TRUSTS
Trusted Secure Data Sharing Space

D2.1 'Definition and analysis of the EU and worldwide data market trends and industrial needs for growth'

Authors: **Natalia Simon (IDSA), Ioannis Markopoulos (FNET), Stefan Gindl (RSA), Bert Utermark (G1), Martin Kaltenböck (SWC), Antragama Ewa Abbas (TUD), Hosea Ofe (TUD), Montijn van de Ven (TUD), Rômy Bergman (TUD), Anneke Zuiderwijk (TUD), Mark de Reuver (TUD), Nora Gras (IDSA), Antonia Kuster (IDSA), Julia Jakuzzi (IDSA), Sebastian Emons (IDSA) Gerrit Rosam (LUH), Michael Fribus (LUH), Alina Brockob (LUH)**

Additional Information: **overview on current states and relevant trends on the topic of data marketplaces to ensure that the projects' results are targeting market needs and working with up-to-date requirements.**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871481

TRUSTS Trusted Secure Data Sharing Space

D2.1 'Definition and analysis of the EU and worldwide data market trends and industrial needs for growth'

Document Summary Information

Grant Agreement No.	871481	**Acronym**	TRUSTS
Full Title	TRUSTS Trusted Secure Data Sharing Space		
Start Date	01/01/2020	**Duration**	36 months
Project URL	https://trusts-data.eu/		
Deliverable	D2.1 'Definition and analysis of the EU and worldwide data market trends and industrial needs for growth'		
Work Package	WP2 'Requirements Elicitation & Specification'		
Contractual due date	30-06-2021	**Actual submission date**	30-06-2021
Nature	Report	**Dissemination Level**	Public
Lead Beneficiary	IDSA		
Responsible Author	Natalia Simon (IDSA)		
Contributions from	Ioannis Markopoulos (FNET), Stefan Gindl (RSA), Bert Utermark (G1), Martin Kaltenböck (SWC), Antragama Ewa Abbas (TUD), Hosea Ofe (TUD), Montijn van de		

	Ven (TUD), Rômy Bergman (TUD), Anneke Zuiderwijk (TUD), Mark de Reuver (TUD), Nora Gras (IDSA), Antonia Kuster (IDSA), Julia Jakuzzi (IDSA), Sebastian Emons (IDSA) Gerrit Rosam (LUH), Michael Fribus (LUH), Alina Brockob (LUH)

Revision history (including peer reviewing & quality control)

Version	Issue Date	% Complete[1]	Changes	Contributor(s)
V0.05	27.04.2020	10 %	Initial deliverable structure	Natalia Simon (IDSA)
V0.1	11.05.2020	14 %	Adaption of the initial deliverable structure	Natalia Simon (IDSA)
V0.15	18.10.2020	25 %	Added Chapter on financial and telecom industry	Ioannis Markopoulos (FNET)
V0.2	15.11.2020	36 %	Added Chapter 2.2 - data marketplaces	Antragama Ewa Abbas, Montijn van de Ven, Rômy Bergman, Anneke Zuiderwijk, Mark de Reuver (TUD)
V0.25	19.11.2020	47 %	Added technology area	Martin Kaltenböck (SWC)
V0.3	06.01.2021	58 %	Added Chapter on the political context	Natalia Simon (IDSA)
V0.35	21.12.2021	69 %	Reviewed Chapter on data marketplace context, added Chapter on micro analysis and derived recommendations	Hosea Ofe (TUD), Bert Utermark (G1)
V0.4	16.02.2021	80 %	Finalised technology Chapter and reviewed political and economic Chapter as well as Financial Industry Chapter	Martin Kaltenböck (SWC)

1

V0.5	11.03.2021	85 %	Quality review and general additions	Gianna Avgousti (EBOS)
V0.6	29.03.2021	88 %	Implementation of the first quality review comments	Natalia Simon (IDSA)
V0.7	20.05.2021	90 %	Implementation of the World Café Workshop results	Stefan Gindl (RSA), Ioannis Markopoulos (FNET), Bert Utermark (G1)
V0.8	02.06.2021	95 %	Submitted for internal review	Gerrit Rosam, Michael Fribus, Alina Brockob (LUH)
V0.9	21.06.2021	98 %	Implementation of feedback from reviewers	Natalia Simon (IDSA), Antragama Ewa Abbas (TUD), Hosea Ofe (TUD), Bert Utermark (G1)
V1.0	30.06.2021	100 %	Final version for submission	Natalia Simon (IDSA)

Disclaimer

Copyright message

Table of Contents

List of Figures

List of Tables

Glossary of terms and abbreviations used

Abbreviation / Term	Description
AI	Artificial Intelligence
AMDEX	Amsterdam Data Exchange
AML	Anti-Money Laundering
API	Application Programming Interfaces
B2B	Business to Business
BDVA	Big Data Value Association
BMVIT	Bundesministerium für Verkehr, Innovation und Technology (Austrian Ministry for Innovation)
C2B	Consumer to Business
CA	Certificate Authority
CMS	Content Management System
CO2	Carbon Dioxide
CSV	Comma-separated values
DAPS	Dynamic Attribute Provisioning Service
DCAT	Data Catalogue Vocabulary
DIN	Deutsches Institut für Normung
DIO	Data Intelligence Offensive
DMA	Data Market Austria
EC	European Commission
ELG	European Language Grid
EOSC	European Open Science Cloud
EU	European Union
EU27	27 European Union Countries
FAIR	Findability, Accessibility, Interoperability, and Reuse (of digital assets)
FCAI	Finnish Center for Artificial Intelligence
FFG	Forschungfsörderungs Gesellschaft (National Funding Agency Austria)
GA	Grant Agreement

GDPR	General Data Protection Regulation
GUI	Graphical User Interface
HTML	Hypertext Markup Language
HTTPS	Hypertext Transfer Protocol Secure
ICT	Information and Communication Technology
IDC	International Data Cooperation
IDS	International Data Space
IDSA	International Data Spaces Association
IEEE	Institute of Electrical and Electronics Engineers
INSPIRE	Infrastructure for Spatial Information in the European Community
IoT	Internet of Things
IOTA	IOTA is funded by the non-profit IOTA Foundation. IOTA focuses on the IoT market with the goal to enable secure data transactions between data sellers and buyers.
IPR	Intellectual Property Rights
ISST	Institute for Software and Systems Engineering (IIST) of Fraunhofer
IT	Information Technology
JS	JavaScript
JSON	JavaScript Object Notation
kWh	Kilowatt hour
ML	Machine Learning
MSP(s)	Multisided Platforms
NL AIC	The Netherlands AI Coalition
P2B	Platform to Business
ParIS	Participant Information System
PoC	Proof of Concept
PSD2	Payment Service Directive 2;
RAM	Reference Architecture Model
SME	Small and Medium-sized Enterprises
SOLID	Social Linked Data
SSHOC	Social Sciences and Humanity Open Cloud

SSO	Single Sign On
STOF model	The STOF model consists of the service domain (S), technology domain (T), organisation domain (O), and finance domain (F). The STOF model is a framework that provides the logic of business and its ecosystem[2].
TFEU	Treaty on the Functioning of the European Union
TNO	Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek (Netherlands Organisation for Applied Scientific Research)
TRUSTS	Trusted Secure Data Sharing Space
UC	Use Case
UI	User Interface
US	United States
USP	Unique Selling Proposition
VM(s)	Virtual Machine(s)
W	Watt
W3C	World Wide Web Consortium
WCAG	Web Content Accessibility Guidelines
WP(s)	Work Package(s)

[2] *Bouwman et al., 2008*

Executive Summary

This study is meant to provide an overview of the current state and relevant trends on the topic of data marketplace to ensure that the projects' results are targeting real market needs and working with up-to-date requirements. Therefore, we have analysed the current state and created a holistic overview on both, the academic view on the current state of data marketplaces and different facets of data marketplaces in their "natural habitats" – considering the circumstances data marketplaces are embedded in. During our work, we have already collaborated closely with corresponding work packages and provided them with the latest results directly, so that the requirements have been updated continuously. Nevertheless, we will ensure that all recommendations of this final documentation have been updated and provided to the respective work packages as well. Further, we will communicate and promote the results of the study together with the TRUSTS team responsible for community engagement (work package 7, task 7.2) and the communication team (work package 8) to external stakeholders with the aim to foster the community around TRUSTS.

Within the analysis of the **academic landscape** (Chapter 3) we formulated a profound definition of data marketplaces from the academic perspective (Chapter 3.1.2), valid not only for the TRUSTS project but also as a valuable addition to the current academic state of the art on data marketplaces. Here, we defined a data marketplace as "a digital system where data is traded as an exchangeable economic good. It connects data providers and data buyers and facilitates data exchange and financial transactions. It has mechanisms to enforce laws, rules, and regulations to coordinate transactions, so that the trust of data marketplace users can be enhanced. (…)"[3]. In addition, we are providing a combined framework to classify data marketplaces based on their orientation and ownership (Chapter 3.2.2) as well as examples for such a classification and their matching mechanisms. This framework will then be re-used to elaborate the TRUSTS business model (in the respective work package - WP7).

In addition, we collected 35 functionalities by analysing relevant data marketplace literature (Chapter 3.3) which should be assessed by the TRUSTS group deciding on the TRUSTS functionalities. Most functionalities touch the category of "dataset discovery", "trading arrangements, "data governance", "data transformation" and "user management". This Chapter also shows the growth trend in the value of the data market and data economy in the European Union (EU) (Chapter 3.4.1). Furthermore, the number of academic publications in this area is rapidly increasing, which indicated the start of a *take-off phase* of data marketplace research. The primary research topic of data marketplaces can be divided into six clusters. These are *pricing mechanisms, privacy themes in personal data markets, general context of data markets, technical literature, service offerings, and data markets in IoT* (Chapter 3.4.2). This work also highlights data marketplace challenges identified within literature. These challenges were categorised using the STOF model (Chapter 3.5) and will function as guidance during the development of TRUSTS. We also touched upon the topic of data marketplace fragmentation. In general, fragmentation triggers multiple aspects of data marketplaces (e.g., business models, governance arrangements, and technical standards) to diverge uncontrollably, leading to a decrease of trust in the concept of data marketplaces (Chapter 3.6). Therefore, a federated approach to overcome the fragmentation issues can be considered as a potential solution.

[3] See Chapter 3.1.2 for the complete definition.

Next to this, we elaborated an overview on data marketplaces' macroenvironment (Chapter 4) by addressing five different areas data marketplaces are influenced by and which therefore may have a strong impact on the success of the project. Thus, we have monitored current developments in these areas and evaluated their relevance for TRUSTS to align the TRUSTS project results with current circumstances and occurrences. Those five areas are: political, economic, social, technical, legal, and environmental.

We found out, that on a **political** level (Chapter 4.1), TRUSTS should be aligned with the GAIA-X project, as well as approach the BDVA i-Spaces as a strong and relevant community in this area and consider the recently published results of the Horizon 2020 project OPEN DEI "Aligning Reference Architectures, Open Platforms and Large-Scale Pilots in Digitizing European Industry" as a guidance for future development. Considering current **economic** trends in the field of data marketplaces (Chapter 4.2), one can say that the data-driven economy will continue to grow over the next years which is why the potential for data marketplaces will also grow as central locations for sharing and trading data. Thus, privacy concerns and security issues still need to be addressed to increase adoption, which is exactly where TRUSTS is aiming at. Also, pricing strategies and economic incentives need to be clarified - currently, companies are reluctant to share data because the own benefit or value is unclear. Regarding **social** aspects that touch data marketplaces (Chapter 4.3) one can say that the requirements for skilled data professionals will grow over the next years and that work must be done to explain and make clear the benefits for data providers in order to raise the frequency of and willingness for data sharing. Hence, the adherence to a **legal** framework (Chapter 4.5) is crucial which is why it has been recommended here that TRUSTS processes should ensure compatibility with the Legal and Ethical Recommendations presented in the publicly available TRUSTS deliverable[4], to ensure legal and also ethical compliance. When it comes to **environmental** aspects (Chapter 4.6) a main message we found is that data-driven economy has a significant impact on our environment. Technologies such as the blockchain (e.g. for smart contracting), server centre, and training machine learning (ML) models require significant amounts of energy. Data markets might contribute to this problem but might also help to alleviate it: a) sharing data and models helps to re-acquiring data or retrain the models and b) research initiatives and business models working against climate change might benefit from the shared data and can create innovative solutions.

In the field of **technical** developments (Chapter 4.4) one of the messages is that TRUSTS' technical team take into account the areas of (i) Smart Contracting and (ii) the federated architecture approach and evaluated it in detail. They may also have a look at the overall concept of Semantic Data Fabrics. Many parts and components of such a Semantic Data Fabric are used in different data markets already today, but the integrated combination could, from a technology perspective, be the breakthrough for data markets and data spaces. In addition, deployment methods based on virtualisation should be considered to provide a stable and future proven deployment and operation environment, and finally TRUSTS should be built on top of existing standards and/or standards under current development like DCAT-AP or International Data Spaces (IDS) and GAIA-X.

Analysing the **competitive environment** of data marketplaces, it becomes clear that data marketplaces should always be understood as ecosystems consisting of a set of aligned value-creating capabilities and technology layers / components, with the data marketplace as its central anchor point. Platform

[4] https://www.trusts-data.eu/wp-content/uploads/2020/10/D6.2-Legal-and-Ethical-Requirements.pdf (March 2021)

users and ecosystem components should therefore be attracted to TRUSTS by using and promoting an open, component-based, standards-based architecture to optimise interoperability while supporting the incorporation of and compliance with existing and emerging European standards. To achieve multilateral network effects in a rich ecosystem of data assets and interoperable data services, ecosystem federation is required. In doing so, data market federation connects European data markets and data aggregators and links the accessibility of semi-public cloud systems, as envisaged in the European Data Strategy. TRUSTS can therefore focus on selected data asset domains, providing the depth and temporally stable context required for commercially viable data trading and related services, while providing a portal into the broader European data landscape. A future TRUSTS operator should explore a range of auxiliary services beyond the scope of the project to create a steadfast business moat against the competitive landscape. These include but are not limited to (1) public data harvesting and preparation, (2) support for data provider onboarding through data integration and orchestration, and meta data quality assurance, (3) enablement of co-creation of orchestrated data sets through 3rd party Data Circles as introduced in the Data Market Austria project. In addition, a TRUSTS platform operator should aim to seamlessly internalise essential services, namely the commissioning/brokerage of the computing infrastructure to facilitate SMEs and data-driven start-ups.

Our study did also analyse the current state of the **Financial** (Chapter 6), **Telecom** and **IC(T)** (Chapter 7) Industries with regard to data marketplaces. Here, we conclude that the involvement of financial enterprises in the data marketplace ecosystem can function as a catalyst for the success of this field. Financial industries are digitally transformed and bringing in a flexible and contemporary way to target new economic challenges with easily customised financial instruments. The analysis in the Telecom domain demonstrated that companies in this respective business mainly act as data providers and operators of data sharing platforms. Their value addition in the process of service delivery and sustainability is significant since the respective operators have vast experience in the quality operation processes, market penetration in all segments and they are considered and trusted entities by the whole community. We therefore formulated the recommendation that TRUSTS should define a straightforward commercial model addressing real enterprise data trading and data analysis needs. Business sustainability, transactions integrity and privacy preservation processes are key for the success of the endeavour.

1 Introduction

This study is a result of the TRUSTS project and linked to the specific project task (T2.1) called "*EU, and worldwide data market".* It is meant as an instrument linking the project's activities to current market needs by providing a comprehensive overview on the EU and worldwide data market, its trends, and recent developments. This study builds on the state-of-the-art provided within the project proposal and is considering latest activities on the market, from which suitable recommendations will be derived for the TRUSTS project, serving the project's objective to analyse trends and define requirements for the provisioning of a suitable data marketplace service.

The study is divided into five main topics, that have been targeted:

1. the object "data market" itself (Chapter 3),
2. the macro- (Chapter 4) and
3. the microenvironment of the data markets (Chapter 5) as well as domains specifically of relevance for the TRUSTS project, namely
4. the financial (Chapter 6) and
5. the telecom operator industry (Chapter 7).

First information on these topics has been gathered by a desk research and has been enriched by results of a *world café workshop* (Chapter 2.1) that took place with external experts in order to get insights from the market and industry directly. Afterwards, the information collected in both ways has been evaluated and assessed internally so that suitable recommendations have been derived for relevant work packages (WP) and tasks of the TRUSTS project, addressing functional and non-functional requirements. The aim of this study is to provide suitable recommendations in order to ensure a high market relevance of the TRUSTS outcome.

Within the first topic (Chapter 3), the definition of data marketplaces, we present and analyse current academic material which then mounts into a definition for data marketplaces, also relevant beyond TRUSTS. This Chapter is shedding light on functionalities a data marketplace should consider having and what challenges data marketplaces are currently facing.

The second topic (Chapter 4), the macro-environmental analysis, is focusing on relevant developments and trends in the direct surrounding of data marketplaces. Here, we are touching topics currently of high relevance from a political perspective on data marketplaces, as well as economic, legal, environmental, and social. We also analysed technical trends that could be of interest for TRUSTS. Summarizing the results of the analysis we are providing guidelines for the TRUSTS project, as well for technical issues as for business.

Within the third topic (Chapter 5), the micro-environmental analysis, we are setting the foundations for the business development and market positioning of TRUSTS. Here, we worked together closely with the project's team dedicated to this topic (WP7) while analysing the direct competitive environment and discussing relevant formats for TRUSTS.

In the fourth (Chapter 6) and fifth topic (Chapter 7), we examine the financial as well as telecom and ICT industry with regard to their touching points with data marketplaces and identify issues that TRUSTS can address.

In the end, we summarise the results of the above-mentioned topics and fields and are pointing out relevant recommendations and guidelines we formulated.

1.1 Mapping Projects' Outputs

Table 1: Adherence to TRUSTS GA Deliverable & Tasks Descriptions

TRUSTS Task		Respective Document Chapter(s)	Justification
T2.1	*This task will complement and update the State-of-the-Art presented in Section 1.3 (cf. proposal) and will reconfirm the innovative potential of the TRUSTS Platform.* *Furthermore, the financial and operator industries growth needs having data marketplaces as catalysts including respective regulatory and standardization state of the art and trends will be evaluated and analysed to align target industry needs with the overall data space requirements.* *In addition, established and confidentiality/privacy preservation will be given particular attention best practices in terms of secure data sharing as well as their positioning vs. applicable standards and regulations.* *This market survey and benchmarking of existing solutions and in development state is important to inform the analysis of requirements at T2.2 and establishment of competitive and/or complementary specifications at T2.4 and guide the demonstration work.* *This task, coordinated by IDSA, will be performed by each partner in their field.*	*Chapter 3* *Chapter 6 and Chapter 7* *Chapter 3.7* *(Noted throughout the document as "recommendation")*	*Additions regarding the existing data market ecosystems (IDSA and DMA) and comprehensive overview on the data market landscape and definitions, as well as taxonomies.* *IDSA & DMA as best practices* *Recommendations for TRUSTS*

TRUSTS Deliverable
D2.1: Definition and analysis of the EU and worldwide data market trends and industrial needs for growth *Report containing the detailed analysis of the EU & worldwide data (market) economy (regulatory framework, trends, growth rates, best practices et al.) especially in financial and operator industries where the growth requires the establishment of data marketplaces as catalysts as well as the respective regulatory and standardization state of the art and trends.*

1.2 Deliverable Overview and Report Structure

This deliverable is structured according to the different topics that have been analysed within. Every Chapter is having a closer look at a specific topic about data markets in order to provide a holistic depiction of the current state. In the end of each Chapter, recommendations towards respective project tasks and WPs are indicated. The scope of this deliverable is therefore to identify current needs in the area of data marketplaces and to direct them to suitable WPs of the project, where they should be assessed with regard to their utilisation. In this sense, it is not in the scope of this deliverable to decide whether the needs identified are to be implemented in TRUSTS or not. This deliverable should function as an overview on the current state and therefore as a guideline.

The first Section (Chapter 2) is dedicated to the **methodology** that has been used and is describing the approach that has been followed. In order to achieve a common understanding, the second Section (Chapter 3) is explaining the deliverables view on **the objective "data marketplace"**, defining what is a data marketplace, what types and functions are existing, what are the current trends from the academical perspective and how can TRUSTS leverage this in order to achieve a sustainable output of TRUSTS. The following Section (Chapter 4) is embedding data marketplaces into a **general environmental analysis**, considering political, economic, social, legal, environmental, and technological aspects, whereas the legal aspects are referring to the TRUSTS project deliverable 6.2 "Legal and Ethical Requirements". This analysis will be completed by mapping the ***competitive*** **environment** of data marketplaces (Chapter 5). In the end, the industries relevant for TRUSTS (**Financial** and **Telecom industry**) will be considered (Chapters 6 and 7). The study then closes with a summary of relevant recommendations targeting the above-mentioned WPs (Chapter 8).

2 Methodology

To provide a comprehensive study on the current state of data markets and their relation regarding different fields, like for instance its influence on economy or even social aspects, this deliverable has focused on two methodologies: a secondary research, and the World Café workshop as a practical approach for gathering information. These two approaches will be described in the following.

The secondary research has been chosen in order to collect, process and interpret existing data to provide a presentation of the current state of research and a critical examination of the presented approaches in the literature.

2.1 World Café Method

In order to obtain insight from experts in the data-driven economy, a World Café was held on March 17, 2021. The World Cafe is a structured approach of brainstorming in medium-sized groups, i.e. ~ 20 people and 4 Table chairs. Typically, a set of four different topics is up for discussion. The group of participants is divided into four groups, ideally of the size of 3-5 people. Each group starts discussing one topic at a dedicated "table", moderated by a Table chair. After 15 minutes, the groups switch to the next Table and discuss the next topic for another 15 minutes. This procedure is repeated four times in total, so that each group has discussed each topic.

Subsequently, the insights gained from the four brainstorming rounds are summarised and presented to the entire audience by the Table chairs. The advantage of the World Cafe is that the entire group of participants provides their knowledge, ideas, and expertise on all four topics. The iterations after the initial round build on top of the insights gained in the first round, which means that more and more insights are gained throughout the process. Furthermore, this helps to get insights from a potentially highly diverse set of participants and illuminates the topics from different perspectives.

The concept of the World Café is a highly professional and effective approach and is usable in many different domains and application areas[5].

2.2 Bibliometric analysis

To identify academic publication trends in existing literature (see Chapter 3.4.2 Academic publications), we employ the bibliometric analysis. The bibliometric analysis is a method to quantitatively analyse existing literature of a specific topic. "The main objects of bibliometric are the amount of literature (various publications, especially journal papers and citations), the number of authors (individual group or group), the number of vocabulary (various literature marks, among which the majority are descriptors)."[6]

We start to conduct the bibliometric analysis by identifying literature from the academic perspective. We performed a literature search in the *Scopus database* using keywords of "**data market***" and "**data marketplace***". We found 497 documents, and then we added ten other papers that we considered as key literature[7]. We conducted the literature search on **6 July 2020,** meaning any publications after this date were not considered for the analysis.

[5] http://www.theworldcafe.com/key-concepts-resources/world-cafe-method/ (February 2021)
[6] https://encyclopedia.pub/2024 (February 2021)
[7] The database of literature can be accessed here: https://doi.org/10.4121/14673813.v1

Next, we read the articles' titles and abstracts to include only relevant literature. We excluded 225 papers due to the following reasons (refer Table 2).

Table 2: Reasons for articles' exclusion

No	Reason for exclusion	Number
1	The study is not about the data marketplaces or data markets	160
2	The study merely focuses on data marketplaces or data markets as the core of the research	33
3	The study is a workshop/proceeding description, not a research paper (journal or conference paper)	20
4	The article is not in English	7
5	No abstract available	5
Total		225

After collecting the literature, a network visualisation of the keywords used in articles was created to explore the topics discussed by scholars. A map was created based on bibliometric data that only analyses the terms that occur more than four times in the articles. From 1836 keywords, 89 meet the threshold. The VOSviewer was used to conduct such visualisation. The result of these visualisations can be seen in Chapter 2.2.3.2 - academic publications.

3 Definition of EU and Worldwide Data Marketplaces

The main objectives of this Chapter are fivefold:

a. To define data marketplaces and the key actors involved in them (Chapter 3.1);
b. To explore types of data marketplaces (Chapter 3.2). This Chapter will distinguish data marketplaces based on the determinant of a) orientation and ownership and b) different types of data traded in data marketplaces.
c. To identify data marketplace features (Chapter 3.2.2). It summarises features found in existing data marketplace literature.
d. To explore data marketplace trends (Chapter 3.4). It discusses data marketplace trends from both industry and academic perspectives; and
e. To discuss the main challenges of data marketplaces (Chapter 3.5).

This Chapter's findings will be beneficial to state the stage and position of TRUSTS in the European data economy by providing specific recommendations to the architectural design (WP2), the platform development (WP3) and the formulation of sustainable business models (WP7).

3.1 Data marketplace definitions and actors

This Chapter describes the history of data marketplaces to explore data marketplace evolvement. It is especially beneficial to understand current situations and to recommend future research agendas. We will also define data marketplaces and its associated actors.

3.1.1 History of data marketplaces

The annual amount of data on our planet, data created or copied, is expected to rise from 33 zettabytes in 2018 to 175 zettabytes in 2025. The EU's data economy value is predicted to reach €829 billion; the number of data professionals will hit 10.9 million in 2025.[8] Studies such as that conducted by W. Thomas and Leiponen[9] have shown that there is an ever-increasing interest in industrial datasets commercialisation, which goes beyond its primary use to a secondary one: data sharing between organisations. As an organisation may not always possess the required data to carry out or improve their processes and services, they may wish to purchase data from other organisations. Data marketplaces are a new phenomenon, of which instances have recently emerged in the market and seem to be one approach to enable industrial dataset commercialisation. In general, data marketplaces provide digital systems through which individuals and organisations can exchange data. [10]

Despite the potential benefits of data marketplaces, in practice very little data is shared or traded via platforms.[11] Many of the data marketplaces that have been set up have failed or are shut down. Swivel.com, a commercial data platform that offered visualisation services, closed because there were less than ten customers on the platform.[12] Kasabi, a data marketplace for published linked data, shut down in 2012 because at the time, the growth of the market for data was too slow for the business to

[8] European Commission, "A European Strategy for Data," Brussels, 2020; D. Reinsel, J. Gantz, and J. Rydning, "Data age 2025: the digitization of the world from edge to core," IDC White Paper, pp. 1-29, 2018.
[9] L. D. W. Thomas and A. Leiponen, "Big data commercialization," IEEE Engineering Management Review, vol. 44, no. 2, pp. 74-90, 2016, doi: 10.1109/emr.2016.2568798.
[10] F. Stahl, F. Schomm, G. Vossen, and L. Vomfell, "A classification framework for data marketplaces," Vietnam Journal of Computer Science, vol. 3, no. 3, pp. 137-143, 2016, doi: 10.1007/s40595-016-0064-2; F. Schomm, F. Stahl, and G. Vossen, "Marketplaces for data: an initial survey," ACM SIGMOD Record, vol. 42, no. 1, pp. 15-26, 2013.
[11] P. Koutroumpis, A. Leiponen, and L. D. W. Thomas, "Markets for data," Industrial and Corporate Change, vol. 29, no. 3, pp. 645-660, 2020, doi: 10.1093/icc/dtaa002
[12] Kosara, R. 2010. 12 October 2010. "The Rise and Fall of Swivel.com." https://eagereyes.org/criticism/the-rise-and-fall-of-swivel accessed on May 12, 2020

be sustainable.[13] Microsoft Azure DataMarket, one of the first movers to enter the data marketplaces, was closed six years after its launch due to "a lack of customer interest".[14]

Schomm et al.[15] are among the first scholars who discussed data marketplaces. They conducted a survey and identified dimensions and categories of data marketplaces. Following this, a considerable amount of literature has been published in this domain. Core publications about data marketplaces focus on data marketplace trends[16], a classification framework for data marketplaces[17], the potential of data marketplaces[18], business model taxonomies[19], and requirements for data sharing[20].

Data marketplaces emerged in EU policymaker documents around 2017 when IDC and Open Evidence[21] reported the EU data economy trends. The European Commission continues this study and produces updates on the latest report of "The European data market monitoring tool".[22] Data marketplaces are one of the key instruments to accomplish the EU vision to create a single European Data Market by 2030, as mentioned in another EC report "A European Strategy for Data". It aims to release the full potential of data flow and use across Europe.[23]

In Chapter 4.1, the topic of political conditions and activities will be discussed in more detail.

3.1.2 Data marketplace definitions and actors

To understand the data marketplace phenomenon, this Section discusses the data marketplace definitions used by various acknowledged scholars in recent literature, i.e., in the past two years (see Table 3).

[13] Johnson, B. 2012. "Kasabi shuts down, says data marketplace 'too slow'". https://gigaom.com/2012/07/09/kasabi-shuts-down-says-data-marketplace-too-slow/ accessed on May 12, 2020; Dodds, L. 2012. "Kasabi – Shutting Down Kasabi." https://kasabi.wordpress.com/2012/07/09/shutting-down-kasabi/ accessed on May 12, 2020

[14] Ramel, D. 2016. "Microsoft Closing Azure DataMarket." https://adtmag.com/articles/2016/11/18/azure-datamarket-shutdown.aspx accessed on May 12, 2020

[15] F. Schomm, F. Stahl, and G. Vossen, "Marketplaces for data: an initial survey," ACM SIGMOD Record, vol. 42, no. 1, pp. 15-26, 2013.

[16] F. Stahl, F. Schomm, and G. Vossen, "Data Marketplaces: An Emerging Species," in DB&IS, 2014, pp. 145-158.

[17] F. Stahl, F. Schomm, G. Vossen, and L. Vomfell, "A classification framework for data marketplaces," Vietnam Journal of Computer Science, vol. 3, no. 3, pp. 137-143, 2016, doi: 10.1007/s40595-016-0064-2.

[18] P. Koutroumpis, A. Leiponen, and L. D. Thomas, "The (unfulfilled) potential of data marketplaces," ETLA Working Papers, 2017.

[19] M. Fruhwirth, M. Rachinger, and E. Prlja, "Discovering Business Models of Data Marketplaces," in Proceedings of the 53rd Hawaii International Conference on System Sciences, 2020; M. Spiekermann, "Data Marketplaces: Trends and Monetisation of Data Goods," Intereconomics, vol. 54, no. 4, pp. 208-216, 2019, doi: 10.1007/s10272-019-0826-z.

[20] P. Koutroumpis, A. Leiponen, and L. D. W. Thomas, "Markets for data," Industrial and Corporate Change, vol. 29, no. 3, pp. 645-660, 2020, doi: 10.1093/icc/dtaa002.

[21] IDC and Open Evidence, " European Data Market Study," 2017.

[22] European Commission, "The European Data Market Monitoring Tool," Brussels, 2020.

[23] European Commission, "A European Strategy for Data," ed. Brussels, 2020.

Table 3: Data marketplace definitions and actors

No	Source	Data marketplace definition	Data marketplace actors
1	Fruhwirth et al. (2020), p. 5739	"…electronic platforms that facilitate the exchange of data."[24]	Data providers, data buyers, third-party service providers, marketplace owners
2	Hayashi and Ohsawa (2020a), p. 2	"…platforms where data are traded as exchangeable economic goods."[25]	Data providers, data users
3	(Hayashi and Ohsawa, 2020b), p. 35469	"…online marketplaces where companies that want to buy and sell data (buyers and suppliers) from different areas can participate freely."[26]	Data buyers, data suppliers, data platformers
4	Iwasa et al. (2020), p. 66	"…a platform to buy/sell data, works as a method for accelerating data exchange."[27]	Data holders, data users
5	Koutroumpis et al. (2020), p. 647	"…multisided platforms, where a digital intermediary connects relevant actors. Such platforms could generate value through lower transactional frictions, resource allocation efficiency, and improved matching between supply and demand."[28]	Data providers, data purchasers, other complementary technology providers
6	Sharma et al. (2020), p. 39	"…a trading platform where the data producers can sell the data and others can buy it."[29]	Data producers, data buyers
7	Zheng et al. (2020), p. 769	"…centralized platforms, where data vendors can upload and sell their data,	Data vendors, data consumers

[24] M. Fruhwirth, M. Rachinger, and E. Prlja, "Discovering Business Models of Data Marketplaces," in Proceedings of the 53rd Hawaii International Conference on System Sciences, 2020.
[25] T. Hayashi and Y. Ohsawa, "TEEDA: An Interactive Platform for Matching Data Providers and Users in the Data Marketplace," Information, vol. 11, no. 4, p. 218, 2020.
[26] T. Hayashi and Y. Ohsawa, "Understanding the Structural Characteristics of Data Platforms Using Metadata and a Network Approach," IEEE Access, vol. 8, pp. 35469-35481, 2020, doi: 10.1109/access.2020.2975064.
[27] D. Iwasa, T. Hayashi, and Y. Ohsawa, "Development and Evaluation of a New Platform for Accelerating Cross-Domain Data Exchange and Cooperation," New Generation Computing, vol. 38, no. 1, pp. 65-96, 2020, doi: 10.1007/s00354-019-00080-0.
[28] P. Koutroumpis, A. Leiponen, and L. D. W. Thomas, "Markets for data," Industrial and Corporate Change, vol. 29, no. 3, pp. 645-660, 2020, doi: 10.1093/icc/dtaa002.
[29] P. Sharma, S. Lawrenz, and A. Rausch, "Towards Trustworthy and Independent Data Marketplaces," 2020 2020: ACM, doi: 10.1145/3390566.3391687.

		and data consumers can discover and purchase the data needed."[30]	
8	Agahari et al. (2019), p. 2	"…a multi-sided platform that enables organizations to share and sell datasets."[31]	Data providers, service providers, application developers, infrastructure and tool providers, data brokers, application users, data consumers
9	Spiekermann (2019), p. 4	"…a digital platform on which data products are traded. These platforms must act like a neutral intermediary and allow anyone (or at least a large number of potentially registered customers) to upload and sell their data products."[32]	Data providers, data buyers, third party service providers, data marketplace owners
10	Truong et al. (2019), p. 178	"… marketplace where owners sell and consumers buy data"[33]	Data owners, data customers

At first, the main actors involved in data marketplaces will be discussed and an emphasis will be given to the term 'marketplace', as this term distinguishes data marketplaces with similar terms in the literature (e.g., data exchange services). The marketplace is defined in the Cambridge Dictionary as a place to shop, where things are sold.[34] It is a trading system where buying and selling activities occur. Koutroumpis et al.[35] also uses the term 'marketplace' to emphasise "commercial exchange takes place as a result of buyers and sellers being in contact with one another." This deliverable keeps in mind the 'marketplace' understanding as this also fits with the two actors (i.e., data providers and data buyers) mentioned in the previous elaboration.

Moreover, based on the scholars' definitions, there is a strong indication that data marketplaces possess *platform* characteristics. Platforms function as a mediator between user groups.[36] Platforms aim to match these user groups and facilitate the exchange of goods or services.[37] Therefore, data

[30] Z. Zheng, Y. Peng, F. Wu, S. Tang, and G. Chen, "ARETE: On Designing Joint Online Pricing and Reward Sharing Mechanisms for Mobile Data Markets," IEEE Transactions on Mobile Computing, vol. 19, no. 4, pp. 769-787, 2020, doi: 10.1109/tmc.2019.2900243.

[31] W. Agahari, M. d. Reuver, and T. Fiebig. D. U. o. Technology. (2019). Understanding how privacy-preserving technologies transform data marketplace platforms and ecosystems: The case of multi-party computation.

[32] M. Spiekermann, "Data Marketplaces: Trends and Monetisation of Data Goods," Intereconomics, vol. 54, no. 4, pp. 208-216, 2019, doi: 10.1007/s10272-019-0826-z.

[33] H. T. T. Truong, M. Almeida, G. Karame, and C. Soriente, "Towards Secure and Decentralized Sharing of IoT Data," 2019 2019: IEEE, doi: 10.1109/blockchain.2019.00031.

[34] Cambridge Dictionary, https://dictionary.cambridge.org/dictionary/english/marketplace, accessed on November 05, 2020

[35] P. Koutroumpis, A. Leiponen, and L. D. W. Thomas, "Markets for data," Industrial and Corporate Change, vol. 29, no. 3, pp. 645-660, 2020, doi: 10.1093/icc/dtaa002.

[36] J.-C. Rochet and J. Tirole, "Platform competition in two-sided markets," Journal of the european economic association, vol. 1, no. 4, pp. 990-1029, 2003.

[37] G. G. Parker, M. W. Van Alstyne, and S. P. Choudary, Platform revolution: How networked markets are transforming the economy and how to make them work for you. WW Norton & Company, 2016.

marketplaces attempt to help the two most mentioned actors in Table 3 (i.e., *seller*- and *buyer*-side) accomplishing each other's needs. On the seller side, the term ***data providers*** is used (although other terms like data buyers, data owners, and data producers are often interchangeable) to define those who provide and offer raw, refined, or analysed data. [38]On the buyer side, ***data buyers*** (often used interchangeably with data customers, data users and data purchasers) refer to those interested in buying datasets provided by data providers.[39]

Beyond these two core users, data marketplaces also consider the importance of complementary actors, or what both Fruhwirth et al. and Spiekermann termed as ***third-party service providers***, whose primary role is to add additional services, for example, to provide application or algorithm to ease data access & usage.[40] Both scholars also mention the role of ***data marketplace owners***, which is closely related to platform sponsors' concept*:* single or multiple firms who own the property rights and are responsible for developing platform technology. Platform sponsors design the platform components and rules and decide who may participate in the network.[41] The ***data marketplace operators*** **are also identified** as a core actor*,* for example*,* commercial entities that operate the platform when running as a business.

Other actors of data marketplaces include ***data brokers*** (i.e., the matchmaker, independent traders, and trader facilitators of data providers and data sellers) and ***infrastructure providers***, that are responsible for providing infrastructures (e.g., storage space and computing power).[42] More than two actor categories are identified, meaning that data marketplaces can further be specified and have *multisided platforms (MSPs)* characteristic as it mediates multiple user groups.[43]

In general, scholars' data marketplace definitions are in line with EU policymakers. They define a data marketplace as "an online platform uses a trusted third-party intermediary for data transactions. Data access may be provided through bilateral contracts."[44] In another report, the EC defines data marketplaces as an infrastructure where "digital data is exchanged as *products* or *services* as a result of the elaboration of raw data."[45]

Based on the previous elaboration, we identify four core functionalities that data marketplaces should fulfil:

---

[38] W. Agahari, M. d. Reuver, and T. Fiebig. D. U. o. Technology. (2019). Understanding how privacy-preserving technologies transform data marketplace platforms and ecosystems: The case of multi-party computation; M. Spiekermann, "Data Marketplaces: Trends and Monetisation of Data Goods," Intereconomics, vol. 54, no. 4, pp. 208-216, 2019, doi: 10.1007/s10272-019-0826-z.

[39] M. Fruhwirth, M. Rachinger, and E. Prlja, "Discovering Business Models of Data Marketplaces," in Proceedings of the 53rd Hawaii International Conference on System Sciences, 2020.

[40] A. Muschalle, F. Stahl, A. Löser, and G. Vossen, "Pricing approaches for data markets," in International workshop on business intelligence for the real-time enterprise, 2012: Springer, pp. 129-144

[41] T. R. Eisenmann, G. Parker, and M. Van Alstyne, "Opening platforms: how, when and why?," Platforms, markets and innovation, vol. 6, pp. 131-162, 2009.

[42] W. Agahari, M. d. Reuver, and T. Fiebig. D. U. o. Technology. (2019). Understanding how privacy-preserving technologies transform data marketplace platforms and ecosystems: The case of multi-party computation.

[43] K. J. Boudreau and A. Hagiu, "Platform rules: Multi-sided platforms as regulators," Platforms, markets and innovation, vol. 1, pp. 163-191, 2009.

[44] European Commission, "Work stream on Data," 2020, p. 36

[45] European Commission, "Work stream on Data," 2020, p. 37

1. Data has become the essential exchangeable economic good for the trading commodity;
2. Data marketplaces should match data providers and data buyers;
3. Data marketplaces should facilitate transactions: mechanisms for logistics and settlement should lead to the transportation of the sold product, and transfer of payment via digital infrastructure; and
4. Data marketplaces should have mechanisms to enforce laws, rules, and regulations to coordinate transactions, so that the trust of data marketplace users can be enhanced.

Thus, the generic definition from data marketplaces from the academic perspective is elaborated as follows:

> **Data Marketplace Definition:** A digital system where data is traded as an exchangeable economic good. It connects data providers and data buyers and facilitates data exchange and financial transactions. It has mechanisms to enforce laws, rules, and regulations to coordinate transactions, so that the trust of data marketplace users can be enhanced. Key actors that provide data marketplace functionalities include data marketplace owners, operators, and third-party providers (TPPs). Other actors to support data marketplaces are infrastructure providers and independent data brokers.

Figure 1 provides a general illustration of a data marketplace.

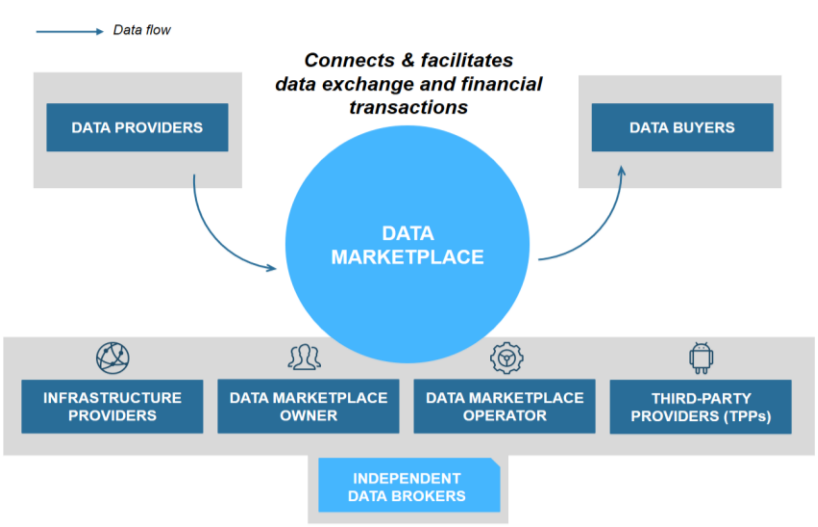

Figure 1: Data marketplace definition and actors

The data marketplace definition specifically built for TRUSTS will be elaborated in WP7. Nevertheless, this generic definition is in line with TRUSTS vision. For example, TRUSTS as a data marketplace also has a paradigm to trade data as economic goods. Beyond that, it also provides services and applications to support and facilitate data trading between two core actors, which are data providers and data services.

3.2 Types of data marketplaces

According to Spiekermann (2019), in theory, data marketplace participants trade data on multilateral basis. However, in reality, companies rarely trade industrial datasets on multilateral data marketplaces and preferably trade data on bilateral basis (Koutroumpis et al., 2017). To better understand data marketplaces that go beyond theoretical concepts, the business models of data marketplaces that occur in practice are offered. This indicates the need to consider different types of data marketplaces.

Thus, in this Section, the types of different data marketplaces are distinguished based on the determinant of:

a) orientation and ownership, or
b) data types.

3.2.1 Data marketplaces based on types of data traded

The data marketplace based on the different types of data traded is also analysed. Firstly, the desk research was conducted in order to compile a database of existing data marketplaces to achieve this objective. Several different sources that link to data marketplace websites were considered to identify data types:

1. Sixty-five websites of data marketplaces that were mentioned and analysed in existing studies of data marketplaces were included in the database. It concerns the following studies.[46]
2. The data discovery platform datarade.ai[47] was accessed, a website that provides an overview of 1800+ data providers, 200+ data platforms, and 200+ data categories.
3. To complement the database with data marketplaces that were not included by following steps 1 and 2, the search engine Google was utilised to conduct desk research.

This desk research resulted in a final database consisting of 178 cases of data marketplaces[48]. Afterwards, the data marketplaces based on data types were labelled. The labels were selected from datarade.ai as a starting point, then complemented with the work of Fruhwirth et al., and Spiekermann.

Table 4 provides an overview of the selected data marketplaces based on their data type.

[46] P. Koutroumpis, A. Leiponen, and L. D. W. Thomas, "Markets for data," Industrial and Corporate Change, vol. 29, no. 3, pp. 645-660, 2020, doi: 10.1093/icc/dtaa002; P. Koutroumpis, A. Leiponen, and L. D. Thomas, "The (unfulfilled) potential of data marketplaces," ETLA Working Papers, 2017; E. Prlja, "Discovering Business Models of Data Marketplaces," Graz University of Technology, 2019; M. Spiekermann, "Data Marketplaces: Trends and Monetisation of Data Goods," Intereconomics, vol. 54, no. 4, pp. 208-216, 2019, doi: 10.1007/s10272-019-0826-z; F. Stahl, F. Schomm, G. Vossen, and L. Vomfell, "A classification framework for data marketplaces," Vietnam Journal of Computer Science, vol. 3, no. 3, pp. 137-143, 2016, doi: 10.1007/s40595-016-0064-2; P. Carnelley, H. Schwenk, G. Cattaneo, G. Micheletti, and D. Osimo, "Europe's data marketplaces—current status and future perspectives,'" European Data Market SMART, vol. 63, 2016.

[47] *https://datarade.ai/, accessed on November 05, 2020*

[48] Readers may access the database here: https://doi.org/10.4121/14679564.v1

Table 4: Labelling of data marketplaces based on data type and number of cases in the database

Segmentation of data marketplaces by type of data	Number of cases in the database (N)
Agriculture data	2
Alternative data	5
Any data	8
Audience data	112
B2B data	9
Connected car data, automotive data	4
Data for AI and machine learning	3
Environmental data	1
Financial data	2
Financial data, alternative data	5
Financial data, market data	2
Healthcare data	8
Location data	6
Parking data	1
Personal data	4
Real estate data	1
Sensor data	2
Satellite data	1

Traffic data, petrol price data, parking data	2
Total	**178**

Recommendation for TRUSTS

The combination of the classification frameworks, depending on their orientation and ownership, shown in Figure 2, should be considered in WP7 as a basis for the formulation of sustainable business models. In addition, WP7 should position TRUSTS as a data marketplace in this classification schema.

3.2.2 Data marketplace based on orientation and ownership determinants

Stahl et al.[49] propose a framework that enables data marketplace classification in different types. They make use of two determinants: orientation and ownership.

1. **Orientation** refers to whether the data marketplace owner coordinates data trade in a hierarchical or market trading structure. In data marketplaces with a hierarchical orientation, the data marketplace owner determines the data price and what data providers and buyers are allowed. In data marketplaces with a market orientation, prices are determined by data providers and buyers depending on competitive offerings.
2. **Ownership** indicates whether one private company, a number of companies or an independent party owns the data marketplace.

Koutroumpis et al.[50] maintain a similar classification in which they sort data marketplaces based on their matching mechanism. They distinguish between four types of data marketplaces; one-to-one, one-to-many, many-to-one and many-to-many. First, one-to-one data marketplaces are bilateral marketplaces where two parties are directly connected. One provider trades with one buyer. Second, at one-to-many data marketplaces there is one provider who trades with many buyers for the same data. In this case, standardised terms of exchange through APIs are maintained, because it is too costly to negotiate data individually. Third, many-to-one data marketplaces allow multiple providers and one buyer. Providers usually make their data available to one buyer and receive a service in return for free, as practiced on social media platforms. Fourth, many-to-many data marketplaces are multilateral marketplaces where many providers and buyers trade data. There is often no specific ownership over the data, but transactions to acquire data are facilitated.

[49] F. Stahl, F. Schomm, G. Vossen, and L. Vomfell, "A classification framework for data marketplaces," Vietnam Journal of Computer Science, vol. 3, no. 3, pp. 137-143, 2016, doi: 10.1007/s40595-016-0064-2.
[50] P. Koutroumpis, A. Leiponen, and L. D. Thomas, "The (unfulfilled) potential of data marketplaces," ETLA Working Papers, 2017.

The framework proposed by Koutroumpis et al.[51] and Stahl et al.[52] were combined as illustrated in Figure 2 . This combination shows the spectrum in which different types of data marketplaces can be classified, depending on their orientation and ownership. Stahl et al. identified six types of marketplaces of which three overlap with the one-to-many, many-to-one and many-to-many data marketplaces as defined by Koutroumpis et al. The one-to-one data marketplace defined by Koutroumpis et al. is added to the selection, resulting in a total of seven types.

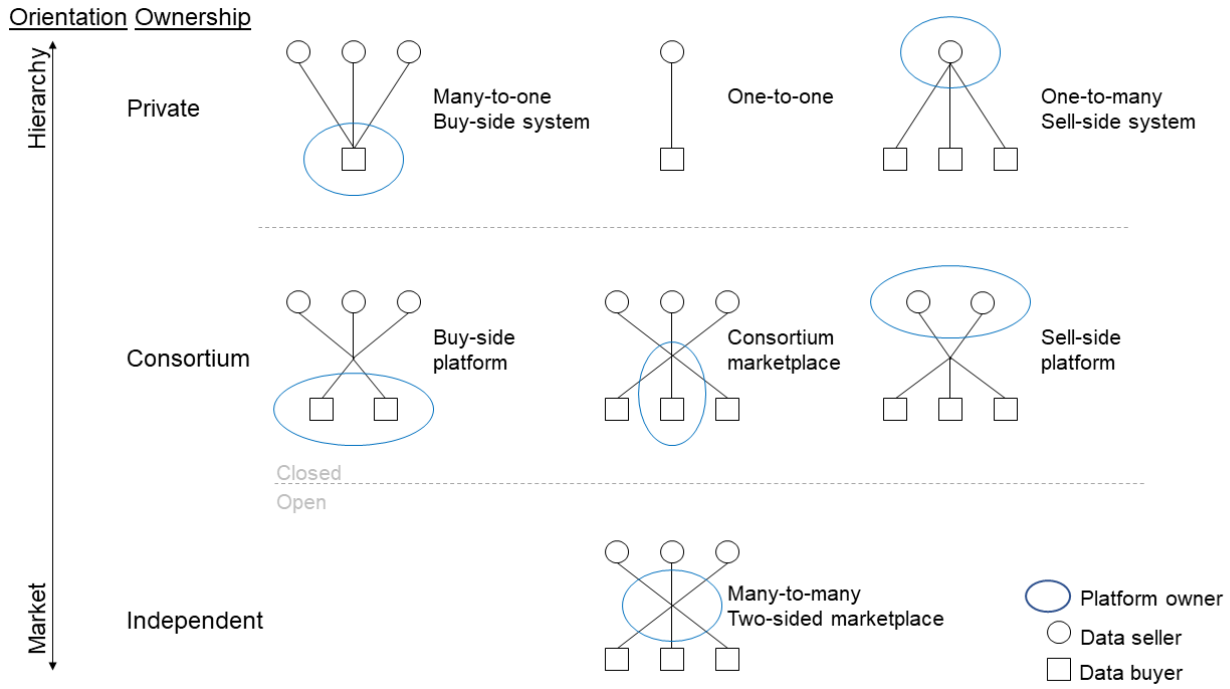

Figure 2: Data marketplace types adapted from Koutroumpis et al. and Stahl et al.

In Table 5, several examples of data marketplaces are provided, based on this classification type. These examples focus on B2B data trade in the automotive industry.

Table 5: data marketplaces examples based on orientation, ownership, and matching mechanism

No	Data marketplace	Founded	Description
1	**TomTom** *Hierarchical, private*	1991	TomTom is a privately owned company that uses location technology to sell mapped data. They trade data in a hierarchically oriented, bilateral market. TomTom is well known for their sale of navigation boxes to end consumers. In this report, the focus is on

[51] P. Koutroumpis, A. Leiponen, and L. D. Thomas, "The (unfulfilled) potential of data marketplaces," ETLA Working Papers, 2017.

[52] F. Stahl, F. Schomm, G. Vossen, and L. Vomfell, "A classification framework for data marketplaces," Vietnam Journal of Computer Science, vol. 3, no. 3, pp. 137-143, 2016, doi: 10.1007/s40595-016-0064-2.

			the B2B segment of TomTom that concerns data trade between TomTom and their commercial data buyers.
2	**INRIX** *Hierarchical, private*	2005	INRIX is also a privately owned company that applies location analytics to make road transportation more intelligent. INRIX trades data bilaterally with their commercial data buyers and serves public organisations. In addition to trading data, INRIX performs research on subjects such as road congestion, commuting time and vehicle carbon emission.
3	**HERE** *Hierarchical/market, consortium*	2015	HERE is formerly known as Navteq and was owned by Nokia. In 2015 the company was acquired by a consortium where Audi, BMW and Daimler are the main shareholders. HERE applies location technology to improve connected driving experiences. The HERE data marketplace has open access for any data provider, data buyer and third-party service provider to foster collaboration and share data among participants.
4	**Caruso** *Hierarchical/market, consortium*	2017	Caruso is funded by TecAlliance, a provider of vehicle data in the automotive industry. Besides TecAlliance, multinationals such as Bosch and Continental are shareholders of Caruso. The data marketplace is closed and only the consortium members and partners are allowed to trade within.
5	**IOTA** *Market, independent*	2017	IOTA is funded by the non-profit IOTA Foundation. IOTA focuses on the IoT market with the goal to enable secure data transactions between data sellers and buyers. The IOTA data marketplace has open access that allows many participants to trade data. IOTA is currently in the proof-of-concept phase.
6	**Ocean Protocol** *Market, independent*	2017	Ocean Protocol is a non-profit organisation based in Singapore. Their data marketplace has open access to create an environment in which many data providers and buyers can exploit data. The data marketplace is currently in its beta stage and is planned for a new release in Q3 of 2020. Ocean Protocol particularly focuses on AI. With high data volumes and trained algorithms, they aim to advance AI development.

In the *hierarchical market orientation with private ownership*, a platform owner also acts as a data provider by offering aggregated data. A platform owner usually working closely with potential data buyers, understanding their specific needs, and build strong bilateral relationships. While this mode works in practice, the scaling power is often limited due to lengthy negotiation and technology alignments for buying and selling processes. In the *mixed market orientation with consortium ownership* mode, data marketplaces classified tend to provide brokering and consulting services to data buyers.

They often provide standardised technology stacks to ease the transactions between participants. Furthermore, tight onboarding mechanisms to ensure only trustworthy participants can join data marketplaces are also essential. Moreover, transactions in this mode are still based on trust (between each participant) and negotiated contracts. While this mode provides better scaling power than the previous one, tight onboarding processes often create a high entry barrier for data marketplace participants. Finally, *market orientation with independent ownership* data marketplaces brings a more 'open' environment. These data marketplaces act as a third-party platform to match and mediate data providers and data buyers. They aim for high scaling power. For instance, data marketplace in this type appears to use standardised contract (e.g., embedded within the smart contract) in decentralise architecture to speed up the transaction processes. Data providers and data buyers do not have to rely on prior bilateral relationships to build trust (because data marketplaces in this type emphasise data sovereignty). However, buying and selling raw datasets automatically without intervention is still challenging due to the nature of data. For instance, knowing the 'right' data value and price for both providers and buyers is difficult.

T7.1 report entitled "D7.1 Sustainable Business Model for TRUSTS Data Marketplace I" will discuss the implication of *orientation* and *ownership* in detail. It will also position TRUSTS within these determinants. Finally, business model archetypes will be constructed.

3.3 Data marketplace features

Table 6 summarises features found in the existing data marketplace literature. The summary is useful to enhance our understanding of what functions data marketplaces possess and how they distinguish data marketplaces with other marketplaces known so far (e.g., amazon, eBay). These features are most relevant to data marketplaces categorised under the market orientation, independent ownership, and many-to-many matching mechanisms. Other data marketplace categories may only provide these features partially.

The term "feature" is loosely used, referring to "a distinguishing characteristic of a software item" (e.g., performance, portability, or functionality) described in IEEE's 829 Standard for Software and System Test Documentation. The features were extracted from four data marketplaces studies that discuss the high-level view/business model of data marketplaces.[53]

[53] P. Koutroumpis, A. Leiponen, and L. D. W. Thomas, "Markets for data," Industrial and Corporate Change, vol. 29, no. 3, pp. 645-660, 2020, doi: 10.1093/icc/dtaa002; M. Fruhwirth, M. Rachinger, and E. Prlja, "Discovering Business Models of Data Marketplaces," in Proceedings of the 53rd Hawaii International Conference on System Sciences, 2020; L. Meisel and M. Spiekermann, "Datenmarktplätze: Plattformen für Datenaustausch und Datenmonetarisierung in der Data Economy," in "ISST Work Report," 2019; M. Spiekermann, "Data Marketplaces: Trends and Monetisation of Data Goods," Intereconomics, vol. 54, no. 4, pp. 208-216, 2019, doi: 10.1007/s10272-019-0826-z.

Table 6: Data marketplace features

No	Feature	Category
1	Provide **user verification and certification.**	Onboarding mechanism
2	Uploading **(meta)-dataset** for data providers.	Dataset discovery
3	**Electronic catalogue** to show existing available dataset, product listing, and provider directories.	
4	**Advanced search** for data buyers to enter selection criteria (e.g., search keyword/query). Dataset discovery is also equipped with sorting and **filtering** functions.	
5	Stable **matching algorithms** to connect data providers and data buyers.	
6	Provide **visibility management** to enable data providers deciding who can see or buy their data products.	
7	***Pricing models*** *to determine a data marketplace's strategy to gain profit;* ***Pricing discovery*** *determines how the monetary value of data (data price) is set prior to a transaction.*	Trading arrangements
8	Create **contractual conditions** regarding the data (license or contract), for example, data ownership and data usage.	
9	Provide **communication channels** to facilitate the negotiation and communication between data providers and data buyers.	
10	Data assets can be accessed prior to purchase to evaluate the value of the data (**pre-purchase testability**).	
11	**Smart contract** mechanisms as a validation means of contractual condition agreements.	
12	**Billing mechanisms** for enabling data buyers to pay data providers.	Transaction workflow
13	Execute **transaction workflow** to move datasets from data providers to data buyers. In case of interruptions, the transaction can be continued in the same place.	

14	Actors of the data marketplaces can evaluate data assets and also show the conclusion of overall data provider reputation (via **review system**).	Review system
15	Provide **transaction security** infrastructure to enable secure data transactions, such as HTTPS, encryption protocols, etc.	Security
16	Using **blockchain** technology, for example, *smart contracts* to mitigate strategic behaviour in data trading.	
17	Provide **profile security** to ensure only verified and authorised users can access their profiles and roles using technology such as two factor-authentication.	
18	Protect individual data through **anonymisation** (removal of personal identity/information).	Privacy-preserving mechanisms
19	Provide **encryption** to protect data assets.	
20	Provide access type to trade data between data providers and data buyers via **API** and provide clear interoperability guidance (e.g., API description).	Interoperability
21	Provide either static or dynamic **data streams**.	
22	**Data provenance mechanisms** that track data from its origin to its destination.	Data governance
23	(Verifiable) **Metadata management** to inform data origin, content, collection, right, etc.	
24	Provide models for **semantic representation**.	
25	Additional services for dataset analysis, for example, **data visualisation**. This helps to simplify complex datasets into graphical representations (e.g., diagrams, scorecards, heat maps). Data visualisation algorithms are useful to evaluate data quality and relevance quickly.	Data analysis
26	Provide **data normalisation** to check traded data to predefined data models, formats, and attributes.	Data transformation
27	Provide **data aggregation** to combine multiple datasets.	
28	Support **multiple formats of data output** types.	

29	Trust management tools for monitoring data consumers' **contractual compliance.**	Data Monitoring
30	**Data quality** guarantee.	
31	Provide data buyers' **purchase history**.	User management
32	Provide **dashboards** to summarise users' overall activities, for example, manage offerings, requests, transactions, etc.	User management
33	Provide **notification control** and support feature (e.g., Help Center, Chatbots).	
34	Provide the trade leads and also the **purchase trend**.	News Service
35	Provide **notification channels** to notify users about certain news (e.g., social networks, email, etc.)	

Figure 3 provides a summary of the identified data marketplace features in Table 6. This figure maps the four generic building blocks of data marketplaces, i.e., *the core process, technological backbone, data service ecosystem, and account management,* into the category presented in Table 6. For instance, the *core process* building block incorporates the category of *onboarding mechanism, dataset discovery, trading arrangements, transaction workflow, and review system.*

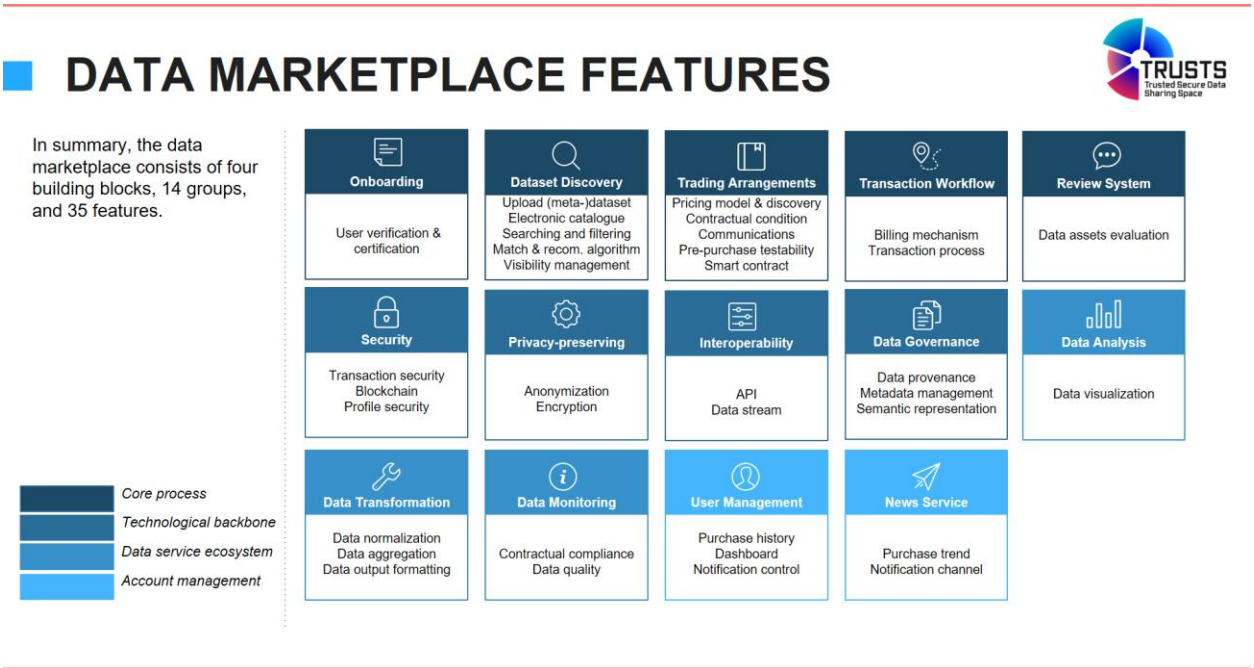

Figure 3: Generic data marketplace features

Recommendation for TRUSTS

The Table above (Table 6) constitutes a collection of functionalities that are used in the state-of-the-art data marketplace initiatives. Analysing the 35 features presented in the above Table, we derive that they are fully compatible with the 44 Functional Requirements for the TRUSTS end-to-end platform defined in the deliverable D2.2.

Nevertheless, each of the functionalities above will be fully analysed in order to augment the description of the current TRUSTS functional requirements. The updated functional requirements will be reported in the deliverable D2.3.

It is recommended that the 35 features presented above are analysed within the scope of D2.3 to feed the updated functional requirements should they are compatible with the TRUSTS vision.

3.4 Data marketplace trends

The Chapter below describes data marketplace trends from industry and academic perspectives.

3.4.1 Industry perspective: a closer look at the market trends

A recent study by the European Commission[54] examines trends of data markets. The study measures **the value of data market**, i.e., "the marketplace where digital data is exchanged as *products* or *services* as a result of the elaboration of raw data" and **the value of data economy**, i.e., "measures the overall impacts of the Data Market on the economy as a whole".

The study compares the value of the data market and data economy from 2018 to 2019. It also projects the facts and figures for the year 2025 based on three scenarios. The scenarios are summarised in Table 7.

Table 7: The 2025 scenarios for data market and data economy

Scenario / Characteristic	Challenge scenario	Baseline scenario	High growth scenario
Data innovation	Low level	Healthy growth	High level
Concentration of power	A moderate level due to digital markets fragmentation	Moderate by data providers	Low data power concentration

[54] European Commission, "The European Data Market Monitoring Tool," Brussels, 2020.

Data governance model	Unclear	Protecting personal data rights	Open and transparent
Distribution of data innovation benefits in the society	Uneven	Uneven but rather wide	Wide

The EU27 data market value is likely to increase after 2019 (see Figure 4). It will reach €72 billion and €83 billion in the 2025 challenge scenario and 2025 baseline scenario, respectively. In the most optimistic scenario, it will grow by 10.7% compared to 2019 (i.e., to reach €107 billion).

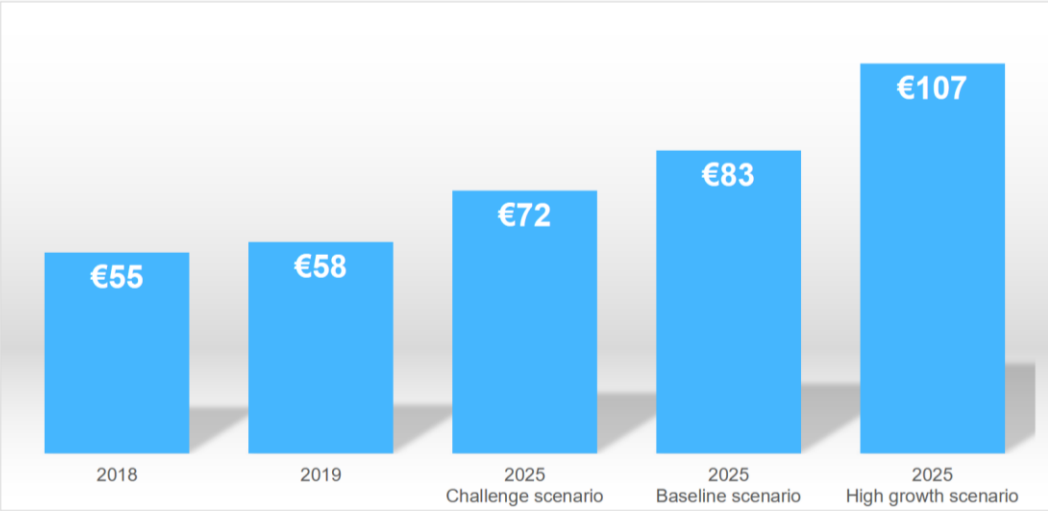

Figure 4: The EU27 data market value adapted from European Commission

Similar to the general trend of the EU27 data market value, the data economy's value is also expected to grow positively between 2020 and 2025, as shown in Figure 5. It will reach a value of €432 billion in the *2025 challenge scenario*. With a compound annual growth rate of 9,2%, the EU27 data economy value will increase to €550 billion in the *2025 baseline scenario*. *In the 2025 high growth scenario,* it will reach a value of €827 billion.

The growth trend in the data market and data economy brings several implications. According to the European Commission[55], for instance, the total number of data professionals (i.e., those who deal with data endeavours as their primary task) will also consistently rise. Many opportunities will open in data-related jobs, and more knowledge workers are needed. Despite its positive trend, there is still a potential lack of data professionals' supply in the high-growth scenario. Following this, companies taking the role as data providers and data buyers will also grow in overall number and share. The impact on data marketplaces will be addressed more detailed in Chapter 3.5.

[55] European Commission, "The European Data Market Monitoring Tool," Brussels, 2020.

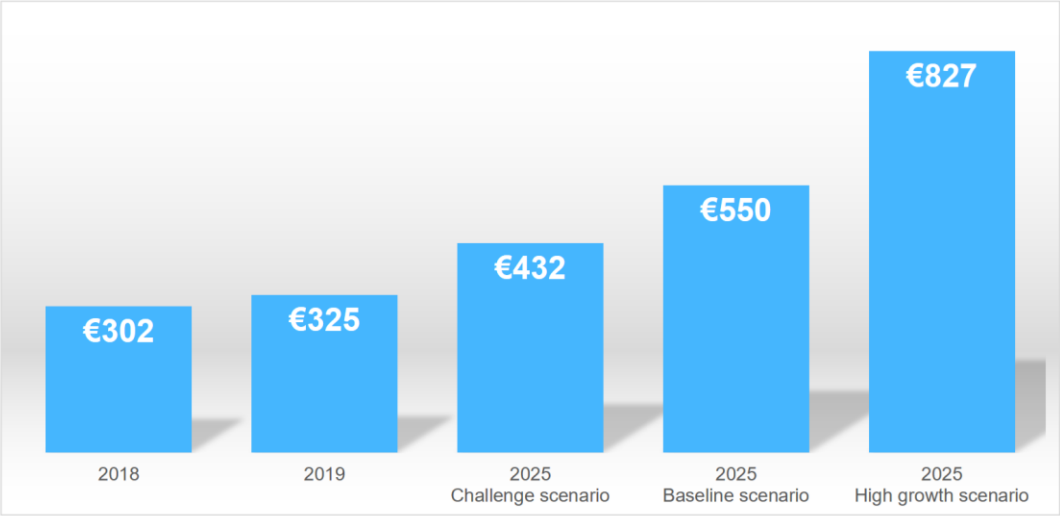

Figure 5: The EU27 data economy value adapted from European Commission

3.4.2 Academic publications

The trend of data marketplace publications in our database is as illustrated in below Figure 6:

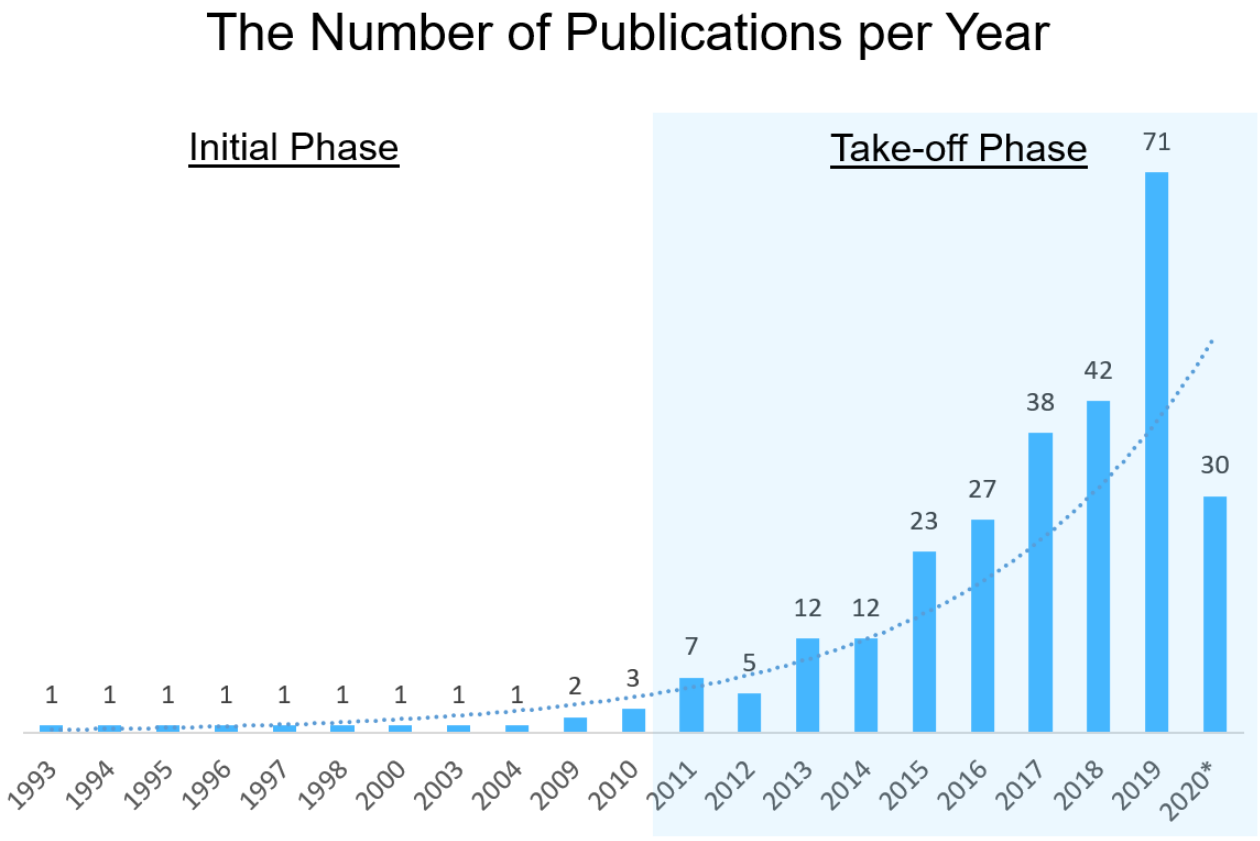

Figure 6: The number of Publications per Year

The trend of publication can be divided into two clusters: the initial phase and the take-off phase. Data marketplaces have emerged in the take-off phase (i.e., after 2010), when the number of publications rapidly increased. Publications regarding stolen data markets[56], data markets in the cloud[57], data as-services[58], survey in data marketplaces[59] are among ones that triggered the hype of data marketplaces. Considering the EU investment in data marketplaces, and the projected trends in the data economy, the increasing trend of data marketplace publications is also predicted to continue in the future.

We are now moving to consider publication trends based on co-authorship countries. Figure 7 presents the top 10 countries that actively publish data marketplace articles. Authors from the United States published the most, followed by authors from Germany and China. From the regional perspective, i.e., continent level, authors from the EU27 and the UK published the most. The trend may correlate to the EU vision to keep increasing the data economy's value, resulting in many grants being available in the data marketplace domain.

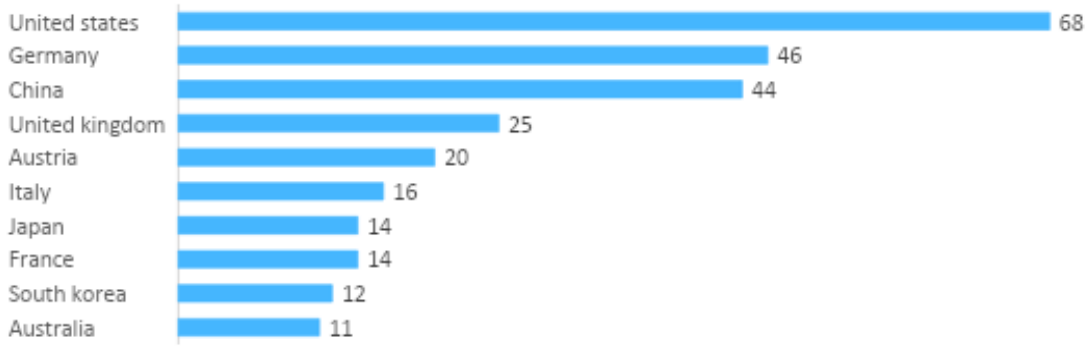

Figure 7: The Numbers of Publication by Co-Authorship Country (Top 10)

A network visualisation of the keywords used in articles was created as a beneficial illustration to explore the topics discussed by scholars. A map was basically created based on bibliographic data that only analyses the terms that occur more than four times in the articles. From 1836 keywords, 89 meet the threshold. The VOSviewer[60] was used to conduct such visualisation that can be seen in Figure 8 below. The bubble scale indicates the occurrences, whereas the line indicates the link between the keywords. The most used keyword is "commerce", followed by "data marketplaces", "big data", and "data privacy".

Each colour shows the cluster that each keyword belongs to. The first recognition is that the scholars discuss pricing mechanisms as the primary research theme (refer to the red cluster). Scholars use advanced technology (e.g., machine learning, query processing) and mathematical concepts (e.g., polynomial approximation) to determine price or budget for data. Second, represented in the green

[56] T. J. Holt and E. Lampke, "Exploring stolen data markets online: products and market forces," Criminal Justice Studies, vol. 23, no. 1, pp. 33-50, 2010.
[57] M. Balazinska, B. Howe, and D. Suciu, "Data markets in the cloud: An opportunity for the database community," Proceedings of the VLDB Endowment, vol. 4, no. 12, pp. 1482-1485, 2011.
[58] Q. H. Vu, T.-V. Pham, H.-L. Truong, S. Dustdar, and R. Asal, "Demods: A description model for data-as-a-service," in 2012 IEEE 26th International Conference on Advanced Information Networking and Applications, 2012: IEEE, pp. 605-612.
[59] F. Schomm, F. Stahl, and G. Vossen, "Marketplaces for data: an initial survey," ACM SIGMOD Record, vol. 42, no. 1, pp. 15-26, 2013.
[60] *https://www.vosviewer.com/ accessed on 11 November 2020*

colour, the privacy theme in personal data markets is discussed. Next, scholars discuss the general context of data marketplaces in relation to, for instance, commerce, data economy, data sharing, data ecosystem, etc. (refer to the blue colour). This cluster, for instance, discusses market mechanisms for data trading[61], and an investigation of data ecosystems.[62]

The fourth cluster, i.e., the yellow, represents the technical theme. Topics such as access control, cryptography, metadata, ontology, smart contacts, etc., are explored. Scholars also explore the services data marketplaces offer, such as profit maximisation and profitability in big data, data brokerage, data service, information management, etc. (as shown in the purple cluster). Lastly, the data marketplaces in the IoT domain are intensively discussed. It is mostly related to smart city and transportation contexts.

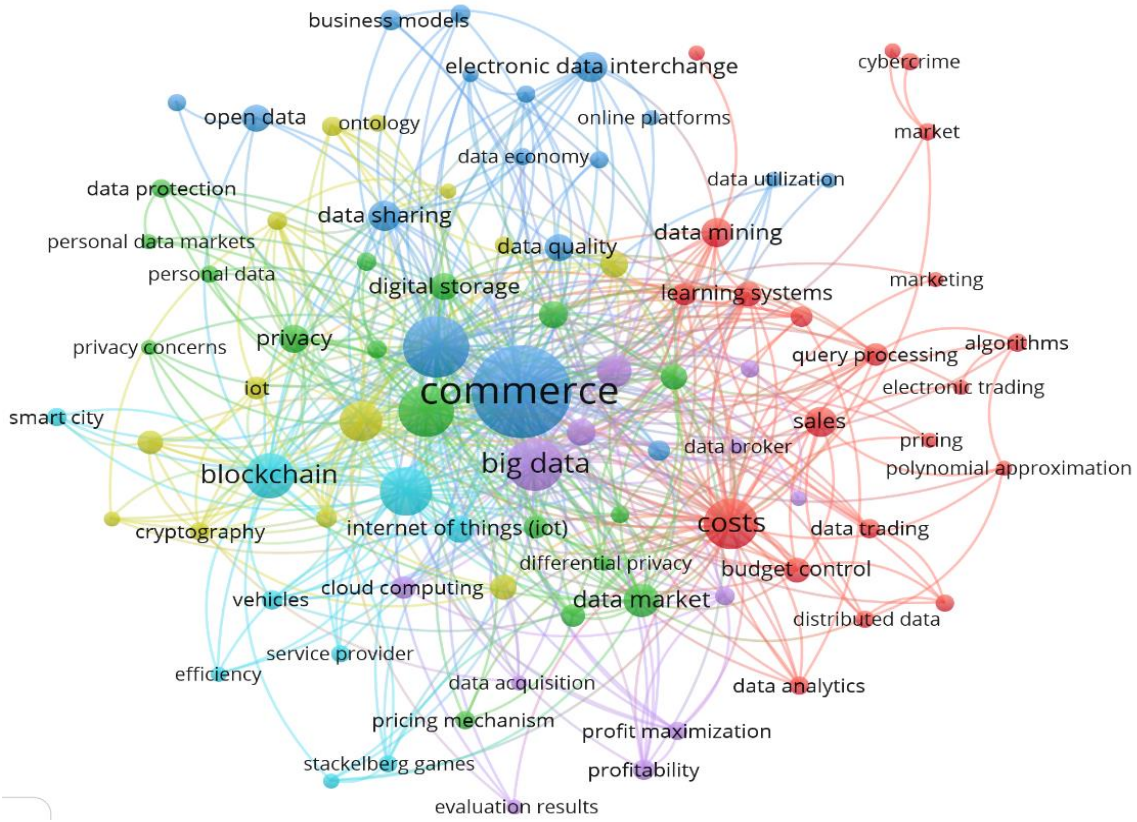

Figure 8: Network visualisation of keywords

Figure 9 below, presents the network visualisation of keywords by years where the trend discussed in data marketplaces is detectable by analysing the keywords used in recent years, i.e., 2016-2019. Recently, scholars heavily discuss blockchain implementation in data marketplaces, especially in the

[61] P. Koutroumpis, A. Leiponen, and L. D. W. Thomas, "Markets for data," Industrial and Corporate Change, vol. 29, no. 3, pp. 645-660, 2020, doi: 10.1093/icc/dtaa002.

[62] M. I. S. Oliveira, G. d. F. B. Lima, and B. F. Lóscio, "Investigations into Data Ecosystems: a systematic mapping study," Knowledge and Information Systems, pp. 1-42, 2019.

domain of IoT and Smart Cities. Security aspects, such as access controls, are also heavily discussed by scholars. Pricing mechanisms using economic theories, e.g., the Stackelberg model, also gain significant attention.

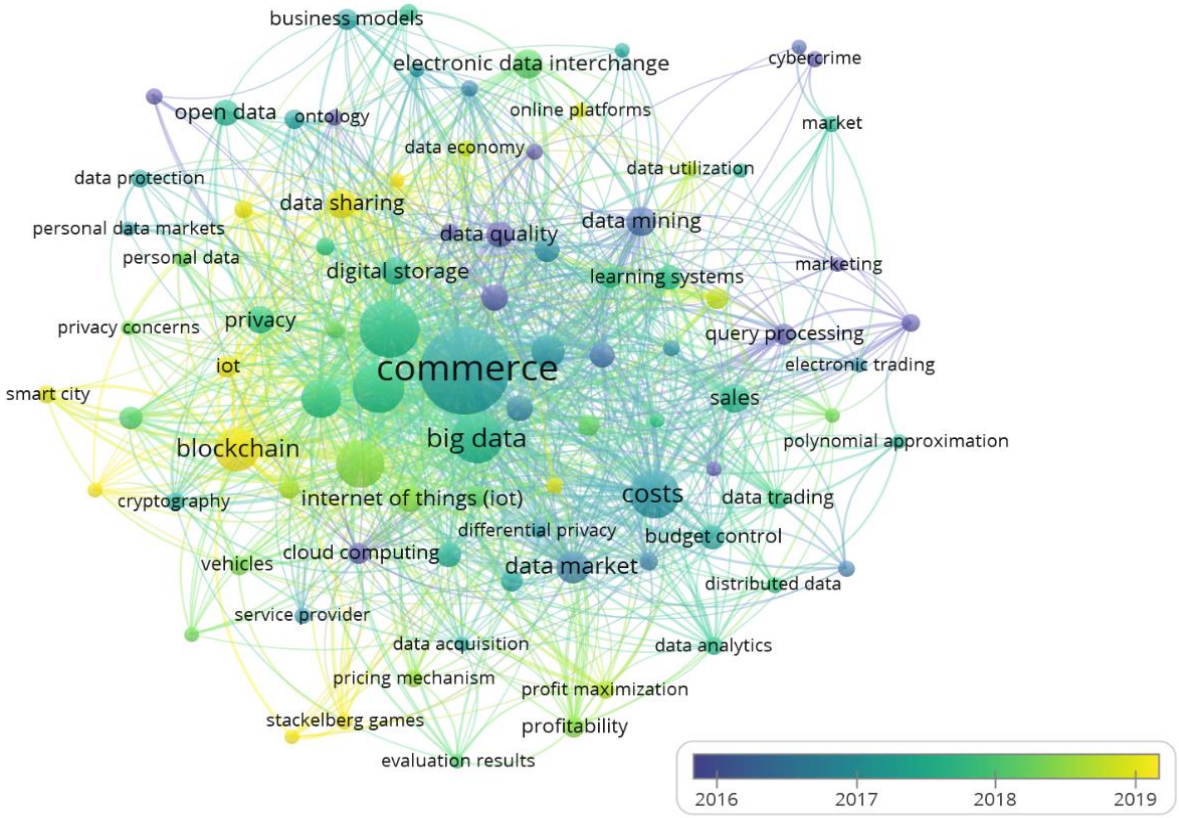

Figure 9: Network visualisation of keywords by years

To complement the keyword analysis, a conduction of the term analysis (i.e., analysis of a term occurrence based on text found in titles and abstracts) was done. The minimum number of occurrences of a term is 10. 186 terms were found to meet the thresholds. Further, the visualisation can be seen in Figure 10.

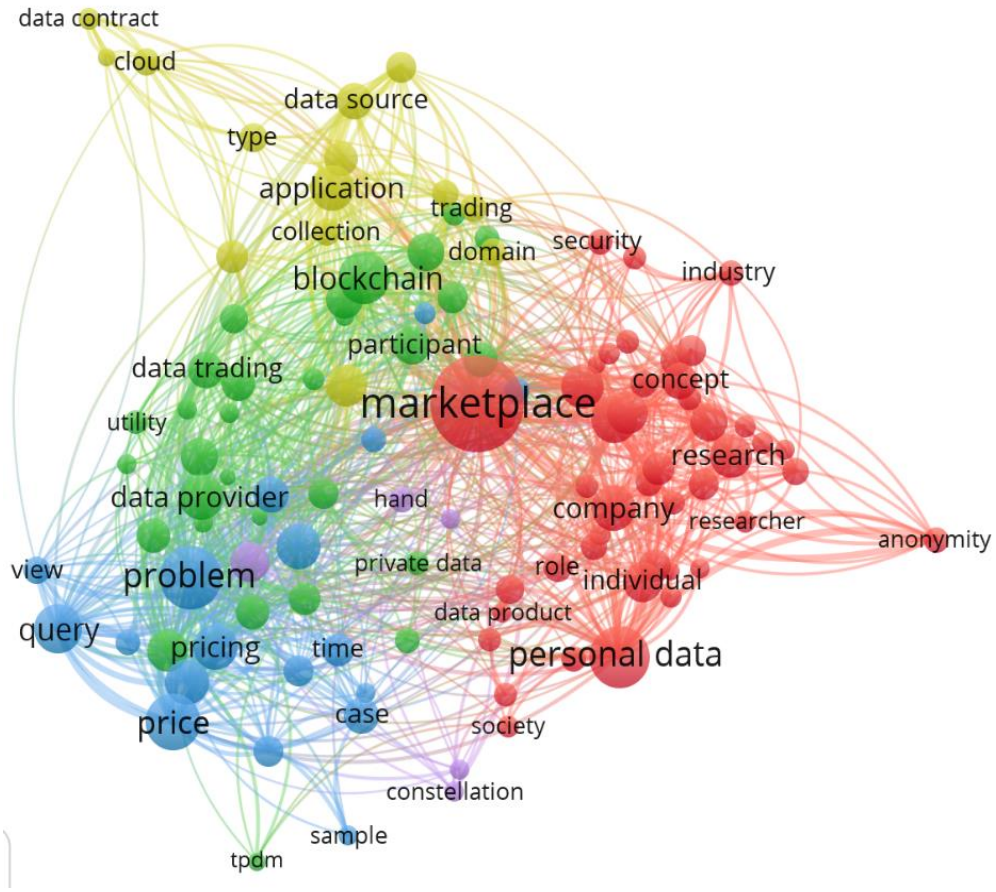

Figure 10: Visualisation of Term analysis

The term analysis confirms the themes discussed in data marketplaces, such as blockchain technology, pricing & query mechanisms, and securities. It also confirms key actors in data marketplaces, such as data providers/owners and data buyers. The term analysis also reveals interesting insight, e.g., by showing how personal data marketplaces correlate with the *anonymity* feature. It also points out that *data contracts* and *utility aspects* have become an important topic in the data marketplace domain.

3.5 Challenges of data marketplaces

In this Section, a discussion of the challenges of data marketplaces is provided. The challenges were categorised using the STOF model (see Table 8) which is a framework that provides the logic of business and its ecosystem (Bouwman et al., 2008). The STOF model consists of the service domain (S), technology domain (T), organisation (O) and finance (F).

Table 8: Challenges of data marketplaces

Category	Challenge	Short description	Source	Perspective

Service	Data ownership definition	Who should own the data? The idea of data ownership is still debatable. Some scholars suggest that data (especially personal data) should be owned by individuals, while others argue it is not feasible or conceptually flawed.	(Koutroumpis et al., 2020)[63]	Data providers
	Ensuring data integrity	Ensure data integrity, i.e., data is not altered during the lifecycle of data trading processes.	(Lawrenz et al., 2019)[64]	Data providers, data buyers
	Assessing data quality	Data buyers do not know how to assess the quality of the data or evaluate the data before purchasing it.	Koutroumpis et al., (2020)[65]	Data buyers
	Ensuring contractual compliances	Data buyers may violate data access and usage restrictions.	Koutroumpis et al., (2020)[66]	Data providers
	Loss of control over data	Data providers are unable to track down data usage and ensure compliances towards data sharing agreements. In consequence, they are afraid that competitors may benefit from their data in unanticipated ways. It also brings potential privacy risks.	(Spiekermann , 2019)[67]	Data providers
	Lack of transparency	Little transparency between data providers and data brokers. In some cases, data providers do know how data brokers assess their data value and how the assessment occurred in fair processes.	(Oh et al., 2019)[68]	Data providers

[63] P. Koutroumpis, A. Leiponen, and L. D. W. Thomas, "Markets for data," Industrial and Corporate Change, vol. 29, no. 3, pp. 645-660, 2020, doi: 10.1093/icc/dtaa002.
[64] Lawrenz, S., Sharma, P., & Rausch, A. (2019, March). Blockchain technology as an approach for data marketplaces. In Proceedings of the 2019 International Conference on Blockchain Technology (pp. 55-59).
[65] P. Koutroumpis, A. Leiponen, and L. D. W. Thomas, "Markets for data," Industrial and Corporate Change, vol. 29, no. 3, pp. 645-660, 2020, doi: 10.1093/icc/dtaa002.
[66] P. Koutroumpis, A. Leiponen, and L. D. W. Thomas, "Markets for data," Industrial and Corporate Change, vol. 29, no. 3, pp. 645-660, 2020, doi: 10.1093/icc/dtaa002.
[67] Spiekermann, M. (2019). Data marketplaces: Trends and monetisation of data goods. Intereconomics, 54(4), 208-216.
[68] Oh, H., Park, S., Lee, G. M., Heo, H., & Choi, J. K. (2019). Personal data trading scheme for data brokers in IoT data marketplaces. IEEE Access, 7, 40120-40132.

Technology	Privacy protection	Personal data trade can lead to unintended disclosure of personal information, especially when data providers are individuals and data buyers are huge corporations or governments, causing an imbalance of power.	Virkar et al., (2019); Charitsis et al., (2018)[69]	Data providers
	Security	In general, data marketplaces must provide security technologies to protect data trading processes from hacking, counterfeiting, and other unwanted behaviours.	Lawrenz et al., (2019), Virkar et al., 2019)[70]	Data providers, data buyers
	Technical efficiencies and scalabilities	Technical efficiencies and scalabilities, especially in data marketplaces employing a distributed-ledger technology or decentralised architecture, are known as a general problem. In general, data marketplaces consume high computation and communication cost.	Liu et al., (2019); Ishmaev, (2020)[71]	Data providers, data buyers
	Data placement cost	Data placement and replication cost (i.e., after the purchase) is considerably high and consumes both bandwidth and latency.	(Ren et al., 2018)[72]	Data providers, data buyers
	User-friendly applications	User-friendly applications and interfaces are required for advancing data marketplaces.	Ramachandran et al., (2018)[73]	Data providers,

[69] Virkar, S., Pereira, G. V., & Vignoli, M. (2019, September). Investigating the Social, Political, Economic and Cultural Implications of Data Trading. In International Conference on Electronic Government (pp. 215-229). Springer, Cham; Charitsis, V., Zwick, D., & Bradshaw, A. (2018). Creating worlds that create audiences: Theorising personal data markets in the age of communicative capitalism. tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society, 16(2), 820-834.

[70] Lawrenz, S., Sharma, P., & Rausch, A. (2019, March). Blockchain technology as an approach for data marketplaces. In Proceedings of the 2019 International Conference on Blockchain Technology (pp. 55-59); Virkar, S., Pereira, G. V., & Vignoli, M. (2019, September). Investigating the Social, Political, Economic and Cultural Implications of Data Trading. In International Conference on Electronic Government (pp. 215-229). Springer, Cham.

[71] Liu, K., Chen, W., Zheng, Z., Li, Z., & Liang, W. (2019). A novel debt-credit mechanism for blockchain-based data-trading in internet of vehicles. IEEE Internet of Things Journal, 6(5), 9098-9111; Ishmaev, G. (2020). The ethical limits of blockchain-enabled markets for private IoT data. Philosophy & Technology, 33(3), 411-432.

[72] Ren, X., London, P., Ziani, J., & Wierman, A. (2018). Datum: Managing data purchasing and data placement in a geo-distributed data market. IEEE/ACM Transactions on Networking, 26(2), 893-905.

[73] Ramachandran, G. S., Radhakrishnan, R., & Krishnamachari, B. (2018, September). Towards a decentralized data marketplace for smart cities. In 2018 IEEE International Smart Cities Conference (ISC2) (pp. 1-8). IEEE.

	and interfaces			data buyers
Organisation	The absences of legal frameworks	No IPR (e.g., digital right management models) is attached to data. In addition, no clear liability rules can be asserted to contracts or violations thereof.	Sørlie and Altmann, (2019); Spiekerman (2019)[74]	Data providers
	Lack of resources and technical knowledge	Lack of resources (e.g., operating cost) and technical knowledge to manage technical complexities.	Oliveira et al., (2019)[75]	Data marketplace owners and operators
	Unclear organisational structure	The absence of well-defined models regarding actor definitions, their roles, and interactions between actors.	(Oliveira et al., 2019)[76]	Data marketplace owners and operators
	Ethical concern	Giving monetary incentives to individuals to share their sensitive personal data e.g., health data raises ethical concerns such as undue influence.	Ishmaev, (2020), Ahmed and Shabani, 2019)[77]	Data providers
Finance	Pricing mechanism	Data providers have no clear and standardised mechanisms to price data assets.	Niu et al., 2020, Chen et al., (2019),	Data providers

[74] Sørlie, J. T., & Altmann, J. (2019, September). Sensing as a Service Revisited: A Property Rights Enforcement and Pricing Model for IIoT Data Marketplaces. In International Conference on the Economics of Grids, Clouds, Systems, and Services (pp. 127-139). Springer, Cham; Spiekermann, M. (2019). Data marketplaces: Trends and monetisation of data goods. Intereconomics, 54(4), 208-216.

[75] Oliveira, M. I. S., Lima, G. D. F. B., & Lóscio, B. F. (2019). Investigations into Data Ecosystems: a systematic mapping study. Knowledge and Information Systems, 1-42.

[76] Oliveira, M. I. S., Lima, G. D. F. B., & Lóscio, B. F. (2019). Investigations into Data Ecosystems: a systematic mapping study. Knowledge and Information Systems, 1-42.

[77] Ishmaev, G. (2020). The ethical limits of blockchain-enabled markets for private IoT data. Philosophy & Technology, 33(3), 411-432; Ahmed, E., & Shabani, M. (2019). DNA data marketplace: an analysis of the ethical concerns regarding the participation of the individuals. Frontiers in genetics, 10, 1107.

			Mao et al., (2019)[78]	
	Data valuation	Data providers and data sellers do not recognise the value of data because of the limitations in calculating potential benefits/revenues.	Spiekermann, 2019, Ha et al., (2019)[79]	Data buyers, data sellers
	Profit maximisation	Data providers struggle to find a strategy to optimise profit by finding the balance between revenue maximisation strategy and cost structure for data acquisition.	Zheng et al., (2020), Jiao et al., (2018)[80]	Data providers

In the *service* category, the majority of challenges have been discussed adequately. Scholars attempt to find solutions *to ensure contractual completions, retain control over data, and provide transaction transparency* via technological enforcements. For instance, initiatives to implement blockchain, smart contracts, and access controls have flourished. The attempt to discuss *data quality* and *data protection* have also been conducted. So far, however, there has been still little discussion about *defining data ownership* in data marketplace research.

In the *technology* domain, some challenges such as *privacy protection* and *security technology* (e.g., cryptography) have been major themes in data marketplace research. Moreover, efficient *data placement* via cloud computing or digital storage has attracted much interest. What we still lack in the literature is the discussion that focuses on technical efficiency and scalability. Previous studies also have not dealt with user-friendly aspects of data marketplace applications and interfaces sufficiently.

The keyword and term analysis revealed only a few studies attempting to solve the organisation domain's challenges, such as the *absence of legal frameworks, lack of resources and technical knowledge, unclear organisation, and ethical concern.* Further research in this area will be done by academia to foster the development of data marketplaces.

[78] Niu, C., Zheng, Z., Wu, F., Tang, S., & Chen, G. (2020). Online pricing with reserve price constraint for personal data markets. IEEE Transactions on Knowledge and Data Engineering; Chen, L., Wang, H., Chen, L., Koutris, P., & Kumar, A. (2019, June). Demonstration of Nimbus: Model-based Pricing for Machine Learning in a Data Marketplace. In Proceedings of the 2019 International Conference on Management of Data (pp. 1885-1888); Mao, W., Zheng, Z., & Wu, F. (2019, April). Pricing for revenue maximization in iot data markets: An information design perspective. In IEEE INFOCOM 2019-IEEE Conference on Computer Communications (pp. 1837-1845). IEEE.

[79] Spiekermann, M. (2019). Data marketplaces: Trends and monetisation of data goods. Intereconomics, 54(4), 208-216; Ha, M., Kwon, S., Lee, Y. J., Shim, Y., & Kim, J. (2019). Where WTS meets WTB: A Blockchain-based Marketplace for Digital Me to trade users' private data. Pervasive and Mobile Computing, 59, 101078.

[80] Zheng, Z., Peng, Y., Wu, F., Tang, S., & Chen, G. (2019). Arete: On designing joint online pricing and reward sharing mechanisms for mobile data markets. IEEE Transactions on Mobile Computing, 19(4), 769-787; Jiao, Y., Wang, P., Feng, S., & Niyato, D. (2018). Profit maximization mechanism and data management for data analytics services. IEEE Internet of Things Journal, 5(3), 2001-2014.

Lastly, the challenges in the *financial* cluster have gained a lot of attention. The financial aspect of data marketplaces has always been a major topic in data marketplaces. Scholars use advanced technology (e.g., machine learning, query processing) and mathematical concepts (e.g., polynomial approximation) to determine price or budget for data.

Figure 11 summarises the above-mentioned challenges of data marketplaces.

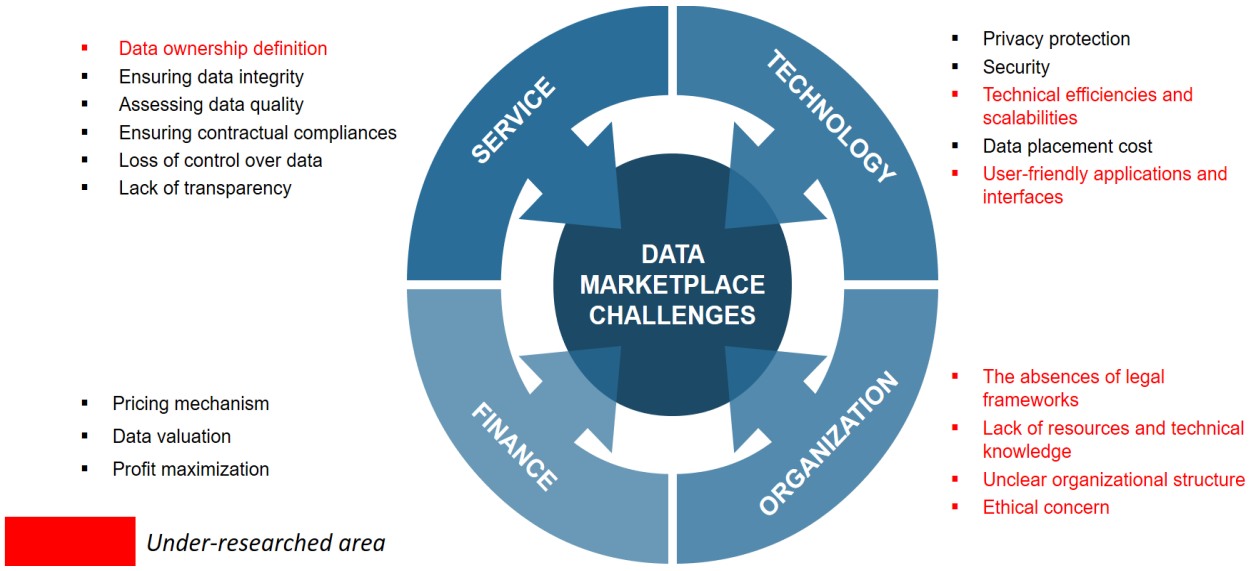

Figure 11: Challenges of data marketplaces

Recommendation for TRUSTS

TRUSTS aims at providing a sound, expandable and federated technological platform with a consistent set of processes towards sustaining a user friendly, trusted, and reliable data exchange environment. To define the business model for TRUSTS' operation, the respective USPs to assist sustaining business need to be analysed. In addition, business, technological, environmental, and operational challenges must be analysed towards defining the TRUSTS business model. It is recommended that the challenges above will be analysed in WP2, WP3, WP5 and WP6 as described in the following to implement TRUSTS' value proposition.

From the academic perspective, TRUSTS can also contribute to literature by publishing and filling the gap in the under-researched area as such:

- Discussion about **legal frameworks**, **ethical concerns**, and **data ownership** definitions have been conducted. The **WP6** Legal & Ethical Framework has discussed these topics profoundly.
- Our technical partners that work on the a) **WP3** (TRUSTS platform implementation), b) **WP4** (privacy-preserving technologies), and c) **WP5** (three use cases), are recommended to include the topics of a) **technical efficiencies** and **scalabilities** and b) **user-friendly**

applications and **interfaces**. This is especially in line with TRUSTS challenge related to advancing the state-of-the-art with regard to scalability and computational efficiency.

- The **WP7** Business Model, Exploitation & Innovation Impact Assurance is advised not to overlook internal ecosystem dynamics in the discussion, such as **lack of resources** (e.g., operating cost) and **technical knowledge** to manage technical complexities in daily operations. This is especially in line with TRUSTS challenge regarding lack of ICT and Data skills, limiting Europe's capacity to respond to the digitization challenge.

3.6 Towards a single European Data Market: Data marketplace fragmentation

In response to the current trends in the data-driven economy and to release the full potential of data flow and use across Europe, the EU announced the 2030 vision to create a *single European Data Market,* the ecosystem where data can flow within the EU and across sectors (European Commission, 2020b). The European Commission mentions data marketplaces as one of the key instruments to accomplish this vision (European Commission, 2020c). In this Section, a discussion on how the current data marketplace initiatives struggle to achieve the EU vision is offered.

For the policymakers involved in providing regulations and laws in the context of data marketplaces, one of the main obstacles in data marketplaces' growth toward achieving the single European Data Market is **fragmentation** (European Commission, 2020b). Currently, a large heterogeneity of data marketplace initiatives exists (i.e., focusing on regional-level or domain-specific industry). From a national initiative perspective, for instance, Austria, has its data marketplace project, the Data Market Austria (DMA)[81]. From a city-level perspective, for example, Amsterdam, has started its development of the Amsterdam Data Exchange (AMdEX)[82]. Another example of heterogeneity are the variations of small-size domain specific data marketplaces such as Caruso Dataplace[83] and Xignite[84] that focus on the automotive industry and financial data, respectively.

In general, fragmentation triggers multiple aspects of data marketplaces (e.g., business models, governance arrangements, and technical standards) diverge uncontrollably, leading to a decrease of trust in the concept of data marketplaces as a whole (TRUSTS, 2019). For potential data buyers, the fragmentation triggers difficulties in data discovery processes (i.e., finding appropriate data they need) since they do not know which data marketplaces trade such datasets. Overall, overheat in discovery processes lead to data buyers' dissatisfaction. Besides, data providers and buyers also suffer from vendor lock-in.

Data marketplaces often apply strict onboarding mechanisms (e.g., a certification) to ensure only trusted actors join the ecosystem. Data providers and buyers also need to align with sophisticated data sharing technologies (such as API integration and blockchain). Thus, switching from a data marketplace to another is an extensive effort. The fragmentation then indirectly impacts data marketplaces owners

[81] *https://datamarket.at/ accessed on May 1, 2020*
[82] *https://www.towardsamdex.org/ accessed may 18, 2020*
[83] *https://www.caruso-dataplace.com/ accessed on May 1, 2020*
[84] *https://www.xignite.com/ accessed on May 1, 2020*

since their platforms cannot attract both potential data providers and data buyers. To sum up, the fragmentation has slowed down the platforms' emergence due to a lack of users (i.e., data providers and buyers) (Basaure et al., 2020).

The federated approach of the IDS that has been chosen for TRUSTS is meant to overcome some of the above-mentioned hurdles regarding the fragmentation. The concept foresees common rules and standards for the infrastructure set-up which avoids the vender lock-in and lowers technological entry barriers. Further, according to the concept, not every company has to build their own components necessarily to participate in the IDS, but technology providers can offer ready and/or customised solutions, based on the same and open sources. Since IDS is a blueprint for building data sharing ecosystems it can also be applied on different domains, such as health care, mobility, or energy once the domain specific features have been considered and implemented.

The following Section will provide a short overview on the **market of data sharing initiatives**, that has been elaborated by the Netherlands AI Coalition, and will help to frame the IDS about its applicability.

Analysis of Data Sharing Initiatives

The Netherlands AI Coalition (NL AIC) is currently considering the IDS Reference Architecture as their blueprint for the development of a Trust Framework to share data for AI, as also the TRUSTS project did, next to the DMA. To validate this choice, the NL AIC conducted in December 2020 in cooperation with the organisations TNO and Innopay a scan on the data sharing market and current data sharing initiatives. A reference to the NL AIC's analysis is given in this document since the sharing of data for AI is as sensitive as sharing personal data, according to Frans van Ette, coordinator of the AI Coalition's working group Data Sharing and therefore requires similar approaches[85].

Figure 12: Data sharing initiatives in the market (source)

[85] *https://datasharingcoalition.eu/2020/why-facilitating-data-sharing-is-of-great-importance-to-artificial-intelligence/*

Figure 12 depicts the identified different data sharing initiatives on the market, clustered by their own specific domain focus. It turned out that there are at least 56 initiatives active in the development of suitable data sharing architectures or principles for AI. During the work carried out for this document, additional initiatives were found:

1. The X-Road, which is a Finnish-Estonian initiative for data exchange, providing an open-source architecture that is used in public services in corresponding countries.
2. The FCAI - Finnish Center for Artificial Intelligence, which is similar to the NL AIC.

The NL AIC also undertook a "quick scan" of the data sharing market itself and stated that it is still in a divergent phase, in the sense that there are many existing and evolving technological designs for data sharing. However, the analysis states that several elements (e.g., data sovereignty, federated and distributed data access through data brokerage, coupled to open and inclusive interoperability) appear to emerge in multiple designs, indicating that the period of technological convergence is near.

They went even farther and compared the most promising initiatives that are a) suitable for data sharing for AI and b) mature enough (see Figure 13) and found out that even those initiatives that have been assessed as most mature are not tackling all challenges for data sharing of sensitive data. One of the study's results is that the IDS framework is one of the most promising ones, but also other approaches have to be observed, such as SOLID, FIWARE, Ocean Protocol or BDVA i-Spaces.

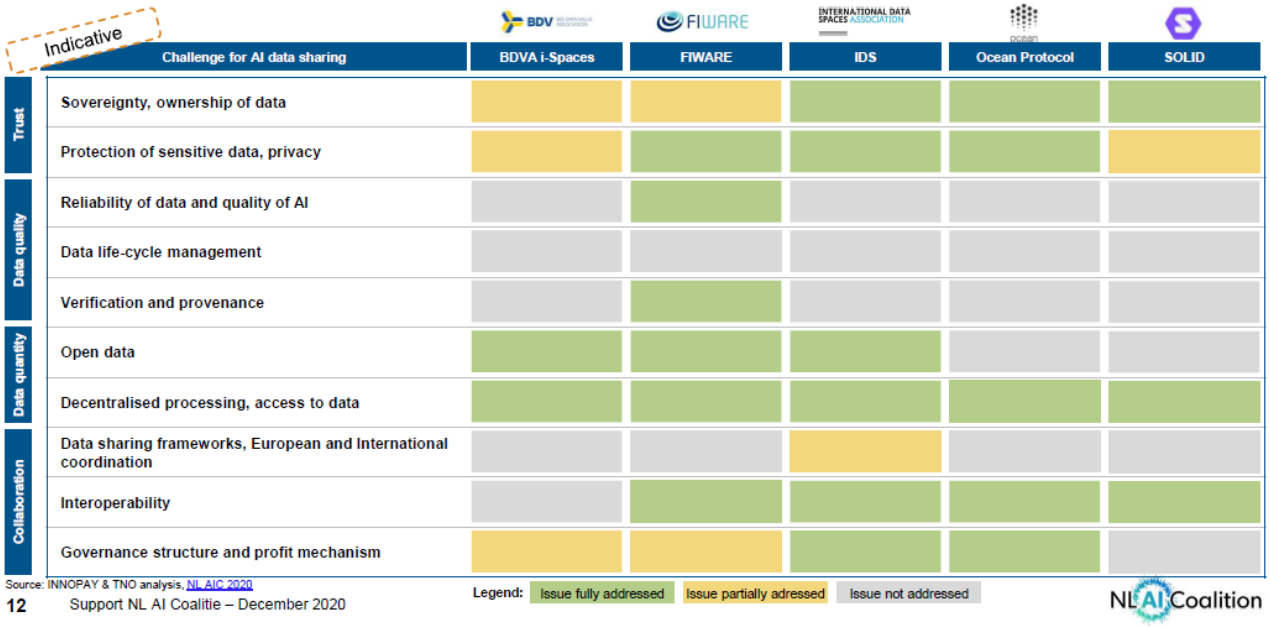

Figure 13: Assessment of data sharing initiatives with regard to their suitability for AI data sharing

Recommendation for TRUSTS

Other data sharing approaches such as SOLID, FIWARE, Ocean Protocol or BDVA iSpaces should be scanned for relevant requirements to enrich the architecture (**WP2**).

3.7 Data Market Austria (DMA) and International Data Spaces (IDS)

The objective of this Section is to explain some concrete examples of data sharing mechanisms in terms of the International Data Space (IDS) and the Data Market Austria (DMA);

The TRUSTS project will build upon existing knowledge and experience of one large national data marketplace endeavours, the Data Market Austria (DMA) and one international initiative dealing with the topic of sovereign data sharing mechanisms, the International Data Spaces Association (IDSA).

In this subsection, we will briefly discuss therefore both the DMA and IDSA, as examples for data sharing mechanisms.

Data Market Austria (DMA)[86]

In 2016, a consortium of 17 Austrian partners started the Data Market Austria (DMA) project. The DMA is partially funded by the Austrian Ministry of Transport, Innovation and Technology (BMVIT) and the Austria Research Promotion Agency (FFG). The DMA project is a pioneer of the data services ecosystem in Austria aiming to provide a data innovation environment by improving technology for secure data marketplaces and cloud interoperability. The ecosystem is supporting a full spectrum of data, i.e., both open and business data, with a sufficient level of access control (Höchtl and Lampoltshammer, 2017).

The DMA does not host the data, but it provides a catalogue for registered datasets and facilitates data exchange between core actors (data providers and data customers). The DMA developed technological innovations for data exchange, such as blockchain mechanisms to ensure data provenance and security, recommender systems based on brokerage technology and semi-automated data quality improvement (Höchtl and Lampoltshammer, 2017). The DMA pilots have demonstrated the reuse of data and services by showcasing innovative applications built on multiple open data sources (i.e., Kaggle and the European Data Portal) within the *Earth Observation* and *Mobility* domains.

The target group of DMA customers comprises a broad range of data-driven organisations and individuals, including non-profit & civic society, scientific & research, public sector government & admin, and private sector (Virkar et al., 2019). To further develop the DMA project and establish a full-operating data marketplace, the Data Intelligence Offensive (DIO) has been established as a direct spin-off of the DMA project.

The DMA[87] was an Austrian key project to conceptualise, implement, and establish an Austrian ecosystem for data services. From a technical perspective, the DMA was a federated platform leveraging a microservices architecture, data and metadata harvester and ingestion services, a recommendation engine, and distributed ledger technology for smart contracting.

The DMA accomplished a lot of basic research work and hands-on activities - kind of ground work - to elicit and specify the requirements of a data market including its objectives, its role in an ecosystem of demand and supply in regards to data trading and exchange, as well as in specification of relevant roles and stakeholders for a data market and in respect to components, features and technologies, as well

[86] *https://datamarket.at/ accessed on November 05, 2020*
[87] *http://www.datamarket.at*

as the overall architecture of a data market. Findings are well documented and publicly available[88] and also, several software components are available as open-source software.

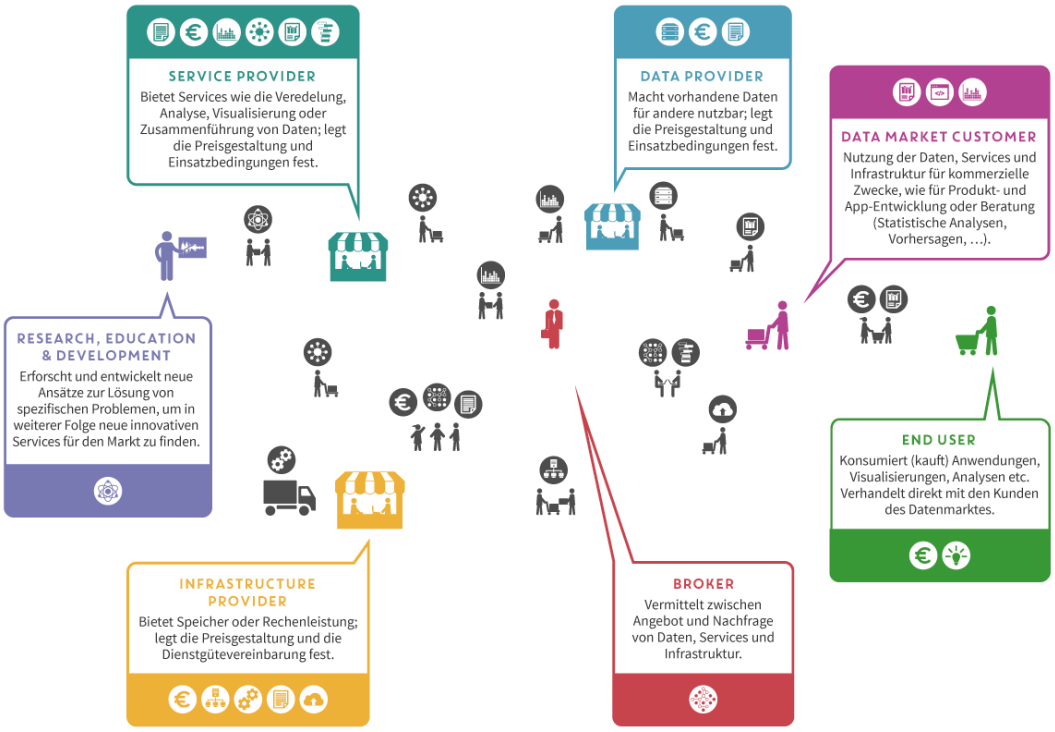

Figure 14: Roles and Stakeholders in a data market (DMA). Source: https://datamarket.at/, accessed 13/01/2021

With regards to smart contracting, the DMA explored the legal framework for smooth data exchange and developed a tool to assemble licenses from license templates in a semi-automatic way. The DMA featured distributed ledger technology for data provenance and lineage. This technology was chosen to comply with the project's demand for a federated architecture and to avoid single points of failures, as they could be introduced by a centralised architecture. The Ethereum[89] project became the basis of the DMA blockchain implementation. Decisive reasons for selecting Ethereum were its maturity, the large audience of contributors, its better support for smart contracting as compared to Bitcoin, and its open license (GPLv3[90]).

Figure 15 illustrates the architecture of the blockchain. Data market participants exchange contracts for datasets and services as well as their terms of use and terms of service via the member nodes of the network. The inner circle depicts the actual blockchain, which is cloned on all nodes in the network.

[88] *https://datamarket.at/ergebnisse-berichte/*
[89] https://ethereum.org/en/, accessed 18.04.2021
[90] *GPLv3: https://www.gnu.org/licenses/gpl-3.0.en.html*

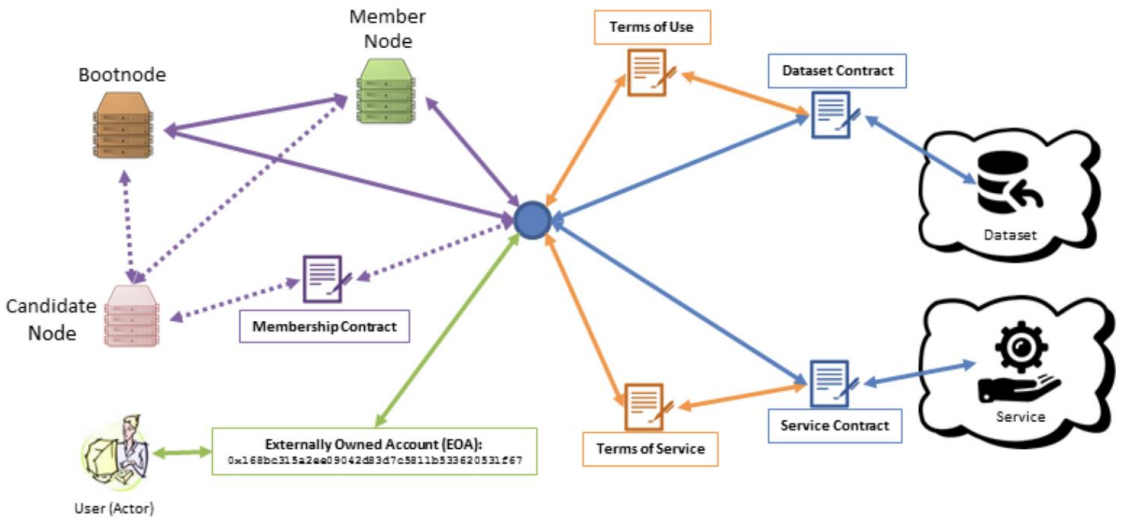

Figure 15: The architecture of the DMA blockchain[91]

The DMA had two pilots, one in the mobility area and the other one in the energy sector. The first pilot concerned taxi fleet management and combined mobility data, weather data, and open data such as traffic jams or events to generate a taxi demand heatmap. The second pilot was the prediction of energy demand based on energy and mobility data.

After project lifetime, the DMA came to an end. The reasons were the lack of activity in the created data ecosystem as well as the expiration of hardware licenses for hosting the federated platform and thereby the lack of an operator of the DMA, that was willing to take over and exploit the DMA after the project duration[92].

International Data Spaces Association (IDSA)

The International Data Spaces Association (IDSA)[93] is a coalition of more than 130 member companies that share a vision of a world where all companies self-determine usage rules and realise the full value of their data in secure, trusted, equal partnerships; and we are making that vision a reality.

IDSA's goal is nothing less than a global standard for international data spaces (IDS) and interfaces, as well as fostering the related technologies and business models that will drive the data economy of the future across industries.

The International Data Spaces Association (IDSA) aims to unlock the data economy of the future by providing the blueprint for secure, self-determined data exchange among trusted partners. "Data

[91] *DMA Deliverable D5.2, DMA Blockchain Design: https://datamarket.at/wp-content/uploads/2017/10/DMA_D5.2_DMA_Blockchain_Design_PUBLIC-1.pdf*

[92] *https://datamarket.at/ accessed on November 05, 2020*

[93] https://internationaldataspaces.org/, accessed 18.04.2021

sovereignty" is what it refers to and it is vitally important, in light of the fact that data access and exchange are rapidly becoming critical success factors for both companies and entire economies.

The IDSA has defined a reference architecture model (IDS-RAM) and a set of agreements that can be used to create virtual data spaces which establish trust among partners and a basis for innovative, new business models, products, and services.

The central technical component for secure and trusted data exchange is the *IDS connector* (see Figure 16). It acts as a security gateway and sends data directly to the recipient from a device or database in a trusted, certified data space, so the original data provider always maintains control over the data and sets the conditions for its use. The connector uses technology that puts data inside something like a virtual container, which ensures that it is used only as agreed upon per the terms set by the parties involved.

The "requirements and reference architecture of a security gateway for the exchange of industry data and services" are defined in the DIN SPEC 27020 which has been published by the German Institute for Standardization (DIN) in February 2020. IDS is not a platform, but an architecture that allows and enables interconnectivity with other systems so that the principle of data spaces can be applied to all domains and is therefore not domain specific.

Next to the connector, an IDS ecosystem consists of several other mandatory and optional components. Among them are the *Dynamic Attribute Provisioning Service* (DAPS), the *Participant Information System* (ParIS) and the *Certificate Authority* (CA). To add **formal trust** to this architecture, the IDSA has elaborated a concept for a *certification* of IDS components as well as the participants' operational environment. With the help of approved and independent Evaluation Facilities, the IDS ensures trust in a manner that is transparent and equal for all. Optional components are the Clearing House, a Vocabulary Provider, a Broker, and an App Store, to prepare the system for use as a data marketplace.

The Chapter "System Architecture and Infrastructure" (Chapter 4.4.2) will provide an overview on the architectural principles of the IDS technology. The different roles can be found in the current version of the IDS RAM[94].

Since the IDSA is heading towards the implementation of the IDS as an open-source community several developments are already available in the IDS' Git-Hub repository[95].

[94] *IDS RAM https://internationaldataspaces.org/download/16630/*
[95] *IDS' Git-Hub repository (https://github.com/International-Data-Spaces-Association*

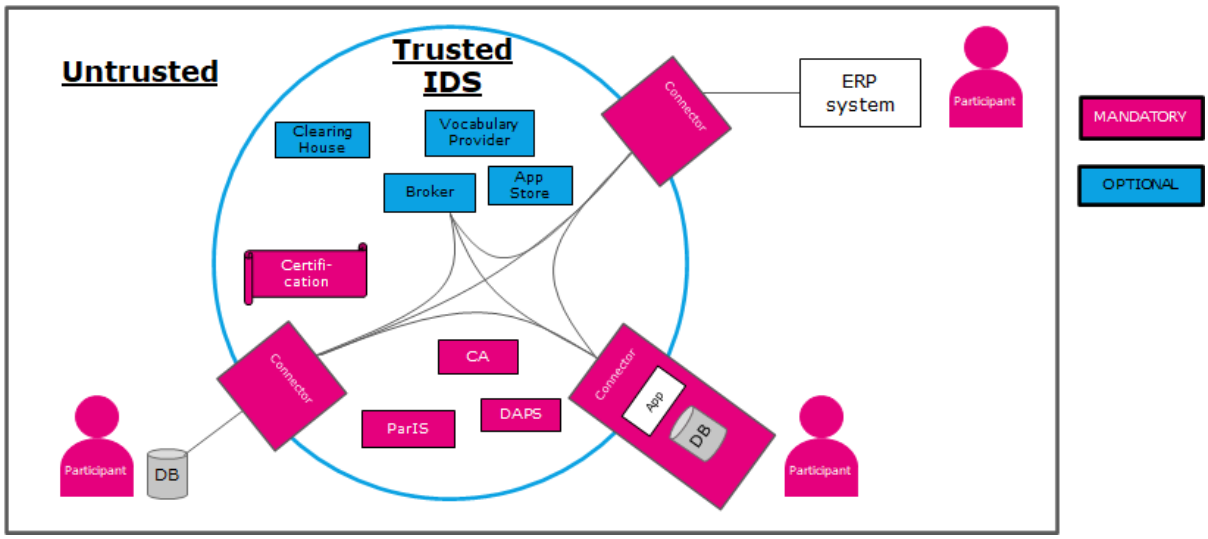

Figure 16: Required components for the IDS ecosystem, IDSA

In October 2015, the Fraunhofer Society initiated the IDS project, former IDS, funded by the German Federal Ministry of Education and Research. This endeavour is supported by the non-profit organisation IDSA, which is actively contributing. In 2021, the IDSA consists of more than 120 members from all over the world who together define the IDS standard for data sovereignty. The members of the IDSA come from different industries and provide use cases where the IDS architecture is applied in their corresponding domain.

The following Section will turn back to a broader view on different types of data marketplaces, on the one hand based on their orientation, ownership, and matching mechanism and on the other hand based on the data traded.

4 Macro Analysis – Mapping the External Environment of Data Marketplaces

The following Chapters will provide an overview on the current external surroundings data marketplaces are embedded in. Six different perspectives will be considered, shedding light on the a) political activities affecting data marketplaces, proceed with the b) economical and c) social aspects, followed by d) technology areas and technological principles used for data marketplaces while stressing the e) political conditions and finalise with f) environmental implications of data marketplaces (see Figure 17).

This Chapter is meant to complement and update the State-of-the-Art presented in Chapter 1.3 (cf. proposal) and will reconfirm the innovative potential of the TRUSTS Platform.

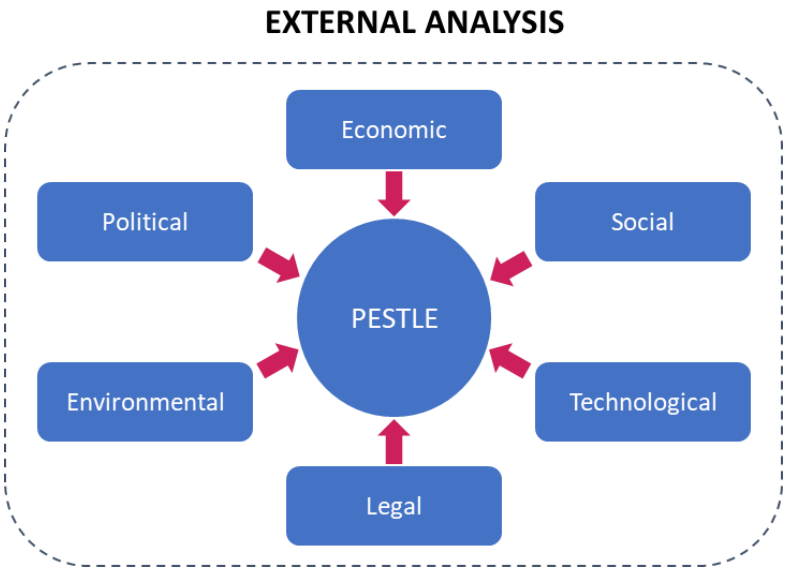

Figure 17: PESTLE framework, own representation

4.1 Political Context of Data Marketplaces

As described in the previous Chapter (3.1.1), data marketplaces emerged in the EU policymaker documents around 2017, within the Report of the European Data Market SMART 2013/0063 study entrusted initially in 2013 to IDC and Open Evidence by the European Commission (2017)[96]. In this periodically updated report and in the related Final Study Report (2020)[97], the data market, i.e., "the marketplace where digital data is exchanged as *products* or *services* as a result of the elaboration of raw data" (European Commission, 2020a) has been assessed as possessing an enormous socio-economic potential for the EU data economy. Chapter 3.4.1 "Industry perspective: a closer look at the market trends" has already emphasised the high economic potential that data marketplaces hold.

Unleashing as much of the data marketplaces' potential as possible, the EC is working towards a European Single market and has published in this regards the European Data Strategy on the 19th of February 2020, describing a vision of a common European Data Space where data can be used irrespective of its physical location of storage in the Union in compliance with applicable law[98]. This

[96] *https://a2528ba5-a-c3c32646-s-sites.googlegroups.com/a/open-evidence.com/download/repository/SMART20130063_Final%20Report_030417_2.pdf?attachauth=ANoY7cpIKvWmlRLX4olm1mEkdg14h1tnYkUV4vK0TwNQUACgO1uKgp-Jh0BNTyvWNJDkyzz3xAFlWGMj_SUyl21V9i2hawnDyeKIK7AVgqmPv6MqwGMkhwRBJ9JktnWAKjvwdbh62zaz4H44uwAtL4w95Rw41P9KzF0mCqN46-192FXk8DZ6OdH3g3-T9-XvbqZGMA6i2G7x5GjhT_Vk-w7KOKgPaHwfclQGZuB-6F5s164up6SD1Sc_zXCvxx0-CKAlsHL2nyn6&attredirects=0*

[97] *Final Study Report (Deliverable D2.9) of the Update of the European Data Market Study (SMART 2016/0063), entrusted in 2016 to IDC and the Lisbon Council. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68015*

[98] *https://eur-lex.europa.eu/legal-content/DE/TXT/?qid=1593073685620&uri=CELEX:52020DC0066 European commission (February 2020) Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions. available at https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf*

results in the goal to create an attractive economic policy framework, so that by 2030 the EU's share of the data economy is at least equal to its economic weight in the world. The EC is committed to the creation of a single European data space, open to data from all over the world, which would also support the growth of data marketplaces in a long term.

To achieve this, several key strategy points have been defined. This includes a cross-sector governance framework for data access, exchange, and use. To turn this strategy into a credible initiative, the EC expects a financial framework of 6 billion Euro from the private sector and 2 billion from the Budget of the EC. It is planned to invest this money, to foster the development of data processing infrastructures, tools, architectures, standards, and mechanisms for data sharing[99].

Furthermore, the focus lies on empowering individuals, investing in skills and in SMEs. This is intended to overcome the before mentioned challenge of fragmented data marketplaces (3.5.1 Towards a single European Data Market: Data marketplace fragmentation) by creating common "thematic" data spaces in strategic sectors and areas of public interest, considering the following nine domains:

1. Industrial (manufacturing),
2. Green Deal,
3. Mobility,
4. Health,
5. Financial,
6. Energy,
7. Agriculture,
8. Public Administration, and
9. Skills.

The open but proactive international approach, as it is called by the EC, aims to create data spaces that require an open but decisive approach to international data traffic, based on European values, such as data protection and security, equal opportunities through a federated design and the guarantee of data sovereignty for the creator of the data and trust between participants[100]. In doing so, the approach of a common European data space aims to stimulate a higher availability of data pools, technical tools and infrastructures that address domain-specific challenges and legislations.

The first instrument to realise the goals of the EU Data Strategy is the "Proposal for a Regulation of the European Parliament and of the Council of the European Union of European data governance" (Data Governance Act[101]). This instrument is meant to increase the availability of data for use by strengthening trust in data intermediaries and reinforcing data sharing mechanisms across the EU. It aims especially at establishing governance structures and mechanisms that lead to a coordinated approach for the use of data across sectors and member states which would help data economy actors take advantage of the scale of a single European data market.

[99] *https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2103*
[100] *https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf*
[101] *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767*

With the new legal framework, the Commission wants to ensure, among other things, that the data stream does not only flow neither via US corporations, such as Amazon, Google, or Facebook, nor the Chinese Social Credit System, but via independent data trustees, that act as intermediaries between data owners, data providers and data users in the data exchange[102].

To implement this, the instrument has defined four scenarios that will be addressed. The first scenario is to make public sector data available for reuse in cases where it is subject to the rights of others. As well as the sharing of data by companies in return for payment in any form (second scenario). The third scenario is to enable the use of personal data with the help of an "intermediary for the sharing of personal data". This will assist individuals in using their rights under the General Data Protection Regulation (GDPR). As a final point, the proposal also mentions the goal to enable the use of data for altruistic reasons (such as presented in the decode project[103], where citizens of Barcelona collected, among other things, data on noise, air pollution, temperature and humidity through environmental sensors inside and outside their homes)[104].

Data sharing between domains and sectors is already being done in different initiatives and forms. However, they are currently not interoperable by default due to their own technological and conceptual approaches. To tackle this challenge (amongst others), the GAIA-X[105] project has been brought to life "initiated by Europe for Europe"[106] and is now supported and carried out by representatives from politics, business and science from France and Germany, together with other European partners. GAIA-X is meant to pay also into the implementation of the European Data Strategy goals. Public funding is thus only part of the total investment that is to flow into project GAIA-X.

The GAIA-X project works not only towards interoperability of data ecosystems but also towards the establishment of data and infrastructure ecosystems according to European values and standards, to enhance the development of federated, trusted and a user-friendly digital ecosystem. In detail, the project is focusing on addressing the following challenges:

- decentralised processing locations,
- multiple technology stacks,
- lack of transparency and sovereignty over stored and processed data and infrastructure,
- insufficient clarity about the applicable jurisdiction,
- sector-specific data spaces and lack of ontology,
- absence of widely accessible application programming interfaces (APIs),
- multiple stakeholders and difficult accessibility of existing data and
- infrastructure services.

The implementation of GAIA-X is not intended to create a competing product to existing offerings. Rather, GAIA-X is intended to network various elements via open interfaces and standards in order to

[102] *https://www.wiwo.de/politik/europa/datenschutz-eu-will-zum-weltweiten-datenkontinent-nummer-eins-werden/26657952.html (February 2021)*
[103] *https://decodeproject.eu/ (February 2021)*
[104] *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767 (February 2021)*
[105] *https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html (February 2021)*
[106] *https://www.bmwi.de/Redaktion/EN/Dossier/gaia-x.html (February 2021)*

link data and create an innovation platform and is therefore leveraging existing approaches. With this endeavours, GAIA-X aims to provide an alternative data ecosystem approach to the handful of currently dominating global platforms, establishing its own data sharing and storage service mechanisms, preserving the digital sovereignty of the data owner and at the same time forming the basis for smart services and innovative business processes[107] [108].

Another important European initiative in the field of data sharing and data marketplaces is the Big Data Value Association (BDVA), which is an industry-driven international non-profit organisation strongly collaborating with the European commission to position Europe as the world leader in the creation of Big Data Value. The main targets are therefore to strengthen competitiveness and ensure industrial leadership of providers and end users of Big Data Value technology-based systems and services. Further, BDVA's aim is to promote the widest and best uptake of Big Data Value technologies and services for professional and private use, which is closely related to the development of data marketplaces. One of the BDVA's efforts culminated in the setup of the so called *i-Spaces*[109], which is a European level quality label for cross-sectoral data innovation hubs. Those are gathering data sources, AI technologies, competencies and other aspects required to allow SME's and start-ups to get their data-driven and AI-related services up and running. This endeavour is a promising instrument to foster and accelerate the uptake of a large-scale data sharing and therefore also for TRUSTS, as it offers a trusted and secure environment allowing Research, Education, and Innovation stakeholder to innovate with data. In this way, relevant stakeholders can approach the topic and become familiar with it, so that on the one hand a mind-shift takes place and on the other hand a location is created for developing and trying out innovative technologies. It also makes the economic benefits of data sharing more apparent and therefore desirable.

Worth mentioning in the context of a common approach for data sharing spaces is also the achievement of a position paper on "Design Principles for Data Spaces"[110], elaborated by the Open Dei Task Force 1 of the Horizon 2020 project "OPEN DEI – Aligning Reference Architectures, Open Platforms and Large-Scale Pilots in Digitizing European Industry", published in May 2021. Here, data space experts teamed up to define for the first time cross-sectoral and across initiatives the fundamental design principles to build data spaces. This document is summarizing the state of the art on what data spaces are and helps to understand the relation towards data marketplaces such as TRUSTS. It paves the way towards the set-up of the European Data Spaces foreseen in the above-mentioned EU Data Strategy by providing an overview on fundamentals of data spaces such as offering a definition and a high-level architecture for a data spaces in the mobility sector. Further, it defines common technical building blocks (Hardware, software, middleware, networking, etc.) as well as business, organisational and operational building blocks (artefacts), of which a data space should consist of. The document then presents data spaces of four domains (manufacturing, health, energy, and agriculture), discusses governance and business

[107] *https://www.bmwi.de/Redaktion/EN/Publikationen/gaia-x-driver-of-digital-innovation-in-europe.pdf?__blob=publicationFile&v=8 (February 2021)*
[108] *https://www.handelsblatt.com/politik/international/eu-datenstrategie-altmaier-treibt-sein-projekt-cloud-fuer-europa-weiter/25440478.html (February 2021)*
[109] *https://www.bdva.eu/I-Spaces (March 2021)*
[110] *https://design-principles-for-data-spaces.org/ (May 2021)*

models for data spaces and outlines a roadmap for co-creating a soft infrastructure suitable for the EU data spaces.

Recommendation for TRUSTS

TRUSTS is recommended to be in line with both the EU data strategy and GAIA-X.

It is also recommended to connect with the BDVA initiative to establish a vibrant community around the TRUSTS project (WP7).

4.2 Economical Context of Data Marketplaces

Data marketplaces can become significant stakeholders in a data-driven economy. The following Chapter discusses aspects of value creation from data marketplaces. Value creation with data requires the application of a series of steps starting with the raw data. This step-by-step procedure is the data valuation chain[111],[112] [23, 20], see Figure 18.

Starting from the raw data, the first step is to process data, including pre-processing and clean-up. Subsequently, data needs to be integrated, for example mapped to the business processes within a company. Via the following analysis step, insights and conclusions can be generated from the data, which results in actionable insights for an organisation and might lead to actions generating actual value for the organisation.

The data valuation chain always needs to be completed to gain value. Furthermore, data can and should be merged and blended with different data, i.e., from other sources, domains, or data that the organisation itself does not have at hand. Blended data has the potential to generate new insights and application areas beyond the possibilities of using only singular datasets.

[111] Mawer, C., (2015), "Valuing data is hard," Silicon Valley Data science: www.svds.com/valuing-data-is-hard, last accessed 6 Nov 2020.
[112] Wdowin, J. & Diepeveen, S. (2020). The Value of Data - Literature Review. Report, Bennet Institute for Public Policy.

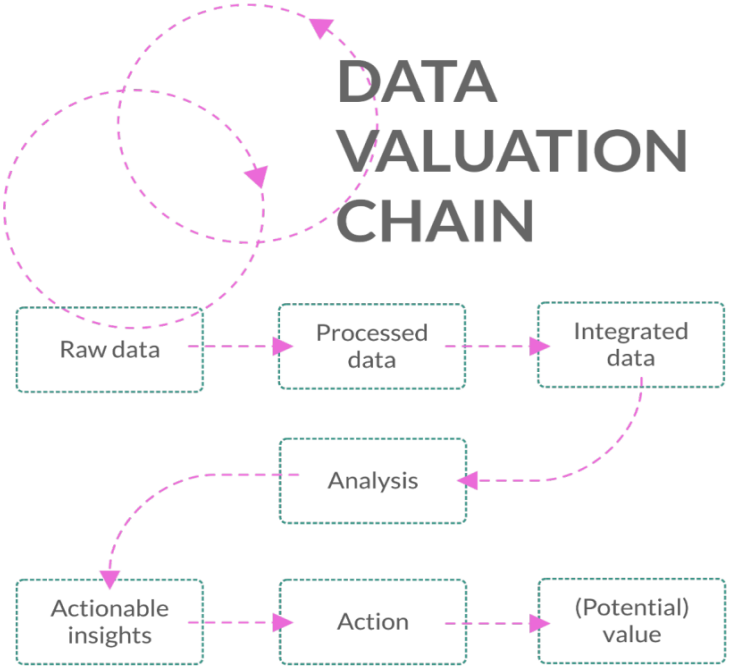

Figure 18: From data to value[113]

.

Fruhwirth et al. analysed 20 marketplaces along the business model archetypes "value creation", "value proposition", "value delivery", and "value capture"[114]. The first archetype, value creation, is represented by the three pillars platform infrastructure (centralised vs. decentralised), data origin (self-generated, user-generated, etc.), and review systems (user reviews, reviews by the marketplace, etc.) in data marketplaces. For value proposition, factors such as privacy, data quality guarantee, or pre-purchase testability are relevant. Value delivery covers, among others, technical aspects, such as the file formats of exchanged data (CSV, JSON, reports, …) or the type of access to data, e.g., via an API. Lastly, they identify the factors pricing model, price discovery, and payment currency as relevant for the business model archetype value capture.

The authors also provide an exemplary application of their taxonomy on four data marketplaces . For example, Dawex, one of the high prominent European data marketplaces, has a centralised platform, and delivers both static and dynamic datasets. Dawex provides access to its datasets both via API and download. However, it seems that Dawex itself does not provide an API. Instead, data providers need to expose a self-made API. Dawex just verifies legitimate data consumers and reroutes them to the API exposed by the data provider. This is in contrast to data marketplaces such as Namara (namara.io), which themselves provide an API a consumer can connect to. Lastly, Dawex, has a usage-based pricing model using fiat money. In contrast, IOTA and Datacoup use cryptocurrencies for billing.

[113] Mawer, C., (2015), "Valuing data is hard," Silicon Valley Data science: www.svds.com/valuing-data-is-hard, last accessed 6 Nov 2020.
[114] Fruhwirth, M., Rachinger, M., & Prlja, E. (2020). Discovering Business Models of Data Marketplaces. In Proceedings of the 53rd Hawaii International Conference on System Sciences | 2020 (pp. 5736-5747)

Table 9: Examples of implemented business model archetypes (replicated based on Fruhwirth et al. (2020)[115] , Table 3, p. 5743).

Data marketplace archetype	Centralised data trading	Centralised data trading with smart contract	De-centralised data trading	Personal data trading
Data marketplace	Quandl	Dawex	IOTA	Datacoup
Value creation	Centralised	Centralised	De-centralised	De-centralised
Value proposition	Anonymised Dynamic datasets	Encrypted Static and dynamic datasets	Encrypted Dynamic datasets	Anonymised Dynamic datasets
Value delivery	API or download Restricted access to data samples B2B No smart contract	API or download Restricted access to data samples B2B Smart contract	API No test data samples B2B Smart contract	Specialised software to access No test data samples C2B Smart contract
Value capture	Freemium pricing Prices set by sellers Fiat currency	Usage based pricing Prices set by sellers Fiat currency	Flat free pricing Price set by sellers Crypto currency	Usage based pricing Fixed prices Crypto currency

[115] Fruhwirth, M., Rachinger, M., & Prlja, E. (2020). Discovering Business Models of Data Marketplaces. In Proceedings of the 53rd Hawaii International Conference on System Sciences | 2020 (pp. 5736-5747)

A study by the EC shows that there are five key aspects enabling a data ecosystem, i.e., the easy availability of datasets, an infrastructure, required skills, security, and trust, see Figure 19 below.

Figure 19: The five pillars for a healthy data ecosystem[116] .

The EC conducted a study with 129 companies in the European Economic Area, i.e., EU member states, as well as Iceland, Liechtenstein, and Norway, to quantify data sharing efforts in the economy[117] . The sample covered all company sizes, i.e., small companies with fewer than ten employees to large companies with more than 250 employees. The study investigated factors such as the perceived benefits of a data-sharing economy, but also potential obstacles.

A data-sharing economy is expected to become more relevant in the next few years, with more and more companies having data sharing as their primary source of income (see Figure 20). In case the trend materialises as expected the potential of data marketplaces as a central point for sharing, selling, and buying of data will increase as well. As shown in Figure 19, a data-sharing infrastructure is one of the five pillars of a data ecosystem, and data marketplaces play a significant role here with their ability to efficiently store, archive, exchange, and settle payments of data purchases.

[116] Arnaut, C., Pont, M., Scaria, E., Berghmans, A. & Leconte, S. (2018). Study on data sharing between companies in Europe, Final report. DOI: 10.2759/354943.
[117] Arnaut, C., Pont, M., Scaria, E., Berghmans, A. & Leconte, S. (2018). Study on data sharing between companies in Europe, Final report. DOI: 10.2759/354943.

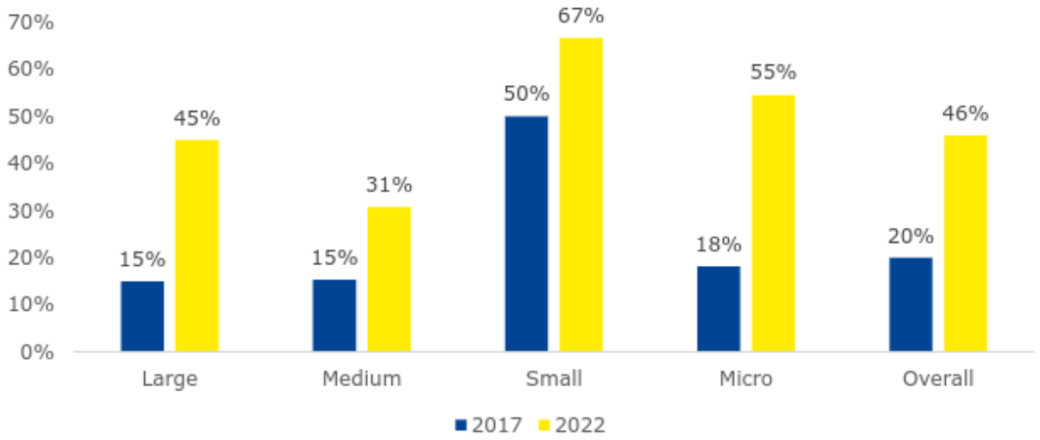

Figure 20: Data sharing as a primary source of income in 2017 vs. 2020[118]

.

Companies mostly shared data related to the IoT and from internal IT systems. The IoT is relevant for sectors such as manufacturing, the automotive industry, or smart home automation. On the other hand, crowdsourcing and data from cookies only play a minor role for sharing activities (see Figure 21). Companies see three major benefits from data sharing. The most important factor is being an entry point to establish relations and partnerships with other companies (see Table 10). Another important aspect is the generation of revenue for the company as well as to drive further innovations. Here, data marketplaces can play a significant role. Since they can potentially provide data from a wide variety of domains from countless sources, be it research, the public, or private companies, they will also drive

[118] Arnaut, C., Pont, M., Scaria, E., Berghmans, A. & Leconte, S. (2018). Study on data sharing between companies in Europe, Final report. DOI: 10.2759/354943.

innovation and allow the invention and realisation of completely new applications and services. When data is combined with additional data, its value increases immensely.

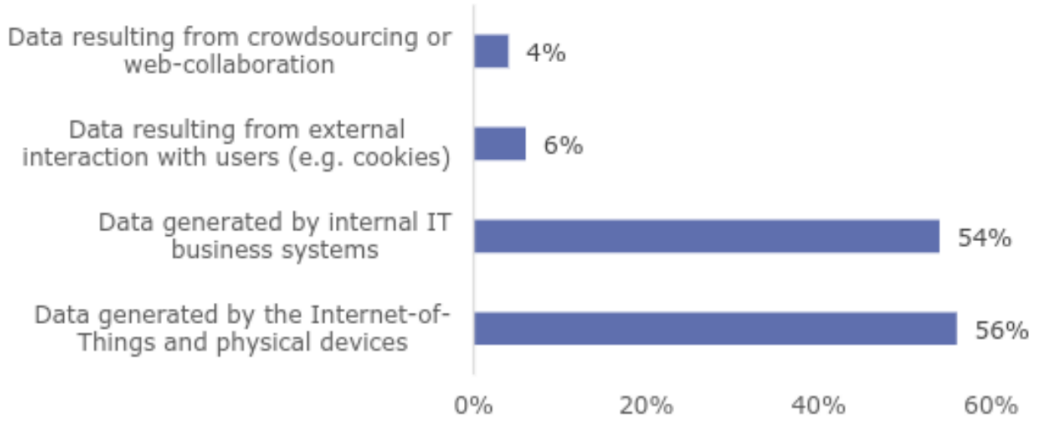

Figure 21: The different sources of shared data[119] [19, p. 40].

Table 10: Benefits that companies see in data sharing (replicated based on Arnaut et al. (2018)[120], Table 5, p. 45)

Benefits	*No. responses*	*%*
Data can allow my company to enter into partnerships with other companies	*28*	*62 %*
Data can be monetised and generate revenues for my company	*27*	*60 %*
Data support the innovation component of my company	*24*	*53 %*

However, there is also a range of concerns relevant to companies, which might curb their willingness to contribute to a data-sharing economy and exchange data, potentially even with competitors. Most important are privacy concerns (49%), followed by the fear that trade secrets might get revealed or others get insight into a company's commercial strategy (33%). Also, a 32% of companies feel that there is no demand for their data.

Another important factor refers to the pillar of "security" of the five pillars required for a data ecosystem: 28% of the companies do not trust the technical procedures used to share data. This is an important aspect that data marketplaces need to target.

On one hand, it is essential for them to adopt state-of-the-art security and privacy technologies. On the other hand, it is crucial to promote knowledge about these technologies and clarify misunderstandings. The perception of the public with regards to data privacy and security is often driven by media coverage

[119] Arnaut, C., Pont, M., Scaria, E., Berghmans, A. & Leconte, S. (2018). Study on data sharing between companies in Europe, Final report. DOI: 10.2759/354943

[120] Arnaut, C., Pont, M., Scaria, E., Berghmans, A. & Leconte, S. (2018). Study on data sharing between companies in Europe, Final report. DOI: 10.2759/354943.

about data leaks and data stolen by malicious perpetrators. An information campaign about the trustworthiness is necessary to restore trust into processes and technologies required in a data-sharing economy.

Another reason why companies are reluctant to share data is merely the lack of incentives to do so. It can be difficult to define prices for data, as it is an intangible good, which can be duplicated at almost no cost as soon as the initial costs of creation of the data are accomplished. Unclear pricing strategies make it difficult for companies to fix prices for data they offer, and also difficult for potential buyers to actually see the value of potentially expensive datasets. Educational investments will help to overcome another obstacle mentioned by companies, which is the lack of skills required to participate in a data ecosystem. Qualified employees are missing, which significantly increases the entrance barrier. The European Data Market Monitoring Tool [24] predicts that in the case of a high-growth scenario of the European data economy there will be a gap of 1.1 million or 10.5% of the required data professionals in the EU27 in the year 2025.

Table 11 shows a detailed listing of reasons preventing companies from entering the data-sharing economy.

Table 11: Entry barriers of the data-sharing economy (replicated based on Arnaut et al. (2018)[121] , Figure 27, p. 44).

Reasons	No. responses	%
Privacy concerns	39	49 %
Trade secrets / fear of misappropriation by others / considerations of commercial strategy	26	33 %
Lack of demand for my company's data	25	32 %
Uncertainty about safety, security and liability conditions related to the technical process of sharing data	22	28 %
Lack of incentives to share data	22	28 %
Lack of data skills inside the company	14	18 %
Economic costs of sharing data (e.g. costs of making the data available in the desired format, infrastructure costs related to data collection, data curation costs, etc.)	12	15 %
High efforts and burden on the company to engage in this activity (e.g. collection, analysis, etc.)	12	15 %
Uncertainty about usage rights on the data and potential reputational costs for the company in case of misuse	12	15 %

[121] Arnaut, C., Pont, M., Scaria, E., Berghmans, A. & Leconte, S. (2018). Study on data sharing between companies in Europe, Final report. DOI: 10.2759/354943.

Difficulties with measuring the value of data	9	11 %
Lack of appropriate licensing conditions	0	0 %

There are several aspects helping to overcome entry barriers to the data-sharing economy. An important aspect is a clearly defined legal framework with regards to data ownership (62%). Another highly relevant aspect is the ability to log data usage after a dataset has been shared (46%). Table 12 gives more details on helping companies to overcome entry barriers.

Table 12: Considerations reducing reluctance to share data (replicated based on Arnaut et al. (2018)[122] , Table 6, p. 45).

Factors that can increase willingness to share data	No. responses	%
Legal clarity about the "ownership rights" of the data	49	62 %
Ability to track the usage of the data once it has been shared	36	46 %
Certainty about how to share data from a contractual point of view	33	42 %
Availability of the necessary technical skills inside my company to ensure the quality and security of the data shared	23	29 %
An improved framework to protect the investments made for the purpose of data collection, curation, anonymisation, etc.	19	24 %
Availability of standards and/or infrastructure to facilitate the adequate storage, transfer and processing of data	13	16 %
A defined framework for liability in case of damage caused by the data that are shared	12	15 %

There are major differences in the willingness to share data depending on the sector of the company. for example, manufacturing companies (39%) and IT companies (34%) show the highest degree of willingness, followed by the automotive and transport industry (27%). The domain least willing to share is, interestingly, the research domain, with only 20% of all companies willing to share their data. A potential reason might be that research data is considered as a crucial aspect for the survivability of those companies and their advance as compared to competitors too important to give it up. Figure 22 shows a detailed chart comparing willingness to share by domains.

[122] Arnaut, C., Pont, M., Scaria, E., Berghmans, A. & Leconte, S. (2018). Study on data sharing between companies in Europe, Final report. DOI: 10.2759/354943.

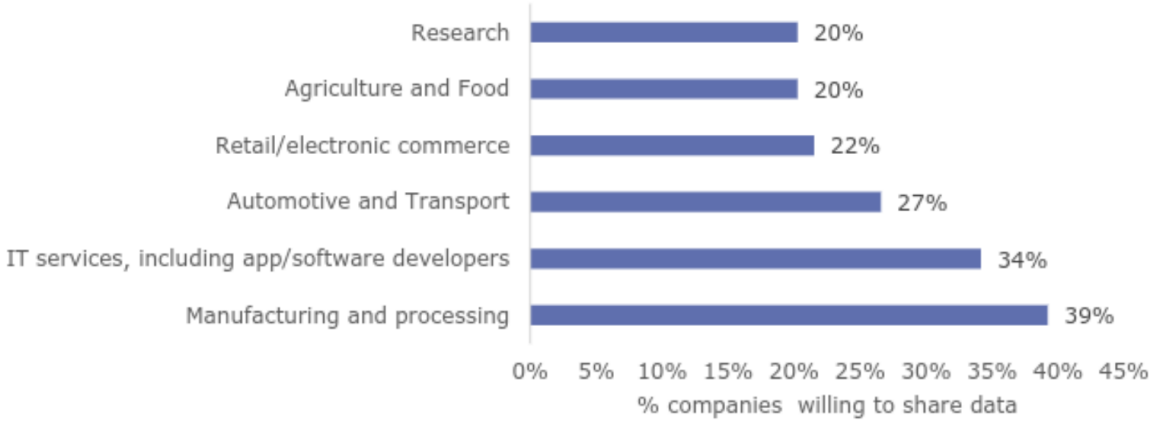

Figure 22: Economic sectors and their willingness to share data[123] [19, p. 46].

Recommendation for TRUSTS

The growing economic significance of a data-driven economy bears huge potential for data marketplaces. The following recommendations for the TRUSTS platform can be derived from the economic analysis:

- The provision of all necessary means to complete the data valuation chain, i.e., either by built-in services or space for third-party apps (see Figure 18),
- The provision of facilities to combat the fears of potential data market customers, i.e., privacy concerns, fear of disclosure of trade secrets, lack of trust in the technical procedures of the platform (see Table 11), **[WP4]**
- The implementation of a persuasive promotion strategy to inform stakeholders about TRUSTS trustworthiness, (**WP7** and **WP8**)
- The launch of an information and awareness campaign to convince organisations about demand for their data and consequently increase their willingness to share, (**WP7** and **WP8)**
- The development of a clear legal framework to guide organisations through the process of developing appropriate business models as well as the setup of the technology stack required to make their data exchange ready (WP6).

4.3 Social aspects of data marketplaces

Data marketplaces shape the way how data is used in society. The ubiquity and availability of data from an extensive variety of domains can be a driver of innovation and create new and unforeseen business models. A challenge data marketplace is facing an ironic twist in the perception of the value of data: data is being considered as the new oil, gold, or renewable energy, however, potential buyers share an

[123] Arnaut, C., Pont, M., Scaria, E., Berghmans, A. & Leconte, S. (2018). Study on data sharing between companies in Europe, Final report. DOI: 10.2759/354943

unwillingness to pay for data, as shown in the case of Swivel[124],[125] . The case of Swivel is a decade ago, but still relevant today. The societal perception that everything on the Internet has to be free needs to make room for an appreciation of the value of data and a willingness to pay (fair) prices to those providing it. In parallel, the awareness of potential use cases for data-driven business needs to be raised. There is still a gap between the potential of data and the knowledge companies possess about it.

Insights from the TRUSTS World Café

The discussions in the TRUSTS World Café with regards to social aspects showed the relevance of appropriate incentivization schemes for data providers need to be combined with a solid legal framework. The benefits of sharing data need to be clear for data providers. Issues such as the **right to be forgotten** are especially relevant for de-centralised data markets such as TRUSTS. Appropriate means to enforce these rights have to be conceptualised and implemented. Differences in legal regulations by countries add another layer of difficulty. A concrete suggestion to combat such problems was the introduction of a data trust, i.e. an entity dedicated to adherence and execution to existing law and ethical rules. **It is impossible for individuals sharing their data within a data market to find out if all laws and regulations with regards to their data rights have been met. Thus, an entity has to take over this task, and, for example, make sure that data gets removed across nodes if demanded by the data owner.**

The increasing availability of data, especially when served on data marketplaces, is also going to shape the profile of data science jobs. Currently, data scientists need to be hired at a company to work with the company's data. In the future, when data marketplaces are established on the market and allow data scientists to offer self-written applications and services offered as products on data marketplaces, data scientists can become self-employed and create business models around the existing data.

Similar to app stores, data marketplaces can become a place where new applications and services are developed and offered to interested and potential buyers. This can create an entirely new ecosystem of data-driven applications connecting data of multiple stakeholders in ways that could not have been developed with the variety of available data that companies alone usually do not have. The European Data Market Monitoring Tool[126] predicts growing numbers of data professionals in the upcoming years. Historical growth rates in 2019 of 6.1% in the EU27 will be followed by annual growth rates of 7.2% and 6.5% annually in the years until 2025. In the best-case scenario, this would result in a total of data professionals of 10.9 million in the EU27.

Despite the continuously and significantly growing numbers, the European Data Market Monitoring Tool predicts a skill gap in the upcoming years. The gap between demand and supply of data skills will grow by 8.2% and 10.5% in the baseline and high-growth scenarios for the EU27. Only the challenge scenario predicts a gap of 3.3% and thus manageable by the market. The challenge scenario is combined

[124] Kosara, R.: The rise and fall of swivel.com (2010). Last Accessed: 2014-11-20

[125] Stahl, F., Schomm, F., Vossen, G. (2016). A classification framework for data marketplaces. Vietnam J Comput Sci 3, 137–143. https://doi.org/10.1007/s40595-016-0064-2

[126] Cattaneo, G., Micheletti, G., Glennon, M., La Croce, C., Mitta, C. (2020). The European Data Market Monitoring Tool: Key Facts & Figures, First Policy Conclusions, Data Landscape and Quantified

with a significant throttling of demand in the data-driven economy in the years until 2025. In absolute numbers, the baseline scenario reveals a shortage of 759,000 positions, while the high-growth scenario struggles with 1.1 million unfilled positions. In the challenge scenario, there are still 484,000 skilled laborers missing.

Data marketplaces will also change the way people are willing to share their data. There has been a dramatic increase in sensitivity about the dangers of data sharing and analysis with detrimental effects on the willingness of people to share their data. However, when data marketplaces become established and trusted places where data sovereignty is guaranteed and fair prices are paid for data, this might result in a rethinking of people. They are able to see the value of their data clearly and are aware of what can be done with it. If driven correctly and in a fair way, the increased trust will persuade people that sharing data can be beneficial. Privacy-ensuring technologies such as anonymization, de-identification, or federated learning guarantee that no personal data is shared. Trust in these technologies must be raised by informing and persuading people of their efficiency, trustworthiness, and usefulness. This requires a change in societal thinking that will ultimately also be crucial for the sustainability of data marketplaces and their future businesses.

Data marketplaces are strongly dependent on the availability of players willing to share their data. This requires a solid and healthy ecosystem of organisations actively collecting data and/or willing and are ready to leverage data for their business purposes. The European Data Market Monitoring Tool evaluates the growth rates of data companies, hereby differentiating between data suppliers and data users. Data suppliers are defined as producing and delivering data-related products, services, and technologies. Data users represent the demand side of the data-driven economy and are represented by companies using the insights gained from data analysis to proactively improve their business processes. The tool predicts solid growth rates of 1.3% (the challenge scenario), 2.5% (the baseline scenario), and 4.8% (the high-growth scenario) in data suppliers. On the demand side, there is an equally significant growth of 0.7% (the challenge scenario), 1.4% (the baseline scenario), and 2.9% (the high-growth scenario) among data users.

Recommendation for TRUSTS

The previous Chapter identified a significant gap in the demand and supply of data professionals in the upcoming years until 2025. Despite growing numbers of data professionals, actual demand will outweigh supply. Data marketplaces mitigate this challenge in their role as connectors between organisations requiring data skills and professionals seeking to apply their knowledge and become active in the data-driven economy. They can become the incubators for innovative business ideas since they allow the combination and merging of data and data services in new and unforeseen ways.

TRUSTS can support these efforts by:

- directing dissemination activities explicitly to relevant stakeholders, i.e., data companies seeking data skills and data professionals alike. (WP8)
- putting effort into the creation of a UI that attracts skilled labour and facilitates matching between involved stakeholders. This recommendation is not within the scope of the TRUSTS

project itself, but strongly recommended to the TRUSTS operator taking over after the project ends. (WP3)

- making the platform ready to support the creation of innovative business models to attract organisations. (WP7+WP8)
- providing a legal framework and playground where data exchange is accomplished in a trusted environment. This recommendation is at the core of TRUSTS. Our research has confirmed the relevance of this aspect.

4.4 Technological aspects of data marketplaces

With the rise of the data economy and the respective business models, regulations but also funding opportunities, as well as technologies used for such data marketplaces are becoming more and more important and different approaches for the use of technology in this field have been evaluated and used in real-world implementations. This Chapter provides an overview of the technology areas and technological principles used for data marketplaces. Furthermore, it provides an outlook on the current and future trends of data marketplace technologies.

The below Section focuses on the technologies and standards used for data markets and 'data markets enabling technology activities and projects' to realise secure data exchange and enabling sovereignty in data exchange.

4.4.1 Types of Data Markets

Looking at the history of online data marketplaces, several phases and types of data marketplaces have been mainly specified by their respective objective as well as by the types and sources of data.

In a first phase, the development and operation of data catalogues were prevalent, mainly in the field of open data. These were systems providing mainly search and browse mechanisms over the metadata[127] of openly/publicly available data sets. Such data catalogues rarely provide data itself but all metadata and links to the datasets that are hosted at the data publisher side (e.g., a government website or similar).

In a second phase we have seen government and industrial (business) data marketplaces coming into existence providing mechanisms for secure data exchange by also providing data sovereignty for the data owner.

Afterwards, in a **third phase,** we see (i) personal data markets appearing that provide solutions for the management and provision/sharing of individual personal data, as well as open science marketplaces entering the stage, that provide huge amounts of data via the European Open Science Cloud (EOSC, https://www.eosc-portal.eu/ and https://marketplace.eosc-portal.eu/).

[127] *https://en.wikipedia.org/wiki/Metadata, accessed March2021*

An additional type - mentioned by https://rubygarage.org/blog/big-data-marketplaces - is the Sensor Data Marketplace, i.e., marketplaces providing features to exchange and trade data coming from IoT devices and sensors.

The EU Data Landscape[128] lists 14 active data marketplaces, but this seems to be a small selection of what is available.

4.4.2 System Architecture and Infrastructure

This Section provides an overview of the architecture approaches as well as the infrastructure principles that data marketplaces follow today. There are two different directions that have been discussed and used in regards to data marketplace architecture so far:

- centralised approach and
- federated approach.

Centralised Approach

In a centralised approach all metadata and data are stored and processed centrally in one single system and infrastructure. This means all components are running on a dedicated infrastructure (e.g., cloud-based server cluster)

Figure 23 below shows a centralised architecture approach of the SSHOC Marketplace[129].

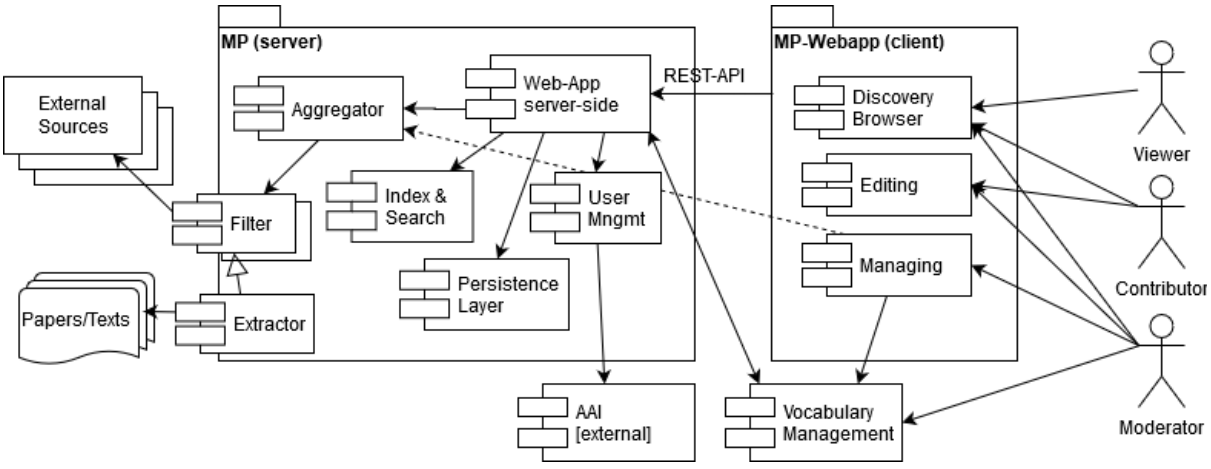

Figure 23: Centralised architecture approach of SSHOC Marketplace

[128] *https://datalandscape.eu/eu-data-landscape, accessed 18.04.2021*
[129] https://www.sshopencloud.eu/, accessed 18.04.2022

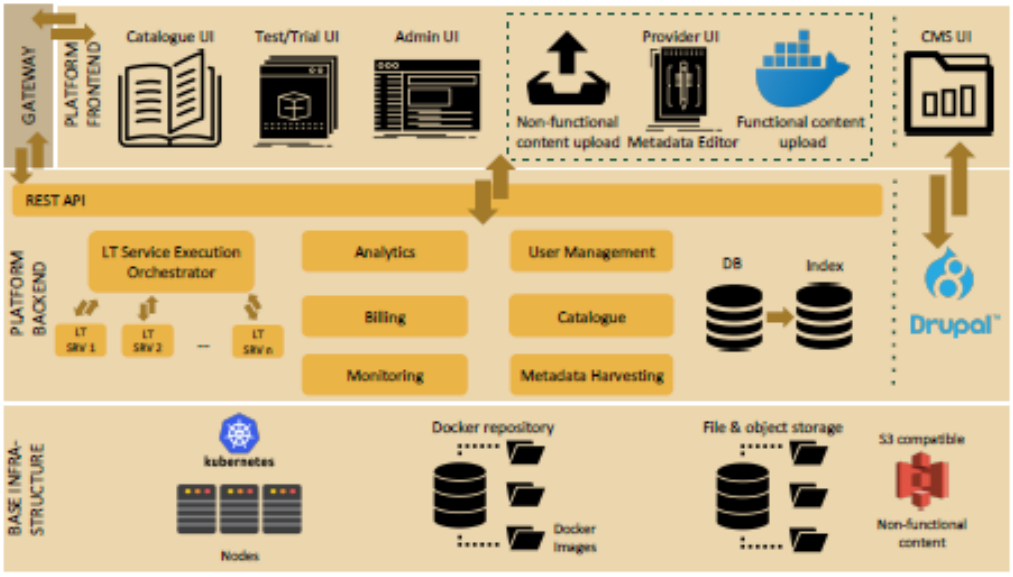

Figure 24: Centralised architecture approach European Language Grid.
Source: ELG, https://www.aclweb.org/anthology/2020.lrec-1.413.pdf

Federated Approach

This approach has been strongly promoted by IDS in its RAM, taken from: ISST Report[130], and it is as follows. See also Figure 27: Federated Approach of IDS Architecture.

In the course of digitalisation, a number of initiatives have been launched to guide the way towards the design of a networked and digital world. To identify relevant partners for the exchange of data, domain-specific communities arise and promote a federated approach to share knowledge and develop technological solutions. IDS is an initiative that promotes a virtual data space leveraging existing standards and technologies, as well as governance models to facilitate secure and standardised data exchange and data linkage in a trusted data ecosystem. It thereby provides a basis for creating smart-service scenarios and facilitating innovative cross-company business processes, while guaranteeing data sovereignty for data owners.

Figure 25 is depicting the architecture of the IDS connector.

[130] *ISST Report: Data Ecosystems - Conceptual Foundations, constituents and recommendations for action, October 2019, Boris Otto, Dominik Lis, Jan Cirullies, Jan Jürjens.*

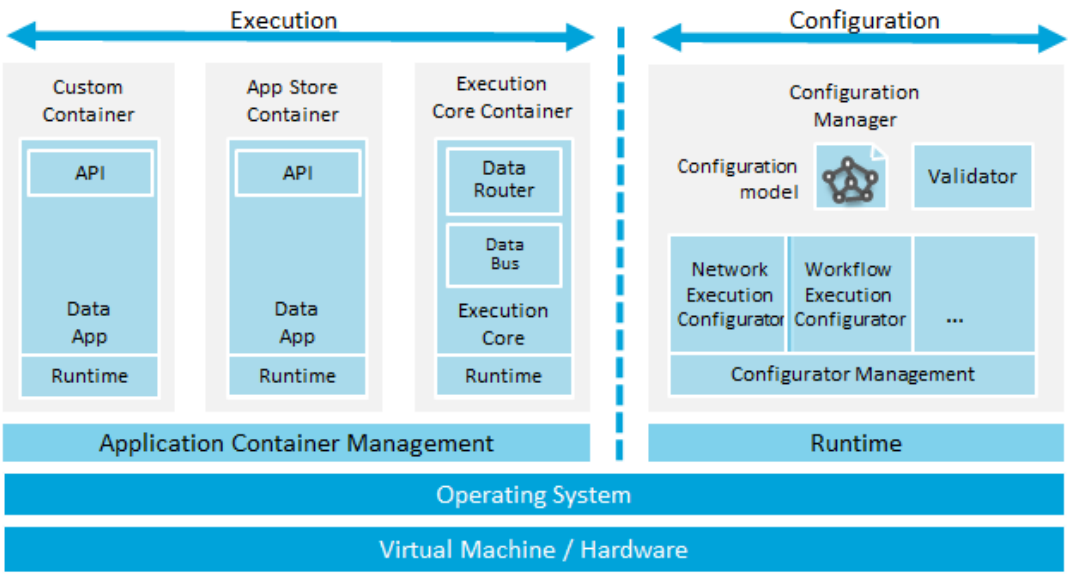

Figure 25: Reference Architecture. Source: IDS RAM3.0 page 62 https://internationaldataspaces.org/download/16630/, accessed 17/02/2021.

The IDS is open and non-profit and sets standards for the secure and trustworthy use of data in the platform economy. Market participants - software companies, technology providers, etc. - take up the standards and develop innovative business models based on them.

Data sovereignty is a central aspect of the IDS. It can be defined as a natural person's or corporate entity's capability of being entirely self-determined regarding its data. The IDS initiative proposes a RAM for this particular capability and related aspects, including requirements for secure and trusted data exchange in data ecosystems. An overview of the components' interactions according to the IDS-RAM can be found in Figure 26.

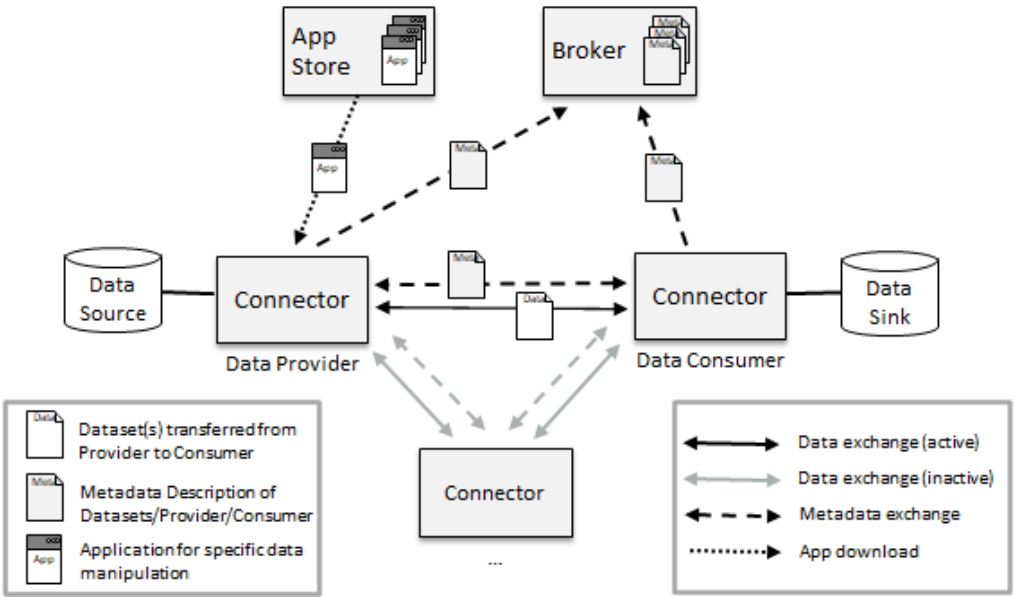

Figure 26: Overview of the component interactions in IDS Reference Architecture. Source: IDS RAM3.0 page 62 https://internationaldataspaces.org/download/16630/, accessed 17/02/2021[131]

The development of the IDS architecture was guided by the following principles and requirements:

- **Trust**: Trust is the basis of the IDS. Each participant is evaluated and certified before being granted access to the trusted business ecosystem.
- **Security and data sovereignty**: All components of the IDS rely on state-of-the-art security measures. Apart from architectural specifications, security is mainly ensured by the evaluation and certification of each technical component used in the IDS. In line with the central aspect of ensuring data sovereignty, a data owner in the IDS attaches usage restriction information to their data before it is transferred to a data consumer. To use the data, the data consumer must fully accept the data owner's usage policy.
- **Ecosystem of data**: The architecture of the IDS does not require central data storage capabilities. Instead, it pursues the idea of decentralisation of data storage, which means that data physically remains with the respective data owner until it is transferred to a trusted party. This approach requires a comprehensive description of each data source and the value and usability of data for other companies, combined with the ability to integrate domain-specific data vocabularies. In addition, brokers in the ecosystem provide services for real-time data search.
- **Standardised interoperability**: The IDS Connector, being a central component of the architecture, is implemented in different variants and can be acquired from different vendors. Nevertheless, each Connector is able to communicate with any other Connector (or other technical component) in the ecosystem of the IDS.
- **Value adding apps**: The IDS allows to inject apps into the IDS Connector in order to provide services on top of data exchange processes. This includes services for data processing, data

[131] *IDS RAM3.0 page 62 https://internationaldataspaces.org/download/16630/, accessed 17/02/2021.*

format alignment, and data exchange protocols. Furthermore, data analytics services can be provided by remote execution of algorithms.

- **Data markets**: The IDS enables the creation of novel, data-driven services that make use of data apps. It also fosters new business models for these services by providing clearing mechanisms and billing functions, and by creating domain-specific broker solutions and marketplaces. In addition, the IDS provides templates and other methodological support for participants to use when specifying usage restriction information and requesting legal information.

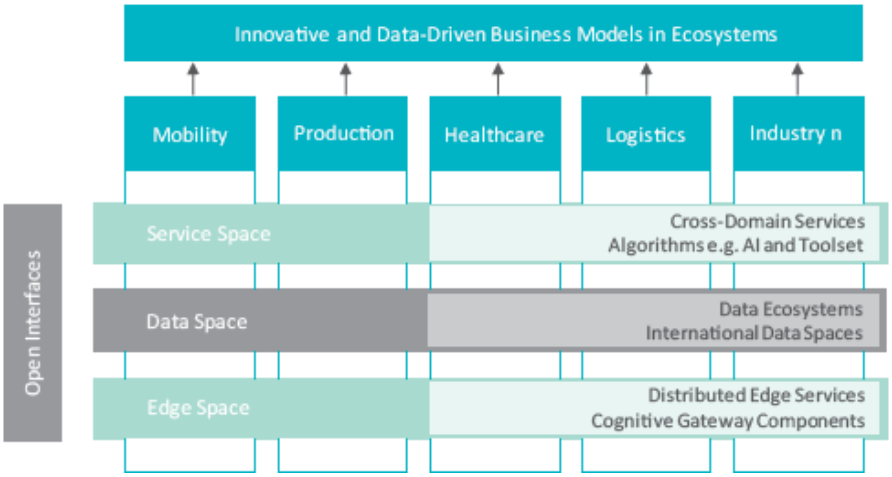

Figure 27: Federated Approach of IDS Architecture[132]

4.4.3 Components and Standards

This Section provides an overview of (i) the most important technical components used in data markets as well as of (ii) standards relevant in the field.

The technical components Section is aligned with the list of features in Chapter 2.2.3 'Data marketplace features' of this document but keeps the focus on the technical components required and used for such features.

Overview of the technical components

- **Data/metadata harvester**: software components that can harvest metadata and data from 3rd party sources along a specified schema/format. Sometimes such a harvester component is for metadata only (federated architecture approach) as metadata needs to be acquired, enriched, and stored centrally. Such components often include mapping features (to map incoming metadata/data to a specified data model) and sometimes also quality assurance mechanisms (data cleansing, and aggregation features.

[132] *ISST Report: Data Ecosystems - Conceptual Foundations, constituents and recommendations for action, October 2019, Boris Otto, Dominik Lis, Jan Cirullies, Jan Jürjens*

- **Catalogue components**: for metadata and/or data where such metadata and data can be browsed and searched regarding the user's requirements. Such catalogues often include sub-components such as:
    - metadata store
    - data ingestion component and data store
- **Authentication and Authorization components**: allow user management, potentially single sign on (SSO) and access rights management for users and groups (and roles).
- **Billing:** provides billing and payment features for the marketplace.
- **Smart Contracting** component: allows two or more parties to make use of contract templates for data trading or the commercial use of data driven services. Such smart contracting systems are often based on blockchain technology as of the nature of contracts to allow no changes and see the whole history of versions after execution.
- **Analytics**: includes the metadata and data analytics features and services that such a system requires to provide a useful and powerful marketplace.
- **Data Model & Metadata Management**: in the form of a component that allows e.g., vocabulary or taxonomy or ontology management, to develop and maintain the semantic model of the data market/data space.
- **Frontend component**: provides ability to develop and maintain graphical user interfaces (GUIs) and views.
- **Security Layer**: one or more components that ensure security through the whole system and all components in place.
- **Orchestration Component**: component that orchestrates all relevant services of a data market that are required to enable operation of the system.

In regard to the infrastructure and deployment of data markets and data spaces, a clear trend in the direction of making use of cloud infrastructure and containerization technologies (also called OS Level Virtualisation)[133] is perceptible.

Whereby a **Cloud Infrastructure** can be defined as a term used to describe the components needed for cloud computing, which includes hardware, abstracted resources, storage, and network resources. Think of cloud infrastructure as the tools needed to build a cloud. In order to host services and applications in the cloud, you need cloud infrastructure[134].

[133] *https://en.wikipedia.org/wiki/OS-level_virtualization , accessed 12/2020*
[134] *https://www.redhat.com/en/topics/cloud-computing/what-is-cloud-infrastructure , accessed 12/2020*

For **Containerization/Virtualisation Technologies**, different approaches are in place at the moment like for application containerization Docker[135] and for orchestration Kubernetes[136] or OpenShift[137].

In regards to the **frontend technologies** (for GUI/UI) is specified as the use of state-of-the-art web technologies with a strong focus on JavaScript (JS)[138], often ReactJS[139], and HTML but also in combinations with an out-of-the-box Web Content Management System ((W)CMS) like Drupal[140] or others.

List of relevant standards

- DCAT (AP), The DCAT Application Profile for data portals in Europe (DCAT-AP), https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe
- schema.org, https://schema.org/
- INSPIRE Data Specifications, https://joinup.ec.europa.eu/collection/inspire/solution/inspire-data-specifications/about
  and the related directive: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32007L0002
- FAIR principles, https://www.go-fair.org/fair-principles/
- RDA FAIR data maturity model, https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/rda-fair
- GAIA-X - A Federated Data Infrastructure for Europe, https://en.wikipedia.org/wiki/GAIA-X, https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html
- DIN SPEC 27070 "Requirements and reference architecture of a security gateway for the exchange of industry data and services", https://www.beuth.de/de/technische-regel/din-spec-27070/319111044 and https://www.internationaldataspaces.org/ids-officially-a-standard-din-spec-27070-is-published/
- W3C Web Content Accessibility Guidelines, https://www.w3.org/WAI/standards-guidelines/wcag

[135] *https://www.docker.com/ , accessed 12/2020*
[136] *https://kubernetes.io/ , accessed 12/2020*
[137] *https://www.openshift.com/ , accessed 12/2020*
[138] *https://en.wikipedia.org/wiki/JavaScript , accessed 12/2020*
[139] *https://reactjs.org/ , accessed 12/2020*
[140] *https://www.drupal.org/ , accessed 12/2020*

4.4.4 Comparison: examples of data marketplaces technology

This Section contains a comparison of four data marketplace systems in different areas along a specified set of attributes that are as follows.

- Infrastructure
- Architecture
- Standards (used)
- Type | Industry

The following data markets and data markets enablers have been identified for this comparison (see Table 13 Comparison in key technological areas of 4 data marketplaces).

1. International Data Space (IDS; https://www.internationaldataspaces.org/)
2. Data Market Austria (DMA, https://datamarket.at/)
3. SSHOC Marketplace (https://www.sshopencloud.eu/)
4. European Language Grid (see Figure 24) (ELG, https://www.european-language-grid.eu/)

Table 13: Comparison: examples of data marketplaces technology

	Infrastructure / Deployment method	**Architecture**	**Standards used**	**Type | Industry**
IDS	Docker, information about orchestration et al not available	federated	DIN SPEC 27070 IDS Standards like IDS RAM (also used for GAIA-X).	Focus on manufacturing industry (data and services)
DMA	Docker, plus: using OpenShift for orchestration	federated	DCAT-AP schema.org INSPIRE	All types of open and commercial data (data and services)
SSHOC	Docker, no orchestration technology used	centralised	FAIR Principles	Research and Scientific Data (data and services)
ELG	Docker, plus using Kubernetes for orchestration	centralised	W3C WCAG	Language Resources and Language Technology relevant data and services

4.4.5 Additional findings from the TRUSTS World Cafe

In March 2021 the TRUSTS consortium organised a two-hour virtual World Cafe with ~30 invited experts to discuss and collect additional input for this deliverable.

The statements and discussions with the 4 groups of participants took place along 2 questions in the field of technologies, as follows:

- What are the most important features for a data marketplace?
- How important is standardisation, which standards are you aware of / are important?

Insights from the TRUSTS World Café

An important expectations for a data marketplace is to **act as a broker between supply and demand** (sellers and buyers) and thereby offer a useful range of features and functionalities for this like **a) rich metadata** that informs the user about the data and its attributes, the data quality (data quality was mentioned as one of the most important attributes), potentially also user ratings, etc that is used for **b) search and discovery mechanisms** as well as **c) smart assistants** as recommender that guide users to relevant data available based on their needs and profiles and industry (e g in the form of data circles).

Such metadata, the data itself and the features (e.g., APIs) shall be **based on existing standards where possible** (data models, metadata schemas, formats, etc. for certain industries as well as cross-industry) **to ensure interoperability** with source systems (e g in data provision by data providers) but also to other data marketplaces and ensure compliance. **Standardisation as a feature.**

Security and trusted data sharing are key enabler. Secure data storage, and data exchange (for instance data is secure before transferred to an end customer) by encryption and secure protocols, clearly specified access rights as well as information about provenance and lineage of data are important factors to ensure trustful data exchange.

Support the valuation of data for an end user. Such valuation in different context can be supported - beside the topics mentioned above - by the provision of data samples, the possibility to experiment with data in a secure space and/or by enabling Proof of Concepts (PoCs) on top of the data, means to validate if data can clearly support an end customer use case. Potentially AI mechanisms could be used to suggest pricing for data for certain users.

4.4.6 Trends and Outlook

Taking all information of this Section into account and looking at current data and metadata management trends there is a trend towards a (Semantic) Data Fabric (Gartner, Top 10 Data and Analytics Technology Trends for 2019), that combines elements of a traditional Data Warehouse and a Data Lake with advanced feature sets of (i) connectivity and (ii) reusability. For details on this approach please see the graphics and explanations below.

For example, Gartner Identifies Data Fabrics as a clear trend in their Top 10 Data and Analytics Technology Trends for 2019[141].

Data fabric enables frictionless access and sharing of data in a distributed data environment. It enables a single and consistent data management framework, which allows seamless data access and processing by design across otherwise siloed storage.

Through 2022, bespoke data fabric designs will be deployed primarily as a static infrastructure, forcing organisations into a new wave of cost to completely re-design for more dynamic data mesh approaches.

The Data Fabric approach

Gartner Research provides specification of and insights into the Data Fabric approach[142] as follows:

Summary:

Data management teams are under constant pressure to provide faster access to integrated data across increasingly distributed landscapes. Data and analytics leaders must upgrade to a data fabric design that enables dynamic and augmented data integration in support of their data management strategy[143].

Impacts and Recommendations[144]

- ML-Augmented Data Integration is making active metadata analysis and semantic knowledge graphs pivotal parts of the data fabric,
- Data Fabric must have the ability to collect and analyse all forms of metadata,
- Data Fabric must have the ability to analyse and convert passive metadata to active metadata,
- Data Fabric must have the ability to create a knowledge graph that can operationalise the data fabric design,
- Data Fabric must enable business users to enrich the data models with semantics,
- Extreme levels of distribution, scale and diversity of data assets add complexity to Data

[141] *https://www.gartner.com/en/newsroom/press-releases/2019-02-18-gartner-identifies-top-10-data-and-analytics-technolo, accessed 12/2020*

[142] *https://www.gartner.com/en/documents/3978267/data-fabrics-add-augmented-intelligence-to-modernize-you, accessed 12/2020*

[143] *Gartner Research: Data Fabrics Add Augmented Intelligence to Modernize Your Data Integration, 17 December 2019, Ehtisham Zaidi, Eric Thoo, Guido De Simoni, Mark Beyer*

[144] *Gartner Research: Data Fabrics Add Augmented Intelligence to Modernize Your Data Integration, 17 December 2019, Ehtisham Zaidi, Eric Thoo, Guido De Simoni, Mark Beyer*

Integration Design and Delivery,

- A strong Data Integration Backbone is necessary for versatile Data Sharing in support a Data Fabric Design,
- Core Data Fabric functionalities now appear in many separate data management tools; Distinction among them is blurring,
- Delivering the Data Fabric with a combination of tools and capabilities.

Towards a Semantic Data Fabric

A Semantic Data Fabric[145] *(see*

Figure 28: The Semantic Data Fabric, Gartner Research 2019) combines the respective advantages of Data Lakes and Data Warehouses and complements them especially with the advanced linking methods that Semantic Graph Technologies bring with them (Blumauer et al, 2020).

[145] *https://www.poolparty.biz/what-is-a-semantic-data-fabric*

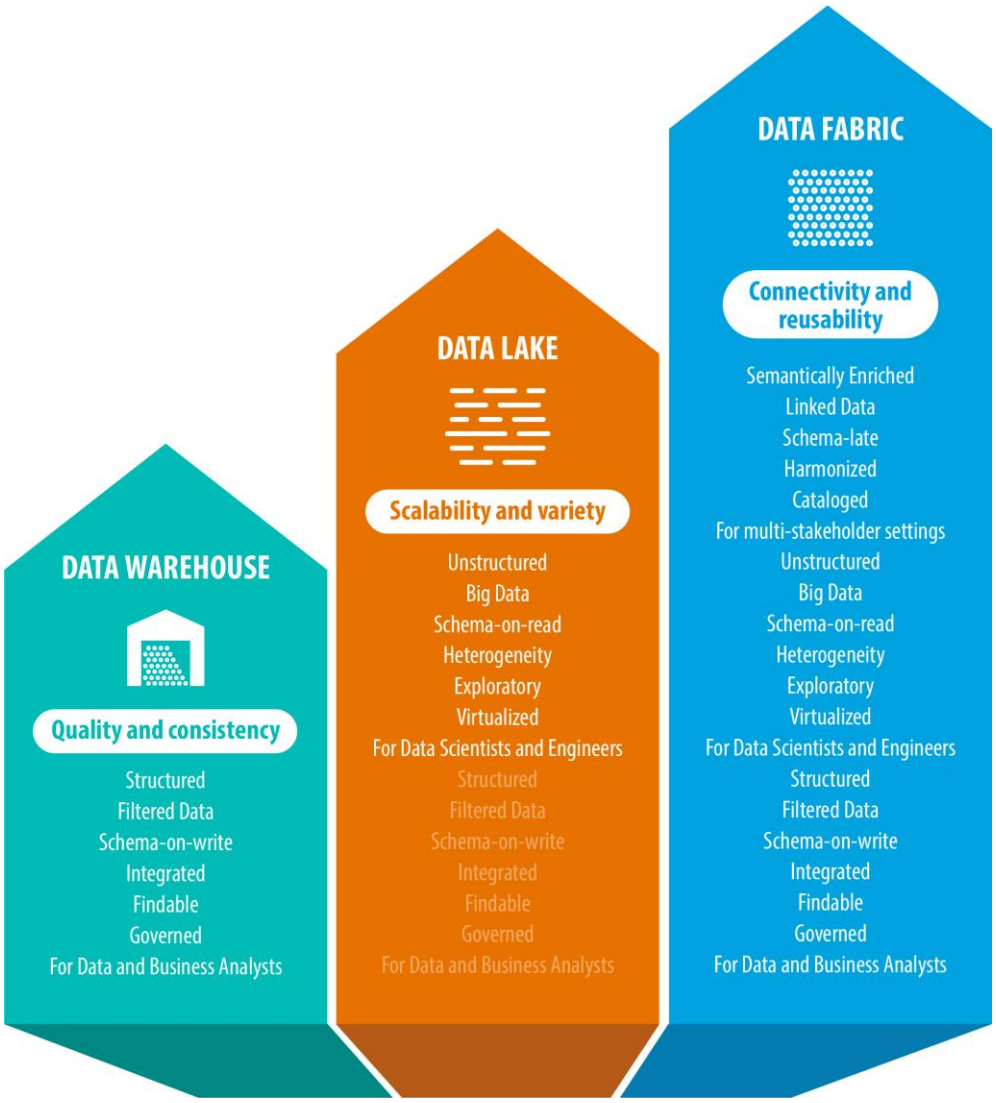

Figure 28: The Semantic Data Fabric, Gartner Research 2019

Data.world[146], a US based data catalogue vendor and Semantic Web Company[147], an Austrian semantic middleware vendor (of PoolParty Semantic Suite) has adapted and expanded this approach to an integrated view as depicted in Figure 29.

[146] *https://data.world/*
[147] *https://www.semantic-web.com*

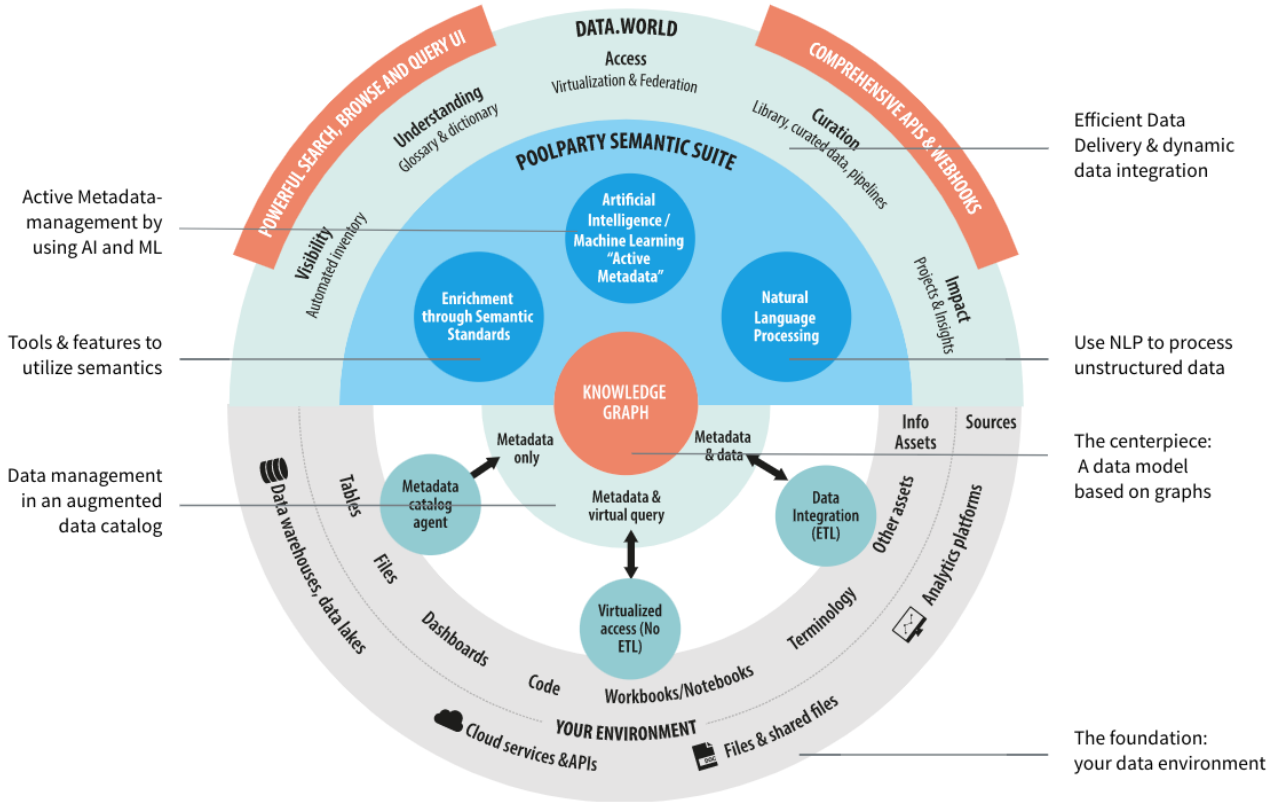

Figure 29: The Semantic Data Fabric, White Paper, A New Solution to Data Silos. Source: data.world & PoolParty Semantic Suite.[148]

As this figure above shows the main cornerstones of the Semantic Data Fabric approach are:

- Active Metadata: by making use of machine learning and AI to enable dynamic metadata over time and not only once when importing a metadata set manually.
- Tools and features to use semantics: to provide context for data sets and data objects
- Use NLP to process unstructured data: the main part of data in organisations is still unstructured data like documents. Thereby powerful analysis mechanisms are required to get the full benefit of such (unstructured) data.
- A data model based on graphs: graph technologies can bring more benefits to data management than traditional data models, as of the organisation of data and data objects in a graph-based structure, that allows complex querying as well as easy and fast querying.

To make use of the Semantic Data Fabric ideas, and principles for data markets and data spaces should be the next step in the evolution in this area. Many parts and components are used in different data markets already today, but the integrated combination could - from a technology perspective - be the breakthrough for data markets and data spaces.

[148] White Paper, A New Solution to Data Silos (data.world & PoolParty), Andreas Blumauer, Bryon Jacob, https://data.world/resources/reports-and-tools/new-solution-data-silos/

Beside the trend towards a Semantic Data Fabric, the following technological trends are identified based on the evaluation of the sources mentioned in this Section and above:

- **Smart Contracting**: a technical component that allows sellers and buyers to easily configure a contract for their data exchange/business, whereby the framework of the Smart Contracting is provided by the data marketplace to its customers and these customers (seller and buyer) decide on the terms and conditions for their dedicated agreement. From a technical point of view such Smart Contracting is often based on blockchain technology, as such technology allows to create a contract (also including data samples or attributes) that cannot be changed afterwards and thereby provides a secure and trustful mechanism for legal agreements.
- **Containerization / Virtualisation Technologies**: to provide both, an easy to use and efficient deployment method for components and micro-services as well as to enable scalability of such services, there is a clear trend in regards of containerization/virtualisation technologies like Docker, as well as several orchestration technologies.
- **Federated Architecture Approach**: although still lots of data markets are being developed and deployed following a centralised architecture approach, a clear trend occurs towards a federated architecture for data marketplaces, that allows to connect various data sources and/or data marketplaces between each other and run marketplace core services (e.g., search et al) over the nodes of such federated approach.

Recommendation for TRUSTS

TRUSTS technical team should have a deeper look at the Semantic Data Fabric idea in WP2 (architecture) and WP3 (technical implementation), including the previous approaches of data lakes and data warehouses and double check all relevant attributes and features regarding their relevance to the TRUSTS technical design and implementation. For the Semantic Data Fabric, the concept of active metadata as well as ML-augmented data ingestion is key and could provide a clear USP and value for the TRUSTS marketplace.

Furthermore, the areas of Smart Contracting (WP3) and the federated architecture approach should be taken into account and evaluated in detail, whereby TRUSTS already follows a federated architecture design as well as smart contracting (WP2 & 3). Deployment methods based on virtualisation (like for instance Docker) should be considered to provide a stable and future proven deployment and operation environment, and finally TRUSTS should be built on top of existing standards and/or standards under current development like DCAT-AP or IDS/GAIA-X.

4.5 Legal aspects of data marketplaces

This Chapter provides an overview of the most important legal aspects of data marketplaces and data spaces in Europe.

In the TRUSTS project, a dedicated deliverable (D6.2) has been developed by the partner KU Leuven on the Legal and Ethical Requirements (authors: Ducuing Charlotte, Dutkiewicz Lidia, Miadzvetskaya

Yuliya), submitted in October 2020 as a public report. To avoid duplication, this Chapter provides only a summary, as well as a rough overview of the most important areas of the legal and ethical requirements for the data marketplaces.

The deliverable 6.2 titled 'Legal and ethical framework' identifies the relevant EU legal frameworks applicable to various data transactions that are envisaged in TRUSTS. More specifically, it provides insight into the privacy and data protection legal framework supporting the data sharing compliance with the EU rules. It determines the issue of "controllership" over personal data, the ensuing allocation of data protection responsibilities and the legal basis for processing personal data. It informs project partners on the main concepts of the ePrivacy legal frameworks and their relationship with the GDPR. Furthermore, the deliverable is a continuation of the work done in WP9 of TRUSTS with regard to pseudonymization and anonymisation of personal data. It provides further conceptual legal information on privacy preserving techniques that might be relevant for TRUSTS partners and the public.

The deliverable provides an overview of the legal frameworks and ethical principles that may be applicable to data marketplace ecosystems such as TRUSTS. By doing so, the aim is mainly to provide guidance for partners in the research project to elaborate on the business and technical aspects of TRUSTS, but also to provide useful insights for the public and interested parties.

The areas identified as being most important for data marketplaces (and thereby for TRUSTS) are explained in detail in the mentioned deliverable, and are as follows:

1. **Privacy and data protection**
    a. Controllership in a data marketplace context
    b. Legal basis for processing personal data
    c. The E-Privacy Directive and forthcoming Regulation
    d. Anonymisation and pseudonymization as privacy preserving techniques
2. **Regulation of data as an (economic) asset**
    a. The Free Flow of Non-Personal Data Regulation
    b. Data sovereignty
    c. Towards a Data Law, Patterns for future data regulation
3. **Law applicable to online platforms and intermediaries**
    a. Introduction – data sharing platform as (an) intermediary(ies)
    b. The Platform to Business Regulation ('P2B Regulation')
    c. Intermediary liability for data sharing platforms
4. **Economic law applicable to data transactions**
    a. Regulation of B2B unfair commercial practices in data-driven ecosystems (with a focus on Germany, France and Belgium plus a conclusion for EU27)
    b. B2B data sharing principles and contractual terms
5. **Competition law and access to data**
    a. Introduction - the role of data for competition law analysis
    b. Article 101 TFEU
    c. Article 102 TFEU
    d. Connections to data protection law
6. **Financial law applicable to data transactions**
    a. The Anti-Money laundering (AML) Directive

   b. The Second Payment Services Directive (PSD2)
7. **Blockchain and law**
   a. Why are blockchain technology and smart contracts considered for data markets?
   b. EU policies towards blockchain legal and regulatory framework
   c. Overview of legal issues related to blockchain technology
8. **Ethical challenges in data sharing**
   a. Ethics requirements for Trustworthy AI
   b. Data-driven discrimination and data bias

The deliverable 6.2 'Legal and Ethical Requirements' can be downloaded via the following link as a PDF file:

https://www.trusts-data.eu/wp-content/uploads/2020/10/D6.2-Legal-and-Ethical-Requirements.pdf.

Recommendation for TRUSTS

TRUSTS processes should ensure compatibility with the Legal and Ethical Recommendations presented above (WP6). This is already considered in TRUSTS WP2 (D2.2), targeting the non-functional requirements.

4.6 Environmental aspects of data marketplaces

Considering environmental aspects is crucial when discussing data marketplaces especially with regards to their potential future increase of relevance. Research indicates that data-driven technologies might have a significant impact on the environment. It is important for data marketplaces to have an eye on such issues right from the beginning on, reducing the risk of having to come up with green strategies at a stage where change is hard or when media coverage is negatively affecting the market.

Insights from the TRUSTS World Café

The environmental impact of a data market such as TRUSTS was discussed in opposing directions in the World Café. Data markets impose challenges on society when they generate an additional burden for the CO2 footprint of the economy. For instance, the infrastructure required to sustain a large-scale data market consumes significant amounts of energy, resulting in the emission of an equivalent amount of CO2. Furthermore, the blockchain, used to digitise legally binding contracts within TRUSTS and often data markets in general, is known to consume significant amounts of energy. The Bitcoin blockchain is the most prominent example of a publicly used blockchain known to consume significant amounts of energy, surpassing the energy consumption even of countries[149]. Another aspect are the data assets shared on data markets themselves. Training machine-learning models requires significant amounts of energy. Incentivizing society to create increasing amounts of machine-learning models might thus lead to an increased emission of CO2 in that area. However, data markets have the potential to alleviate this problem, because they allow the sharing of trained models. Models are trained once, and potentially

[149]Cambridge Bitcoin Electricity Consumption Index: https://cbeci.org/cbeci/comparisons, last accessed April 08, 2021.

reused by multiple consumers. Exactly this is an aspect where data markets can help to reduce unnecessary emission of CO2. They support the reuse of data sets or trained models, and thus reduce redundancy and duplications. Another benefit of data markets with regards to environmental aspects is that they harbour a variety of complementary and yet different data assets. This supports the creation of new research initiatives or business models aiming to tackle the problem of CO2 emission.

Data marketplaces serving both datasets and services, or applications must, for example, keep the CO2 footprint of training machine-learning models in mind. Table 14 investigates the effort that was required to train a range of well-known ML models and compares them in terms of energy consumed, released CO2, as well as the training costs.

Table 14: CO2 estimates of well-known machine learning models (replicated based on Strubell et al.[150] , Table 3).

Model	Hardware	Power (W)	Hours	kWh PUE	CO2e	Cloud compute cost
Transformerbase	P100x8	1415.78	12	27	26	$41–$140
Transformerbig	P100x8	1515.43	84	201	192	$289–$981
ELMo	P100x3	517.66	336	275	262	$433–$1472
BERTbase	V100x64	12,041.51	79	1507	1438	$3751–$12,571
BERTbase	TPUv2x16	-	96	-	-	$2074–$6912
NAS	P100x8	1515.43	274,120	656,347	626,155	$942,973–$3,201,722
NAS	TPUv2x1	-	32,623	-	-	$44,055–$146,848
GPT-2	TPUv3x32	-	168	-	-	$12,902–$43,008

[150] Emma Strubell, Ananya Ganesh, Andrew McCallum (2019). Energy and Policy Considerations for Deep Learning in NLP. In the 57th Annual Meeting of the Association for Computational Linguistics (ACL). Florence, Italy. July 2019. arXiv:1906.02243

In addition to the CO2 footprint of ML models, data markets also need to consider the cost and energy consumption of running and operating data centres where infrastructure and tradable assets are hosted. Data marketplace operators need to be aware of techniques to reduce energy consumption and release of CO2. On an algorithmic level, this includes a range of techniques such as "Efficient allocation of VMs in servers", "Energy efficient dynamic resource management", or "Dynamic energy-aware scheduling for parallel task-based application".

Data marketplaces might also be helpful to combat the negative effects of climate change. When more data about natural events, climatic changes, and industrial processes is readily available on data marketplaces, this might spark innovation to tackle environmental problems. For example, such data could help to predict dangerous climatic activities or to handle pests.

Recommendation for TRUSTS

Eco-awareness is likely to increase in the upcoming years and probably decades. Consequently, data marketplaces will also be under the scrutiny of their ecological perspectives. For future operation and initiatives implemented on the TRUSTS platform this aspect might become crucial. For instance, data stored, or models trained in data centres of countries with cold climate produce a smaller carbon footprint as cooling of processing facilities is reduced. Clever routing techniques, e.g., routing training jobs to colder countries to reduce the carbon footprint, could prove an innovative data marketplaces of the future (even though beyond the scope of TRUSTS).

5 Micro Analysis – Mapping the Competitive Environment of Data Marketplaces

With the growing value expected from data marketplaces, it is useful to understand the industry's competitive landscape and to position TRUSTS in the landscape. This Section of the report aims to provide a competitive analysis of data marketplaces in order to support the development of a sustainable project outcome. Porter Five Forces as a broad analysis framework for competitive analysis and strategic considerations that shape the competitive landscape is used to examine the data market industry's current landscape and thus help in the positioning of TRUSTS.

5.1 Methodology for Competitive Analysis

The framework aims to guide a current and plausible trend that prospective firms can encounter as they consider participation in an industry. By understanding these five forces, it could position TRUSTS and thus provide key aspects in the positioning of TRUSTS. The forces are described in Figure 30 and discussed subsequently.

1. Threat of new entry,
2. Buyer power,
3. Threat of substitutes,
4. Supplier power, and
5. Competitive rivalry.

Porter's five forces have been used across other studies analysing the competitive landscape of industries, for example, in the adoption and outsourcing of IT services[151], mining industry[152], the educational sector[153], and the banking industry[154]. While Porter's five forces have been used to analyse competition across sectors, its application when analysing data marketplaces is limited. Most data marketplaces, especially multi-lateral data exchanges, remain conceptual or still in their relatively early stages[155],[156]. As earlier mentioned, data marketplaces are, similar to other phenomena such as transaction platforms or two-sided markets; however, data marketplaces are distinct. They solely trade on data, which is a highly liquid resource. The next Section provides an overview of Porter's five forces.

Figure 30: The Five forces that shape industry competition[157] (p. 27)

[151] FUNG, H. P. 2013. Using porter five forces and technology acceptance model to predict cloud computing adoption among IT outsourcing service providers. *Internet Technologies and Applications Research ITAR,* 1**,** 18-24.
[152] ALRAWASHDEH, R. 2013. The Competitiveness of Jordan Phosphate Mines Company (JPMC) Using Porter Five Forces Analysis. *International Journal of Economics and Finance,* 5**,** 191-200.
[153] PRINGLE, J. & HUISMAN, J. 2011. Understanding Universities in Ontario, Canada: An Industry Analysis Using Porter's Five Forces Framework. *Canadian Journal of Higher Education,* 41**,** 36-58.
[154] SIAW, I. & YU, A. 2004. An analysis of the impact of the internet on competition in the banking industry, using Porter's five forces model. *International Journal of Management,* 21**,** 514.
[155] Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2017). ETLA Working Papers The (Unfulfilled) Potential of Data Marketplaces. Retrieved from http://pub.etla.fi/ETLA-Working-Papers-53.pdf
[156] SPIEKERMANN, M. 2019. Data marketplaces: Trends and monetisation of data goods. Intereconomics, 54, 208-216.
[157] PORTER, M. E. 2008. The five competitive forces that shape strategy. Harvard business review, 86, 25-40.

5.2 Summary of Porter Five Forces

The threat of New Entry refers to the ease of entry competitors can enter and disrupt an industry's existing competitive landscape. Multiple factors, e.g., capital requirements, governmental policies, switching cost, supply-side economies of scale, and demand-side benefits of scale, shape the threat of entry. Porter, for example, argues that an industry with low capital requirements is likely to be associated with low entry barriers. A high threat of entry negatively affects firms in an industry (e.g., through low-profit margins) as firms resort to costly tactics that reduce profit margins. However, as Porter points out, the existence of a high threat of entry does not necessarily translate to actual entry into the industry. However, its presence makes competition in the sector fragile and intense, impacting long-term profitability.

Buyer Power: According to [158], buyers are said to be powerful if they have and can exercise considerable leverage relative to industry participants. Buyers' power might be gained from the volume of their purchases or cost structures in an industry (i.e., fixed, or variable cost). For example, an industry associated with high fixed implies that sellers are likely to be in a relatively weaker bargaining position with buyers as there is a constant need to ensure products are sold to cover fixed cost incurred. Powerful buyers impact competition in multiple e.g., through demand of prices or quality of services favourable to their terms.

The threat of substitutes: The competitive landscape is affected by the existence of substitutes. The presence of substitutes means that consumers have alternatives that they can choose from. For example, social media platforms such as Facebook, Instagram, Twitter serve as substitutes for users since they perform slightly similar roles. According to Porter, the availability of more substitutes means that firms in the industry will seek to differentiate themselves from rivals. Such distinction might involve rebranding and marketing research. A resultant outcome being fierce competition in the industry and lower profit margins.

Supplier Power: According to Porter, suppliers have a significant impact on an industry's competitive landscape. Suppliers with much power can influence prices and the extraction of rent from participating in the industry. In general, industries that are characterised by large market share are concentrated among a few suppliers; the suppliers are likely to have an enormous impact in setting prices and arranging for more suitable outcomes for their interest.

Competitive Rivalry: According to Porter, rivalry among existing competitors limits the profitability margin in an industry. Rivalry can be in the form of price wars or product differentiation.

[158] PORTER, M. E. 2008. The five competitive forces that shape strategy. Harvard business review, 86, 25-40.

5.3 Critique/limitations of applicability

While Porter's five forces provide relevant insights into industry dynamics, its use in performing analysis related to data marketplaces needs to be taken with care. Porter's five forces, as earlier mentioned, emerged out of the need to consider strategy and competitiveness at the level of the industry. Thus, it implicitly assumes that firms can be categorised and placed in distinct sectors. One of the critical limitations of using Porter's five forces as a framework for analysing data marketplaces is the relatively fluid nature of data marketplaces that crosses different industries.

As earlier mentioned by [159] points out that data has other characteristics compared to tangible products. For example, [160] argues that people are generally less willing to pay for data due to difficulties predetermining its value. Furthermore, data is a non-rival good, meaning its usage does not exclude other actors from use — the value of data increases in combination with a web of other resources. Another vital critique when considering Porter's five forces for analysing data marketplaces is that data market federations remain mostly conceptual.

5.4 Application of Porter Forces on Data marketplaces

Figure 31, summarises an overview of the competitive forces in relation to data marketplaces. overall, the idea is to provide an understanding of these forces which could help in the positioning of TRUSTS.

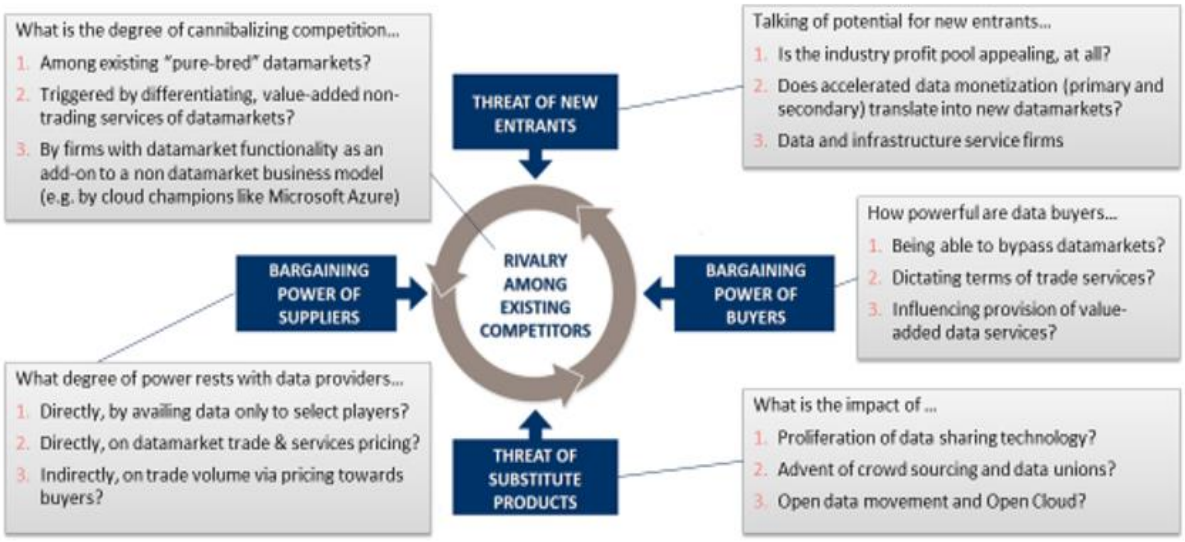

Figure 31: Positioning of Data Marketplaces based on Porter five Forces

The threat of New Entry: According to the EU, data is a key resource that many organisations depend on and rely on data to grow with the vast amount of data generated from different technologies. This development suggests that the threat of entry into the data marketplace industry is high. However,

[159] SPIEKERMANN, M. 2019. Data marketplaces: Trends and monetisation of data goods. Intereconomics, 54, 208-216.
[160] SPIEKERMANN, M. 2019. Data marketplaces: Trends and monetisation of data goods. Intereconomics, 54, 208-216.

according to Spiekermann [161], a data marketplace is relatively hard to establish as many cases have failed. Presumably for reasons such as lack of data demand, trust, security, etc., it thus appears that while data is continuously being generated, there seems to be a relatively low threat of entry into the industry. Furthermore, considering that such marketplaces are generally characterised by network effects (demand-side and supplier side), first-mover advantages pose entry barriers. How as seen in Figure 31, above, the threat of new entry in the data market industry would be highly dependent on aspects such as the profitability of the industry, the speed at which data could effectively be monetised, and the involvement of data infrastructure firms that could provide added services for the trading and bundling of data services.

Buyer power: In data marketplaces, buyers are users or other companies that buy data from the market. Porter argues that buyers with significant power can leverage their positions to bargain for prices favourable to their terms. While such can be true, it is essential to know that data buyers' bargaining power is equally impacted by the number of data suppliers, willingness to participate, and the perceived expectation of value generated from using data traded. However, with the evolving landscape of regulations surrounding data use across different organisations,[162] suggest that it might be increasingly difficult to ascertain data value. [163] equally points out that buyers are often unwilling to pay for data because of the uncertainty of a priori determining the value of data and the associated costs involved in the data processing. Accordingly, the data buyer's data marketplace can have bargaining power — also, the data marketplaces exhibit characteristics of two or multi sided markets[164]. Same-sided network effects and cross-side networks characterise these forms of markets. Same-side network effects are the value that users gain from the platform with the same type of users' participation. In contrast, cross-sided network effects are the value users of the opposite gain from users' participation of the other type. For example, data buyers' power can be low if few data providers participate in the data marketplace. This implies that data providers' power will be high if they perceive many data suppliers to participate in the market since more data suppliers mean more alternatives for data buyers in the marketplace to choose from. This dynamic can be difficult for the data marketplace since suppliers might be reluctant to participate in the marketplaces if they are unsure of many buyers' participation. As seen in Figure 31, the ability of data markets to exercise power in relation to data buyers would depend on the existence of alternative channels through which data buyers can resort to expose and sell their data.

The Threat of Substitutes: The threat of substitutes can be high. Data marketplaces are competing with alternative channels such as open data portals. The presence of such alternatives will mean that data marketplaces need to provide additional services that distinguish them from alternative channels. While close substitutes might not necessarily perform all the functions of data marketplaces, the convergence of industries and the relative affordability of cloud computing services pose significant threats to data marketplaces' viability. Also, with the growth of IoT, sensors, and other technologies that facilitate data processing and capture, data marketplaces might, in the long term, compete with

[161] SPIEKERMANN, M. 2019. Data marketplaces: Trends and monetisation of data goods. Intereconomics, 54, 208-216.
[162] Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2017). ETLA Working Papers The (Unfulfilled) Potential of Data Marketplaces. Retrieved from http://pub.etla.fi/ETLA-Working-Papers-53.pdf
[163] SPIEKERMANN, M. 2019. Data marketplaces: Trends and monetisation of data goods. Intereconomics, 54, 208-216.
[164] Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2017). ETLA Working Papers The (Unfulfilled) Potential of Data Marketplaces. Retrieved from http://pub.etla.fi/ETLA-Working-Papers-53.pdf

individual and more specialised data markets. A resulting outcome might be the need for marketplaces to trade data and provide a web of services such as building competencies, knowledge sharing activities, and visualisation capabilities that can serve as entry barriers for competitors.

Supplier Power: Major providers of data to a marketplace can have high bargaining power. Data Suppliers are likely to have high power for at least two reasons. Data marketplaces need to show relevance as a viable alternative to existing approaches, such as internal and bilateral data exchanges between organisations. [165] suggest that while many organisations are typically comfortable with bilateral exchanges of data, data suppliers tend to be hesitant to trade data in multi-lateral data marketplaces. [166] equally points to that data providers are generally sceptical about providing data to other organisations due to the lack of trust and fear of disclosing critical insights of their activities to competitive rivals. The power of data can further curtail due to increasing requirements such as GDPR and legal requirements. Data providers are expected to use data generated within the EU (European data strategy document). EU cloud-based services make up a relatively small share of the cloud-based market data (strategy document). This makes the EU data market vulnerable to external influence. Second, because data supplied to the marketplace is a by-product rather than data suppliers' sole aim, data suppliers are probably in a stronger bargaining position to negotiate plausible arrangements or expect much more favourable incentives before providing their data in the marketplace. Suppliers' high bargaining power could also be reinforced if the revenues or profits generated from the marketplaces makes up a relatively small portion of their earnings. A resultant outcome being that a data marketplace might emerge with significant erosion of its power to suppliers. Second, at the early stage of the data marketplace, suppliers can have high power since the switching cost at that early stage could be significantly lower due to the absence of path dependencies that have been created to reduce the motivation for switching.

Competitive Rivalry: According to Porter, competitive rivalry can be in price wars or competition in terms of various services/products. In terms of the data market industry, competitive rivalry can equally manifest in the rivalry between US Vs. EU. For example, control over data and its storage will increasingly be important since control over data could provide a vital competitive advantage for companies. However, as the EU pushes forward with a strategy that seeks to address data protection and control of data within the EU, data marketplaces are likely to be in a good position to compete.

Considerations and hypothesis:

1. Buyer's power
    a. The ability of data markets to exercise power in relation to data buyers depends on the existence of alternative channels from which buyers can expose their data
    b. The ability data markets to dictate the terms of trade on the data marketplace depends on the existence of alternative channels available for data buyers
2. Threats of substitute products

[165] Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2017). ETLA Working Papers The (Unfulfilled) Potential of Data Marketplaces. Retrieved from http://pub.etla.fi/ETLA-Working-Papers-53.pdf

[166] SPIEKERMANN, M. 2019. Data marketplaces: Trends and monetisation of data goods. *Intereconomics,* 54**,** 208-216.

   a. The vulnerability of data markets to the threats of new substitute products depends on the speed and scale of data related technologies and expansion of alternatives such as open data platforms.

3. Bargaining power of suppliers
   a. The bargaining power of data market in relation to data suppliers depends on the volume and scale of the user-base of data buyers rely on the data marketplace

5.5 Discussion of Data market Federators

While it has been recognised that data will increasingly constitute a vital part of the economy, data marketplaces are negatively affected by fragmentation within the EU resulting from different standards as already mentioned in chapter 3. According to the EU data strategy document, there is equally a lack of available data specifically suited for innovative use. Furthermore, imbalances exist between different actors in the data market. For example, fears of significant data providers or suppliers gaining a high bargaining power. Attributing much power to the major data providers and suppliers can make the competitive landscape not favourable to small and medium-sized industries since they lack initial resources. The imbalance issue is propagated because network effects and lock-ins characterise such markets formed around platforms. This development calls for the need to rethink the current landscape of the industry.

It is suggested that data market federators can address some of these issues. A marketplace federation could provide a level playing for different actors in the industry, making data assets available across a multitude of aggregators and marketplaces available to users. Another key role of a marketplace federation could be to bridge and encourage collaboration across different actors in domains. In that way, TRUSTS becomes a relevant actor with a distinct value proposition for different participants and industries.

5.6 Insights from the TRUSTS World Café

In March 2021, the TRUSTS consortium organised a two-hour virtual World Cafe with ~30 invited experts to discuss and collect additional input for this deliverable.

The statements and discussions with the 4 groups of participants took place as an open discussion which can be clustered in 3 business-related topics:

- How to kick-start a new data marketplace?
- What are prevalent barriers to market entry of a new data marketplace?
- Which data market value proposition are deemed worth exploration?

We have summarised the key messages as follows:

Insights from the TRUSTS World Café

Vis-à-vis competitive landscape dynamics and success factors of a new data marketplace, World Café participants viewed the following approaches as most promising *starting points to kick-start a new data marketplace* with an enveloping ecosystem:

1. Focus on a niche domain in a specific data / application space, then expand it.
2. Explore high-growth domains to carve out and occupy a niche within a wider market, e.g., machine-generated / sensor data.
3. Incorporate public data, e.g., statistical data, because this type of data is less sensitive. In turn, avoid proprietary trading personal data due to potential regulatory implications.
4. Bootstrap from peer-to-peer networks so the value can extend to others.

Barriers to successful market entry of a new data marketplace, modulating insights from the user sentiment findings of chapter *4.2 Economical Context of Data Markets* were seen as follows:

1. Data marketplaces run danger of unclear value propositions and opaque end-user needs. Other than data aggregators (or publishers) with proven business models, data marketplaces facilitate data asset trading between data asset providers and data asset buyers and must continuously test and tailor their business model accordingly.
2. Data onboarding is a key impediment, as data is often not well organised and structured to the requirements of a data market and connected data buyers.
3. Lack of trusts between data providers towards data buyers and their onward exploitation of data assets. It is noteworthy that there is not a lack of trusts towards the actual data market.
4. 'Fear of missing out' of data asset providers, which precedes the oft highlighted conceptual, real problems of data asset pricing. Even if a specific data asset exploitation would never be pursued by a data provider itself and the price of would be based on Cost+, value capture from data usage by a 3rd party is often perceived as not equitable.
5. Fear of risk exposure and principal-agent behavioural complications drive reluctance of potential data asset providers. Particularly business-side staff approached for data assets shies away from the personal cost of internal coordination efforts, and due diligence and at best unclear, at worst bureaucratic approval processes. Additionally, it is not always self-evident from the outset, in what far availed data could have competitive implications.
6. Absence of specialised ecosystem services for data bundling services and aggregation creates concentrated risk exposure to data uptake, particularly in cases where the data marketplace embarks on proprietary aggregation and trading.
7. During start-up, data marketplaces are at risk to focus on the provision of sophisticated data exchange and trading infrastructure but fall short of holistically addressing supply and demand in a holistic manner. That is, the prevalent chasm between the supply-side paradigm of "Data-as-a-Product" is not bridged to the demand-side paradigm of "Data-as-a-Problem-Solution".

Next to addressing the inversion of the highlighted barriers to successful market entry, the following *value propositions* were perceived by the World Café participants as promising research hypothesis when designing the data marketplace towards a favourable competitive positioning:

1. When defining its value propositions, a new data market should consider a triple-A model: availability, accessibility (efficient, easy), analytics (e.g., aggregation).
2. In light of growing compliance concerns related to emerging data privacy regulation, a data market and its enveloping ecosystem should provide support mechanisms for anonymisation of personal data and encrypted processing of sensitive data.
3. Different UI/X and corresponding functionalities are required to address varying requirements for different user groups.
4. Research of the Data Market Austria (DMA) project indicates the emergence of a new actor linked to data sharing platforms and data marketplaces: the data brokers. Data brokers are intermediaries between data providers and data buyers. Data marketplaces can hire/build partnerships with data analysts to promote and enable the interaction of data marketplace environments.
5. If data marketplaces actively facilitate and partially automates match-making, contracting and price negotiations to avoid lengthy negotiation contracts to get datasets (because these will take months), data marketplaces could reduce coordination and search costs for data providers and buyers. In other words, efficiency and speed of acquiring required data is key.
6. Data marketplaces that act as meta-platforms (platform federator) can create distinct value by coordinating activities of specialised aggregators and data marketplaces. Ultimately, end-users would not have to subscribe to a multitude of platforms with different policies, standards, processes and interfaces. This partial disintermediation could be offset by federated platforms by the ability to more strongly focus more on creation of time-stable data contexts and domain-related Unique Selling Propositions (USPs).
7. A data market federator could identify and provide non-differentiating shared services o specialised platforms at cost-advantaged economies of scale and take this as a springboard to later on branch out into innovative value-added services (VAS).

Recommendation for TRUSTS

The marketplaces' commercial viability will require USPs that make TRUSTS relevant along a multitude of dimensions as well as in their interdependencies, when compared to alternative data trading mechanisms. Thus, TRUSTS must be developed to perform multiple roles to make its value offering attractive. g attractive. g attractive. g attractive. g attractive. g attractive. g attractive.

A prototypical industry / competitive analysis does not appear to be fully applicable to data markets such as TRUSTS. Within the data economy, competition appears to occur more on the level of the enveloping ecosystem of a data market, which can be depicted as a set of concerted

value-creating capabilities and technology layers / components, whereby the data market acts as central anchoring point. This reconfirms TRUSTS' approach to enrich the core mandate of (1) creating a stand-alone data market with the additional two mandates of (2) creating an ecosystem facilitator and (3) enabling federation of data markets. Ecosystem facilitation in this regard predominantly refers to the establishing of a rich ecosystem of data assets and interoperable data services, and data market federation refers to TRUSTS exploring the viability of become a nexus for a future federation of interlinked European data markets and data aggregators in conjunction with accessibility of semi-public cloud systems as envisioned in the European Data Strategy.

Data can be shared and re-used infinitely. But contrary to common believe, for data markets, both-sided network effects do not scale exponentially with increased supply and demand, as the match of supply with demand is infinitely harder as compared to marketplaces for physical goods. This can be likened to buying running shoes on an online marketplace vs. buying weather data on a data market. Whereas most shoe brands will ultimately do the job for most buyers, weather data not only needs to comply with a wider array of exact requirements, e.g., scope and granularity, but is exposed to a number of binary discriminators, such as right time frame or location. Weather data from Vienna cannot simply be substituted with weather data from Berlin, whereas running shoe brand A will for most use cases do as fine as those of brand B.

A derived recommendation for TRUSTS is accordingly to not rely on scaling of data sources across domains based on ease of provider enlistment, and to avoid the pitfall of all-purpose, domain-unspecific data markets. Rather, when attracting data providers / sellers, it should strive to create pockets of data domains with time-stable contexts. This may be linked to an industry focus, e.g. based on its use-cases such in financial services and telecommunications, or to a solution focus, e.g. cross-industry customer data for onward analytics. This in turn will leverage network effects and may even lead to cross-fertilisation on the demand side, as data consumption in turn generates additional supply of domain-relevant processed data and / or pertinent applications and services.

However, in domains where the cost of data source enlisting and data preparation support is low in comparison to the number of data uses and scope of monetization potential, a TRUSTS data market operator may consider investments. For example, data from open / public sources can be made accessible through the recommender system, or – where warranted by unit economics – could be harvested and enriched for (limited scope) proprietary trading of such quality-assured data. This way, a key role of a TRUSTS marketplace federation could be to bridge and encourage collaboration across different actors in domains. On a macro-economic perspective, this also contributes to making "hidden" and "siloed" data at large more easily accessible for the European data economy, hereby amplifying TRUSTS mandate for ecosystem facilitation and federation.

Most successful data markets have embarked on value chain / value web integration to create their business moat and to amplify and stabilise value creation and value capture. Data aggregators – which are not data markets as per the specific data markets definition within the TRUSTS project – use data integration, curation, and internalisation of data transformation services to this end. Data markets like DAWEX also use supply-side value chain expansion through technical and non-technical services to support potential data providers with much needed data governance and stewardship for data cleansing, integration and orchestration, which not only eases onboarding but represents a value in its own right to its clients. A TRUSTS operator will be

well advised to build and offer according to capabilities, providing additional value creation, creating customer intimacy, and overcoming a major supply-side hurdle for data exchange and data trading. Furthermore, mechanisms for co-creation of orchestrated data sets such as the Data Circle concept introduced in the Data Market Austria can provide a key differentiator, supporting business sustainability.

Lastly, a review of big tech offerings inspires further recommendations to a future TRUSTS platform operator. Existing big tech ecosystems are vast and comprehensive and leverage multi-sided network effects and economies of scale across ecosystem components, which at this stage already more than substitute for the lowering of abstract switching cost in the wake of the current push for open architectures and interoperability. Accordingly, introduction of any new functionality or component becomes a highly risk-contained investment bet with the potential for outsized returns. Case in point is Microsoft's Azure Data market, which linked to data application / services market originally failed and shut down, but now sees a reincarnation with the cautious new addition of data trading functionality to Azure. Whilst not at the breadth and sophistication of data trading functionality of a stand-alone data market, adding this simple functionality to its long established, rich technology stack and data ecosystem induces easy uptake of data trading at scale from the get-go (economies of scope) whereas new, stand-alone data markets will struggle to scale. This particularly holds true in B2B data sharing / trading, which often takes the form of point-to-point exchange between parties where contractual relationships and threshold trust levels are already established.

TRUSTS will be well advised to attract platform users and ecosystem constituents by utilising and promoting an open, componentised, standards-based architecture to optimise interoperability whilst promoting incorporation of and adherence to existing and emerging European standards. Additionally, a TRUSTS platform operator should seek seamless integration of essential services to ease deployment – particularly by SMEs and data-driven start-ups. In particular, commission and / or brokering of infrastructure services, namely computing and storage capacity, which is availed as a foundation layer of competing big tech functionalities and offerings, should be pursued.

6 Financial Industry

The finance sector is an important aspect within the TRUSTS project and respective trends will need to be taken into account since it may impact the ecosystem development and the envisaged sustainability.

Data marketplaces choose business models based on their data type and distribution model[167]. The market for big data technology in the financial and insurance domains is one of the most promising[168].

[167] *https://rubygarage.org/blog/big-data-marketplaces*

[168] *Big Data in the Finance and Insurance Sectors: https://link.springer.com/Chapter/10.1007/978-3-319-21569-3_12, authors: Kazim Hussain, Elsa Prieto*

The finance sector by nature has been an intensively data-driven industry, managing large quantities of customer data and with areas such as capital market trading having used data analytics for some time[169]. The advent of big data in financial services can bring numerous advantages to financial institutions. Benefits that come with the greatest commercial impact are highlighted as follows:

- enhanced levels of customer insight, engagement, and experience through the digitization of financial products and services and with the increasing trend of customers interacting with brands or organisations in the digital space. With the digitization of financial products and services and the increasing trend of customers interacting with brands or organisations in the digital space, there is an opportunity for financial services organisations to enhance their level of customer engagement and proactively improve the customer experience. Many argue that this is the most crucial area for financial institutes to start leveraging big data technology to stay ahead, or even just keep up with competition.
- enhanced fraud detection and prevention capabilities using big data is now possible to use larger datasets to identify trends that indicate fraud. Financial services institutions have always been vulnerable to fraud. There are individuals and criminal organisations working to defraud financial institutions and the sophistication and complexity of these schemes is evolving with time. In the past, banks analysed just a small sample of transactions in an attempt to detect fraud. This could lead to some fraudulent activities slipping through the net and other "*false positives*[170]" being highlighted. Utilisation of big data has meant these organisations are now able to use larger datasets to identify trends that indicate fraud to help minimise exposure to such risk.
- *enhanced market trading analysis*, where trading strategies which make the use of sophisticated computer algorithms to rapidly trade the financial markets. Trading the financial markets started becoming a digitised space many years ago, driven by the growing demand for the faster execution of trades. Trading strategies that make use of sophisticated algorithms to rapidly trade financial markets are a major benefactor of big data.

In terms of data strategy, financial services organisations are taking a business-driven approach to big data. Business requirements are identified in the first place and then existing internal resources and capacities are aligned to support the business opportunity.

The findings, after analysing the requirements and the technologies currently available, show that there are still research challenges to develop the technologies to their full potential in order to provide competitive and effective solutions.

In the last few years, the financial sector seems to move towards services digitization in order to meet the needs of contemporary businesses. Fintech is key to this trend since it bases its existence on services digitization while doesn't bear the burden of traditional core banking processes[171].

[169] *Big Data in the Finance and Insurance Sectors: https://link.springer.com/Chapter/10.1007/978-3-319-21569-3_12, authors: Kazim Hussain, Elsa Prieto*

[170] *a false positive is a test result which wrongly indicates that a particular condition or attribute is present. In AML processes, a false positive result in flagging a customer or a transaction as suspicious or fraudulent when it is actually legitimate.*

[171] https://howbankswork.com/core-banking-processes-and-recent-strategies/

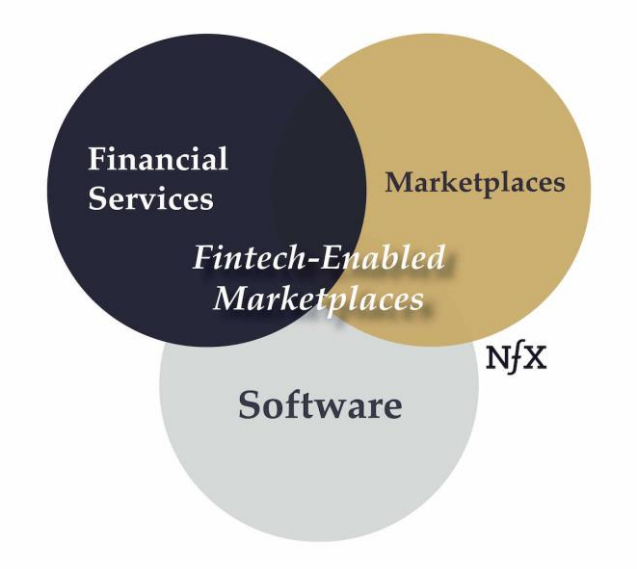

Figure 32: Fintech-Enabled Marketplaces

To expand on what it is introduced above, fintech-enabled marketplaces are marketplaces with tech-enabled financial services built directly into the platform. Recently, the marketplaces begin to offer services like:

- **Insurance:** In-house insurance products made possible by better underwriting models.
- **Financing:** Non-traditional financing options such as rent-to-own or income-sharing,
- **Banking:** Novel and customer-specific solutions to manage transactions, deposits and payments.

Marketplaces containing one or more components that fall into the three buckets described above can be taken as a sign of the way things are headed. Some examples are:

- *Opendoor* — is an online real estate company based in San Francisco that makes as-is cash offers to property sellers through an online process, improves and repairs the properties it purchases, and relists them for sale[172]. Its leverages debt capital to provide instant liquidity for home-sellers.
- *Lambda School* — is an online coding program that's free until you finish and get a job. The central conceit is an income-share agreement: students pay nothing while attending the school and then pay a portion of their earnings once they're employed[173]. It offers income-sharing for tuition financing as an alternative to student loans. It has been heralded by some as a market-based solution to student debt.

[172] *https://opendoor.io/join/*
[173] *Lambda School's For-Profit Plan to Solve Student Debt | WIRED*

- *Fair.com* — buys used vehicles and rents them out with a low-price month-to-month lease. Lease an affordable used car, from right where you are and for as long as you choose. Get a car from the comfort of any device[174].
- *Airbnb* — is an Online Vacation Rental Platform, that began offering a $1M insurance product not long after they launched, mitigating the risk of guests damaging their property for hosts
- *Apple Card* — in 2019, embedded financial services deeper into their iOS marketplace with the launch of Apple Card. A complete reinvention of the credit card where the "users" personal information lives on their iPhone, beautifully laid out and easy to understand. Apple Card eliminated fees and built tools to help the "user" pay less interest. Advanced technologies (like Face ID, Touch ID, and Apple Pay) give a new level of privacy and security and with every purchase the "user" gets Daily Cash back. Which all adds up to a healthier financial life[175].

Incorporating fintech elements like the above, might seem just like an incremental change, but the effect can be revolutionary. By adding innovative financial services, marketplace start-ups can reduce the friction involved in (especially high-value) transactions for purchasers, and can improve incentive alignment amongst all parties with the removal of financial intermediaries between sellers and buyers. Importantly, this has the potential to lead to radical and breakthrough product experiences and let start-ups find a wedge into one or both sides of the market.

Based on broader trends in the marketplace evolution, it can be seen that at each stage in the evolution of marketplaces, improved user experience has been a constant and unstoppable theme[176].

TRUSTS UC1, as part of this project and according to the GA, envisions to leverage the power of the TRUSTS Platform in view of securely sharing data between organisations, applying smart big data analytics for AML compliance purposes as well as fairly trading the resulting data to end-users such as FIs, internal / external auditors, fiduciaries, audit firms, etc. Acknowledging the significance of AI/ML and smart analytics in providing better efficiency in combating money laundering, the purpose will be to securely share closed-loop data so as to feed a next generation advanced AI/ML-based AML solution. These algorithms and models will be able to understand malicious behaviour through data analysis, work with metadata and provide deeper insights about existing and prospective customers.

TRUSTS UC2 proposes a key innovative process advancement which enables the agile marketing correlating siloed data owned by enterprises with significant privacy and security constraints to create advanced market insights and new targeted products for the benefit of both market and economy. The objective of this use case is to demonstrate the capabilities of the TRUSTS Platform as a 'Trusted Secure Data Sharing Space' for advanced marketing activities through correlating anonymised banking and telecommunications data. Enterprises use demographics and life-style data to extract meaningful information on their target customer base. However, most of them use ad-hoc methods, which provide results of high uncertainty since data are collected via different methods in different time periods. A promising approach would be to correlate up-to-date financial and telecommunication customers' data at a local level. Such information could give timely valuable insight on the actual disposal income and

[174] *Fair: The Used Car Leasing Platform*
[175] *Apple Card - Apple*
[176] https://www.nfx.com/post/fintech-enabled-marketplaces/

spending habits of the targeted region. Despite the benefits, current methods impose significant obstacles to the analysis: a) there is no established process in place that enables sharing of up-to-date data between telco and financial operators and b) trusted external environments able to provide compliance and anonymisation services are non-existent or their activities are limited. Using the TRUSTS platform, both Banks and Telecommunication enterprises are able to demonstrate and validate sharing anonymised up-to-date data in a sustainable and GDPR compliant manner in order to target marketing actions to specific local areas or even individuals.

The TRUSTS Data Marketplace vision as UC3, is also to create an out-of-the-box analytics solution for the anonymisation and visualisation of Big Financial Data, specifically to advance new ways of human-computer interaction currently in their infancy, such as chatbots that can act as automated assistants to allow customers to converse about the management of their debt at their own pace and with a personalised experience, through the integration of Big Data. The integration of cognitive computing and financial services will transform the finance industry in a variety of ways.

Advancing the state of the art of human-computer interaction with advanced techniques will lead to an improved experience when it comes to personal finance management as one of the key objectives of TRUSTS UC3.

Recommendation for TRUSTS

FINTECs is one of the digital economies' driving forces bringing in a flexible and contemporary way, new economic challenges with targeted and easily customised financial instruments. It is recommended that when defining TRUSTS business model and commercialisation actions, collaboration with appropriate financial institutes should be investigated and sought.

In addition, APIs towards payment and other financial services should be explored for implementation in the platform in order to facilitate each collaboration with players in the financial sector (WP2).

7 Telecom and IT (ICT) Industries

Telecom operators participate in the data marketplace ecosystems mainly facilitating platform operations or value-add services (e.g., sensor data exchange). Also, this domain is in the focus of TRUSTS and will be considered in the following Section.

Sensor or IoT data marketplaces provide buyers with data collected from smart devices. The main feature of sensor data marketplaces is the ability to purchase real-time data feeds and datasets from remote devices. Real-time data from sensor data marketplaces helps companies understand consumer behaviour, improve sales, and build better marketing strategies. Streamr, Dawex, and QueXopa are examples of sensor data marketplaces where buyers can purchase data collected with IoT devices.

In any case the data marketplace model is still in its infancy which business models, regulations, processes, standards and platforms are still under development. The telecom and IT sector are key to the proliferation of such platforms due to the knowledge of operating data systems and cloud services as well as due to the tendency to enter into the value-added services domain beyond the operation of network operation and connectivity services.

The following Table 15 is indicative to reveal the ICT sector initiatives in the data marketplace domain:

Table 15: ICT sector initiatives in the data marketplace domain

Data marketplace	Company	Country	Source
Advaneo	Advaneo GmbH	GER	https://www.advaneo.de/de/data-marketplace https://www.crunchbase.com/organization/advaneo-gmbh https://www.internationaldataspaces.org/use-case-01-advaneo/
DAWEX	Dawex Systems SAS	USA	https://www.dawex.com/en/ https://www.crunchbase.com/organization/dawex https://www.linkedin.com/company/dawex
Caruso	Caruso GmbH	GER	https://www.caruso-dataplace.com/ https://www.linkedin.com/company/carusodataplace https://www.tecalliance.net/en/official-green-light-for-the-caruso-data-marketplace/
DIH	Deutsche Telekom	GER	https://dih.telekom.net/en/
Streamr	Streamr Network AG	CH	https://www.streamr.com/ https://www.crunchbase.com/organization/streamr https://www.streamr.com/whitepaper https://coinmarketcap.com/currencies/streamr-datacoin/
Qlik DataMarket	Qlik Technologies	USA	https://www.qlik.com/us/products/qlik-data-market https://www.crunchbase.com/organization/qlik-technologies
xDayta	xDayta	USA	http://www.xdayta.com/ https://www.crunchbase.com/organization/xdayta
Kasabi	Kasabi	GBR	https://www.crunchbase.com/organization/kasabi https://www.slideshare.net/ldodds/kasabi-linked-data-marketplace https://gigaom.com/2012/07/09/kasabi-shuts-down-says-data-marketplace-too-slow/

InfoChimps	Infochimps Inc.	USA	https://www.crunchbase.com/organization/infochimps http://radar.oreilly.com/2012/03/data-markets-survey.html http://cloudofdata.com/2013/02/is-infochimps-running-from-the-data-market-business/
IOTA	The IOTA Foundation	GER	https://www.iota.org/ https://data.iota.org https://blog.iota.org/part-1-iota-data-marketplace-update-5f6a8ce96d05
Databroker DAO	SettleMint NV	BEL	https://databrokerdao.com/ https://www.crunchbase.com/organization/databroker-dao https://www.linkedin.com/company/databroker-dao
Microsoft Azure Data Market	Microsoft Corp.	USA	https://msdn.microsoft.com/en-us/magazine/gg309173.aspx https://adtmag.com/articles/2016/11/18/azure-datamarket-shutdown.aspx http://radar.oreilly.com/2012/03/data-markets-survey.html
Otonomo	Otonomo	ISR	https://otonomo.io/platform/ https://www.crunchbase.com/organization/otonomo https://europe.autonews.com/article/20180220/ANE/180229998/israeli-startup-takes-on-google-in-rush-to-process-car-data
Datafairplay	Data Fairplay GmbH	GER	http://www.datafairplay.com/ https://www.handelsblatt.com/technik/it-internet/cebit2014/neue-plattform-data-fairplay-geld-her-fuer-meine-daten/9565908.html

The value proposition indicates which core offer the data marketplace provides in order to instil added value for platform users. Here, a distinction is made between transaction-centred and data-centred trading platforms. The former focuses on the switching function of data goods and data services, i.e., the platform brings two parties together, either by providing the necessary infrastructure or by direct switching. The data-centric marketplace also provides tools for data analysis, visualisation and preparation within the platform infrastructure in order to gain new insights from the data goods. IOTA, Streamr, and DAWEX are examples of transaction-centric marketplaces. In contrast, the Telekom Data Intelligence Hub, Advaneo, and Caruso are exemplary data-centred trading platforms.

Insights from the TRUSTS World Café

The discussions in the TRUSTS World Café with regards to data sharing versus data trading demonstrated the need for data marketplaces related services to provide **concrete processes addressing privacy issues and concerns.** Such processes should be visible to the users both business and private reassuring for the compliance with the data privacy regulations e.g. GDPR.

The digital ecosystem is expanding while bilateral collaboration had to become global. This can only be achieved if **global data marketplace rules** apply thus enabling federation of the digital marketplaces and the facilitation of the respective ecosystem.

Large enterprises and SMEs start to consider that **data is an infrastructure** that needs to be exploited. To this end, i**nitiatives towards facilitating the data marketplace ecosystem** is perceived positively from the market.

In addition, to privacy, quality and integrity issues have to be addressed by the data marketplaces. To this end respective **processes that ensure data and transaction quality and integrity** should be established and offered as integral part of the service. This will constitute a **key differentiator vs competition**.

Current efforts to analyse data between different enterprises in an GDPR complaint manner exist but they are limited, and they enable only bilateral collaboration. **Data marketplaces are well placed to establish common rules and services thus enabling easy multiparty collaboration**. (Relevant TRUSTS use case is UC2 entitled "Agile Marketing").

While trading **data valuation methods** need to be defined. This is a complicated problem since subjective attributes have to be taken in account as well.

Various business models exist without having clearly a winning one at this stage. **TRUSTS should offer a straightforward commercial propositio**n addressing data trading/exchanging needs of enterprises and individuals. **Business sustainability is key** to create trust and willingness to use the data marketplace services.

Recommendation for TRUSTS

Analysis in the Telecom domain demonstrated that these respective businesses mainly act as data providers and operators of data sharing platforms. Their value addition in the process of service delivery and sustainability is significant since the respective operators have vast experience in the quality operation processes, market penetration in all segments and they are considered and trusted entities by the whole community. It is recommended that when commercialisation actions are defined within TRUSTS, collaboration with telecom companies should be explored regarding their support for end-to-end platform operation and additional value creation.

8 Conclusions and Next Actions

To sum up, this study has condensed the current state of the dynamic topic of data marketplaces and created a wholesome overview on both, the academic view on the current state of data marketplaces and the different facets of data marketplaces in their "natural habitats" – considering the circumstances data marketplaces are embedded in.

Within Chapter 3 on "Definition of EU and Worldwide Data Marketplaces" we derived a profound definition of data marketplaces, valid not only for the TRUSTS project but also as a valuable addition to the current academic state of the art on data marketplaces. According to Chapter 3.1.2, we defined a data marketplace as "a digital system where data is traded as an exchangeable economic good. It connects data providers and data buyers and facilitates data exchange and financial transactions". The Section also provides a framework to classify data marketplaces based on their orientation and ownership as well as examples for such a classification and their matching mechanisms. This framework will then be re-used to elaborate the TRUSTS business model (in the respective work package - WP7).

Further, this study has summarised 35 functionalities that have been found in the state-of-the-art data marketplace literature (Chapter 3.3) and has been recommended to be considered for TRUSTS. Most functionalities touch the category of "dataset discovery", "trading arrangements, "data governance", "data transformation" and "user management". This Chapter also shows the growth trend in the value of the data market and data economy in the European Union (EU) (Chapter 3.4.1). Moreover, the number of academic publications in this area is rapidly increasing, bringing us to the *take-off phase* of data marketplace research. The primary research topic of data marketplaces can be divided into six clusters. These are *pricing mechanisms, privacy themes in personal data markets, general context of data markets, technical literature, service offerings, and data markets in IoT* (Chapter 3.4.2). This study also highlights the challenges of data marketplaces identified in the literature. The challenges were categorised using the STOF model (Chapter 3.5). Another highlighted point from the discussion is the data marketplace fragmentation issues. In general, fragmentation triggers multiple aspects of data marketplaces (e.g., business models, governance arrangements, and technical standards) to diverge uncontrollably, leading to a decrease of trust in the concept of data marketplaces as a whole (Chapter 3.6). Therefore, a federated approach to overcome the fragmentation issues can be considered as a potential solution.

On the other hand, Chapter 4 "Macro Analysis – Mapping the External Environment of Data Marketplaces" provides an overview on five different areas, data marketplaces are influenced by. These five areas are closely related to the topic of data marketplaces and can therefore have a strong impact on the success of the project. Thus, we have monitored current developments in these areas and evaluated their relevance for TRUSTS to align the TRUSTS project results with current circumstances and occurrences. In this regard we have detected, that on a **political** level (Chapter 4.1), TRUSTS should be aligned with the GAIA-X project, as well as approach the BDVA i-Spaces as a strong and relevant community in this area and consider the recently published results of the Horizon 2020 project OPEN DEI "Aligning Reference Architectures, Open Platforms and Large-Scale Pilots in Digitizing European Industry".

Considering current **economic** trends in the field of data marketplaces (Chapter 4.2), one can say that data-driven economy will continue to grow over the next years which is why the potential for data marketplaces will also grow as central locations for sharing and trading data. Thus, privacy concerns and security issues still need to be addressed to increase adoption, which is exactly where TRUSTS is aiming at. Also, pricing strategies and economic incentives need to be clarified - currently, companies are reluctant to share data because the own benefit or value is unclear.

In the area of **social** aspects touching data marketplaces (Chapter 4.3) one can say that the requirements for skilled data professionals will grow over the next years and that work must be done

to explain and make clear the benefits for data providers in order to raise the frequency of and willingness for data sharing. Hence, the adherence to a **legal** framework (Chapter 4.5) is crucial which is why it has been recommended here that TRUSTS processes should ensure compatibility with the Legal and Ethical Recommendations presented in the publicly available TRUSTS deliverable[177], to ensure legal and also ethical compliance.

In the field of **technical** developments (Chapter 4.4) one of the messages is that TRUSTS' technical team could have a look at the overall concept of Semantic Data Fabrics. Many parts and components of such a Semantic Data Fabric are used in different data markets already today, but the integrated combination could, from a technology perspective, be the breakthrough for data markets and data spaces. For the Semantic Data Fabric, the concept of active metadata as well as ML-augmented data ingestion is key and could provide a clear USP and value for the TRUSTS marketplace. However, this has to be assessed by the TRUSTS Team responsible for elaborating a suitable business model (WP7). Furthermore, the areas of (I) Smart Contracting and (ii) the federated architecture approach should be taken into account and evaluated in detail as well. In addition, deployment methods based on virtualisation should be considered to provide a stable and future proven deployment and operation environment, and finally TRUSTS should be built on top of existing standards and/or standards under current development like DCAT-AP or IDS and GAIA-X.

When it comes to **environmental** aspects (Chapter 4.6) a main message is that data-driven economy has a significant impact on our environment. Technologies such as the blockchain (e.g. for smart contracting), server centre, and training ML models require significant amounts of energy. Data markets might contribute to this problem but might also help to alleviate it: a) sharing data and models helps to re-acquiring data or retrain the models and b) research initiatives and business models working against climate change might benefit from the shared data and can create innovative solutions.

When analysing data marketplaces' **competitive environment**, it becomes apparent that any data market offering needs to be understood in the context of it enveloping ecosystem of a data market, which can be depicted as a set of concerted value-creating capabilities and technology layers / components, whereby the data market acts as central anchoring point. TRUSTS will be well advised to attract platform users and ecosystem constituents by utilising and promoting an open, componentised, standards-based architecture to optimise interoperability whilst promoting incorporation of and adherence to existing and emerging European standards. Ecosystem facilitation is required to achieve multi-sided network effects in a rich ecosystem of data assets and interoperable data services, whereas data market federation interlinks European data markets and data aggregators in conjunction with accessibility of semi-public cloud systems as envisioned in the European Data Strategy. In doing so, TRUSTS can focus on select data asset domains, thereby creating the depth and time-stable context required for economically viable data trading and related services, whilst providing a portal into the wider European data landscape. A number of auxiliary services should be explored by a future TRUSTS operator, beyond the scope of the project, to create a defendable business moat vis-à-vis the competitive landscape. These encompass inter alia (1) public data harvesting and preparation, (2) support for data provider onboarding through data integration and orchestration, and meta data quality assurance, (3) enablement of co-creation of orchestrated data sets through 3rd party Data

[177] https://www.trusts-data.eu/wp-content/uploads/2020/10/D6.2-Legal-and-Ethical-Requirements.pdf (March 2021)

Circles as introduced in the Data Market Austria project. Additionally, a TRUSTS platform operator should seek internalisation seamless essential services, namely computing infrastructure commissioning / brokerage to ease deployment by SMEs and data-driven start-ups.

Our study did also analyse the current state of the **Financial** (Chapter 6), **Telecom** and **IC(T)** (Chapter 7) Industries with regard to data marketplaces. Here, we conclude that the involvement of financial enterprises in the data marketplace ecosystem can function as a catalyst for the success of this field. Financial industries are digitally transformed and bringing in a flexible and contemporary way to target new economic challenges with easily customised financial instruments. The analysis in the Telecom domain demonstrated that companies in this respective business mainly act as data providers and operators of data sharing platforms. Their value addition in the process of service delivery and sustainability is significant since the respective operators have vast experience in the quality operation processes, market penetration in all segments and they are considered and trusted entities by the whole community. We therefore formulated the recommendation that TRUSTS should define a straightforward commercial model addressing real enterprise data trading and data analysis needs. Business sustainability, transactions integrity and privacy preservation processes are key for the success of the endeavour.

During the elaboration of this study, we have already collaborated closely with the respective work packages touched by the recommendations, so that the requirements have been updated continuously. Nevertheless, we will ensure that all final recommendations have been updated and provided to the respective work packages in the upcoming month. Further, we will communicate and promote the results of the study together with T7.2 (Community Engagement) and WP8 (Communication) to external stakeholders with the aim foster the community around TRUSTS.

As our final remark, we would like to inform that the part of the study in Chapter 3.4.2 (academic trends) has been accepted for publication in the proceedings of the 34th Bled eConference (https://bledconference.org/). This paper, entitled ***"Business Data Sharing through Data Marketplaces: A Systematic Literature Review,"*** will be presented during the conference held on June 27 – June 30, 2021. This Systematic Literature Review (SLR) study provides an overview of the state of the art of data marketplace research. We studied 137 articles from the Scopus database and structured our analysis using the Service-Technology-Organization-Finance (STOF) model. We find that the extant data marketplace literature is primarily dominated by technical research. To move past the first stage of the platform's lifecycle (i.e., platform design) to the second stage (i.e., platform adoption), we call for empirical research in non-technological areas, such as *value networks* and *organizational arrangements.*