V's of big data, https://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data

IDW, 18/05/2021, *Vernetzer Verkehr: acatech gründet Trägergesellschaft „DRM Datenraum Mobilität GmbH" als Non-Profit-Organisation*, https://idw-online.de/de/news769008;

iHub, 14/09/2021, *Smart Building Data Clearing House, https://www.ihub.org.au/ihub-initiatives/smart-building-data-clearing-house/*

iHub, 15/12/2020, *DCH2 reports: learnings from the use of FDD, analytics and demand response applicatiosn in the 3 trial buildings (hosted on the digital layer software platform DCH 1.2), https://ihub.org.au/wp-content/uploads/2021/01/DCH2-Technical-Report-Learnings.pdf*

IHUB, 18/10/2020, *i-Hub Switch Data Clearing House, https://ihub.org.au/wp-content/uploads/2021/01/DCH2-Lessons-Learnt-Final.pdf*

iHub, Report #001: Final DCH2 Knowledge Sharing Report, *15/12/2020*, https://ihub.org.au/wp-content/uploads/2021/01/DCH2-Final-Sub-project-KS-Report.pdf

IMDA (2019), *Guide to data valuation for data sharing*, https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Guide-to-Data-Valuation-for-Data-Sharing.pdf

Industrial internet consortium & platform industrie 4.0, *Digital Twin and asset administration shell concepts and application in the industrial internet and industrie 4.0,* https://www.iiconsortium.org/pdf/Digital-Twin-and-Asset-Administration-Shell-Concepts-and-Application-Joint-Whitepaper.pdf

Insee, 2018, Produits intérieurs bruts régionaux et valeurs ajoutées régionales de 1990 à 2015, https://www.insee.fr/fr/statistiques/1893220

INSPIRE, 10/03/2002, Inspire architecture and standards position paper, https://inspire.ec.europa.eu/reports/position_papers/inspire_ast_pp_v4_2_en.pdf

Institute of Science and technology, sin dato, *Chemistry of Electronic Materials,* https://www.sathyabama.ac.in/sites/default/files/course-material/2020-10/UNIT-5_2.pdf
IPC, 21/09/2021, *What is digital thread?, https://www.ptc.com/en/blogs/corporate/what-is-a-digital-thread*

iSHARE website, https://www.ishareworks.org/en/ishare

ISO, ISO and agriculture, https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100412.pdf and ISO, 2020, Getting big on Blockchain, https://www.iso.org/news/ref2540.html

ITEA3, 01/12/2020, *Machinaide, Deliverable D1.2, State-of-the-art description*, https://itea4.org/index.php/project/workpackage/document/download/7292/D1.2%20State-of-the-Art%20Description_V1.pdf

ITU, 2018, *Assessing the economic impact of artificial intelligence,* https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-ISSUEPAPER-2018-1-PDF-E.pdf

ITU, 2018, Assessing the economic impact of artificial intelligence, https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-ISSUEPAPER-2018-1-PDF-E.pdf

Järvenpäa Eeva et all, 14/11/2017, *The development of an ontology for describing the capabilities of manufacturing resources, https://d-nb.info/1165373912/34*

JCMB, sin dato, *Data Quality Assessment Process*, http://www.jcmb.com/jcmb2/images/PDF/smart_data_smart_grid_whitepaper.pdf

JDA Journal, sin dato, *Airbus' and Delta's SKYWISE is missing a key focus*, http://jdasolutions.aero/blog/airbus-deltas-skywise-missing-key-focus/

Jha Dipendra et all, 06/11/2018, *ElemNet : Deep learning the chemistry of materials from only elemental composition*, https://d-nb.info/1175382124/34

Journal of Cheminformatics, 2020, An open source chemical structure curation pipeline using RDkit, https://europepmc.org/article/pmc/pmc7458899

JRC Science for policy report, 2020, Smart Grid Interoperability Laboratory, https://publications.jrc.ec.europa.eu/repository/bitstream/JRC119721/kjna30114enn.pdf

K Q Wang, 2014, Analysis of Data Quality and Information Quality problems in digital manufacturing, https://www.researchgate.net/publication/251860789_Analysis_of_Data_Quality_and_Information_Quality_Problems_in_Digital_Manufacturing

Kearney, K., sin dato., 7 Ways to Improve Maintenance Costs with the Connected Aircraft, https://aerospace.honeywell.com/en/learn/about-us/blogs/2019/03/7-ways-to-improve-maintenance-costs-with-the-connected-aircraft

Kerber Wolfgang and Scweitzer Heike, 2017, Interoperability in the Digital Economy, 8 (2017) JIPITEC 39 para 1, https://www.jipitec.eu/issues/jipitec-8-1-2017/4531

Kerber Wolfgang, Scweitzer Heike, 2017, *Interoperability in the digital economy*, https://www.jipitec.eu/issues/jipitec-8-1-2017/4531

Kleinberg Jon et all, 01/01/2018, *Advances in Big Data Research in Economics — Algorithmic fairness*, https://www.cs.cornell.edu/home/kleinber/aer18-fairness.pdf

KMENDIS, sin dato, *Electronic Health Records, https://kmendis.net/electronic-health-records/*

Kotagiri Sunil, 06/03/2019, *Data Quality and Governance Critical for Utilities, https://www.tdworld.com/smart-utility/data-analytics/article/20972300/data-quality-and-governance-critical-for-utilities*

Lanz Minna et all, 06/06/2012, *Towards adaptive manufacturing systems — knowledge and knowledge management systems, https://www.intechopen.com/chapters/36408*

Libretext Chemistry – Belford Robert, 23/06/2019, *Cheminformatics*, https://chem.libretexts.org/Courses/University_of_Arkansas_Little_Rock/ChemInformatics_(2017)%3A_Chem_4399_5399

LibreTexts Chemistry, 12/08/2020, *Line Notation Smiles and InChI, https://chem.libretexts.org/Courses/Fordham_University/Chem1102%3A_Drug_Discovery_-_From_the_Laboratory_to_the_Clinic/05%3A_Organic_Molecules/5.08%3A_Line_Notation_(SMILES_and_InChI) and on* Institute of Science and technology, sin dato, *Chemistry of Eletronic Materials,* https://www.sathyabama.ac.in/sites/default/files/course-material/2020-10/UNIT-5_2.pdf

Lockheed Martin, 14/05/2021, *Usability of manufacturing data for analytics, https://www.nist.gov/system/files/documents/2021/04/26/Wednesday%20-%2001%20-%20Usability%20of%20Manufacturing%20Data%20for%20Analytics.pdf*

LOTAME, 2019, Why Is Data Quality Important?, https://www.lotame.com/why-is-data-quality-important/#:~:text=Improved%20data%20quality%20leads%20to,in%20consistent%20improvements%20in%20results

Lucht- en Ruimtevaart nederland, 01/03/2018, *Roadmap Aeronautics 2018-2025, https://www.nlr.nl/wp-content/uploads/2018/03/Roadmap-Aeronautics-2018-2025.pdf*

Lundholm et all, 01/12/2016, *Integration of Digital Factory with smart factory based on internet of things, https://www.researchgate.net/publication/306023791_Integration_of_Digital_Factory_with_Smart_Factory_Based_on_Internet_of_Things*

Marketandmarkets, 2020, Big Data Market, https://www.marketsandmarkets.com/Market-Reports/big-data-market-1068.html ,

Markets and markets, 2017, Lab Automation Market by Equipment and Software (Robotic Arms, Microplate Readers, LIMS), Applications (Drug Discovery, Clinical Diagnostics, Genomics), End Users (Biotechnology and Pharmaceuticals, Diagnostic Labs) - Global Forecast to 2022, https://www.marketsandmarkets.com/Market-Reports/lab-automation-market-1158.html

McKinsey & Co., 2017, Smartening up with Artificial Intelligence (AI) - What's in it for Germany and its Industrial Sector? https://www.mckinsey.com/~/media/McKinsey/Industries/Semiconductors/Our%20Insights/Smartening%20up%20with%20artificial%20intelligence/Smartening-up-with-artificial-intelligence.pdfMedium, 2019, How to ensure data quality for machine learning and AI projects?, https://medium.com/vsinghbisen/how-to-ensure-data-quality-for-machine-learning-and-ai-projects-c8af1fe18c57

McKinsey Global Institute,2017, Artificial intelligence the next digital frontier?, https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.pdfMedium, 2017, Data quality in the era of Artificial intelligence, https://medium.com/innovation-machine/data-quality-in-the-era-of-a-i-d8e398a91bef

Mejía Jorge, Ontology Localization, PhD Thesis, 2014, Saragosa University, https://zaguan.unizar.es/record/15513/files/TESIS-2014-058.pdf

Michael J Madison, 2014, Commons at the Intersection of Peer Production, Citizen Science, and Big Data: Galaxy Zoo, http://arxiv.org/abs/1409.4296

Michalpoulos Sarantis, 2020, EIB official: €200 billion needed to build broadband infrastructure in EU rural areas, https://www.euractiv.com/section/agriculture-food/news/eib-official-e200-billion-needed-to-build-broadband-infrastructure-in-eu-rural-areas/

Mobilitätsdatenmarketplatz, sin dato,*about the MDM*, https://www.mdm-portal.de/was-uns-aktuell-bewegt/

Mobility Data Space, 15/09/2021, Architecture and components, https://www.mobility-data-space.de/en/architecture-and-components.html

MordorIntelligence, 2020, Lab robotics market, https://www.mordorintelligence.com/industry-reports/laboratory-robotics-market

Mouzhi Ge et all, 22/03/2019, *Data Quality Management Framework for Smart Grid Systems*, https://www.researchgate.net/publication/331951606_Data_Quality_Management_Framework_for_Smart_Grid_Systems

Mussman Andrea et all, 24/10/2016, *Data quality challenges and future research directions in threat intelligence sharing practices,* http://library.usc.edu.ph/ACM/SIGSAC%202017/wiscs/p65.pdf

Nalin Marco et all, 01/06/2019, *The European cross-border health data exchange roadmap: Case study in the Italian setting, https://www.sciencedirect.com/science/article/pii/S1532046419301017*

Napcore, 16/09/2021, National Access Point Organisation for Europe, https://www.gov.si/en/registries/projects/napcore-koordinacija-nacionalnih-tock-dostopa-eu/

National Center for Ontological Research, November 2017, The industry ontology foundry, https://emmc.info/wp-content/uploads/2017/12/EMMC-IntOp2017-Cambridge_Smith_Buffalo.pdf

Natsiavas Pantelis et all, 18/06/2021, *Developing an infrastructure for secure patient summary exchange in the EU context: lessons learned from the KONFIDO project,* *https://journals.sagepub.com/doi/full/10.1177/14604582211021459*

Nature, 2017, Unique identifiers for small molecules enable rigorous labelling of their atoms, https://www.nature.com/articles/sdata201773

Nature, 2020, Machine learning for chemical discovery, https://www.nature.com/articles/s41467-020-17844-8

Nature, 2020, Representation of molecular structures with persistent homology for machine learning applications in chemistry, https://www.nature.com/articles/s41467-020-17035-5

NHS England, 01/02/2016, *Joint report on international success factors for adoption and use of digital health in the US and NHS England, https://www.healthit.gov/sites/default/files/adoptionreport_-branded_final4.pdf*

Niedergassel, 2011, Knowledge Sharing in Research Collaborations

OECD, 2018, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption, https://doi.org/10.1787/sti_in_outlook-2018-en

Office of the National Coordinator for Health Information technology, 2020, *Connecting health and care for the nation, a shared nationwide interoperability roadmap*, https://www.healthit.gov/sites/default/files/nationwide-interoperability-roadmap-draft-version-1.0.pdf

Ontology Alignment Evaluation Initiative: https://oaei.ontologymatching.org/ One can also find an example of ontology mapping http://www.ontobee.org/tutorial/ontobeep

Open Data Institute, 2018, What is a data trust?, https://theodi.org/article/what-is-a-data-trust/

Open Source Initiative website, https://opensource.org/licenses

Osma Hafizah, 12/12/2018, *Australia leads the world in personal control of electronic health records,* https://www.healthcareit.com.au/article/australia-leads-world-personal-control-electronic-health-records

Ostrom Elinor, 1990, Governing the Commons: The Evolution of Institutions for Collective Action, Cambridge University Press 1990

Pachl Christian et all, sin dato, *Overview of chemical ontologies*, https://arxiv.org/abs/2002.03842

Paper not yet published whose anticipated access was granted to the research team by the authors: Wigh Daniel, Goodman Jonathan, Lapkin Alexei, An updated review of Molecular representations in the Age of Machine Learning.

PIPAME, 2010, Maintenance et réparation aéronautique, https://www.entreprises.gouv.fr/files/files/directions_services/etudes-et-statistiques/etudes/aeronautique_maintenance/aeronautique_maintenance.pdf

Portail Européen de Donnée, sin dato, Open Energy data on the European data portal, https://www.europeandataportal.eu/fr/highlights/open-energy-data-european-data-portal

Power Grid International, 01/01/2010, *Maintaining Intelligent Networks — Smartgrid & Data Quality*, https://www.power-grid.com/news/maintaining-intelligent-networks-smartgrid-data-quality/#gref

Power, 01/01/2020, *Which country's smart grid is the smartest,* https://www.powermag.com/which-countrys-grid-is-the-smartest/

PROGNOCIS, 14/06/2019, *HER Interoperability — Why is it important*, https://prognocis.com/ehr-interoperability-why-is-it-important/

PROGNOCIS, 29/09/2016, *Is HER interoperability possible with a national patent identifier*, https://prognocis.com/is-ehr-interoperability-possible-with-a-national-patient-identifier/

PwC, 2015, Healthcare: A digital divide? Insights from PwC's 2015 Global Digital IQ Survey, https://www.pwc.com/gx/en/healthcare/pdf/healtcare-digital-divide.pdf

PwC, 2018, An introduction to implementing AI in manufacturing, https://www.pwc.com/gx/en/industrial-manufacturing/pdf/intro-implementing-ai-manufacturing.pdf

PWC, sin dato., Predictive maintenance for airlines, https://www.pwc.com/us/en/industries/transportation-logistics/airlines-airports/predictive-maintenance.html

ResearchGate, 01/06/2019, *IoT and data interoperability in agriculture, https://www.researchgate.net/publication/333853181_IoT_and_data_interoperability_in_agriculture_A_case_study_on_the_gaiasense_TM_smart_farming_solution*

Riemenschneider Rolf, 2020, Principles & Architectures for Data Infrastructures, European Commission DGCNECT

RoccoResearch, 03/04/2019, Results: The leading vendors in Data Clearing 2019, https://www.roccoresearch.com/2019/04/03/results-the-leading-data-clearing-houses-of-2019/

Rubinstein H., 2018, The Benefits of Applying Behavioural Science to Business, https://doi.org/10.1007/978-3-030-01698-2_9

S. Martínez-Fernández, 2015, A Survey on the Benefits and Drawbacks of AUTOSAR, https://ieeexplore.ieee.org/document/7447213

Safe-DEED, 2020, Legal Requirements for Processing of Non-Personal Data, Horizon 2020 project, https://safe-deed.eu/wp-content/uploads/2020/06/Safe-DEED_D3_3.pdf.

SILA, Sin dato, SILA 2 is now officially released, https://sila-standard.com/

Skybrary, sin dato, *Aviation Safety Information Analysis and Sharing*, https://skybrary.aero/index.php/Aviation_Safety_Information_Analysis_and_Sharing_(ASIAS)

Smart Grid Task Force, https://ec.europa.eu/energy/topics/markets-and-consumers/smart-grids-and-meters/smart-grids-task-force_en?redir=1

Smart Specialisation Platform, Chemical, sin dato, https://s3platform.jrc.ec.europa.eu/chemicals

SmartDataModels, Website, https://smartdatamodels.org/

Steger, U., 2018, The secret to keeping secrets secret: The German Trade Secrets Act, https://paytechlaw.com/en/trade-secrets/?pdf=9792

Sundarajan Aditya et all, 24/06/2020, *Adapting big data standards, maturity models to smart grid distributed generation: critical review*, https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/iet-stg.2019.0298

Surgere Autosphere website, https://surgere.com/autosphere/what-is-autosphere/

Szmeja Pawel et al, Declarative ontology alignment format for semantic translation, February 2018, https://ieeexplore.ieee.org/document/8519921

Task force Charter, sin dato, *Semantic Interoperability*, https://www.eosc.eu/sites/default/files/tfcharters/eosca_tfsemanticinteroperability_draftcharter_20210614.pdf

Teaoman Duran Timocin, 15/08/2020, *Data quality in the interface of industrial manufacturing and machine learning*, https://www.diva-portal.org/smash/get/diva2:1469008/FULLTEXT01.pdf

TechScurry, 17/03/2021, *Angela Merkel urges: Car companies should join the mobility data room*, https://techscurry.com/general/angela-merkel-urges-car-companies-should-join-the-mobility-data-room/17995/

The Behavioural Insights Team, Broughton, N. et al, 2019, Boosting businesses: applying behavioural insights to business policy, https://www.bi.team/wp-content/uploads/2019/10/BIT_Boosting-Businesses_Report_Final.pdf

Thomas, J. and Wendehorst, C., 2020, Response to the public consultation on "A European strategy for data" COM(2020) 66 final, https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Projects/Data_Economy/ELI_Response_European_Strategy_for_Data.pdf

UMATI brings together several standard from VDMA and OPC to enable machine to machine communication. See UMATI Website, https://umati.org/about/#vision

UZA, 2020, *Integrated HER's in Belgian hospitals, what about Antwerp, http://www.karva.be/wp-content/uploads/2019/09/Reinhart-Maertens-Integrated-EHR%E2%80%99s-in-Belgian-hospitals_20190830.pdf*

VDMA, 2017, Mechanical engineering demands a decade of innovation, https://bayern.vdma.org/en/viewer/-/v2article/render/16646489

Verbruggen, P., 2019, Tort Liability for Standards Development in the United States and the European Union, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3254864

Vijayaraghavan A, Dornfeld D, 2010, Automated energy monitoring of machine tools, https://core.ac.uk/download/pdf/191543802.pdf

Warren et all, 03/12/2019, *Working with patients and the public to design an electronic health record interface: a qualitative mixed-methods study*, https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-019-0993-7

World Economic Forum, 2020, White Paper - Share to Gain: Unlocking Data Value in Manufacturing, https://www.internationaldataspaces.org/wp-content/uploads/2020/02/WEF_Share_to_Gain_Report.pdf

World Economic Forum, July 2021, White Paper, The Resiliency Compass: Navigating Global Value Chain Disruption in an Age of Uncertainty, https://www.weforum.org/whitepapers/the-resiliency-compass-navigating-global-value-chain-disruption-in-an-age-of-uncertainty

Yeh Pang Toh et all, 01/01/2021, *Developping a Digital Twin and Digital Thread Framework for an "Industry 4.0" Shipyard*, https://www.researchgate.net/publication/348771602_Developing_a_Digital_Twin_and_Digital_Thread_Framework_for_an_'Industry_40'_Shipyard

Yu Cong et al, May 2019, OntoIMM: An ontology for product intelligent master model, https://www.mdpi.com/2076-3417/9/12/2553/htm

Zeid Abe et all, 11/02/2019, *Interoperability in Smart Manufacturing: research challenges,* https://www.mdpi.com/2075-1702/7/2/21/pdf

Zhumagambetov Rustam, April 2021, *Towards automated molecular search in drug space,* https://nur.nu.edu.kz/bitstream/handle/123456789/5495/Thesis%20-%20Rustam%20Zhumagambetov.pdf?sequence=1&isAllowed=y

ZVEI, 01/04/2017, *Example of the Asset Administration Shell for Industrie 4.0 components,* https://www.zvei.org/fileadmin/user_upload/Presse_und_Medien/Publikationen/2017/April/Asset_Administration_Shell/ZVEI_WP_Verwaltungschale_Englisch_Download_03.04.17.pdf

ANNEXES

ANNEX I 9 BUILDING BLOCKS FOR THE DEFINITION OF KEY COMPONENTS IN INDUSTRY AGREEMENTS - TEMPLATES AND EXAMPLES FOR THE DEVELOPMENT OF A STREAMLINED DATA SPACE

A series of opportunities for the development of impactful industry agreements and their economic impact have been analysed in chapter 2. In addition, the main barriers preventing industrial players from developing them were listed. This section provides the draft of 3 template agreements and 3 opportunity cases (based on chapter 2) where the development of these agreements can be first tested. This account for a total of **6** IA, which goes beyond the initial requirement of between 3 to 5 IA to develop. The approach is summarised in the graph below.

Figure 25: Business opportunities and IAs

	Data quality for advanced analytics and artificial Intelligence	Interoperability	Data Asset Value Exchange Mechanism
Area	• Smart Grid • AI for manufacturing	• Chemical Lab Robotics • Agricultural IoT • P&P machine tools • Digital Health Record systems	• Molecule representation • Smart Maintenance
Template IAs	Data Quality and Data Value Exchange	Ontology-driven data Documentation	Shared industrial Data Space
	• IA for translation software (mechanical Engineering) • IA for lab robotics • IA on chemical representation		

Source: CARSA, 2020

The 9 building blocks for the development of an industrial data space were already defined and clarified in Chapter 3 of this document. These building blocks cover all dimensions of the three key elements for the development of a common industrial data space: technical, legal/organisational or data specifications. As addressing all the building blocks in one industry agreement, a coherent set of 3 industry agreements were crafted. If adopted together these agreements tackle all of these 9 elements as summarised in table below:

Table 18: List of building blocks and what IA address them

Building Block	Description[376]	Category	Data Quality and Data Value Exchange	Ontology-driven data documentation	Shared industrial data space
Data Standards	Standards for static data: data structure, semantics, etc.	Technical Specifications	X	X	
Exchange protocol	Standards for data in transit: message formats, API specs, etc.				X
Identification & Authentication	Common practices and tools for identification and authentication of entities involved in initiative				X
Authorisation	Common practices and tools that entitled party uses for authorising other party access to his/her data				X
Metadata	Standards for the structure and semantics of metadata of data that is being shared through initiative	Data Specifications		X	
Business Model	Specification of the business model for both the initiative itself and participants/users of initiative	Legal and organisational Specifications	X (Data Asset Value Exchange Mechanism)		
Governance	Common governance structure that oversees operations, change management, disputes, etc.			X	X
Legal Agreements	Common agreements on all relevant legal matters such as liability, penalties, contracts, etc.		X	X	X
Operational Agreements	Common agreements on relevant operational procedures such as SLAs, change processes, etc.			X	

[376] Riemenschneider Rolf, 2020, *Principles & Architectures for Data Infrastructures*, European Commission DGCNECT.

I.1 Legal analysis for industry agreements

The deployment of new technologies and the development of digital value chains require also to consider the position of IAs in the search for long-time evolutionary compliance.

IAs as incubators for B2B data sharing

In its report ''*FinTech: Regulatory sandboxes and innovation hubs*'' three European regulatory bodies[377] recognise the continuous efforts of the EU on facilitating innovation. It is clearly shown that new technologies - technologies that are capable to learn, adapt and evolve - evoke the need to collect more relevant information, to better understand and to regulate these technologies in a different way. Granted the specific report encompasses the area of financial innovation, its conclusions are equally applicable to other types of innovations in various markets and in a variety of sectors.

Industry agreements are bi- or multi-lateral contractual frameworks, and as such, they are required to be in line with legislative provisions and case law, which lay out the boundaries of what stakeholders can regulate and under what conditions.

When it comes to unlocking new market opportunities at present or in the future, the necessary regulatory background due diligence may determine that the regulatory framework is overly restrictive and furthermore deter stakeholders from engaging in innovative co-operations (e.g., through high compliance costs or concerns about anti-competitive effects). At the same time, industry agreements provide an opportunity to shed light on the sector- and domain-specific contractual limits and affordances which are vital in determining the possible extent of integration and standardisation/interoperability which could be achieved in a sector without violating competition law or infringing third-party rights.

The IAs can help to reveal in a recurrent and future-proof manner what is regarded to be fair, acceptable and desirable when it comes to actual data transfers and dealings. Through an incubation function, IAs will help contextualise what fair, practical, clear and trustworthy data exchanges could mean in a given ecosystem.

On the other hand, the collaborative nature and the market development purpose of the industry agreements make them a valuable resource for regulators in order to establish mutually beneficial B2G relationships. The wide scope and application of the industry agreements render them useful as a source of information for different stakeholders, but especially regulators, on market trends and development, including the establishment of new markets, new technologies and business models, and the approach adopted by business towards the market.

Industry agreements as incubation processes

IAs can be an important source of information for the future-proof regulation of new technologies and applications. As the Commission envisages the creation of a European data space under common rules and efficient enforcement mechanisms – ensuring that personal data protection, consumer protection and competition are safeguarded – industry agreements can provide insights into the data practices by serving as incubation processes for the creation of voluntary B2B data sharing schemes.

[377] The three regulatory bodies being the European Securities and Markets Authority (ESMA), the European Banking Authority (EBA) and the European Insurance and Occupational Pensions Authority (EIOPA).

IAs as innovation facilitators

In order to ensure a balance between innovation and the law and at the same time to preserve the interests of both business and civil society, an anticipatory collaborative approach to regulation could be considered based on close interactions between the various stakeholders.[378]

Innovation hubs

So far, the two officially recognised categories of innovation facilitators are regulatory sandboxes and innovation hubs.[379] The main characteristics of the innovation hubs are providing a point of contact for companies with competent authorities and to seek non-binding guidance on the compliance of their products/services/business models. This form of innovation facilitator is particularly **beneficial** for SMEs who can establish a good contact with the respective regulator and also be incentivised to engage in industry agreements that are developed under the framework of the innovation hub. Such agreements have two major advantages: in the first place they could be synchronised with the regulator to avoid any compliance concerns; in the second place the regulator will have much better insights into the market and possible further regulatory needs.

Despite the benefits of integrating industry agreements incubators as a function to a traditional innovation hub, there are some **potential pitfalls** that need to be addressed. First and foremost, the traditional innovation hub currently offers a well-defined and established scheme for interacting with a regulator. This scheme has three functions: guaranteeing equal access but tailored to the specific needs of a company, creating transparency and predictability of the process, and mitigating the regulator's workload. All of the three need to be taken into consideration if industry agreements are introduced on the hub's level.

Another issue could exist in the context of industry agreements that need to involve entities from different jurisdictions and hence subject to different regulators. A possible opportunity is to position IAs in the context of the establishment of the extensive network/federation of innovation hubs on EU level that could offer such guidance. This solution, even though very challenging to achieve, offers an enormous added value for the single market.

Collective intelligence

The governments' need to better understand technology is at the core of the emerging new types of regulation, engaging numerous tools, most of which are still being developed and tested. In general, the widespread use of this new anticipatory regulation[380] in broad sense creates a better ground for innovation, the anticipatory tools can therefore indeed be considered as innovation facilitators.

One of the anticipatory regulation tools adopted by regulators is the so called 'collective intelligence' approach that utilises information, ideas, and insights from the citizens. The extremely wide scope of the industry agreements allows to evolve the collective intelligence tool and expand it in order to use input from companies. This would prove to be **valuable** with respect to regulations that concern the interest and rights of companies. The proposed expansion would satisfy all three main aims of utilisation of collective intelligence: better understanding of issues, solution-mining and improvement of decision making.

Possible risks that this tool involves are mainly related to the acquiring of a critical mass of industry agreements that could actually be used to generate new knowledge, the development of appropriate

[378] Parenti, R.(2020), *Regulatory Sandboxes and Innovation Hubs for FinTech*, Study for the committee on Economic and Monetary Affairs, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg.

[379] Parenti, R. (2020), *Regulatory Sandboxes and Innovation Hubs for FinTech*, Study for the committee on Economic and Monetary Affairs, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg; See also the joint report by the ESMA, EBA and EIOPA, 'Fintech: Regulatory sandboxes and innovation hubs', https://eba.europa.eu/sites/default/documents/files/documents/10180/2545547/154a7ccb-06de-4514-a1e3-0d063b5edb46/JC%202018%2074%20Joint%20Report%20on%20Regulatory%20Sandboxes%20and%20Innovation%20Hubs.pdf

[380] Anticipatory Regulation: 10 ways governments can better keep up with fast changing industries, https://www.nesta.org.uk/blog/anticipatory-regulation-10-waysgovernments-can-better-keep-up-with-fast-changing-industries/.

methodology to extract the knowledge and transparency concerns that could be related to competition law but also potentially other branches such as consumer law.

Behavioural insights

Even though this tool is usually associated with behavioural science and psychology that apply to human beings and thus not suitable for companies, there are new theories and research that argue that it is possible to apply this approach to non-human behaviour as well.[381]

The analysis of industry agreements by regulators could prove to be **extremely beneficial** for possible application of the Nudging theory[382] especially when the desired outcome might not be a top priority for the business such as corporate social responsibility or sustainability. Applying behavioural insights towards companies certainly has some differences compared to humans but there are common elements as well. For example, in both cases demonstrating 'deep impact' of the desired action/omission is a key component. This means that both humans and companies would be more likely to act/ not act in a certain manner if they believe a big percentage of people/ companies already are acting in the same manner.

As is the case with the collective intelligence tool, using industry agreements for the purpose of facilitating innovation will depend on the establishment of the right methodology, in this case the area of behavioural science in order to deduct what incentives would stimulate certain sectors to act in a way desired and promoted by the respective government. This tool could also be successfully implemented in combination with others such as testbeds, pilots or regulatory sandboxes which would allow more wholesome research and analysis of information from various sources.

IAs and innovation facilitating tools

The positioning and integration of IAs should be further explored in relation to the innovation facilitating tools as elicited above.

Types of industry agreements: legal classification

As explained in the First Interim Report, many types of cooperative initiatives between stakeholders are usually based on a certain contractual arrangement (i.e., contractual agreements or contracts).[383] Those arrangements can be bilateral or multilateral depending on the complexity of the contractual relationship and the network of the stakeholders involved.

Furthermore, cooperation between stakeholders in the framework of industry agreements may take the form of a rather simple single (or continuous) 'transaction' (e.g., data access agreement, IP licensing agreement, confidentiality agreement, consultancy agreement); of a continuous collaborative project (e.g., contract research agreement, collaborative agreement);[384] of a joint management and cooperation project consisting of a large consortium of stakeholders (e.g., consortia agreements); or even entail an establishment of a separate legal entity (e.g., joint venture agreement, agreement establishing a standard setting organisation or a

[381] Rubinstein H. *The Benefits of Applying Behavioural Science to Business*. In: Applying Behavioural Science to the Private Sector. Palgrave Pivot, Cham, 2018 see also Broughton, N. et al., *Boosting businesses: applying behavioural insights to business policy,* Report by The Behavioural Insights Team, 2019

[382] The Nudging theory involves sets of measures that could 'nudge' people into making a certain decision that that would lead to a desired outcome. A typical example involves measure that promote healthy choices.

[383] It should be specified that within the framework of this project the term industry agreements does not refer to a contractual agreement.

[384] For more information on the establishment of R&D partnerships consult Gorbatyuk, A. (2019) 'Rethinking Intellectual Property Ownership in the Context of Open Innovation', KU Leuven Ph.D. Thesis https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS2815296&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US&fromSitemap=1.

platform). The latter is an example of an equity-based industry agreement, the former is non-equity-based.[385] Non-equity-based contractual arrangements allow a higher level of autonomy and flexibility between stakeholders than the equity-based ones.[386] One cannot provide an exhaustive list of contractual arrangements that can be concluded between the stakeholders in the framework of industry agreements due to their complexity.

Below we will analyse essential parent clauses, with an emphasis on IP and IT clauses that many of the industry agreements may require when regulating their relationship contractually. Within a contractual setting, it should be noted that a comparable negotiating power can be essential for B2B data sharing agreements to result in balanced transactions and the creation of fair data spaces.

Core parent clauses

I. Data-related clauses[387]

Sample clause (SC) 1: data availability and quality – The contractual arrangement in the framework of industry agreements has to clearly describe which data is shared (e.g., customer data, diagnostic data) and what the relevant specs are (e.g., format, source and method of collection, legal basis, volume, location, date, etc.). Furthermore, the quality of the data that is provided under the agreement at the moment of contractual negotiations and over time needs to be stated. In particular, it needs to be specified whether the data will be updated, how often and which party is responsible. The source/origin of data needs to be mentioned and how that data was collected or constructed. In this respect, it needs to be specified whether datasets are (partially) composed of or contain public data. It is essential to ensure that rights of thirds parties to the data in question are respected and there are no legal obligations that may prevent the data access and exchange. Furthermore, the data protection legislation needs to be carefully consulted. In particular, among other legislative provisions, the shared data has to be in line with the GDPR provisions.

SC 2: data protection – In the case where personal data are processed in the course of the cooperation, parties in the agreement will have to regulate a number of data protection-related issues, depending on the allocation of responsibilities, to be identified *in concreto.* For example, in the case of a controller / processor relationship, the controller shall instruct the processor in writing about the conditions for the data processing activities. If the parties are found to be joint controllers, they should determine their respective responsibilities for compliance with the GDPR. In the case of complex and multilateral agreements, several different situations can be found with respect to the allocation of data protection law responsibilities (e.g., joint controllership). Sufficient attention should also be paid to the dynamic aspect of (non-)personal data: e.g., provide information with regards to anonymised data (such as an assessment of robustness of anonymisation techniques used) and with regards to potential inference of information on individuals from industrial data.

SC 3: data access – Contractual clauses have to clearly specify in a transparent, clear and understandable manner who has a right to access, right to (re-)use and distribute data and under which conditions. They should also provide information on whether (elements of) a dataset is (are) also available through other means, for instance as publicly available data. The conditions for data re-use and distribution have to be specified (SC 4). The right to access and (re-)use of data can be limited. For instance, it can be limited to a certain group or certain purposes of data use. To achieve this purpose, it should be considered to define different categories of data that are included in a data pool, depending on the access and re-use rights (SC 4) that apply to the specific data (e.g., data that is open and freely accessible to all, data of which access is restricted by law, data that is only accessible to designated entities, data that is accessible through API's, etcetera).

---

[385] Hagedoorn, 2002, Inter-firm R&D Partnerships: An Overview of Major Trends and Patterns since 1960', p. 478, https://doi.org/10.1016/S0048-7333(01)00120-2

[386] Niedergassel, B., 2011, *Knowledge Sharing in Research Collaborations,* Springer 2011, p. 36.

[387] The analysis of the subsection 'Data-related contractual clauses' are based on Commission Staff Working Document – Guidance on sharing the private sector data in the European data economy. (COM (2018) 232 final), p. 7 and on the research conducted in the framework of Horizon 2020 project Safe-Deed, D3.3. "Legal Requirements for Processing of Non-Personal Data", pp. 21-23, available at https://safe-deed.eu/wp-content/uploads/2020/06/Safe-DEED_D3_3.pdf.

SC 4: (re-)use of data – Contractual clauses should clearly specify what the (re-)user is allowed to do with the acquired data. The terms of data usage (the exact usage that can be made of the data and, if applicable, time limits on its use) have to be specified as clear and concrete as possible, including the rights on derivatives of the data. In the case of co-generated data, this could include the respective rights (e.g., access, porting, and right to desistance from further use) of the different parties that contribute to the creation of the data or have a stake in the data creation. If the receiving party does not follow the agreed upon terms of data (re-)use it may lead to a breach of contractual obligations and give the data supplier clear means to start a lawsuit, unless the parties resolve the conflict amicably.

SC 5: security measures and technical means for data access and exchange – The necessary IT security mechanisms should be in place to ensure that data can be accessed and exchanged efficiently. For this reason, the partners could consider establishing a separate (independent) body that would be responsible for monitoring the data access and exchange processes. Contractual parties should ensure that the shared data is protected from any foreseen and unforeseen circumstances, including theft, misuse, technical problem and human error. Failure to provide the necessary level of security measures may lead to liability concerns. Furthermore, if parties exchange trade secrets under the contractual obligations, it is essential that all parties install the necessary secrecy mechanisms to ensure that the trade secret protection is secured (SC 8 and 9).

II. IP-related clauses

SC 6: access to background IP – Within industry agreements, stakeholders may need to obtain access to background IP of one another. The access can be granted by concluding bilateral or multilateral IP licensing agreements or by incorporating a clause, governing access to background IP, into agreements of a broader scope (e.g., contract research agreements, collaboration agreements, and consortium agreements). The conditions on which access to background IP is to be provided can either be negotiated on an individual basis or standard conditions can be applied (e.g., open sources licenses).[388] Generally, the owner of shared background IP remains unchanged. An exception is an IP assignment agreement, which aims at 'selling' the IP. In IP licensing agreements or IP clauses, governing access to background IP, parties specify the details under which conditions the background IP may be used. In particular, in case of a license agreement, the scope of the access rights has to be carefully defined, e.g., whether the license is exclusive or non-exclusive or whether the IP at stake can be sublicensed. The territory, time and the field of use for which the license is given should also be specified.[389]

SC 7: access to foreground IP – Foreground IP consists of IP which is generated by stakeholders in the framework of joint collaboration projects (e.g., collaboration agreements, consortium agreements).[390] If the stakeholders generate new IP in the framework of industry agreements, it is essential that they contractually specify, first, how the ownership of jointly developed IP will be allocated, and, second, how the jointly developed IP will be exploited.[391]

III. Confidentiality and trade secrets-related clauses

SC 8: confidentiality – In the framework of industry agreements, stakeholders may decide to share or develop confidential information or even trade secrets (SC 9). To ensure that this valuable information is not disclosed

[388] https://opensource.org/licenses
[389] For more detailed information see European IP Helpdesk (2019) "Your Guide to IP and Contracts", available at https://www.iprhelpdesk.eu/sites/default/files/2018-12/european-ipr-helpdesk-your-guide-to-ip-and-contracts.pdf.
[390] COM, European Research Area Guidelines on Intellectual Property (IP) Management in International Research Collaboration Agreements between European and Non-European Partners, 2012, https://ec.europa.eu/research/innovation-union/pdf/international_cooperation_guidelines_erac_kt_group.pdf, p. 7.
[391] For more information on the issue of jointly developed IP in the framework of collaborative project, please consult Gorbatyuk, Arina (2020) '*The Allocation of Patent Ownership in R&D Partnerships: Default Rules v. Contractual Practices*', SCRIPTed: A Journal of Law, Technology and Society, Vol. 17; iss. 1; pp. 4 – 53.

to third parties, strict confidentiality clauses or confidentiality agreements (or non-disclosure agreements (NDAs)) must be provided.[392]

SC 9: trade secrecy - To be protected through trade secret protection, confidential information (including data), in essence, has to fulfil three requirements: (1) it has to be kept secret; (2) it has to have commercial value due to its secrecy; and (3) the trade secret holder has to undertake reasonable steps to keep the information secret.[393] If stakeholders intend to safeguard valuable information as trade secrets, a simple confidentiality clause may not be considered sufficient to satisfy the third criteria for protection. To satisfy the third criteria, apart from the general NDAs, the stakeholders could also implement physical access restrictions, firewalls, encryption or secure storage.[394]

IV. Other core clauses

SC 10: liability – In the context of industry agreements, parties may include a clause on liability provisions in the contract governing their relationship for instance for supplying erroneous data. If the requested data is not provided or it is not up to negotiated standards the receiving party may request the partners to pay damages and/or to hold them harmless from all losses, claims and expenses. This clause may be considered important by the clients of for instance a platform, since they may want to ensure that the data that is provided to them is of high (agreed upon) quality. Liability clauses may also be important with regards to limitations set on the (re-)use of data.

SC 11: parties' rights – In general, the scope of parties' rights and obligations should be clearly specified in the contract. If one of the obligations are not met by one of the parties it may give rise to a breach of contractual obligations and lead to a lawsuit, unless parties manage to amicably resolve the dispute.

SC 12: duration and termination of the contract – Parties to an agreement have to specify for how long the negotiated contractual obligations last and under which conditions the contract can be renewed. It is also important to specify under which conditions and circumstances a contract can be terminated prior to the termination date.

SC 13: (alternative) dispute resolution – It is important to specify in a separate contractual clause which law is applicable to the contract. The contract at stake will be governed by the selected applicable law. Furthermore, the parties should specify which dispute resolution mechanism is selected in case a conflict arises. The parties have a number of choices. First, the parties may decide to litigate in court. In this case, the jurisdiction has to be specified, to ensure that parties are aware in which country the dispute will be litigated. Second, instead of following the traditional litigation route, parties may decide to use the alternative dispute resolution (ADR) mechanisms, e.g., mediation and/or arbitration. The alternative mechanisms may be combined. For instance, one may first try to resolve a dispute through mediation and if the mediation process was unsuccessful, turn to arbitration. To ensure that the selection of the ADR institutions is efficient, parties may in advance specify which institution should facilitate the conflict resolution process. There are multiple ADR centres: national, regional, international. One of the renowned ADR centres is the International Chamber of Commerce (ICC).

I.2 Drafted industry agreements templates

This section addresses the main aspects to be considered in the development of industry agreements with high potential market impact. While the specific terms of the agreements may individually differ, it is still possible to define at high level, which are the main aspects that would need to be agreed upon so the IA becomes

[392] Halt, G., 2014, *Intellectual Property in Consumer Electronics, Software and Technology Startups*, Springer Science+Business Media 2014, p. 77.
[393] Art. 39(2) TRIPS; Art. 2 of the Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure (Text with EEA relevance) OJ L 157, 15.6.2016, p. 1–18.
[394] Steger, U., 2018, The secret to keeping secrets secret: The German Trade Secrets Act, https://paytechlaw.com/en/trade-secrets/?pdf=9792; Holzapfel, H. and Königs, M., 2018, New German Law on the Protection of Trade Secrets, https://www.mwe.com/insights/new-german-law-protection-trade-secrets/

successful in supporting the articulation and effective operation of data-driven digital value chains of strategic market importance. Therefore, this study has identified the main topics (high-level specifications) from a legal and technical perspective that need to be addressed by the IA to be considered complete and effective. The legal formulation of such clauses and the technical definitions of the areas identified are still subject to future research and/or specification. Nonetheless, the aim of the high-level specifications is to identify the core elements to be addressed in an IA to be considered effective and actionable in the context of the ecosystem, DVC vertical and/or DVC that is adopting it. We have classified such high-level specifications in two groups:

- ***Technical Clauses (TC)***: Techno-economic dimensions that DVC members need to agree upon.
- ***Sample Clauses (SC):*** That refer to the legal dimensions that need to be considered and agreed upon by the DVC.

Interviews and workshops conducted for this study revealed that there are key areas that the large majority of stakeholders (if not all the verticals) identified as main blocking areas to unveil the full potential of digital transformation across European industries. The IAs were scoped to impact information/knowledge sharing across industrial ecosystems and value chains, and hence, beyond individual companies' application of specific innovation areas (e.g., AI, autonomous things & robotics and data spaces). The objective is to avoid having IAs only impacting industries in a disconnected or "siloed" manner.

The project team has therefore selected three areas of key economic impact where developing industry agreements can accelerate the development of digital ecosystems and value chains. Failing to develop such IAs in the medium or long term would result in loss of competitiveness for the entire European industry. The areas identified in this report and the corresponding IAs are listed as follows:

- **Advanced data analytics and Artificial Intelligence —** Industry agreements on Data Quality and Data Value Exchange;
- **Plug & Play/Interoperability —** Industry agreements on Common Ontology-driven Data Documentation;
- **Data sharing/exchange —** Industry agreements on Shared Data Spaces.

Following the conclusions from the interviews and workshops, it appears that these three types of agreements would have to be developed with different levels of detail to leverage business opportunities by the verticals and associated ecosystems. Please note that the areas and their industry agreements are directly related to: 1) the need for data ecosystems (inherent to DVC) to gain trust in the relevance and appropriateness of the data used; and 2) the pertinence and alignment of the used data models with the supported business processes and with the partners engaged in the data exchange or data sharing. Overall, the return on investment should be sustained by agreed and well-known value exchange mechanisms, independently of individual business models. Developing independent business models is indeed an inherent need to gain a competitive advantage in DVCs.

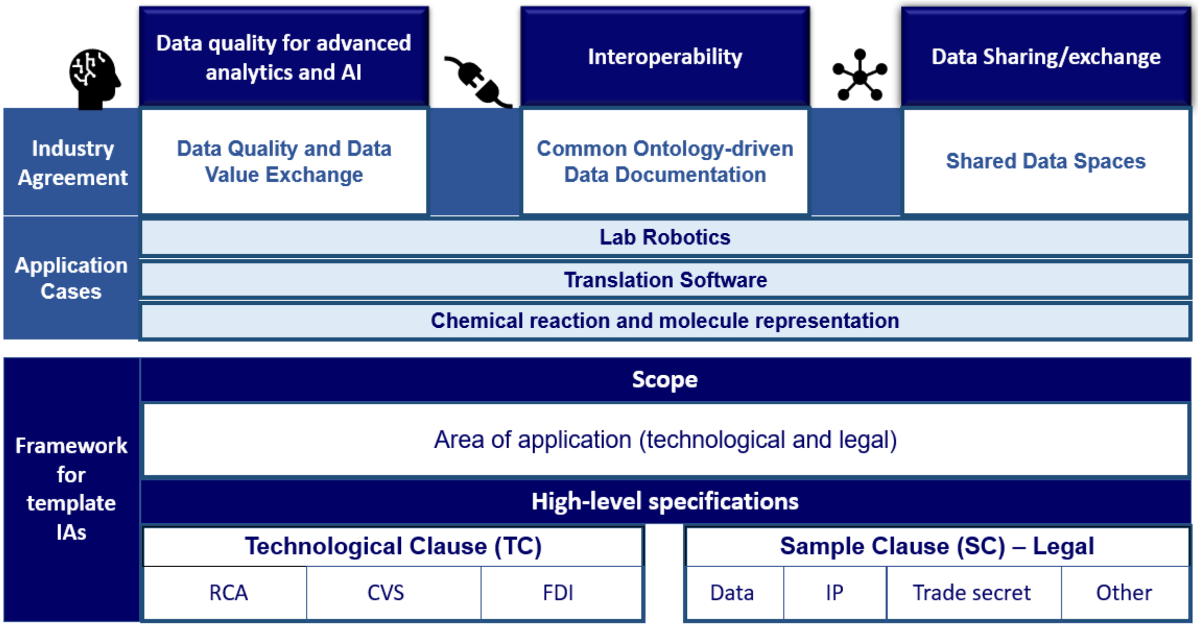
Consequently, each IA drafted in this section is developed based on the three dimensions identified and introduced in Chapter 1: (1) Relational contractual agreements (RCA); (2) Common vocabulary standards (CVS); and (3) Federated digital infrastructures (FDI).

Each of these agreements would individually contribute to the development and faster adoption of trusted AI, autonomous Plug & Play industrial things and common European data spaces.

In addition, when possible, example of specific industrial applications for these IAs were drafted as well to illustrate where this approach could be tested and where it could be impactful.

The graph below summarises the framework and the list of draft IAs and related examples.

Figure 26: Definition of draft industry agreements



Source: CARSA, 2020. This figure was created using resources from Flaticon.com.

Drafted IAs Innovation Area I – Development of trusted data sharing and data exchange practices

Shared data spaces and enhanced platform interconnectivity

Digital platforms offer abundant opportunities to improve industrial competitiveness and create value for the society. The development of trusted data sharing and data exchange practices aims at facilitating all activities in the digital supply chain. In addition, the introduction of digital platforms would remove various barriers across industrial sectors. For instance, cross-border barriers, which may prevent large-scale testing and block the deployment of new technologies (i.e., very relevant for the T&L and manufacturing sectors).

Development of a common standard platform solution for implementation and use in different settings/sectors could be seen as another opportunity to overcome barriers with data exchange and data sharing. Such common standard platforms could be used for sharing agriculture information as well as for sharing smart maintenance information for aerospace.

Technologies used in all industries (e.g., machine tools, agricultural equipment, and environmental sensors) are becoming part of the cloud or use IoT and M2M communications. This transformation of data also requires a new level of interoperability in order to make sure that all connected systems are able to share and exchange the data. Platforms can be interpreted broadly as agreements on functions and interfaces between industrial players, that create markets and market opportunities leading to ecosystems and standards.[395] To create such ecosystems the platforms must encompass reference architecture, interaction protocols and interoperability frameworks. Common standard platforms must therefore ensure an appropriate level of interoperability to facilitate the integration of digital industry technology platforms within and across the value chains.

395European Commission (2016), DEI Working Group 2. Strengthening Leadership in Digital Technologies and in Digital Industrial Platforms across Value Chains in all Sectors of the Economy. Available online: https://ec.europa.eu/futurium/en/system/files/ged/dei_working_group_2_platforms.pdf

To succeed in developing a common standard platform, ontologies and collection of vocabularies for industry are important for seamless data exchange. Semantic model development can provide a common vocabulary for semantic interoperability, which enables systems to manage and exchange data with common understanding by implicit and distributed meaning. Moreover, integrating industrial ontologies will support interconnection among entities, e.g., providing communication capability across various industrial domains. For instance, in manufacturing, it would mean increased communication opportunity between maintenance, production planning, supply chain management and product service systems.

To achieve interoperability for a common standard platform, the industrial players should not only agree on common IT languages, ontologies and collection of vocabularies, but also focus on its reference architectures (e.g., structures, building blocks and requirements), using common and standardised APIs and industrial data models. These aspects are key in developing trusted and enhanced platform interconnectivity. Furthermore, development of standard data components and data models are needed to abstract the differences between machine manufacturers, data sets, models and setups. Optimised inference data models (i.e., in manufacturing machine learning models) can collect specific needed real data to solve the problem, such as predicted failures in manufacturing sector. Lastly, definition of standards on what security processes to use to ensure protection and separation between the different layers is required. E.g., full encryption technology for first layer, distributed ledger for the second, etc.

The industry agreements in the development of data sharing and data exchange platforms are mainly related to the development and retention of trust among participants to the DVC and data space. Therefore, the industry agreements in this case should focus on aspects connected with the technical, operational and functional aspects that will govern the data space. The data space and the scope of the industry agreement should be viewed as a living organism in continuous evolution and therefore the IA is the basis for such organism to grow fit and strong or to die early and sick. Procedures and clear means to deal with inefficiencies in the DVC and data space are crucial, since the IA should cover the initial aspects for the DVC to start operations and adaptation, optimisations, enhancement and enrichment of protocols, procedures and tools must be expected.

Table 19: Scope and high-level specifications – Industry agreements on Shared Data Spaces

Industry agreements on Shared Data Spaces		
	Technological	**Legal**
Scope	• Standardisation of data sharing and exchanging platforms. • Cross-industrial IT language and standards	• Standard agreement of data sharing and exchange for users interested in obtaining data from or through the platform.
High-level Specification	***Technical Clauses (TC):*** TC 1 Functional agreements *(Definition of roles, essential services, additional roles, certification bodies for participation in the ecosystem, interactions).* TC 2 Authentication agreements (*Common practices and tools for identification and authentication of entities involved in the initiative*.) TC 3 Technical agreements (*Common Reference Architecture for data space, interoperability assessment & criteria)* TC 4 Operational agreements *(DVC Data*	***Sample Clauses (SC):*** SC 1 data availability and quality SC 2 data protection SC 3 data access SC 4 re-use of data SC 5 security measures and technical means for data exchange SC 6 access to background IP SC 7 access to foreground IP SC 10 liability SC 11 parties' rights

	Space governance model, shared operational processes including criteria for admission, certification, withdrawal, warnings, suspension, exclusion, incident management, change management, Service Level Agreements and Policies, release management, maintenance.) TC 5 Technical protection of data sovereignty agreement (*Practices and tools that entitle party uses for authorising other party access to proprietary data. This includes protection at the aggregated level and not only of individual data points.)* TC 6 Technical features for secure data portability. TC 7 Tracking of ownership rights. (*Practices and tools that enable to track ownership rights of data assets at every step of the data value chain, including once data are aggregated and exchanged.)*	SC 13 (alternative) dispute resolution ***Additional types of clauses:*** • A certification system - or any other similar system - should be set up in order to guarantee the identity and trustworthiness of the data providers and users. • The platform may also require an embedded payment system.
Points of attention	• IoT, Artificial Intelligence, robotics, cloud and Big Data needs to be reused and integrated with interfaces. • Common digital platform should aim for openness, avoiding lock-ins, preventing dominant positions of individual players and compliance with standards and regulation.	• Significant imbalance of power between the platform and the users may cause competition law and the (mainly national) regulation of B2B unfair commercial practices to be triggered. • It may also have a chilling effect on the willingness of industry players to share "their" data. • Governance structures should take into account the dynamic nature of data and be designed to accommodate evolutionary aspects or developments in terms of IP rights and personal data. • A data sharing platform constitutes an ecosystem and therefore implies a large number of contractual relationships. Attention should be paid to the overall consistency, since contracts may not be standalone.

Example of application: Standards for chemical reaction and chemical molecule representation

This section provides an example of concrete application of the general framework drafted for our industry agreement for the data sharing/exchange area. Drafting more specific IA can only be done by the players active in the industry. The example below should, consequently, only be seen as an illustration of what could be achieved in sectors/segments identified as especially interesting for Data Sharing/Exchange applications by the interviewed stakeholders.

Here we deepen in one particular data space that has already been defined and identified as having great business impact by a particular industrial community during one of the workshops hold in the context of this study. This should help to provide to the interested reader additional information whereas the lines of work would need to be further developed to substantiate the general technical, operation and functional TC discussed in previous pages.

As for other drafted industry agreements, the objective for the development of standards for chemical reaction and chemical molecule representation should be to favour cross-sectoral exchange whenever possible. For example, biotechnological companies and pharmaceutical companies can face similar challenges in their batch processes. Developing exchanges and cross-fertilisation on a cross-sectoral case would favour the emergence and expansion of common solutions instead of a fragmented digital landscape.

The table below presents the draft technological and legal aspects to consider to develop an industry agreement on standards for chemical molecule and reaction representation.

Table 20: Scope and high-level specifications – Standards for chemical molecule and reaction representation.

Data sharing and data exchange: Standards for chemical molecule and reaction representation.		
	Technological	**Legal**
Scope	• Standardisation of chemical reaction and chemical molecule representation	• Interoperability and standardisation of chemical molecule representation/ format. • IP rights.
High-level Specification	***The Technical Clauses (TC) listed here are aimed to complement the TC listed in the template IA.*** TC 1 Defining the type of notation *(Linear or ontology)* TC 2 Definition of the format for data storage and exchange *(e.g. mol., rgf., etc)* TC 3 Rules and requirements for portability TC 4 Differentiation rules *(what molecule, specificities do this standard aim at, what type and amount of information form the core of this IA)* TC 5 Metarules and metadata criteria for cross-sectorial retrievability and possibility to translate (*chemical sectors are vastly different and meta-rules for cross-sectorial usage should be define from the start)*	***Sample Clauses (SC)*:** SC 6 access to background IP SC 7 access to foreground IP SC 10 liability ***Additional clauses:*** • Agreements may set the rules in the area of chemical representation.
Points of attention	It is yet unclear if the definition of a new standard could address all needs in the Chemical data value chain. Special emphasis should be placed on the definition of compatible standards.	• Attention should be put on the potential for dual use: technologies used for civilian purposes vs technologies used for the military.

Drafted IAs Innovation Area II – Interoperable autonomous machines and systems (Plug & Play)

Common Ontology-driven Data Documentation: Shared Taxonomies and Models

The digital transformation of industrial sectors at European level is on one hand increasing the connectivity of systems and platforms. On the other hand, it is increasing complexity and autonomy in Things and Platforms (physical and digital), leading to a system of system scenario.

With increased connectivity comes increased benefits of economies of scale and reduced costs, but at the price of higher integration costs. The development of agreements in the area of data and system models is becoming increasingly more strategic in order to reduce such costs and increase the autonomy and automation of future digital systems (e.g., robotic systems or production facilities). Moreover, as highlighted during the workshops, there is an increased digital integration along the full lifecycle of products and processes with early attempts of digital thread implementation. The cost-effective implementation of the digital thread, connecting the data across the product and process lifecycle, demands industry agreements in the development of common taxonomies powered by shared semantics and ontologies.

The need for such agreements and the access to a common repository of outcomes for the benefit of industry (industry commons) is already identified as a key milestone for the solid development of digital transformation. Hence, in this innovation area, the focus is the need for industry agreements in common taxonomies and more importantly, related to how such taxonomies can be shared, re-used, improved and actioned.

This IA will then leverage additional opportunities in the area of semantic matching and robotic process automation which could greatly benefit from the foundational IA discussed in this section. In fact, the capabilities of robotic process automation, digital twin powered model-based service engineering or cognitive semantic mapping/matching, could profoundly benefit from a more general IA in the area of taxonomy sharing.

Table 21: Scope and high-level specifications – Industry agreements on Common Ontology-driven Data Documentation: Shared Taxonomies and Models

Industry agreements on Common Ontology-driven Data Documentation: Shared Taxonomies and Models		
	Technological	**Legal**
Scope	• Creation of DVC accepted taxonomies to address DVC operations. • Avoid fragmentation and foster cross-sectoral data sharing and exploitation.	• Access to third parties. • Rules pertaining to the adherence to certain standards.
High-level Specification	***Technical Clauses (TC):*** TC 1 Ontologies initially adopted by the DVC *(ontologies and data models initially used. Including a tracking system to monitor properties and ownership of standards to later ensure that adopted standards do not favour (a) specific player(s)).* TC 2 Procedure for classification and selection of ontologies and vocabularies relevant to the DVC *(ontologies and data models will have to be univocally linked to each specific process that the DVC would implement to allow scalable adoption and*	***Sample Clauses (SC):*** SC 6 access to background IP SC7 access to foreground IP SC 10 liability SC 11 parties' rights SC 13 alternative dispute resolution ***Additional types of clauses:*** • Specific provisions should define the conditions for third parties being granted access to these interoperable solutions. Such clauses would determine the rights

	implementation of models as business development rules. It will also have to be linked with existing semantics to ensure compatibility and interoperability). TC 3 Means agreed to access the shared ontologies and vocabularies relevant to the DVC *(Ontologies could be maintained and offered by free by the DVC ecosystem or they may not be freely available; e.g., ontology marketplace. FRAND terms (fair, reasonable, and non-discriminatory) to such shared ontologies need to be defined as part of the TCs).* TC 4 Procedure to update/change ontologies and models *(This would have on legacy systems and investments so backwards compatibility and DVC acceptance procedures should be clear).* TC 5 Procedures to assess performance of adopted standards. TC 6 Procedures to certify DVC taxonomy conformity *(the procedure for continuous assessment of data conformity to the shared ontologies for the DVC operation is defined here).* TC 7 Procedure to trigger the development or elimination of new ontology *(to avoid fragmentation the DVC should agree on the procedure to decide on the need for a new DVC ontology or deprecation of an obsolete one with no further use in the context of DVC operations).* TC 8 Vocabulary and taxonomy governance model *(the DVC should agree upon the individuals, institutions or stakeholders that will be responsible for taking care of specific domain or subdomain ontologies, ensure consistency across domain and subdomains and act as a communication platform between the different actors involved in the data space. The governance model adopted must also ensure the new standards guarantee an even playing field and does not create any market asymmetries).*	and obligations of such third parties. • Agreements could also specify the rules for adhering to one standard or another. • Compatibility strategies can also be regulated contractually, i.e., whether standards should ensure retro-compatibility with other existing standards.
Points of attention	The shared taxonomy does not impose the taxonomies and/or data models to be used	• Contractual agreements cannot cover every domain. (a) On the one hand, i.e.,

	individually by the companies in their internal processes. As part of a DVC, companies could opt to adopt shared taxonomies as part of their digital data and information/knowledge management processes. However, the IA simply establishes the terms in which collaboration and data models will be established in the context of the DVC processes supported. Obviously, companies in the DVC would have to fix the means to adapt their assets to the models and taxonomies agreed.	certain safety critical fields are covered by a stringent set of regulatory frameworks which do establish or rely on certain standards. Compliance with such frameworks is mandatory and cannot be circumvented by agreements. (b) On the other hand, further statutory law may be required so as to set up relevant standards, which industry actors may not be able to set up on their own. • The joint development of standards may give rise to competition law issues, i.e., in case developing parties deny access to the standard to competitors.

Example of applications: translation software and lab robotics

This section provides two examples of concrete applications of the general framework drafted for our industry agreement for the Plug & Play/interoperability area. The examples of **translation software** and **laboratory robotics** were developed as cases of “legal/technical sandboxing” where the template IA can be first developed and implemented using the pathway/guidelines given in the opportunity analysis provided in the Chapter 2.

Standards on alignment and localisation rules for translation software for Plug & Play machine tools

Translation software is moving away from the old paradigm of developing new standards for new problems. These software are developed to make different standards compatible. Therefore, old and upcoming standards would be made compatible without having to add new standards to the existing mix.

However, worrying signs already indicate that translation software is slowly taking the same paths of failed standardisation initiatives of the past. Bridging different ontologies is a complex task that requires a series of localisation rules (identify and map the ontologies that the software can translate). So far, there are no clear agreements on what are these rules or how to rank ontologies. [396] In addition, a translation software — no matter if it’s about translating human languages or machine languages — needs to define alignment rules. These rules establish how an entity in one ontology can be matched with another item in a different ontology. However: “*One of more common operations on ontologies, regardless of their application, is an alignment - a matching between entities from different ontologies. So far, many alignment formats and languages have been proposed, some of them for general use and some particular to a concrete implementation of an alignment software.”*[397]

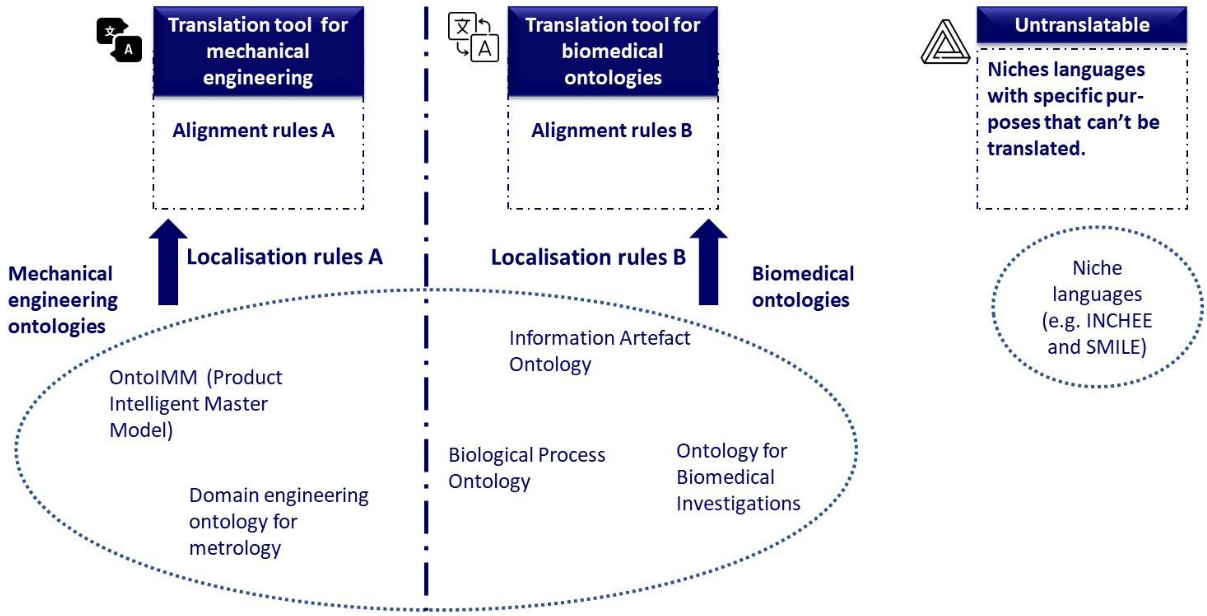
Several tools already exist in this field, for example SAMBO for biomedical ontologies alignment, OPTIMA as a general-purpose alignment tool and AgreementMaker that can map large number of ontologies.[398]

Machine translation is a promising and rapidly growing field that can also cross-fertilise, taking inspiration from the rapidly developing software for human languages translation. The technology is still at a rather young stage for machine communication. In addition, some technical problems remain. For example, alignment is a

[396] Mejía Jorge, *Ontology* Localization, PhD Thesis, 2014, Saragosa University, https://zaguan.unizar.es/record/15513/files/TESIS-2014-058.pdf
[397] Szmeja Pawel et al, *Declarative ontology alignment format for semantic translation*, February 2018, https://ieeexplore.ieee.org/document/8519921
[398] Chondrogiannis Efthymios, Sin dato, *An intelligent ontology alignment tool dealing with complicated mismatches*, http://ceur-ws.org/Vol-1320/paper_16.pdf

complex task and mismatches is an important hurdle that is difficult to address. This is currently being explored using machine learning and other AI-enhanced processes.[399]

Figure 27: Illustration of translation software landscape



Source: CARSA, 2020. The listed ontologies are for illustrative purposes only. They were listed based on interviews and the two sources listed in footnote.[400] This figure was created using resources from Flaticon.com.

As illustrated in the figure above, translation software will not be able to consolidate the entire ontology landscape. **Alignment rules have limits:** some ontologies won't be matched at all as it would imply losing too much of the information that these specific languages can carry. Some ontologies are simply too different as they are applied to very specific fields. Matching them would be too difficult at the present stage of the technology's development and the results would not make any sense.

Nonetheless, it is important to avoid that translation software become yet **another fragmented layer on top of the existing standard landscape**. Translation solutions could **be harmonised within industrial segments and sub-segments.** As already demonstrated, the main gain that industry can expect from greater interoperability is the simplification of system integration and the gain in operational efficiency. While this software would provide greater efficiency no matter what, a fragmented landscape of translation software would **make the standard landscape even more complex**. System integrators would have a new business opportunity, being subcontracted to develop tailored translation solutions that would be expensive, complex, and potentially inaccessible to smaller players. Integration would progressively become more complex as new translation solutions appears and even — in a distant and dystopian future — we can imagine to see translation software created to translate between different translation software.

In conclusion, **translation software holds great promises, but to make them a reality, industry agreements are needed**. The following table outlines a draft list of technological and legal aspects to be considered for the development of such industry agreements.

[399] Chondrogiannis Efthymios, Sin dato, *An intelligent ontology alignment tool dealing with complicated mismatches*, http://ceur-ws.org/Vol-1320/paper_16.pdf

[400] For the listed ontologies, see Stojadinovic Slavenko et al, October 2013, *Developing Engineering Ontology for Domain Coordinate Metrology*, https://www.mas.bg.ac.rs/_media/istrazivanje/fme/vol42/3/12_sstojadinovic.pdf, Yu Cong et al, May 2019, *OntoIMM: An ontology for product intelligent master model*, https://www.mdpi.com/2076-3417/9/12/2553/htm and National Centre for Ontological Research, *November 2017*, *The industry ontology foundry*, https://emmc.info/wp-content/uploads/2017/12/EMMC-IntOp2017-Cambridge_Smith_Buffalo.pdf. SMILE and INCHEE were mentioned as currently impossible to translate during a virtual workshop on the chemical industry.

Table 22: Scope and high-level specifications – Interoperability and translation solutions for interoperable vocabularies.

	Interoperability and translation solutions for interoperable vocabularies.	
	Technological	**Legal**
Scope	• Creation of an arena to discuss and harmonise the emerging translation software within the different sectors. • The objective should be to avoid the fragmentation and have different translation software and alignment languages adding a new layer of complexity.	• Access to third parties. • Rules pertaining to the adherence to certain standards.
High-level Specification	***The Technical Clauses (TC) listed here are aimed to complement the TC listed in the template IA.*** TC 1 Development/choosing the vendor-neutral ontology that will serve as neutral reference (*serves as neutral reference for semantic alignment and transformation)* TC 2 Rules for data ingestion to the data model transformation software (*connectivity and feeding format from the library of ontologies and standards)* TC 3 Rules for syntactic transformation and harmonisation TC 4 Rules for semantic transformation (*terminological method for terms comparison, structural method for similarities calculation, extension method and ontology matching)* TC 5 Definition of neutral format for final output	***Sample Clauses (SC):*** SC 6 access to background IP SC7 access to foreground IP SC 10 liability SC 11 parties' rights SC 13 alternative dispute resolution ***Additional types of clauses:*** • Specific provisions could define the conditions for third parties being granted access to these interoperable solutions. Such clauses would determine the rights and obligations of such third parties. • Agreements could also specify the rules for adhering to one standard or another. • Compatibility strategies can also be regulated contractually, i.e., whether standards should ensure retro-compatibility with other existing standards.
Points of attention	• Identification of niche languages that cannot be translated (ontology mapping). • Development of agreements and framework for localisation rules by segment and sub-segment (one-fit all approach is impossible).	• As explained in this section's introduction, contractual agreements cannot cover every domain. (a) On the one hand, i.e., certain safety critical fields are covered by a stringent set of regulatory frameworks which do establish or rely on certain standards. Compliance with such frameworks is mandatory and cannot be circumvented by agreements. (b) On the other hand, further statutory law may be required so as to set up relevant standards, which industry actors may not be able to set up on their own. • The joint development of standards may give rise to competition law issues, i.e., in case

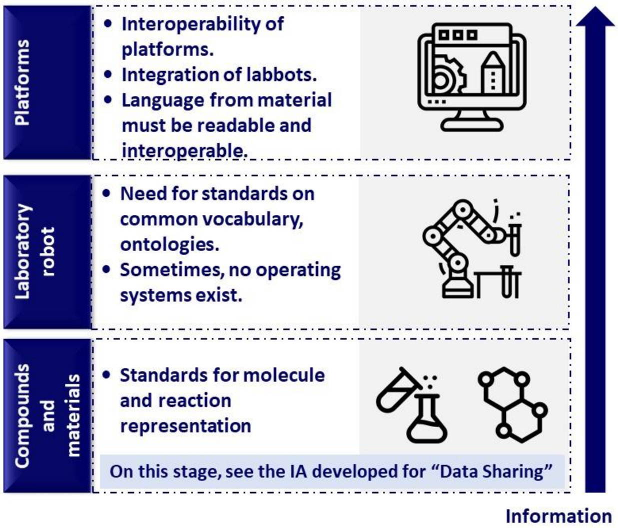
		developing parties deny access to the standard to competitors.

Standards (ontologies and semantics) for lab robotics

For laboratory robots (labbots) to process the information correctly, **standards on compounds, materials and reactions are needed**. Otherwise, the robot cannot make the link between the material and the digital platform. The platform then also needs to be **standardised** to be able to operate the robot, gather and process the information.

As we can see, developing standards for labbots is not an easy task. Initiatives are starting to appear. However, these are focused on specific segments or problems specific to a subset of players. Developing a more encompassing approach bringing together all different players is something that can only be achieved through the support from policy-initiatives. It would bring great benefits to different sectors, such as agronomy, biomedical and chemistry, or any other industry that uses lab robotics. The dimension related to **standard on molecule and reaction representation** is a specific topic that would require a complementary IA, something that is detailed in the example developed in the data sharing section.

Figure 28: Key considerations linked to an IA on Lab robotics



Source: CARSA, 2020. This figure was created using resources from Flaticon.com.

Table 23: Scope and high-level specifications – Lab R&D robotics

Common languages for Lab Robotics		
	Technological	**Legal**
Scope	• Standardisation of chemical reaction and chemical molecule representation	• Agreements on the computer language used by the stakeholders and cross industrial standards.
High-level Specification	***The Technical Clauses (TC) listed here are aimed to complement the TC listed in the template IA.*** TC 1 Rules and techniques to connect robotics/platforms with chemical compound standards (*mapping of existing standards for chemical molecule representation, standards for common machine-readable formats and rules)* TC 2 Ontology development (*definition of a common ontology for robots and platforms*	***Sample Clauses (SC)***: SC 6 access to background IP SC 7 access to foreground IP SC 10 liability SC 11 parties' rights SC 13 (alternative) dispute resolution ***Additional types of clauses:*** • In the absence of common vocabularies and languages for lab robotics, agreements

	in lab robotics) TC 3 Semantics for ChemBots (*definition of a common semantic for robots and platforms in lab robotics)* TC 4 Alignment rules for standards in cognitive factory (*developed ontologies and semantics must be aligned with other processes in the factory, including protocols and machine ontologies)* TC 5 Standards for Electronic Lab Notebooks (*mapping and connectivity solutions with notebooks and rest of the lab equipments)* TC 6 Data threading	could specify the chosen computer language on which stakeholders agree to adhere. • When developing common interfaces, the different stakeholders may agree on the standards and language while assuring that the rights and obligations are clearly defined. Attention must also be brought to intellectual property issues.
Points of attention	It is yet unclear if the definition of a new standard could address all needs in the Chemical data value chain. Special emphasis should be placed on the definition of compatible standards.	• Attention must however be given to the fact that the subject-matter ('Lab R&D robotics') may have to do with several industry sectors and thus with several sector-specific legal frameworks.

Drafted IAs Innovation Area III – Data Quality and Data Value Exchange

Although advanced data analysis, such as machine learning, are already used in many fields, the increased uptake of these and other data intensive technologies have the potential to transform the future value chain of several sectors. For instance, as indicated in Chapter 2, energy and manufacturing are very well positioned to benefit from AI through a series of industrial purposes such as maintenance, quality assurance and process optimisation. However, as highlighted also during the workshops, one of the key barriers to adoption and deployment of these technologies across the value chain is that, in many cases, companies cannot concretely benefit from the huge amount of data available because datasets are "poor in quality or bad", e.g., incorrectly labelled, corrupted or incomplete. For instance, when these datasets are fed into AI solutions, predictive models fail to provide any result, or worse, they can provide erroneous output, e.g., in terms of false predictions. In addition, the required data may be protected by IP and cannot be used without authorisation of the rights holder, which can significantly complicate the 'training' of AI.

To be efficient and reliable, these technologies must be fed with high quality data in terms of accuracy, completeness, relevance, consistency, relevancy and timeliness. In the era of digitalisation, an improved data ecosystem would have far-reaching impact on potentially all industrial activities. However, producing and collecting high-quality data still remains a hot and often complex topic. For this reason, industry agreements that can help companies improve data quality are needed.

To achieve such an agreement, the following technological aspects need to be considered (customised according to the technologies and the sector at stake):

- Development of Data Quality (DQ) Assessment models to better manage and characterise the quality of datasets needed to train and test technologies capabilities. In particular, there is a need to develop models capable of identifying whether the data within the datasets are:
  - Duplicated, counterfeit or stolen;
  - Incomplete;
  - Corrupted or broken;

- Defining common specifications to assess whether the data elements to feed systems can be interpreted correctly;
- Developing a cross-platform metadata management technology as data lineage solution. It will allow to collect and store metadata coming from different analytic platforms in a single, central repository by making available data to everyone. The platform should be able to:
  - Locate all data regardless of where it resides;
  - To use both structured and variably structured data.[401]
- Defining common Data Quality Rules appropriate in order to analyse and evaluate the quality of specific data sets. It will allow to clearly understand if the data set fits the defined requirements; [402]
- Developing common Standard Quality-Assurance Methods that organisations can use for ensuring accuracy and consistency of high-quality training data sets;[403]
- Defining new forms of IT languages. To have a common terminology used by all stakeholders, can enable clear communication and sound decision making. To achieve this, a new approach is needed as language should be developed in a collaborative manner, involving most players of the same value chain. This strategy would allow to move from the development of problem-driven languages to industry-driven languages and standards;
- Defining a common framework to develop interoperable or translatable language. In particular, semantic interoperability is necessary in AI systems development. The semantic interoperability is needed for correct and valid machine interpretation and logic, e.g., to make automated predictions and recommendations.

These general technological elements translate in the following scope and specifications.

Table 24: Scope and high-level specifications – Industry agreements on Data Quality and Data Value Exchange

Industry agreements on Data Quality and Data Value Exchange		
	Technological	**Legal**
Scope	• Definition of common standards and processes to increase quality of data	• Agreements that aim at increasing the quality of data in technical and legal sense (e.g., data ethics considerations), safety of intended functionality.
High-level Specification	***Technical Clauses (TC):*** TC1 Data Asset nature and features to contribute to the DVC *(data assets of interest to the DVC)* TC 2 Data Quality (DQ) framework applied by the DVC *(mechanisms to characterise the obligations related to ensure the “quality” of datasets*). TC 3 Standards & methods adopted by the DVC for Data Quality Assurance (QA) and Conformity Assessment (CA) *(mechanisms*	***Sample Clauses (SC)*:** SC 1 data availability and quality SC 2 data protection SC 3 data access SC 4 re-use of data SC 5 security measures and technical means for data exchange SC 6 access to background IP SC 10 liability

[401] Claudia Imhoff, 2018, Advanced Data Lineage: The #1 Key to Removing the Chaos in Modern Analytical Environments, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE36VUW
[402] Medium, 2017, Data quality in the era of Artificial intelligence, https://medium.com/innovation-machine/data-quality-in-the-era-of-a-i-d8e398a91bef
[403] Medium, 2019, How to ensure data quality for machine learning and AI projects?, https://medium.com/vsinghbisen/how-to-ensure-data-quality-for-machine-learning-and-ai-projects-c8af1fe18c57

	to collectively ensure the high quality and accuracy of data and its continuous consistency- maintenance). TC 4 Industry Commons (IC) metadata models adopted by the DVC *(management technology as data lineage solution for cross-sectoral data sharing – DVC2DVC data sharing).* TC 5 Common, sectoral and shared industrial ontologies & semantics adopted by the DVC *(agreed data models and vocabularies to be respected for publication and sharing of data across the DVC).* TC 6 Data Asset Value Exchange Mechanisms Adopted by the DVC *(framework defining the long-term relationship of the DVC members and the selected monetisation schemes adopted – data licensing, monetary transactions, altruism…. supporting the cost for data quality processes).* TC 7 Data Quality Certification Framework (*systematic approach and framework to assess data quality).*	SC 13 (alternative) dispute resolution ***Additional types of clauses***: • Defining a sociotechnical concept of learning assurance as a counterpart to safety assurance comprising design and development assurances (i.e., the idea being to be able to 'containerise' the learning process itself in machine learning systems). • Defining industry- and case-specific fairness metrics. • Formulate and describe (non-exhaustive) ethical principles and the way in which these principles should be safeguarded, as well as a process for updating these principles according to dependencies (e.g., social and environmental changes). • Defining industry and case-specific transparency and descriptive criteria (i.e., how detailed an explanation is, and the extent to which an explanation relates to established legal standards, such as that of a reasonable person or an expert in the field). • Defining minimum uniform legal requirements that ensure a coherent understanding by an expert in the respective field of what is inside a dataset and what its best use cases could be.
Points of attention	The generation, maintenance and "servicing" of quality data is an expensive task/process. The industry agreements for Quality Data must deeply address the RCA dimension not only the CVS dimension. Clear and agreed value exchange mechanisms are a fundamental part of the IA; be those monetary transaction or "data" licensing schemes. IA and ecosystems should develop suitable dynamics and agreements in the forms of "data asset" exchange & mechanisms. As indicated in the first interim report DVC allow for long term relationships and data-driven alliances and partnerships.	There could be certain limitations when the agreements encompass two or more sectors due to the different sectoral level of regulation. Thus, performance of regulatory due diligence is highly advisable in order to identify any difference between legal obligations of the parties and thus limitations of the scope of a potential agreement.

ANNEX II VALIDATION WORKSHOP ON INDUSTRY AGREEMENTS IN DIGITAL VALUE CHAINS

II.1 List of participants

Table 25: List of participants - Validation workshop | 11.02.2021

#	Name	Organisation Name
1	Alexei Lapkin	University of Cambridge
2	Ana Garcia Robles	Big Data Value Association (BDVA)/Data, AI and Robotics (DAIRO)
3	Andreas Nettsträter	Fraunhofer Institute for Material Flow and Logistics; DigitalHubLogistics
4	Angelo Marguglio	Engineering Ingegneria Informatica Spa, Smart Industry and Agrifood; Industry and Security Technologies, Research and Innovation (IS3) Lab
5	Anthony Staines	Dublin City University, School of Nursing, Psychotherapy and Community Health
6	Carlos Montalvo	TNO-Strategic Analysis and Policy
7	Charlotte Ducuing	KU Leuven
8	Cosmas Vamvalis	ATLANTIS Engineering S.A., EFNMS
9	Davide Dalle Carbonare	Engineering Ingegneria Informatica Spa
10	Dimitris Kiritsis	EPFL (Ecole Polytechnique Fédérale de Lausanne)
11	Emre Bayamlioglu	KU Leuven
12	Ingo Martens	Hanse Aerospace
13	Ioanna Pagoni	Transportation and Decision Making Laboratory (TRANSDEM), University of Aegean
14	Ivo Emanuilov	KU Leuven
15	Jan Debruyne	KU Leuven
16	Jelle Hoedemakers	Agoria, Regulatory & Standardisation expertise centre
17	Katerina Yordanova	KU Leuven
18	Marina Cugurra	ETA
19	Michela Magas	Industry Commons Foundation
20	Olia Kanevskaia	KU Leuven-CiTiP-imec
21	Panos Ilias	ILVO, Flanders Research Institute for Agriculture, Fisheries and Food
22	Patrick Courtney	SILA Standard
23	Ray Walshe	Dublin City University
24	Richard Stevens	International Data Corporation (IDC)
25	Ulrich Seldeslachts	LSEC, International IT - & Information Security cluster
26	Uwe Seidel	VDI/VDE/IT
27	Vivian Kiousi	Intrasoft International SA; BDVA
28	Zoi Kolitsi	the European Institute for Innovation through Health Data (i~HD); Information Technology Laboratory, Aristotle University of Thessaloniki

II.2 Summary of outcomes

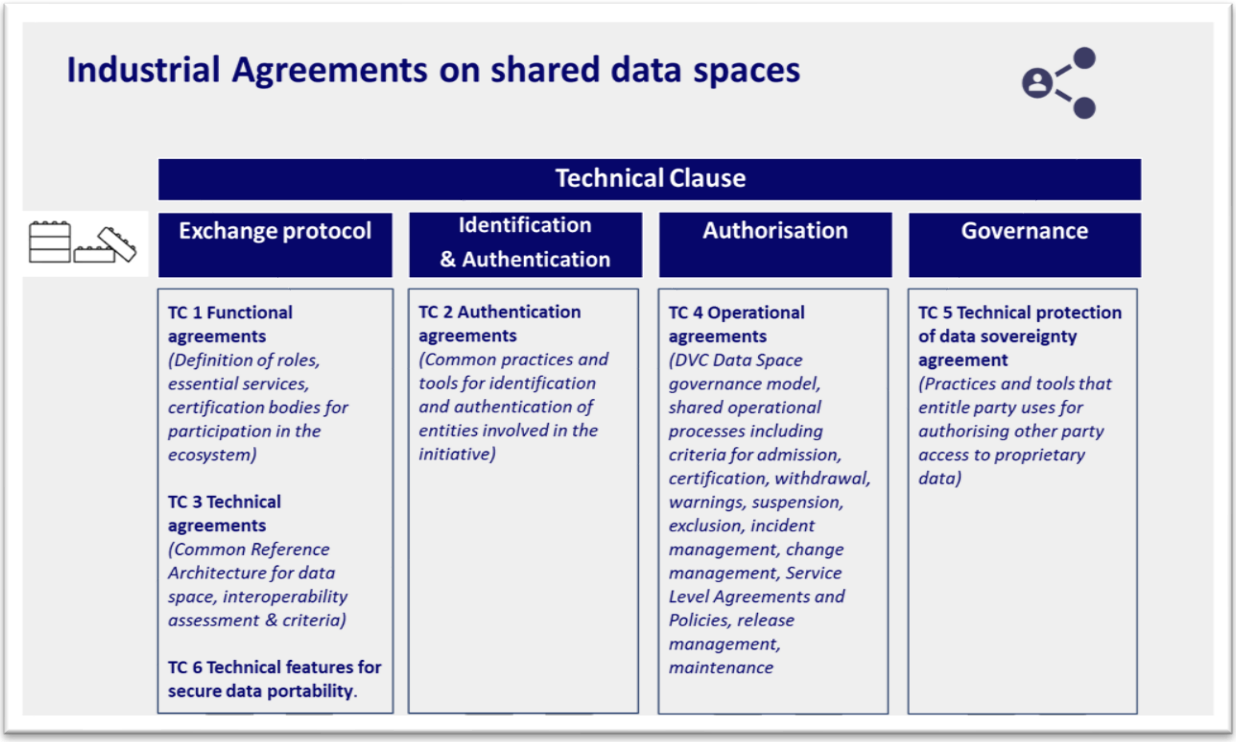
A validation workshop was organised in February 2021, to discuss and validate the preliminary results of the study, including the three draft industry agreement templates. Thanks to a series of polling and short open discussion sessions, the Consortium collected insights and feedback to validate the completeness of the proposed IA templates. The main conclusions are presented below. Legal and technical dimensions are presented separately for each industry agreement.

Data sharing and exchange – Industry agreements on Shared Data Spaces

Technical part

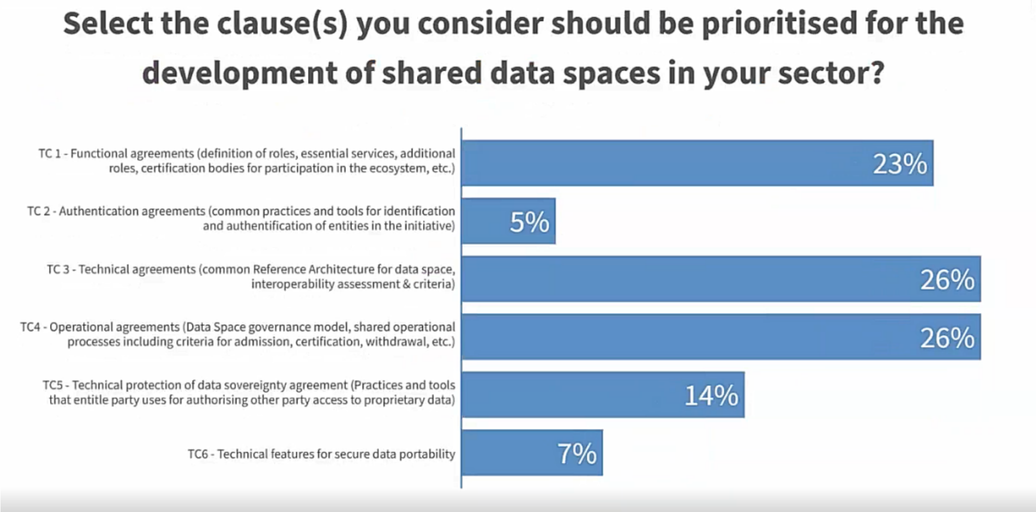
The technical clauses (TCs) for the industry agreements on shared data spaces were presented to the participants and are summarised in the figure below.

Figure 29: IA template on Shared data spaces – Technical clauses



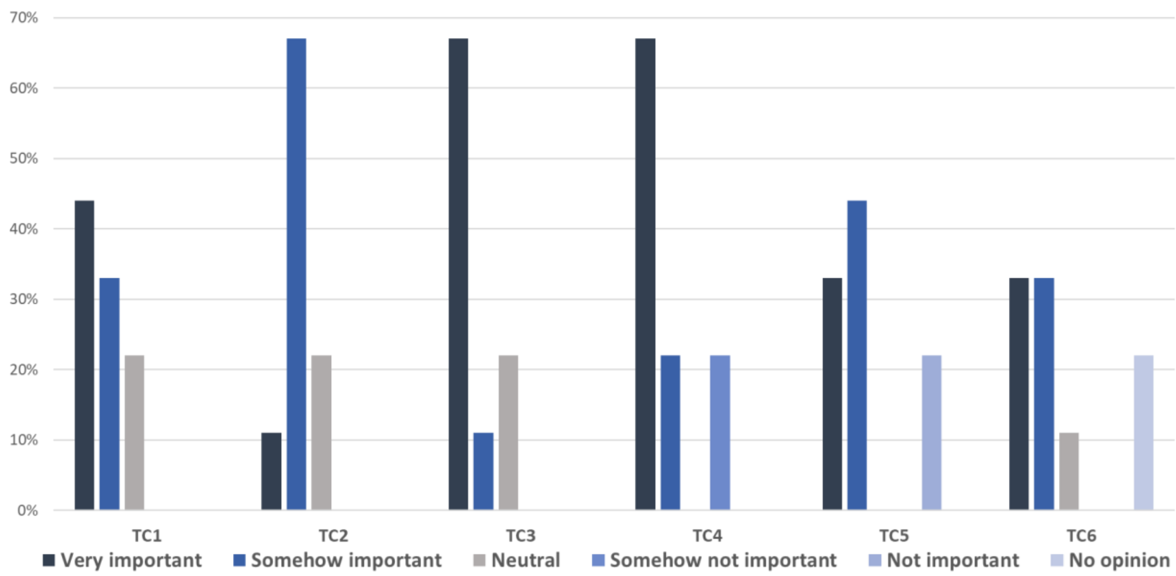
As shown in the figure below, participants identified Technical (TC3), Operational (TC4) and Functional Agreements (TC1), as the three main clauses that should be prioritised for the development of shared data spaces.

Figure 30: Polling on Shared data spaces – prioritised technical clauses



Results that are also confirmed in the bar graph below, which depicts the importance of the different clauses, according to responses to the online survey launched just after the workshop:

Figure 31: Online survey results on Shared data spaces



N=9

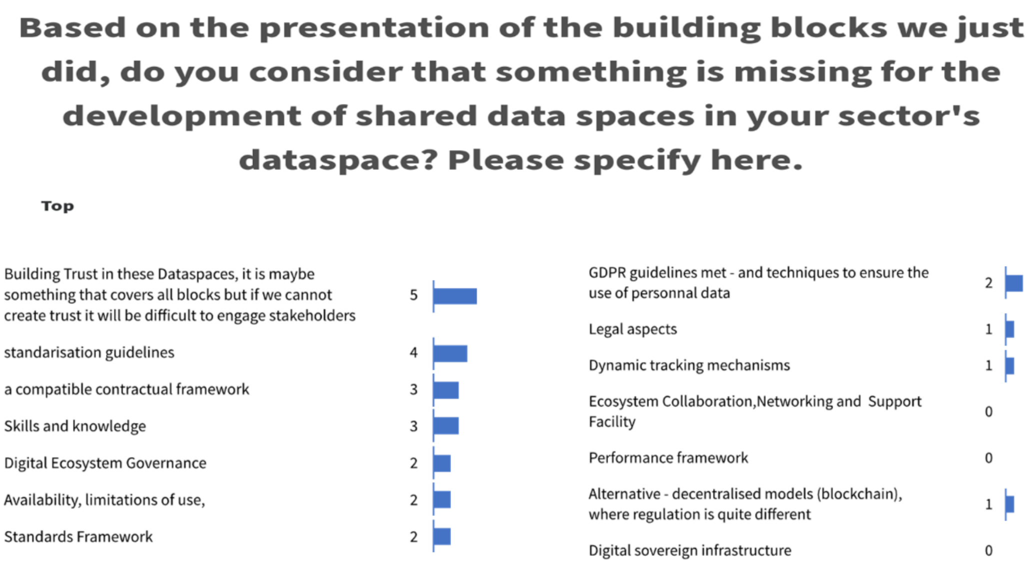
The open discussion led to the identification of several general conclusions, including the following:

- Participants stressed that for stakeholders' security, reference architecture and interoperability are three important aspects to be considered when building technical agreements (TC3). Indeed, it has been observed that in several European Projects and initiatives the levels of maturity and, above all, the interoperability of platforms and architectures, are still low.
- TC 4 (operational agreements, which include governance model, shared operational processes, certification, etc.) was ranked with equal priority as TC 3. It is important to establish trust for data re-use and also to clarify the scope of secondary usage.

- Besides the high importance of TCs 3 and 4, participants indicated that functional agreements (TC1) are also important. When drafting IAs definition of roles, essential services and certification bodies for participation in the ecosystem are very important aspects to be considered.
- TC5 Technical protection of data sovereignty agreements was seen as another valid TC. One participant highlighted that digital sovereignty infrastructure is still missing. Soft infrastructure for decentralised data sharing should be included when boosting data sovereignty.
- Participants stressed out that aspects such as Governance framework, collaboration, networking and support are essential but still missing elements for the development of shared data spaces.

Finally, the participants provided information on missing aspects that had not been taken into account in the drafting of the sector agreement. According to them, building trust is crucial for engaging stakeholders and ensuring the "health" of the data space, followed by (i) developing standardisation guidelines, (ii) strengthening skills and knowledge, and (iii) achieving a compatible contractual framework. Concerning the latter aspect, the participants stressed that in order to carry out a data transmission, it is necessary to agree on a contractual framework. In other words, it is necessary to develop an agreement that is flexible and abstract enough to cover all the different needs that could arise in an industry data ecosystem, together with the definition of adequate mechanisms to address any third-party interventions.

Figure 32: Polling on Shared data spaces – missing building block aspects



Legal part

The sample clauses on industry agreements on Shared Data Spaces were presented to participants, with a main focus on data availability and quality, data protection, data access and re-use of data. Participants were subsequently presented with three open questions;

- "***Within the context of data sharing & exchange, are you confronted with any other legal concerns that you do not see reflected in the template industry agreements?***"

Responses	Upvotes	Downvotes
Mix of "public" and "privately owned" data. Use of such combinations of data should be made clear in IAs.	2	1
Other aspects could be added like: data generation, storage and analytics (use by third parties)	1	0
Responsible data	0	0
Specific use-case related concerns (e.g. the same data can be of high quality in one use case and low in another, even with regard to data provenance)	0	0

- ***"Enforcing data-related contractual clauses (in particular restrictions on data reuse) can prove difficult, which may have a chilling effect on data sharing and collaborations. Could technical reference architectures help enforce such clauses and how?"***

Responses	Upvotes	Downvotes
Yes. Data Security, Immutable Storage, PII, Levels of Sharing, Permission to share, Personal Identifiable Information Tiers,	6	0
the description of the relationships between the different actors (who provides what / who does what) within an ecosystem can be also reflected also in the architecture	3	0
TRAs could include a time limit to the use of a particular set of data.	3	0
flexibility, granularity and alignment with the existing laws and contractual preferences	0	0
the RFA can give an indication of structure, functions, roles and resposibilities of involved parties. But is has limits as the industrial legal contracts must be more detailed, outside the scope of the RAF	0	0

- ***"In your experience, what are the main limitations to using contracts regarding data sharing & exchange? Have you any concrete examples?"***

Responses	Upvotes	Downvotes
Lack of TRUST. "How my data are going to be used" ? "Shall I get the right value for my data" ?	4	0
International, European and National Legislation not harmonised!!	2	0
Concerns were expressed about the missing regulations that would protect the actors from unfair competition	2	0
competition & confidentiality (most of the concerns)	0	0
the limitation is that we can expect that there will be many contracts to manage. We could see that each machinery or equipment supplier is tied to a specific RFA.	0	0
if a competitor stakeholders need that there is no special promotion – fair collaboration and competition	0	0
This is key problem in Chemicals: starting from proprietary databases of reactions, to possibility of sharing process data.	0	0
third party rights (more of a concern)	0	0
lack of legal grounds —no valid entitlement (this is less of a concern)	0	0

On the basis of the responses received and the ensuing open discussion, the following conclusions can be made:

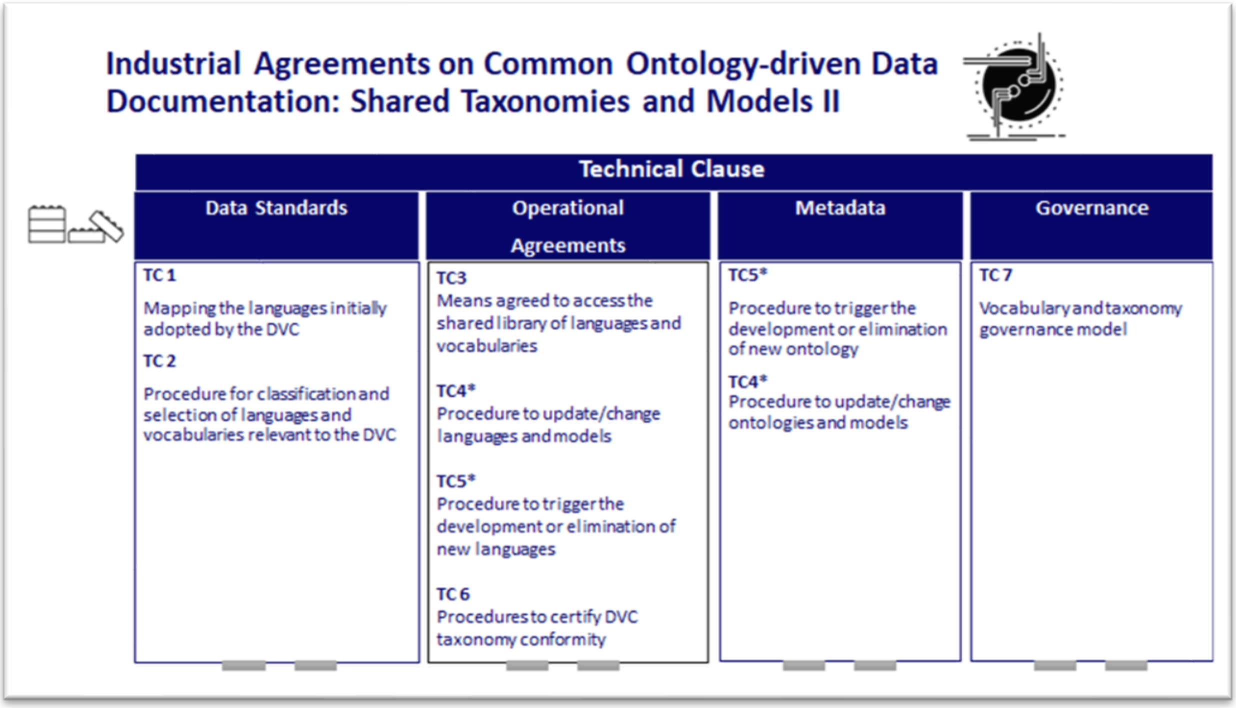
- Within the context of data sharing and exchange, participants feel that the mix of public and privately owned data is not sufficiently reflected in the industry agreement. Also, some aspects such as data generation, storage and analytics (use by third parties) warrant further attention.
- Participants see potential in technical reference architectures to help enforce data-related contractual clauses, for instance through providing data security, immutable storage, PII, providing levels of sharing, setting time limits for the use of a particular dataset, outlining identifiable information tiers and describing the relationship between different actors within an ecosystem.
- The main limitations to using contracts regarding data sharing and exchange that participants indicate relate to the **lack of trust** in how their data will be used and what value will be given to it. Other limitations that are mentioned are a lack of harmonisation of international, European and national legislation form and the absence of regulation that would protect actors against unfair competition.

Plug & Play/Interoperability — Industry agreements on Common Ontology-driven Data Documentation

Technical part

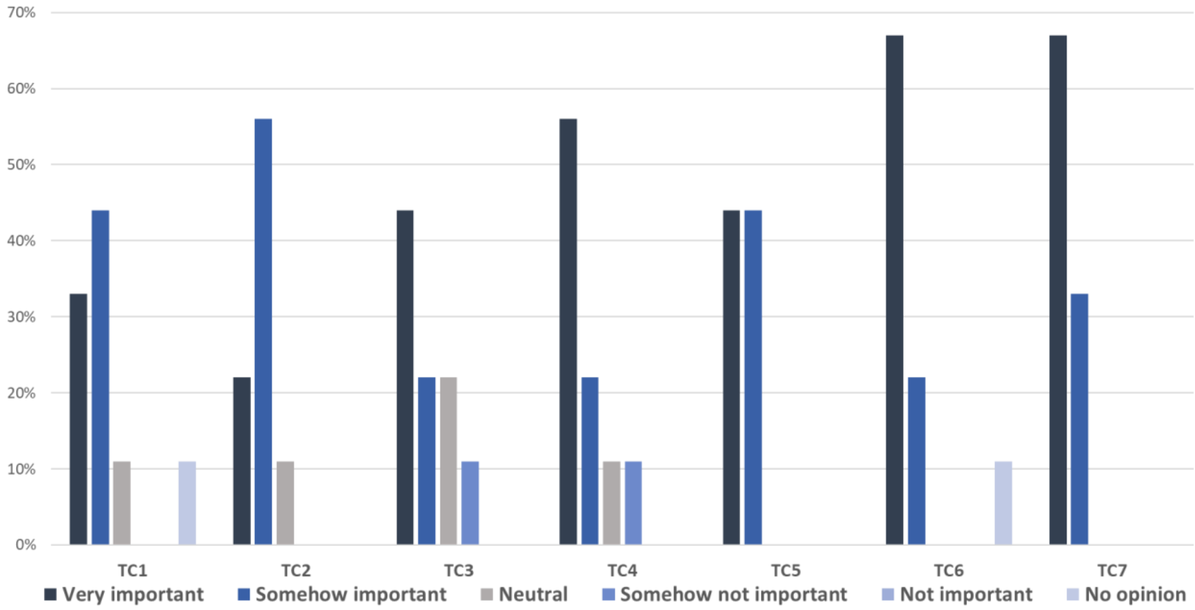
The technical clauses for the industry agreements on common ontology-driven data documentation were presented to the participants and are summarised in the figure below.

Figure 33: IA template on Common ontology-driven data documentation – Technical clauses



As can be seen in Figure 34, among the various clauses included in the IA, the participants responding the survey indicated that **TCs 7, 6, 4, 5** and **3** are the clauses to be prioritised to develop interoperability.

Figure 34: Online survey results on Common ontology-driven data documentation



N=9

Open discussions led to the following conclusions:

- Participants in an agreement need clause(s) to guarantee **fair treatment**. The agreement must provide an equal ground. This is especially important if there are competitors in the same agreements. The agreement must guarantee that there will be no special promotion of a specific actor. Such guarantee(s) would make partners more secure in sharing their data, which they consider as a business asset. Performance obligations and data sharing rules are two possible clauses to include in the agreement;
- Current governance frameworks **lack a model for persistence** of the original data-owner and a dimension taking the **aggregated level** into consideration. The original owner tends to disappear from the final data agreement and there are no norms nor open-source software on this matter. Said differently, there are two elements: the aggregated level and the original level. At the moment, there are a lot of architectures but no management of downstream rights. Once things are reassembled in a new data sets, the rights and conditions of the data provider is simply lost;
- This last point is closely connected to the question of **Intellectual Property management.** In addition to the legal dimension, there is a need for technical specifications to cover the dynamic part of IP. There are use cases where you need rules for track and trace of IP rights – directly in the blockchain. IP is Dynamic and the rules change when data is getting aggregated and when a new data set is created using the data pool.

Figure 35: Polling on Interoperability – additional comments on prioritised clauses

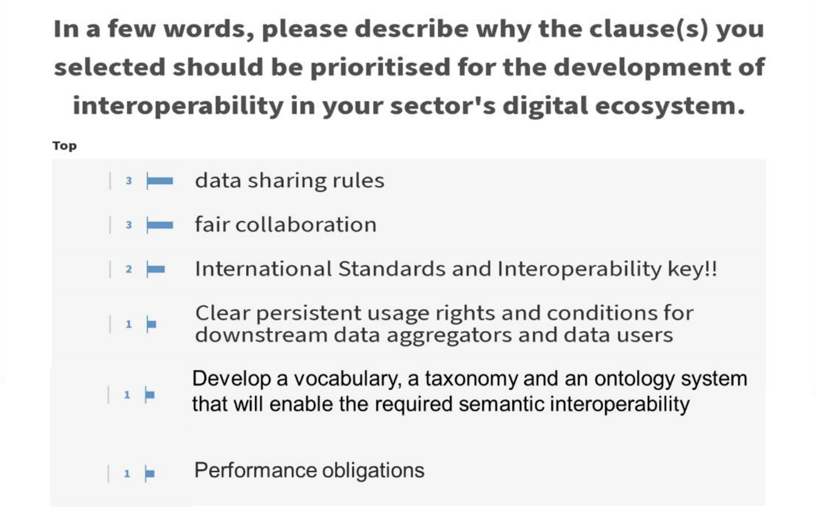


Figure 36: Polling on Interoperability – missing building block aspects

Legal part

The sample clauses on industry agreements on Plug & Play/Interoperability were presented to participants, with a main focus on access to background IP, access to foreground IP and conditions for third-party access. Participants were subsequently presented with four open questions:

- ***"Within the context of industry agreements on Plug & Play/Interoperability, are you confronted with any other legal concerns that you do not see reflected in the template industry agreement?"***

Responses	Upvotes	Downvotes
Responsibility for misuse	4	0
Evolving and unanticipated scenarios (e.g. novel scenarios involving frontier technologies)	1	0
May be, the so-called "data rights" on non-personal data, expected to be introduced by Data Act in Q3 2021	0	0

- ***"In your experience, what are the main limitations to using contracts regarding this topic? Have you any concrete examples?***

Responses	Upvotes	Downvotes
inequality in bargaining power	3	0
it can enabled through procurement requirements for interoperability	1	0

- ***"Should these types of industry agreements by default define the conditions for third parties to be granted access to these interoperability solutions? (Do you consider this feasible in the absence of regulatory intervention?)"***

Responses	Upvotes	Downvotes
it would be necessary to know before hand the permissible types of data re-use	4	0
Yes but they cannot anticipate all third party uses or how useful these can be to the original data proprietor	4	0
Yes, but by taking into considerations the upcoming legislative agenda . EU Data Strategy	1	0
conditions of use/sharing to third parties is necessary to be detailed depending on role	1	0

- ***"Could industry agreements contribute to the development of legal ontologies?"***

Responses	Upvotes	Downvotes
Clear semantics are needed for two sectors to communicate between each other, so yes.	2	0
yes. It enables experimentation and empirical input	0	0
Yes, but taken cautiously.	0	0
An example may be "altuism" organisations in the DGA which do not fall under any standard legal entity form	0	0
the description of the relationships between the different actors (who provides what / who does what) is a subset of an ontology	0	0

On the basis of the responses received and the ensuing open discussion, the following conclusions can be made:

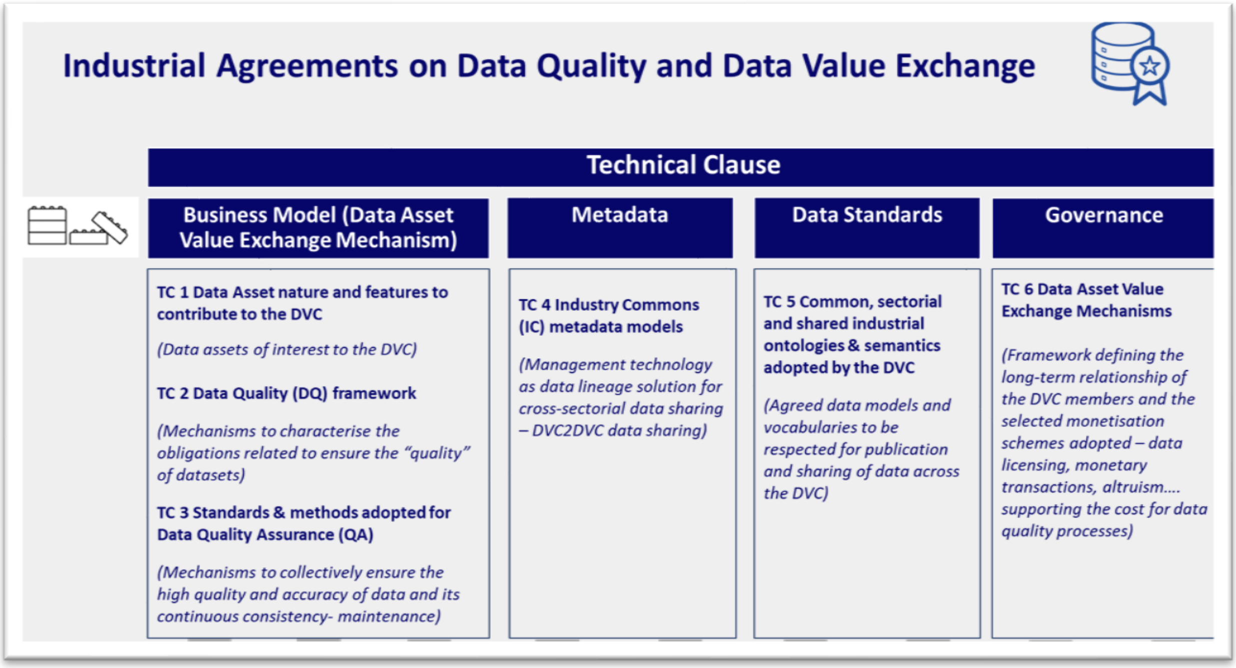
- ❖ The main legal concern that participants consider is not addressed in the template industry agreement is the matter of responsibility for misuse.
- ❖ The main limitation that participants experience with regards to contracts on interoperability is an inequality in bargaining power.
- ❖ Procurement requirements and obligatory use of standards are mentioned as potential ways to combat vendor lock-in.
- ❖ Participants agree that industry agreements should define be default the conditions for third parties to be granted access to interoperable solutions. They indicate, however, that it would be necessary to know beforehand the permissible types of data re-use. Third party uses cannot always be anticipated, nor how useful these can be to the original data holder.
- ❖ Participants agree that industry agreements could contribute to the development of legal ontologies.

Advanced data analytics and Artificial Intelligence — Industry agreements on Data Quality and Data Value Exchange

Technical part

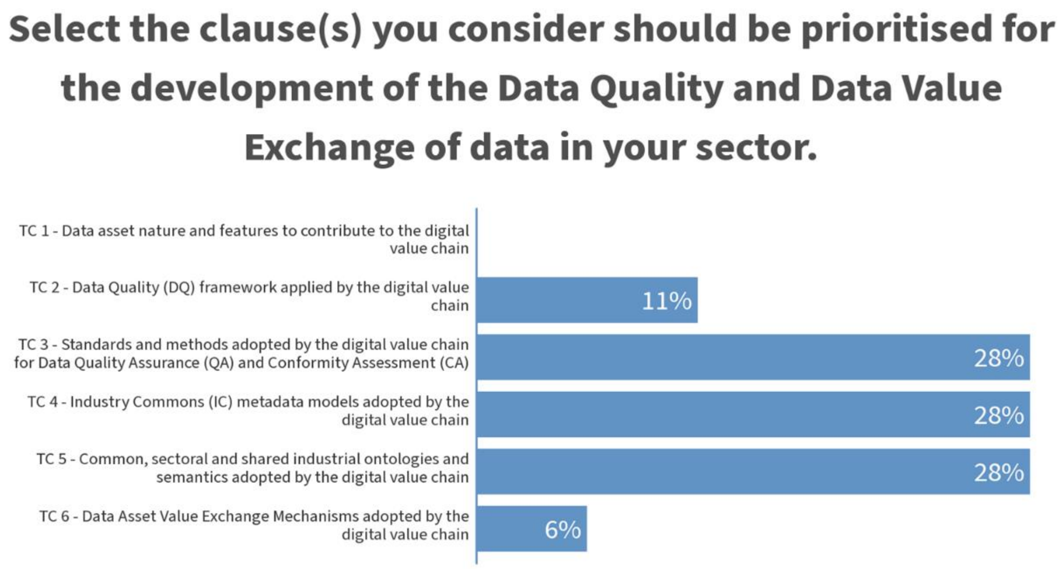
The technical clauses for the industry agreements on data quality and data value exchange were presented to the participants and are summarised in the figure below.

Figure 37: IA template on Data quality and data value exchange – Technical clauses



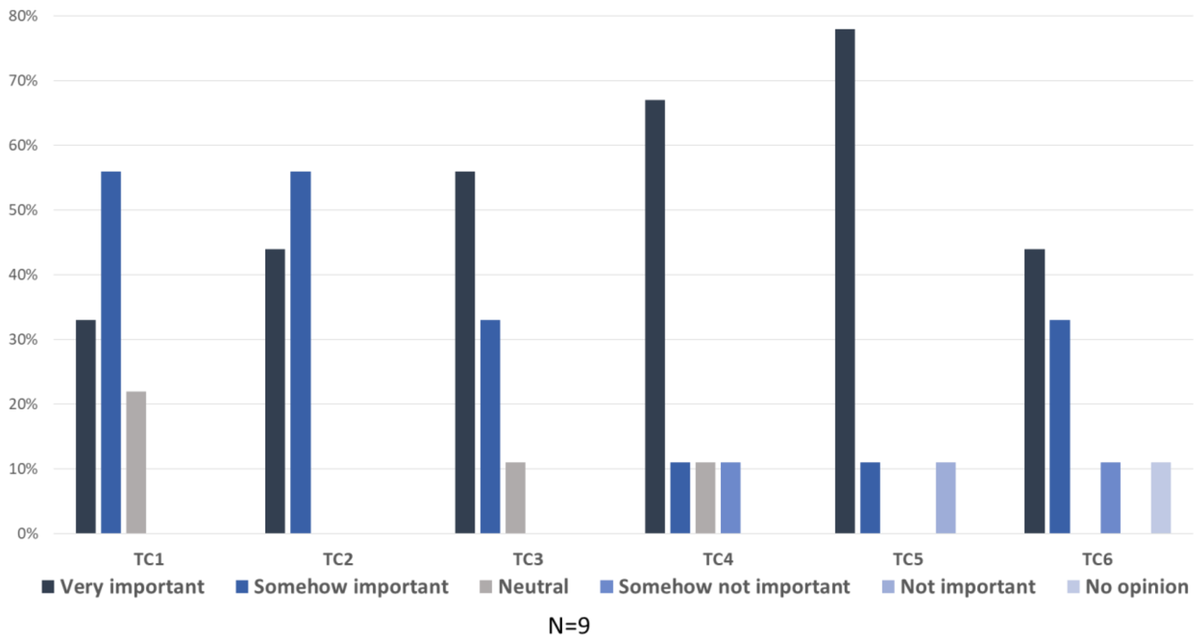
The workshop's participants identified TC3, TC4 and TC5 as the most important clauses for the development of an industry agreement on Data Quality and Data Value Exchange. According to the stakeholders, these clauses are essential to build a common data space where companies can perform rigorous test and control the quality of each data set.

Figure 38: Polling on data quality and data value exchange – prioritised technical clauses



Similar results are represented in the bar chart below, corresponding to the survey launched immediately after the workshop. Although TC3, TC4 and TC5 have been indicated as a priority to be tackled first, the results indicate that all TCs may be considered important, depending on the sector.

Figure 39: Online survey results on data quality and data value exchange



Other important conclusions that emerged from the open discussion are summarised below:

- Developing a ***data quality certification framework*** to provide industry with a systematic approach and framework for assessing data quality. This framework would help to certify that organisations meet data quality and dataset requirements. This framework facilitates achieving data quality, which in turn reinforces trust.
- The content of each TCs needs to be tailored according to the industry;

- Finding mutual agreements on data models and vocabularies to be respected for the publication and sharing of data through the DVC is a necessary step to allow syntactic and semantic interoperability. Supporting both types of interoperability is one of the requirements for efficiently extracting valuable information from the large amount of available data sets;

Legal part

The sample clauses on industry agreements on advanced analytics & AI were presented to participants, with a main focus on liability, transparency and descriptive criteria. Participants were subsequently presented with three open questions:

- "***Within the context of advanced data analytics & AI, are you confronted with any other legal concerns that you do not see reflected in the template industry agreement?***"

Responses	Upvotes	Downvotes
Our experience is that a clear assignment of the exact point of value-creation is extremely challenging	1	0
AI wiht opaque algorithms poses some very specific questions	0	0
AI is not entity but a functionality. Regulatory efforts often ignore this point.	0	0
Liability is often conflated with discussions whether AI should have legal personality	0	0

- ***"Should industry agreements around advanced data analytics & AI include provisions on ethical principles and how these principles should be safeguarded by the different stakeholders (for example, through the appointment of a Chief Ethics Officer)?"***

Responses	Upvotes	Downvotes
Definitely yes, and may be linked to the company's social responsibility programme	3	0
Yes, and they should be dynamically updates according to dependencies (e.g. social, environmental changes)	2	0
Ultimately these need to be manage at social level. Companies tend to take the easy and cheap way out, and this needs to cost them dearly. Strict liability will help.	1	0
On AI, the notoriously opaque alogortihms poses some serious social challenges (Anthony)	1	0

- ***"In your experience, what are the main limitations to using contract regarding advanced data analytics & AI? Have you any concrete examples?"***

Responses	Upvotes	Downvotes
Difficulty in anticipating outcomes	1	0
inequality in bargaining power is an obstacle.	0	0
The work on self-driving cars poses this in a very direct way. Decision support systems in health care have similar issues.	0	0

On the basis of responses received and the ensuing open discussion, the following conclusions can be made:

- Participants agree that industry agreements concerning advanced data analytics and AI should include on ethical principles and how those principles should be safeguarded by the stakeholders. They stress the need for a link with the company's social responsibility program and for dynamic updates of the ethical principles according to dependencies (e.g., social and environmental changes).
- Assurance of data quality raises concerns (for instance, how to measure data quality or support claims on data quality). Legal pathways of direct and indirect liability can be further explored, as well as the capacity of common standards to guarantee data quality.

ANNEX III WORKSHOP ON RECOMMENDATIONS AND IMPACT VALIDATION

III.1 List of participants

Table 26: List of participants - Recommendation workshop| 28.05.2021

#	First name	Last name	Organisation Name
1	Angelo	Marguglio	Engineering Ingegneria Informatica
2	Patrick	Courtney	Sila-standard
3	Ivo	Hostens	CEMA
4	Judith	Kalina	MedTech Europe
5	Irene	Gozalo	ORGALIM
6	Jos	Eikhout	iSHARE foundation
7	Edoardo	Paladini	Safran Aircraft Engines
8	Dimitris	Kiritsis	EPFL
9	Matteo	Bernabei	Edison
10	Guillaume	Joyau	FNSEA
11	Michela	Magas	Industry Commons Foundation
12	Véronique	Brokke	MedTech/Philips
13	Jerome	Bandry	CEMA
14	Sébastien	Picardat	AGDATAHUB SAS
15	Pierre	Constantin	THALES
16	Fabrice	Tocco	Dawex
17	Sebastian	Schlosser	BMW Group
18	Michael	Jochem	Robert Bosch GmbH
19	Sabine	Doerhoefer	Roche Diagnostics International
20	Folco	Ciulli	Regione Lombardia - Presidency
21	Mattia	Adani	Nowal chimica
22	Pieter	van Kooten	HEINEKEN
23	Lars	Nagel	IDSA
24	Boris	Otto	Catena-X, IDSA
25	Davide	Dalle Carbonare	IDSA & SMI Group
26	Matthijs	Punter	SCSN Data Space
27	Wim	Vancauwenberghe	EFNMS / BEMAS
28	Cosmas	Vamvalis	EFNMS / ABE
29	Jordan	Janeczko	ATOS

#	First name	Last name	Organisation Name
30	Mikel	Lorente	AIC Center
31	Ricardo	Lopez	AIC Center
32	Roberto	Perez	GF+
33	Harald	Sehrschön	FILL
34	Fernando	Ubis	Visual Components
35	Massimo	Ippolito	COMAU
36	Oscar	Lazaro	Innovalia
37	Sam	Helmer	AI Coalition
38	Xavier	Gansel	Biomerieux
39	Horst	Pfluegl	AVL/AT
40	Ana	Garcia Robles	BDVA
41	Matthias	Kuom	DG CONNECT
42	Yves	Paindaveine	DG CONNECT
43	Johan	Bodenkamp	DG CONNECT
44	Vasiliki	StergiopoulouTERGIOPOULOU	DG GROW

III.2 Summarised outcomes

Context of the workshop

The workshop on recommendations and impact validation was conducted in the context of the European Commission's (DG CONNECT) study on Technological and economic analysis of industry agreements in current and future digital value chains.

This workshop was organised towards the end of the project and it was targeted towards relevant industry stakeholders as well as EU policy makers. It aimed to validate and further specify the policy recommendations elaborated as a result of the research conducted within the study, as well as the expected impacts of recommended measures on the EU economy's competitiveness in digital technologies. The study team aimed at gathering feedback from selected participants on the recommendations, to validate and further specify them, discuss how industry can make use of them as well as the impacts of the measures proposed. Inputs from participants would hence be used to for the final report of the study.

Agenda and format of the workshop

The workshop was structured around four data-related thematic areas, as presented below, in order to discuss what actions policy makers and industry could take in these areas to facilitate the development and adoption of industry agreements for data sharing. To ensure a fruitful and effective discussion during the workshop, ahead of the event a set of draft policy recommendations was submitted to participants in the form of a short survey, in order to seek their feedback on the effectiveness of the proposed actions.

Feedback from participants during the event was gathered though different ways of interaction, including:

- Teaser presentation of recommendations based on responses to the survey;
- Validation of recommendations by participants;
- Discussion in open forum and in written via the chat function;
- Voting on impacts of proposed recommendations on competitiveness.

A total of 44 participants (excluding the study team) attended the workshop. Participants included representatives from companies operating in the eight sectors analysed in the study, with a particular focus on manufacturing, along with representatives from research/academia and cross-sectoral organisation (e.g., IDSA, BDVA). Representatives from the European Commission (DG CONNECT and DG GROW) were also present. For the detailed full list of participants, please see Annex III.

Table 27 Agenda of the workshop

	Time	Agenda item
Block 1	09:45-10:00	**Registration** - Dial-in and checking technical set-up
	10:00-10:10	**Welcome & Introduction** - Welcome (DG CONNECT) - Introduction to the study *(Ecorys)* - Objectives & outline for today *(Ecorys)*
	10:10–10:35	**Session 1: Building trust** - Proposed recommendations *(Ecorys)* - How can industry put the most effective recommendations into action to build trust? How can industry be best supported? *(Discussion)* - Identification of impacts of recommendations on industrial competitiveness *(Voting)*
	10:35–11:00	**Session 2: Data quality, access and interoperability** - Proposed recommendations *(Ecorys)* - How can industry put the most effective recommendations into action to improve data quality, access and interoperability? How can industry be best supported? *(Discussion)* - Identification of impacts of recommendations on industrial competitiveness *(Voting)*
Block 2	11:00-11:25	**Session 3: Data value** - Proposed recommendations *(Ecorys)* - How can industry put the most effective recommendations into action to better valuate data? How can industry be best supported? *(Discussion)* - Identification of impacts of recommendations on industrial competitiveness *(Voting)*
	11:25-11:50	**Session 4: Technical and regulatory complexity** - Proposed recommendations *(Ecorys)* - How can industry put the most effective recommendations into action to better address technical and regulatory complexity? How can industry be best supported? *(Discussion)* - Identification of impacts of recommendations on industrial competitiveness *(Voting)*
	11:50-12:00	**Conclusion** - Final remarks and summary of discussion

Key outcomes

1. Building trust

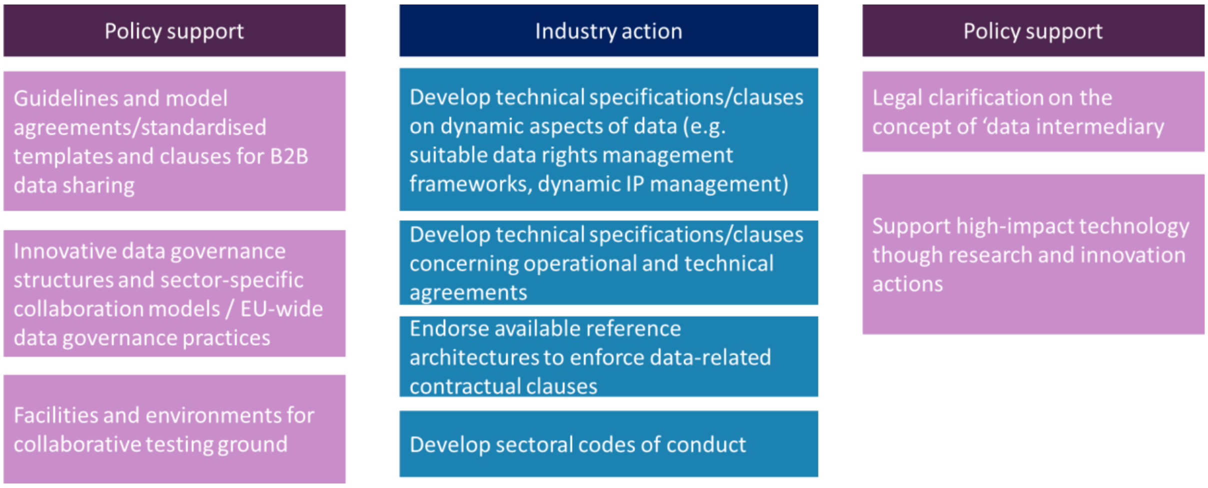
Apart from technical challenges, one of the main obstacles preventing industry agreements (IAs) development and implementation is the lack of trust among industry stakeholders, which comes in different shapes e.g.:

- Doubts on how data provided will be used or reused when they are further aggregated (i.e. uncertainty on the scope of secondary usage);
- Uncertainty on the definitions for roles and processes, business ethics and general principles applied;
- Lack of clarity on the practical implications for data sharing of existing regulations;
- Lack of trust on the integrity of the system for collecting, exchanging and storing data and information (i.e. distrust in the system) that underpins or supports IA implementation;
- Protection of commercially sensitive and personal data, either in terms competitors that are permitted to access the system or illegal access (e.g., unintentional exposure or malicious reverse engineering, data breach and data loss).

Trust is also impacted by different factors, such as data quality, clarity of the regulatory framework, a demonstrated value proposition and transparent financial dynamics. Some of these aspects are also dealt with separately in more detail in subsequent areas.

Proposed actions discussed and validated by the participants are shown in the figure below.

Figure 40 Validated recommendations on how to build trust



Key messages

Based on the discussion during the workshop the following five key messages emerged:

1. **Consent / data sovereignty is critical for building trust**
    - Having the right consent tools in place before exchanging data is crucial to build trust between data owners and users, in order to define the purpose for (re)use, the modalities and time limits. It was also raised that there are other legal bases available beyond consent to be leveraged;
    - To build trust in a data space, there is a need to develop tools to pool, access, use and share all types of data. They should empower rightful data holders to ensure transparency as to where their data is stored and what access and reuse rights apply to it. It was suggested that the technical data infrastructure should integrate the cybersecurity-by-design and privacy-by-design principles;

- Another important building block highlighted is the secure infrastructure which allows a secure communication across participants. At least this should require secure digital identities for all participants;
- It is important to distinguish among different kinds of trust:
  a. Competitors: to develop common protocols, a key concern there relates to the regulatory dimension, notably whether such cooperation could breach competition law;
  b. Customers: a key factor raised relates to the fear that data may go outside Europe, therefore not be protected under EU law. Assurance is needed about where data will be stored / who will have access / what will be done with the data;
- It was highlighted that the GDPR already prohibits transfers to non-EU countries unless a country is considered adequate, or if there is a valid transfer mechanism in place - as it must be ensured the data remains protected in line with EU standards. However, from participants' experience, GDPR is perceived as protecting purely personal data and industrial customers do not feel protected for their manufacturing/process data by GDPR. To that end, the same protection being afforded to non-personal data when it comes to data transfers is perceived as beneficial;

2. **Experimenting and testing are crucial!**
   - Allowing spaces for testing before imposing rules is considered of utmost importance;
   - A specific block could be added for the experimental approach, so trust is built also by experimenting along value chain;

3. **Frameworks need to be strengthened**
   - Requirements are needed for guidelines on security of data / codes of conduct / certification of trustworthiness, but there is a question on how effective these are if on a voluntary basis only;
   - Solutions such as trustworthiness profiles[404] would be needed, as to show capabilities of partners, allow all partners to take informed decision based on trustworthiness profiles. This could be done by setting up standardised trustworthiness profiles which can be used to show that each of the participants is on same trustworthiness level. This could be started with a structure block, including different certifications from the companies themselves, they can provide this info to other participants and check if this is enough for them. Going forward this could but done also automatically, but we need additional rules, combining technical details with legal aspects, standardisation of consent;
   - Sectoral codes of conduct are deemed to be a good solution to address specific situations of each sector, but more work is needed on the condition of implementation of these codes, specifically how to get as many companies as possible on board;
   - When talking about code of conduct it is suggested to discuss it in the context of GDPR, as it provides the opportunity to setup a CoC. Also, we should not have only sectoral codes of conduct, cross-sectoral use cases are too common to be ignored;

4. **Protocols and code of conduct are necessary, but they are not enough**
   - Stakeholders in a value chain will trust each other if, first of all, understand each other when they are talking about data. Therefore, they need to talk about the same "meaning" of each data. This requires using a commonly agreed and structured vocabulary of their domain of discourse. In other words, "Industry Commons" based on the semantics of data are a first order need to enable trust development;
   - To create trust between all participants in a data space, there also needs to be an efficient data governance mechanism, a set of rules of legislative, administrative and contractual nature that determine the rights to access, process, use and share data. All stakeholders in the data space should be represented and engaged in the governance of the data space;
   - The future and harmonized governance for data spaces must rely on industry agreements and be a holistic one - sharing data is not only a technical challenge and an entrepreneurial risk, but it also requires legal conventions and operational guidelines;

[404] https://www.plattform-i40.de/PI40/Redaktion/EN/Downloads/Publikation/IIoT_Value_Chain_Security.html

- An important point raised was to have clear rules for doing data anonymization which are legally accepted. One of the main barriers perceived is the legal uncertainty when it comes to anonymization. To the extent data still qualifies as personal (which includes pseudonymous data), many obligations exist that sometimes can be hard to meet and hamper innovation. Especially considering often, there is no intention to identify individuals;
- A challenge raised relates to the fact that in many instances there are commercial agreements, e.g., on the purchasing of a physical product, whereby the data sharing is happening in the background. It is not always common for procurers/users to think about data sharing upfront and deal with this in the purchasing contract;

5. **No need to reinvent the wheel**
   - There are quite a few initiatives available that already provide some of the building blocks for data sharing, most of them being complementary to each other, it is important to highlight what the possibilities are and bring convergence among the different initiatives;
   - Some initiatives highlighted:
     - www.ishareworks.org/en a framework of agreements that has been developed to overcome most barriers of data sharing. It is not a data space, but rather a framework that enables data spaces;
     - A German public funded project deals with trust and automated contracts. https://legaltestbed.org/en/start/;
     - Dawex Data Exchange Platform technology has solved many issues mentioned by data providers and data users: technical access of the data, right licensing contract formats depending on the need, traceability by the data provider, trust between the data provider and the data user, multiple business models for the exchange, compliance with the rules & regulations, and far more https://www.dawex.com/en/.

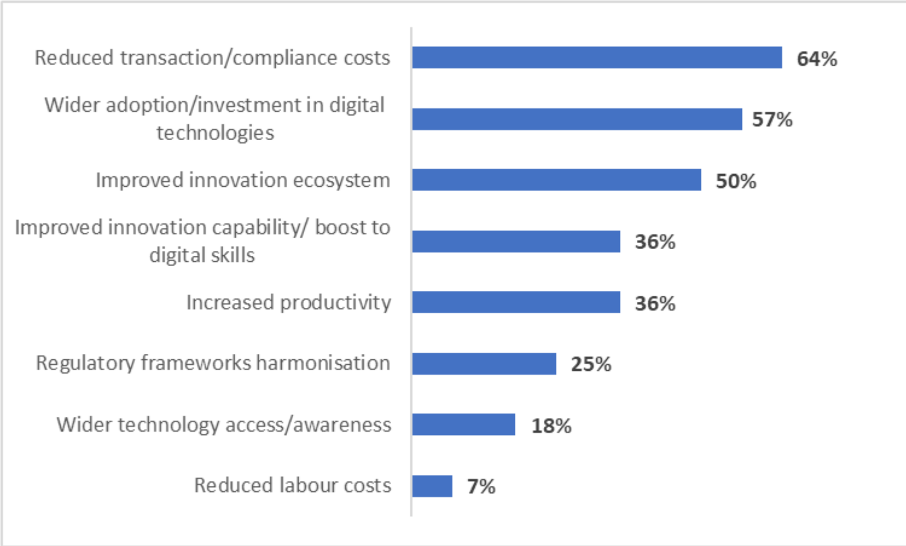
Impacts on competitiveness (poll questions)

Participants were asked to indicate which impacts on industrial competitiveness they expected from selected recommendations (i.e., top ranked for industry and policy), as shown below.

1. Develop guidelines, model agreements and standardised clauses (e.g., technical, operational)

Having standardised agreements/model agreements is expected to give companies orientation but also confidence in the contracts negotiated and lead to less negotiation - which will inevitably reduce transaction cost. Reduced transaction/compliance costs will fix the short-term bottom line. However, an improved innovation ecosystem and wider adoption of digital technologies create novel, both radical and disruptive business models, completely integrated into the system, and generate new market values which ensure future market players' financial security.

Figure 41 Which top 3 impacts on competitiveness for your industry do you expect from the following action?

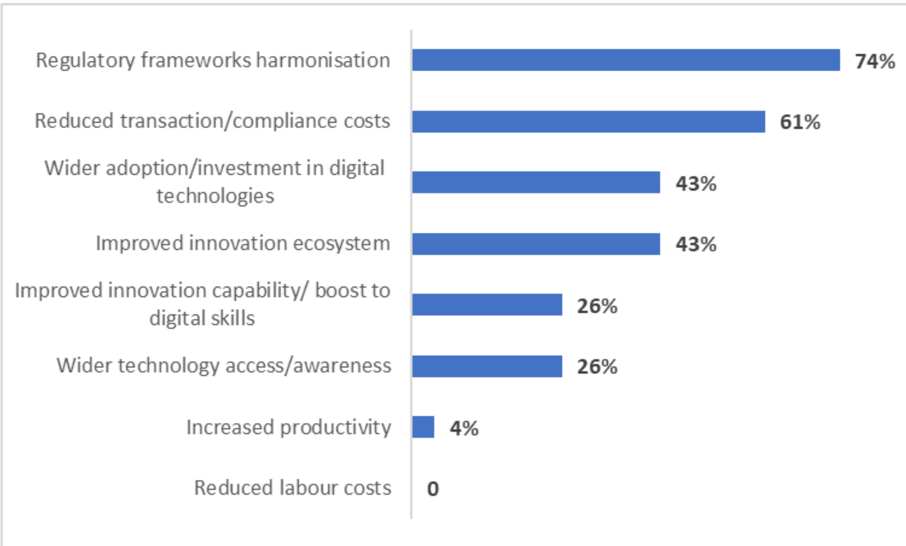


Source: Ecorys based on participants' voting (N=28).
Note: the percentage of participants selecting each option is displayed.

2. Legal clarification on the concept of data intermediaries

Clarifying essential notions of data intermediaries (e.g., type of actors, functioning, ethical standards and accountability) is expected to bring regulatory harmonisation as for these entities to act credibly as a trusted service provider for contractual commitments. This is expected to reduce post-contractual risks and monitoring costs for the participants as to overcome data market failures and enable data sharing transactions that would otherwise not materialise because of perceived risks.

Figure 42 Which top 3 impacts on competitiveness for your industry do you expect from the following action?



Source: Ecorys based on participants' voting (N=24).
Note: the percentage of participants selecting each option is displayed.

2. Data quality, access and interoperability

Data quality requirements can pose a significant challenge for the adoption of collaborative agreements. Data needs to adhere to established quality standards to ensure that the use of data will provide accurate results, meaningful insights and a reliable basis for decisions. Companies are often forced to allocate internal resources to burdensome and time-consuming processes to clean data and make it usable, or they may lack the resources or common approaches/standards to do so. Accessibility to high-quality datasets is also an ongoing challenge for most organisations in the analysed sectors (e.g., due to lack of interoperability).

Moreover, the circumstances under which liability may be incurred for damages due to inaccurate data is a recurring question in all sectors. Lack of clarity on this topic has a significant effect on the willingness of industrial stakeholders to enter into industry agreements.

Proposed actions discussed and validated by the participants are shown in the figure below.

Figure 43 Validated recommendations on how to improve data quality, access and interoperability

Policy support	Industry action	Policy support
Promote the definition of harmonised criteria/guidelines to assess the quality of datasets	Develop contractual clauses to increase the quality and value of data, e.g. standards & methods adopted for data quality assurance, industry commons metadata models, common industrial ontologies & semantics	Promote the development and use of standardised data licensing models /agreements
Support the development of a data quality certification framework or data quality seals	Include procurement requirements and obligatory use of standards in contracts on interoperability, address responsibility for misuse	Promote common spaces allowing for testing of various types of data quality according to diverse scenarios
Support identification of relevant standards e.g. through establishment of a coordination body	Identify a list of standards that need to be supported in the data spaces for each sector and levels of interoperability	Set up independent subjects/bodies that serve as a data quality test engine

Key messages

Based on the discussion during the workshop the key messages emerged:

1. **Data quality requires shared principles and vocabularies**
    - Common industrial ontologies and semantics are considered an important prerequisite regarding data quality. E.g., the Industrial Ontologies Foundry is an industry driven international initiative working along the same lines and collaborating with OntoCommons: https://www.industrialontologies.org/;
    - Important concept of "FAIRification" of data, valuable in order to process industrial commercial data set with a certificate which would increase their value;
    - What is perceived as needed is "Terms and Conditions" for the Data Economy. This is what IDS is aiming at, for instance. These terms and conditions must be understood unambiguously (interoperability), machine-readable, must be able to be negotiated and tracked/logged/enforced. https://doi.org/10.1080/10580530.2015.1044344;
    - An issue raised related to the fact that SMEs have a problem when they should allow data to be exchanged, they would need support in identifying who in the company should approve it, which criteria should be used to make this decision, is not just about the technology, but also a process discussion;

2. **Standards require convergence and harmonisation efforts**

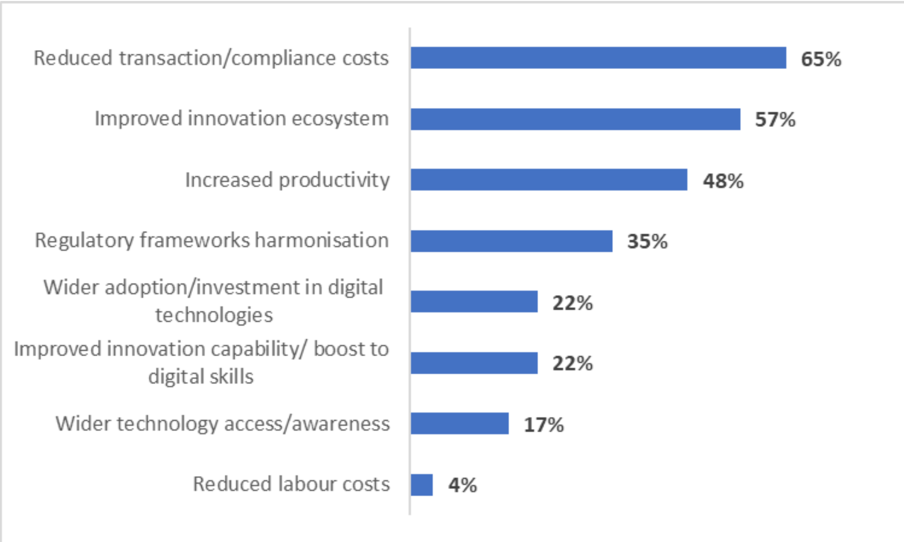
- Standards are crucial, but it also important to put in place capacity building in terms of platforms and functionalities, which are easy to use and adopt in all kinds of scenarios;
- Portability/interoperability are very important: standards required not just for data exchange within platforms but also to:
    - Ease of use of platforms
    - Interoperability between platforms
    - Interoperability/exchange of analytics (shift of emphasis from exchange of data to exchange of analytics)
- There is a need to develop standards but also make them widely available, including with public actors and along the whole ecosystem. Role of public authorities should not be just about promoting standards but also driving the development of standards, bringing the relevant players together in the definition of the standards and taking the drive in this convergence effort towards common standards.

Impacts on competitiveness (poll questions)

1. Promote the development and use of standardised data licensing models /agreements

Having standardised data licensing models/agreements is expected to facilitate trustworthy access to data and reduce transaction costs for companies. Common/standardised contractual provisions (e.g., redistribution obligations) in data license agreements are designed to encourage sharing by limiting the liability of the data provider and ensuring that those downstream can identify where the data came from. This would also facilitate new collaborative approaches and improve the innovation ecosystem, similar to what we have with open-source software.

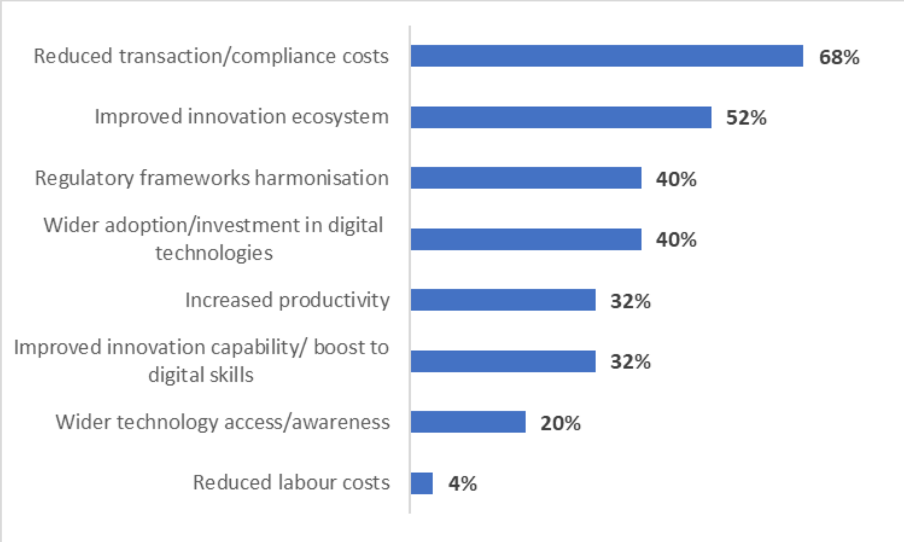
Figure 44 Which top 3 impacts on competitiveness for your industry do you expect from the following action?



Source: Ecorys based on participants' voting (N=25).
Note: the percentage of participants selecting each option is displayed.

2. Develop contractual clauses to increase the quality and value of data (common standards and methods)

Figure 45 Which top 3 impacts on competitiveness for your industry do you expect from the following action?



Source: Ecorys based on participants' voting (N=24).
Note: the percentage of participants selecting each option is displayed.

3. Data value

To give data providers a degree of control over 'their' (industrial) data, the concept of "data sovereignty" has emerged. Data sovereignty involves the technological enforcement of contractual terms to enable data providers to retain some control over the reuse of data. Data sovereignty relies on the premise that a market for data can produce a fair "price for data" and strongly relies on the hypothesis of an appropriate data monetisation mechanism. However, when it comes to the value of data, establish a "fair price for data" may be complicated due to the very nature of data, because its value increases the more it is reused and repurposed.

Sharing data still faces barriers related to business ethics, namely the general principles that are to be adhered to to obtain a proper division of the benefits in sharing the data. The development and implementation of industry agreements (IAs) may also be hampered by a lack of information or uncertainty about the magnitude (value) of expected costs and benefits arising from an IA, or the distribution of costs and benefits across different agents. Notably, the pricing of datasets and valorisation of data are proving to be very complex. Where this is the case, a lack of information (non-transparency) may result in agents under or overestimating the potential gains from an IA, which could discourage efforts to develop or implement an IA.

All of this poses uncertainties on who should reap the economic benefits from data shared in IAs; how data providers can ensure that they receive some of the profits that are realised from the data they share; or how to ensure that data is ultimately not used in a way that would go against the interests of the data providers. Hence, a major concern revolves also around the difficulty of data providers to assess upfront the value in participation - balance their data's perceived value against risks exposed (e.g., perceived loss of control over data, trade secrets loss, and potential data policies breaches).

Proposed actions discussed and validated by the participants are shown in the figure below.

Figure 46 Validated recommendations on how to better valuate data

Key messages

Based on the discussion during the workshop the following key messages emerged:

1. **The right balance needs to be found between data reuse and the costs incurred to make data available**
   - The value of the data depends on the quality, we should not end up in a situation where companies are forced to give data out for free, ensure quality level it comes with a cost. The value can unfold if data is used, we need to make sure we increase usage of data for ecosystem innovation and that individual data providers are not exploited.
   - There should be fairness of compensation for making data available and common assessment models. There are 3 main possibilities currently adopted:
     - Exchange price, see if there is a market and try to sell the date, however this approach is not always possible.
     - In the majority of cases, companies try to identify what is the use value of the data (business benefits). However, it may be hard to identify such value as it depends on the use case and an individual assessment is needed.
     - Third option is to look into the cost as proxy for the value, namely what kind of cost was incurred to make data available.
   - An additional value category raised is when you model two diverse datasets (e.g., from two sectors) for the first time, the process yields a third dataset which may identify a potential value (e.g., innovation IP layer dataset)
   - The notion of data as an asset (future value) vs. resource (being consumed) is very important. Some data definitely have asset potential.
2. **It is crucial to identify win-win situations and define a taxonomy of data/use cases**

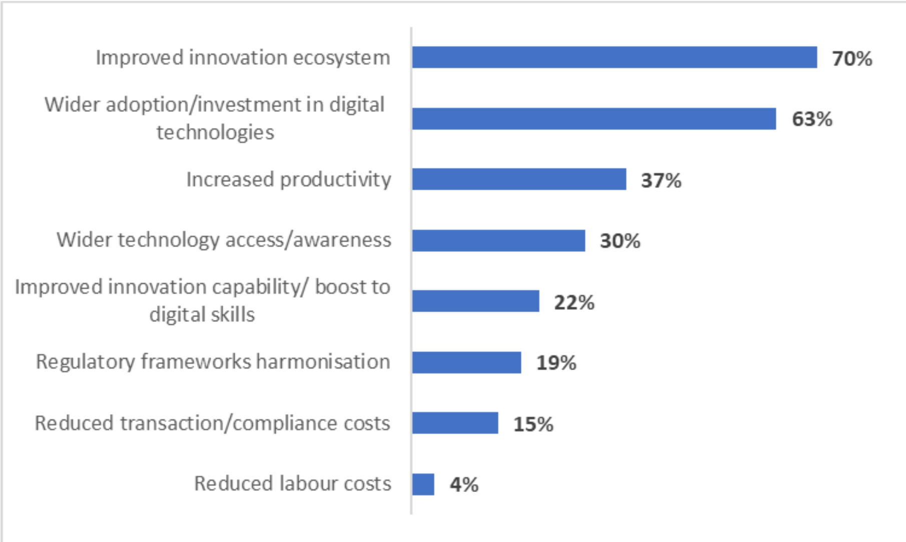
- A standard methodology raised by a participant involves the following process:
  - Ask parties beforehand what their benefits from the agreement would be, as to know where the benefits lie and indicate which parties will benefit directly by having e.g., increase in productivity.
  - Once this is done, a discussion is started on how to ensure that a model is created where everybody benefits, value selling proposition.
- The value of the data depends on how it is used, however when we look into the different types of data, there is a class of data for which a general asset value might be assignable, e.g., master data are used in different processes, but we can identify a baseline value cross use cases.
- Having very different data types and use cases, industry should look into concrete use cases and identify what value a company gets in return when share a certain type of data for a certain period of time, as to develop a taxonomy of data needed to identify certain value classes.

Impacts on competitiveness (poll questions)

1. Dissemination of good practices and guidelines to define common data valuation/ monetisation models

Due to the wide range of potential data and use cases, as well as the relative infancy of data valuation, there are no simple or universally agreed upon methods – making it difficult to assess upfront potential value and benefits. Having guidance and common practices on data valuation models can help create business models that can capture and monetise the benefits, by setting up a mechanism for redistribution of the gains to bring on board the parties that feel unfairly treated or are at risk of losing out in data sharing. This would help companies assess upfront the value in participation, balance data perceived value against risks exposed and drive wider adoption/investment in digital technologies.

Figure 47 Which top 3 impacts on competitiveness for your industry do you expect from the following action?

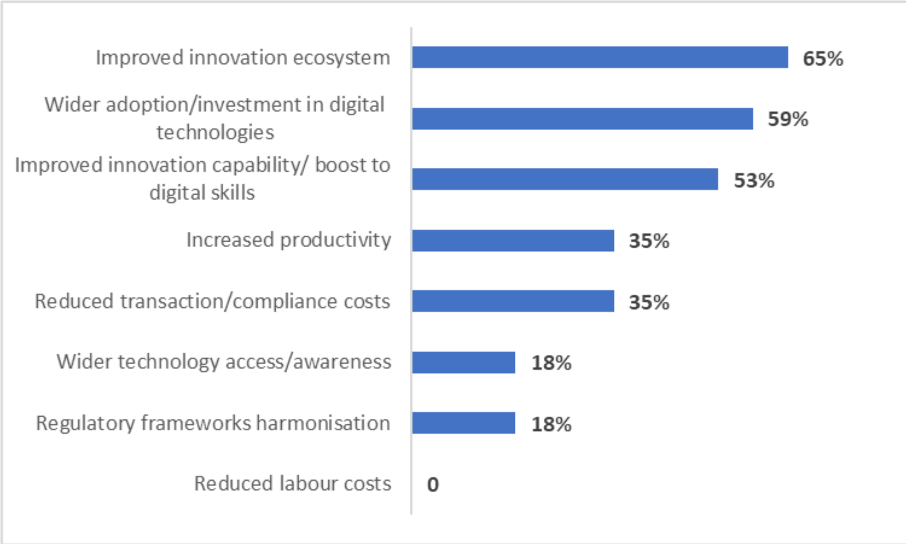


Source: Ecorys based on participants' voting (N=28).
Note: the percentage of participants selecting each option is displayed.

2. *Promotion of innovative collaborative arrangements for the fair allocation of costs and benefits*

Similarly, innovative collaborative arrangements (e.g., data trust, data cooperatives) can carry clear mechanisms for the redistribution of benefits, which would ease the ability of industrial parties to assess upfront the potential value of participation and improve the overall innovation ecosystem.

Figure 48 Which top 3 impacts on competitiveness for your industry do you expect from the following action?



Source: Ecorys based on participants' voting (N=18). Note: the percentage of participants selecting each option is displayed.

4. Technical and regulatory complexity

Technical complexity may represent one of the major barriers to Industry agreements (IAs). High fixed costs may be caused by the complexity of the technical aspects addressed by an IA, or by the complexity of the regulatory environment in which an IA operates. This complexity may actually inhibit IA development. The high costs of IA development may be associated with, for example, the need to involve specialist expertise to help formulate an IA, the volume of effort required (in terms of quantity of inputs and time required), including administration and coordination costs (particularly challenging in fragmented industries).

The regulatory context can also add to the complexity of formulating an IA, especially when regulations are not harmonised across borders. Industry agreements are confronted with a very complex legal framework, comprising a variety of complementary legal instruments that each impose requirements and limitations on the contractual freedom of the parties involved. Industrial companies are faced with considerable doubts, especially when it comes to planned data cooperation with competitors, as to whether the arrangement complies with existing data protection and antitrust laws. The grey areas are large and the legal uncertainty among companies is considerable. This regulatory background may prove to be very restrictive and deter stakeholders from engaging in innovate cooperation (e.g., due to high compliance costs or concerns about anti-competitive effects).

Proposed actions discussed and validated by the participants are shown in the figure below.

Figure 49 Validated recommendations on how to better address technical and regulatory complexity

Key messages

Based on the discussion during the workshop the following key message emerged:

Regulatory harmonisation is needed

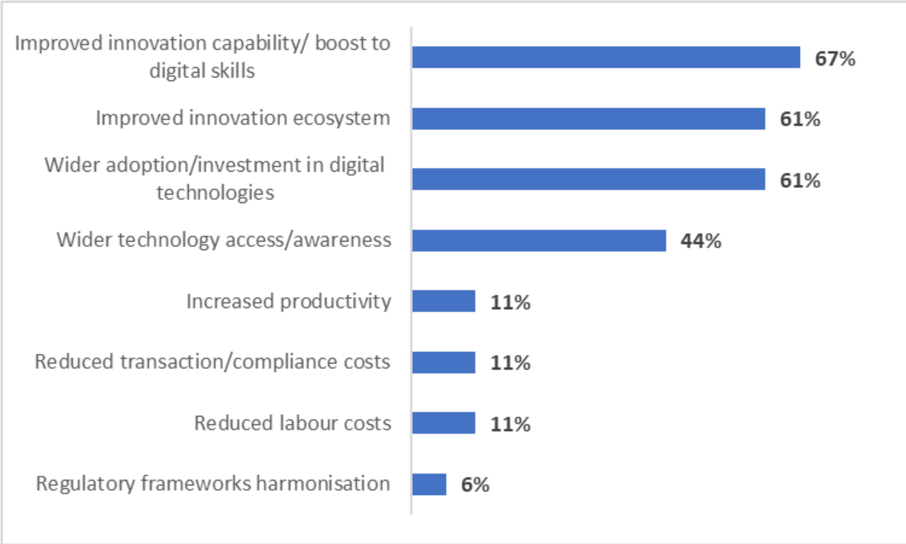
- One of the major challenges is the lack of clarity on how we classify the data, as to know what obligations are applicable, e.g., if it is not clear whether something is either non-personal data or anonymised, especially for organisations who may not have the expertise on board, it could be hard to understand what is applicable for them
- Lack of cross-border harmonisation is an issue, regulator in a Member State may view some data as non-personal, others may see the same data as personal data. There is still a lot of legal uncertainty around the concept of personal data, which impacts also non-personal data to a large extent. Multiplicity of rules at different levels (i.e. regional, national, EU, international) creates some confusion and legal uncertainty, some clarity could be brought by code of conducts or data spaces with clear terms and conditions, as to create common frameworks and rules;
- It is perceived that any further regulatory interventions should take into account existing regulations as not to create contradictory provisions. Even if a company operates in a trusted data space, it may still be subject to same regulatory obligations for paper-based information (disincentive to using digital space). There should be a rule that once data is shared on a trusted space, it then should not require paper trail (avoid data duplication costs).

Impacts on competitiveness (poll questions)

1. <u>Leverage existing entities for safe experimentation and validation, seek tailored advice and technical expertise</u>

Resorting to existing trusted entities (e.g., Big Data Innovation Hubs network) can facilitate stakeholders to connect, seek the technical/legal expertise required and access the enabling technologies, thereby improving connections of stakeholders along the value chain and their innovation capability and digital skills.

Figure 50 Which top 3 impacts on competitiveness for your industry do you expect from the following action?

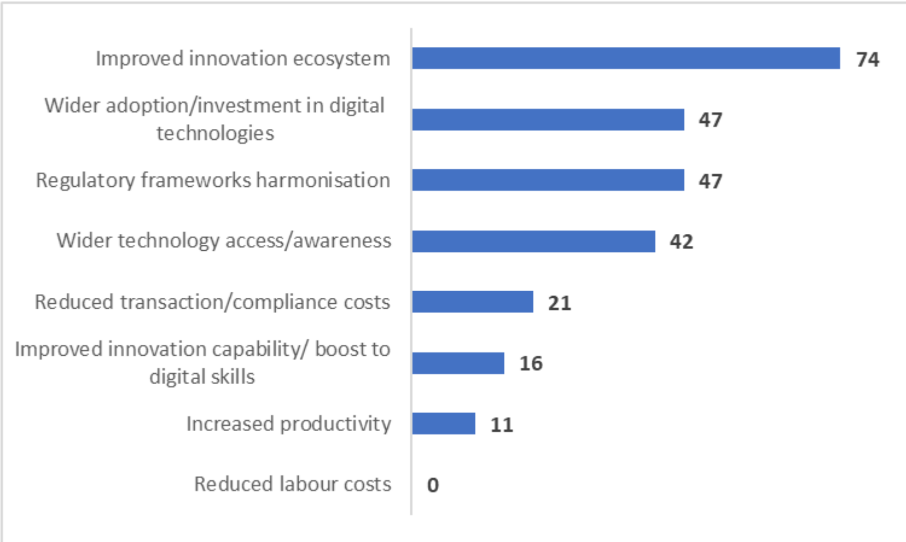


Source: Ecorys based on participants' voting (N=19).
Note: the percentage of participants selecting each option is displayed.

2. <u>Provide additional guidance/practicable guidelines on how relevant laws apply (e.g., competition law)</u>

More clarity and actionable guidelines on how relevant laws apply to the data economy is expected to bring benefits to the overall innovation ecosystems, in terms of increased confidence and trust among stakeholders when sharing their data.

Figure 51 Which top 3 impacts on competitiveness for your industry do you expect from the following action?



Source: Ecorys based on participants' voting (N=19).
Note: the percentage of participants selecting each option is displayed.

European Commission

Study on technological and economic analysis of industry agreements in current and future digital value chains
Brussels, Publication Office of the European Union

2021 – 190 pages

ISBN: 978-92-76-43633-1

doi: 10.2759/495071

KK-07-21-089-EN-N

Publications Office
of the European Union

doi: 10.2759/495071 ISBN 978-92-76-43633-1