Success factors for Federated Data Spaces initiatives

ADRIAN VASILESCU,
University of Twente, The Netherlands; email:
a.vasilescu@student.utwente.nl

Abstract

Federated Data Spaces (FDS) are ecosystems where information is exchanged while ensuring data sovereignty and keeping the information as much as possible at the source. This enables unprecedented benefits in terms of interconnectivity and efficiency. Federated data spaces are considered critical for shaping the future of data economy in Europe. The implementation of FDS is pursued in an increasing number of sectors with varying degrees of progress. The field is new and few initiatives reached deployment. The combination of factors which lead to an efficient new FDS implementation is unknown, due to the lack of research focused on this aspect. Filling a knowledge gap, this study aims to identify common factors across successful FDS and determine those linked to the success of respective initiative. The comparative analysis of use cases from both successful and failed implementations of FDS ensued a list of proposed success factors. Their individual impact was assessed by assigning an arbitrary score. The described factors have a series of economic, technical and social implications that are hereby discussed. The findings need to be validated in the future with larger sets of data once more initiatives will be launched or reach implementation stage. Nonetheless, it is anticipated that this work can potentially guide new implementations of FDS.

Keywords: Federated Data Spaces, Data Spaces, Data sharing Initiatives, success factors

1. Introduction

Digitalization has created tremendous advancements in technology. Along with the technology came an abundance of raw digital information at an increasing rate. The data generated in 2023 will reach 120 zettabytes (1 zettabyte is 1 billion terabytes), an increase of 23.7% over the previous year [1]. This clearly shows that data has become an integral part of every business. Currently, businesses create data (much of it unsorted), store it locally or on the cloud where it's only used by the company and other parties that have direct agreements with that company. This methodology introduces some inefficiencies at an industry level. For example, centralized data sharing creates monopolies, increases dependence and reduces profits for the companies that use them. Moreover, the operation of large server farms requires resources that negatively impact the sustainability and costs of data storage.

The value of data increases with its use. With the evolutions of Big Data analytics and Artificial Intelligence, there is a need for large volumes of data for training and analytics purposes. Nonetheless, the access to data is limited by the existing agreements between companies or organizations or by differences in technical standards. The companies' lack of access to the aggregated information in their specific sector is a limiting factor to their innovation-driven development.

Differences in semantics cause additional inefficiencies in data sharing. The same word used to reference stored data, e.g., "ETA" has different meanings for a company towing a container ship to the dock than for a shipping company transporting the containers from the ship [22].

The issues with traditional data sharing affect all sectors and have besides economic effects, also important social ones. For example, for a hospital, access to patient's history from the family doctor or to medical tests provided in different formats is difficult. Besides the administrative and financial burden for medical institutions, healthcare insurers and patients, this delays treatment or prevents accurate diagnostics.

Recently proposed "Federated Data Spaces" (FDS) have the potential to reduce inefficiencies. FDS facilitate a new way of exchanging data which has at center data sovereignty, i.e., ensuring that the data is accessed and used according to the owner's intentions. A FDS is an ecosystem comprised of data providers, data consumers and a federator where the data is generated and stored locally by multiple providers. The information is accessed by multiple data consumers through protocols enforced by a federator [16,22]. The scope of this system is to provide multiple agents with the possibility to access data produced without having a formal agreement with the data provider. Furthermore, the data provider also has assurances that even without a formal agreement their data is used according to some rules specified in the protocol. The interoperability, sovereignty, trust and security of the entire ecosystem is ensured by clear and detailed protocols. This facilitates safe, wider and faster exchange of data, beneficial to all participants. As a result, FDS help to streamline resources, make processes and business more efficient and accelerate the access to healthcare and to various services.

The safe exchange as enabled by FDS is crucial to the developments envisioned in the European strategy for creating "a single market for data that will ensure Europe's global competitiveness and data sovereignty" [5]. The anticipated technological advances, heightened efficiency and growth have the potential to radically transform the economy and the society. In preparation, strategies and legislation were released to define the European vision for a common data economy [5,20] and to help adapt the industry to the future digital world [28]. A framework was set for the safe use of AI [25] and for implementing FAIR principles to data management [27]. Thus, the framework needed for FDS was created. From a more practical point of view, large initiatives implementing FDS were launched, e.g., IDS [8] and Gaia X [6] and first applications were recently reported [29]. These are currently setting the principles and standards needed for connectivity and data exchanges specific to FDS.

Both Gaia X and IDS put trust and data sovereignty at the center of the developed ecosystem. Their schematic architectures are indicative of the core areas that need to be considered for a functional, successful FDS. Gaia X is a federated data infrastructure that is secure and connects cloud service providers and users in Europe. It has the ambition to "drive the European data economy of tomorrow" [6]. Various ecosystems were established within Gaia X, covering sectors from Aerospace to Agriculture, Finance, Health, Mobility etc. The architecture of FDS within Gaia X [15] considers the providers and users of data and infrastructure. They are connected via a data sharing system whose main pillars are: (i) identity and trust; (ii) sovereign data exchange; (iii) a federated catalogue and (iv) compliance. The implementation of these federation services is enabled by a clear set of policy rules, a well-defined architecture of standards and efficient interconnectivity between the actors in the ecosystem.

The reference architecture of FDS established by International Data Space Association [8, 9] stands at the basis of several initiatives that reached implementation [13]. Thus, it can serve to identify areas to be addressed for successful new implementations. While it gets more detailed according to the targeted application, the basic architecture has several key components (Figure 1).

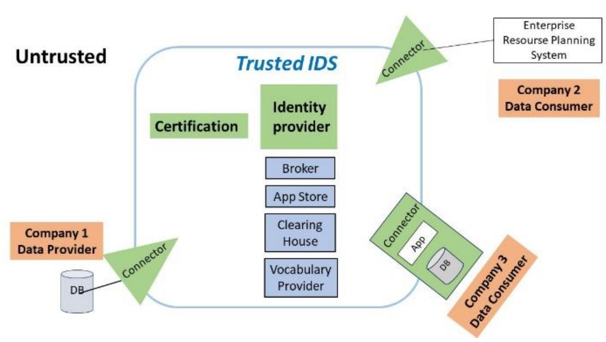

Fig.1 Schematic representation of the IDS architecture. Redrawn from [3].

- *The IDS connector* is the main component that ensures that the data is transferred between different actors (data providers and consumers) via a secure, "trusted" dataspace in a manner that ensures data sovereignty. The IDS connector is a certified software that enables from the technical point of view the access of data users and providers to the trusted dataspace.
- *The IDS broker* functions as a searchable catalogue. It keeps the self-descriptive information of an IDS connector that provides either a service or data and incorporates query and access functions. Thus, it enables to any participant in the dataspace to find it and access a specific IDS connector.
- *Certification:* both the participating actors and the main technical components of the dataspace are certified to ensure a trusted environment and the security of data exchange
- *Identity management*, for managing the identity information of participants to the dataspace and the components of the trusted dataspace
- *Vocabulary provider* delivers elements such as formal names and definitions, reference sets of data or metadata that are used for defining and describing the datasets.
- *Clearing house* keeps track and intermediates the settling of all transactions (financial and data exchanges) within the dataspace in a decentralized manner. A third, neutral party can be involved to build trust in the system.
- *App store* registers the different independent software applications that are used with an IDS connector

Most efforts required to implement FDS were diverted towards the definition of system's structure and protocols. This is reflected by the amount of literature dedicated to the above aspects rather than to the entire concept. There is a variety of FDS initiatives in different stages of development, from simple proposed framework to fully-fledged implementation where the framework functions and information can be written and read [13]. Most of the FDS initiatives are created on frameworks designed by a couple of organizations. Thus, there are few "framework designers" and few successfully implemented projects. The majority of the programs are currently in the testing, pilot phase, e.g., 48.8% versus only 11.2% "live" initiatives as summarized in [13]. A few implementations have outright failed to accomplish their objectives and stopped being pursued [12]. Due to the lack of freely available data, it is unknown if the projects in the pilot phase have encountered difficulties or if their testing was prolonged beyond 1 year. This paper focuses on identifying, presenting and assessing the impact of the factors that may lead to a successful implementation of FDS. Considering the novelty of FDS the paper will also fill a knowledge gap by identifying these "success factors".

2. Problem Statement

The increase of attention towards FDS is fairly recent. While there have been a couple of initiatives by multiple companies and organizations to create a common data space there are few successful implementations. Current studies place a strong emphasis on the techniques and protocols required by FDS. Yet, little [23] to no research focuses upon what can be considered as "best practices" for this field. There is a lack of knowledge related to the success factors of FDS initiatives and their specific impact. This gap in knowledge can be bridged by answering the following **Research Question:**

Which factors lead to a successful implementation of FDS initiatives?

To simply structure the research needed to answer the research question, five sub questions have been derived.

Sub question 1: What is the current state of the art regarding FDS?
Sub question 2: What is a successful FDS initiative?
Sub question 3: What are the current successful and failed FDS initiatives?
Sub question 4: What are the common characteristics found in successful initiatives and to what degree they affect the success of the respective initiatives ?
Sub question 5: What implications have the identified success factors for research and industry?

3. Methodology

Multiple methodologies were used to obtain the answer the research question and to present the findings in this paper.

Literature
Search queries were performed on Google Scholar, Scopus, Web of Science and other databases containing academic literature. The queries included the keywords "Federated Data Spaces", "Data Spaces" and "Data sharing Initiatives". To filter the resulted hits, additional helping words were considered such as "Success Factors", "Failed" and "Delayed". Furthermore, the resulting articles were retrieved and studied. Finally, queries were executed on the internet for reliable websites that contain information on the success factors of FDS. This preliminary literature study produced two results: (1) a set of 103 FDS initiatives was defined ("FDS" dataset) and (2) a definition of "successful FDS" was advanced. Based on this definition, the original dataset was divided in distinct subsets of "Successful" and "Failed" FDS implementations, respectively. A more in-depth study was conducted to detect factors that contribute to the successful implementation of FDS. There was a focus on the "Successful" dataset to identify their common elements. Next, the factors that were also prevalent in the "Failed" dataset were excluded. Finally, the information from the "Successful FDS" dataset was reviewed to determine the impact of each factor on the implementation of the FDS. The impact was assigned a score between 1 and 5, with "1" corresponding to low impact and "5" corresponding to high impact. The scoring was assigned based on: (i) the difference in achievement between an initiative with the success factor present and one lacking it, from the same domain, (ii) the frequency and focus on the success factor in the FDS initiatives, project plans and surveyed literature, (iii) the prevalence of the success factor found in the set of "successful" initiatives and finally (iv) the perceived impact of each factor after reviewing and comparing the original "FDS" dataset (103 initiatives).

Expert Interview.
An expert interview was included to gain a better understanding of the FDS initiatives of the perceived success factors in a new initiative. The expert was presented with the list of identified success factors and asked to confirm or point out flaws with the identified success factors (Appendix 1.1). The feedback from the expert served to refine the answer to the research sub questions and to revise the impact of proposed success factors (Apendix 1.2).

4. Results and discussion

FDS constitute an emerging field of research with most studies achieved in the past four years. A search on Web of Science using the keywords "Data sharing initiatives" produced 649, 746, 896, 937 and 950 hits for each consecutive year between 2018 and 2022. Instead, a query of "Federated Data Spaces" only produced 24, 18, 18, 67 and 91 results (publications) for the same years. From these articles, those discussing specific communication protocols, the architecture and the evolution of dataspaces were filtered out as they fall outside of the scope of this study.
While there is a lot of focus on this subject due to its importance and novelty, the academic literature is limited. It mainly focusses on the definition [9,11, 14], architecture [16,21] and evolution of data spaces [15]. Overall, the academic literature is creating frameworks and defining FDS. One of the first relevant academic sources is the article" Designing a multi-sided data platform: findings from the International Data Spaces case" published in 2019 after 3 years of research by Boris Otto & Matthias Jarke [17]. Most of the related work is found in industry where results are typically announced via websites, whitepapers and company communications, e.g., press releases or brochures. Among these, most are focused on the principles of datasharing, architecture of FDS and data reporting, e.g., [8, 22]. Many initiatives are harbored by large European projects, e., the "Living Labs" under the umbrella of the European project FEDeRATED started in 2019 [18]. The periodic progress reports of these projects provide additional sources of information on the path towards the practical implementation of FDS.

Sub question 1: What is the current state of the art regarding FDS?

FDS provide an efficient tool for various industries and fields of activity to lower the costs and improve the efficiency of data sharing and analysis. As data on FDS builds up, the developers and authors of large initiatives intensify their efforts to define the requirements and practical aspects leading to successful implementations. For example, the World Economic Forum has published an eight-step guide for setting up a federated data consortium for sharing sensitive health information [24]. The process starts with building trust among consortium partners and ends with implementing the application programming interfaces (API) specific for consortium members. The recommendation is to consider the KPI at an early stage and involve technical staff in the initial discussions of the consortium governance model.
In the logistics sector, data sharing in a manner that ensures data sovereignty and wide availability is linked to economic competitiveness, innovation and profit for companies. A roadmap was envisioned to align the data structures in the logistics field in the Netherlands with federated European infrastructures such as IDS [4]. Recent Dutch projects addressed the design of a data sharing structure for logistics (DL4LD, 33], the design of a high level logistics data space architecture (ICCOS, 2019, [34]), federated interoperability aspects related to

the organization of data ecosystems and semantics (CLiCKS, 2020, [35], as well as intra and inter-space connectivity (DASLOGIS, 2020-2023, [36]) and event-driven coordination in logistics (BDI-Business Data Infrastructuur and DIL-Data Infrastructuur Logistiek,[37]). Transforming this vision into reality will require a long-term commitment. Costs, certifications and clearinghouse issues weigh heavily on the speed and success of this endeavor [19].

The challenges are different in the manufacturing sector, faced with the "twin transition", i.e., achieving sustainable production while implementing digitalization. A preliminary report from an ongoing industry-research collaborative project in Finland emphasized the differences in perceived value by the different actors involved in FDS ecosystems. The potential value of solutions offered as a service was highlighted, thus catering to the needs of companies with different sizes to attract their involvement with FDS [10].

There are emerging FDS initiatives from many other sectors, e.g.,. the large Mobility Data Space, a project of the German Federal Government aiming to boost competitivity and innovation in mobility solutions. [38].

In summary, FDS have to potential to have a disruptive impact on the competitiveness, efficiency and sustainability in many sectors. This is mostly due to their main attribute: data sovereignty. FDS solve several problems at once. First, data spaces themselves represent a paradigm shift compared to traditional data sharing via central platforms. Access to central platforms is strictly controlled and allows limited number of participants. FDS are dynamic, open and flexible while retaining rules and standards. This allows the scaling up of data spaces without one-to-one agreements [39]. Secondly, FDS allow for a) trust, b) better data interoperability, c) data value creation, d) better data discoverability and e) machine-to-machine (M2M) communication in harmonized formats. These enable smooth communications between organisations and companies. [3, 30, 40]. Finally, FDS address the semantics issues, harmonizing definitions across the entire ecosystem. This not only vastly helps with data interoperability but also ensures coherence and precision within FDS ecosystem.

Data gathered so far enabled to identify key aspects for implementing FDS and challenges faced in the process. How to exploit best the information to ensure the success of a new FDS initiative? One must first define "success" and then identify the success factors that are common to all industries and fields.

Sub question 2: What is a successful FDS initiative?

The two main contributors to the FDS initiatives dataset are the "FEDeRATED Living Labs" list [18] and the "Data Spaces Radar" report by International Data Spaces [13]. International Data Spaces (IDS) Association classifies FDS initiatives in four stages: "lead in", "case committed", "pilot" and "live". In order to be classified by IDS as "live" the initiative must have sovereign data exchange, the technology of the data space must be fully functional and participants can already access it as a service. Additionally, it should improve processes or solve an issue, be accessible and adopted within a network [13]. Federated Living Labs (FLL) considers that initiatives that "validate the leading principles and the reference architecture"[18] are successful. From these examples it is clear that there is no standardized definition of a successful FDS. However, the definitions of success revolve around proving that information can be exchanged within the FDS ecosystem as intended and the FDS initiative brings added value to a shareholder or process. Therefore, initiatives that have not been completed and put into practice cannot be considered successful. These have not added value to the shareholder or process even if in testing stage data was shared through the entire ecosystem. Furthermore, a determination can be done if an initiative is successful only when the initiatives is live. This is due to the fact that FDS has multiple stakeholders and multiple goals. The impact on all stakeholders and processes can be only measured after the FDS gets live. As the FDS impact is heavily dependent on stakeholder usage, it is impossible to fully predict, in the testing phase how the FDS will be used by stakeholders.

Aggregating the different presented definitions of a successful FDS and extrapolating from literature we come to the following definition:

A Federated Data Space is successful when sovereign data exchange is taking place, the technology behind it is fully functional, participants can fully access it as a service, it has improved its goal process and is fully deployed.

Sub question 3: What are the current successful and failed FDS initiatives?

A set of 103 FDS initiatives was assembled and studied based on resources found on IDS [13], FDS [18] and [12]. From this pool, ten were deemed "successful" based on the hereby adopted definition (In Table 1).

Table 1. Successful FDS according to the definition from sub question 1 taken from a data set of 103 items from IDS [13] , FDS [18] and [12]

No	*Name*	*Source*	*Domain*	*Website*
1	ADVANEO DMP	IDS	Cross Domain	[41]
2	MARKET4.0: 3DFORM	IDS	Manufacturing	[42]
3	ECI Gatewise	IDS	Mobility	[43]
4	Mobility Data Space	IDS	Mobility	[38]
5	Mobilithek	IDS	Mobility	[44]
6	Smart Connected Supplier Network - Market 4.0	IDS	Supply Chain	[45]
7	Benelux transport data sharing facility	FDS	Mobility	[46]

8	Deutche Telekom Data intelligence Hub	IDS	Cross domain	[47]
9	Truzzt port	IDS	Cross Domain	[48]
10	Truzzt box	IDS	Cross Domain	[48]

Within this group, most initiatives (four each) covered Mobility or cross domain while and one each came from Manufacturing and Supply Chain fields and one was cross-domain.
There were 92 FDS initiatives which have not been announced as failed or discontinued. However, since they are not completed and thus don’t pass the definition of a successful FDS (sub question 2) they were considered here as “failed” initiatives. Furthermore, the dataset includes one discontinued FDS initiative, Byzantine-robust federated learning [12]. There, despite the deployment of several strategies to circumvent malicious attacks to the training server, all showed limited efficiency in preserving the global training model. Thus, they failed to accomplish the main objective [12].

Sub question 4: What are the common characteristics found in successful initiatives and to what degree they affect the success of the respective initiatives?

Based on the comparative analysis of the successful FDS and by considering also the findings from literature, a number of 13 common traits were found and proposed to serve as “success factors” (Table 2). These were grouped in four categories according to their relevance for (1) governance, (2) business value, (3) architecture and (4) technology. Most success factors are linked to governance and business value (four factors each), followed by technology (three factors) and finally architecture (two factors).
The common traits of all FDS initiatives that did not affect the success of an initiative were considered excluded factors. The potential impact of identified success factors was evaluated and an arbitrary score was assigned. based on the available information (Table 2), as explained in the “Methodology” section.

Table 2. Common characteristics of successful FDS constituting potential success factors and their impact on the FDS implementation

No	*Success factor (SF)*	*Impact (From 1=low to 5=high)*	*FDS initiatives containing the SF*
	Governance		
1	Narrow focus	4	2,4,5,7
2	Clear Governance Model	4	ALL
3	Iterative approach towards the FDS implementation	2	1,2,3,4,5,7
4	Seamless technical agreements between stakeholders	4	1,6
	Business Value		
5	Shared values & common goals	5	ALL
6	Close collaboration with stakeholders	2	ALL
7	Invested stakeholders and pilot tests	4	ALL
8	Medium and big sized companies/ organizations.	2	7
	Architecture		
9	F.A.I.R compliant	4	ALL
10	Certification of technology	5	ALL
	Technology		
11	Use of multiple data streams.	3	3,5,6
12	Automation of processes	3	ALL
13	Integration and adaptability	2	ALL

A fast analysis of the data summarized in Table 2 shows that some factors were common to all successful implementations. These include all the success factors pertaining to the architecture of the ecosystem, which have also high impact score (IS) . The common factors related to technology (automation of processes and integration and adaptability) have low to medium impact. With regards to governance, a clear governance model weighs heavily on achieving the implementation of FDS. Not in the least, three common success factors of these initiatives were related to business value, namely (i) shared values and common goals (IS= 5), (ii) close collaboration with stakeholders (IS=2) and (iii) invested stakeholders and pilot tests (IS=4).
An important observation is that due to the small set of FDS and successful initiatives in particular, the proposed impact factors and their impact scores need to be validated in future studies. The presented analysis is the result of a first attempt to synthesize the information available so far. Moreover, it represents the starting point for a larger discussion.
The success factors are discussed in more detail in the following.

1) Narrow focus.

All the FDS that are successful focus on just one specific area and have clear, well defined use cases described. A good example is constituted by the successful FDS in the logistics domain. Those FDS solely focus on creating and sharing information regarding either cargo or passengers. Additional areas, e.g., maintenance of the logistics park should be addressed independently to match the requirements of an FDS and integrated in a separate initiative. Domains where avoiding multiple areas is difficult or not desirable are considerably slowed down in their development. One such area is the medical field where patient data is shared but is heavily regulated. There is also insurance and medicine prescriptions. These multiple domains greatly expand the different types of participants in the ecosystem, increase the requirements of the FDS and put additional technological burden on the FDS.

2) Clear Governance Model.

A clear model will translate the shared vision and common values of participants into transparent and fair policies and frameworks agreed by all. These materialize further on into coherent legal and technical agreements and efficient practices within the FDS. The development of FDS is highly dependent on the interest and effort of participating stakeholders (data users and providers). Consequently, ensuring a common vision, transparency and fairness will motivate the stakeholders.
This factor has a high impact since all successful FDS have a clear governance model. Without a clear governance model the stakeholders will be reticent to use it. [12]

3) Iterative approach towards the FDS implementation

The majority of the initiatives in Table 1 were developed in an iterative manner. The failed FDS blockchain initiative [12] attempted to implement the original version of the initiative without any iteration. Moreover, the digital skills and governance ability of FDS stakeholders are constantly changing. As the progress with the FDS initiative advances so do the skills in digital skills and governance prowess of the stakeholders. Thus, the FDS can be improved by making use of those new skills. This results in an iterative process from design to implementation. From analyzed initiatives, as well as the lack of focus in literature on iterative approaches within FDS we conclude that the impact factor is low.

4) Seamless technical agreements between stakeholders and FDS

In order for the FDS to work best it is crucial that the technical agreements between stakeholders and the FDS are coherent and, as much as possible, standardized. Having different technical agreements across the FDS ecosystem increases complexity, delays problem solving and creates disadvantages for some stakeholders. The big impact score reflects the importance of this factor. A score of five was not attributed since this factor was not identified in all successful FDS.

5) Shared values and common goals

FDS are ecosystems where multiple stakeholders interact with each other. It is therefore paramount that all actors share the same goals and values. This will ensure that all participants will be working well with each other and will also build trust for the FDS. Furthermore, by implementing those values into the FDS framework from the beginning ensures that stakeholders are motivated and trust to use the FDS.
The most widespread set of values in the analyzed use cases are:

- Ensuring data sovereignty and data quality
- Creating trust among platforms and participants
- Providing a framework to enable interoperability
- Being open and neutral to any participating party

Without trusting that the data will be handled with care no stakeholder will use the FDS. Therefore, this success factor receives the maximum impact score.

6) Close collaboration with stakeholders

Irrespective of the architecture of the FDS stakeholders play an important role in the FDS ecosystem. Therefore, close collaboration with the stakeholders from the design stages will guarantee that all the requirements from all stakeholders are taken into account. Furthermore, having constant communication with stakeholders enables to take into account during the FDS development all of the issues that they might encounter. This success factor receives a score of two due to the fact that it's only applicable during the development stage.

7) Invested stakeholders and pilot tests .

The support of stakeholders throughout the entire FDS project is crucial. Their involvement allows to shape, better iterate and test the FDS by simulating the use of the FDS as accurately as possible. Consequently, stakeholders that are interested in trying new technologies or new protocols and willing to try pilot programs greatly contribute to a successful initiative.
Proof to that stands the fact that all of the successful FDS initiatives from the analysed data set have had interested stakeholders since the beginning of the project. Some initiatives, not included in Table 1 as they are not yet fully deployed are on a promising path due to invested stakefolders. An illustrative example is the FDS implemented by the American tech company Palantir as a pilot test at Chelsea and Westminster NHS Foundation Trust in the UK in 2022 [32], The pilot test was successful and the results were tangible, including a unified patient digital file and an average two-day reduction in the time required for administrative work, i.e., two days gained for a faster treatment of the patients. Following this success, in June 2023, Palantir was awarded a grant from the NHS for 'transitioning' the NHS projects in the UK to the federated data platform [50].

8) Medium and big sized companies/organizations.

There are no successful implementations done by a small companies (i.e., with fewer than 100 employees) in the analyzed dataset. This is due in part to the fact that small companies concentrate their resources towards operational issues rather expensive research and development. Collaboration is needed across different companies and organizations and that places additional burden on smaller companies. Often small organizations don't have enough resources needed for the participation to a FDS, e.g., to invest in certifications etc. While smaller companies do participate to FDS, the majority of companies involved in FDS initiatives are medium (100-499 employees) to large enterprises (with over 500 employees). These have more financial, human and technical resources to invest in FDS. For small companies it is recommended to partner with bigger companies or use FDS as a pay-per-service if possible. The assigned impact score is two as small companies can join FDS without major problems- although will face difficulties.

9) F.A.I.R compliant data and standardized semantics

The data managed within the FDS has to be FAIR (Findable, Accessible, Interoperable and Reusable) in order for the system to be efficient. The standardization of the semantic language contributes to efficiency of data sharing by enabling stakeholders to refer to the same item by the same name. The high impact factor was given to this success factor due to its prevalence in successful initiatives.

10) Certification of technology

Certifications are crucial for convincing participants to trust the security, integrity and technical soundness of the ecosystem. There are a couple of certification bodies, usually affiliated to the companies that create the architecture of the FDS. One such example is IDS Certification Scheme [50]. The security certifications cover technical components, protocols and data connectors.
All successful initiatives have certifications or plan to certify their technology soon. Therefore, this success factor has an impact score of five.

11) Use of multiple data streams

Data streams are places from where data originates or is collected. Implementing as many data streams as possible increases the amount of information accessible by FDS stakeholders. This results in a more attractive FDS. More data streams also increase reliability by having multiple sources presenting the same information making sure there is no error. Furthermore, more data allows for algorithms to produce better results. A good example of implementing multiple data streams is Mobilithek, enabling access to real time data. All this data is collected through agencies and IoT devices installed on several transport vehicles.The impact score is three. Multiple data stream make the FDS more valuable, yet are not essential to the success of an FDS (this succes factor is present in only 3 initiatives). The impact score is three.

12) Automation of processes

In all industries there are a couple of "middlemen" or middleman processes that cost money and time and decrease the efficiency of the process. If an FDS is created that aims to remove as many middlemen as possible and automating as many middlemen processes as possible then the FDS will increase the efficiency of the process significantly. Furthermore, FDS have a unique advantage to push for automation through the protocols and existing technology implemented in the ecosystem. If this advantage is exploited, not only will the FDS become more useful but it will significantly impact for the better the entire industry where it is applied. An example of such an optimization could be that instead of writing down on a paper a shipment's information from company A inventory manager, transporting the paper along with the shipment and then putting the information from the paper into the inventory manager of company B, the FDS could simply transfer the information from inventory manager A to inventory manager B when the IoT device detects that the shipment arrived. The impact of automation of processes on the success of an FDS implementation is medium.

13) Integration and adaptability.

In order to speed up the process of implementing FDS it is important to not "re-invent the wheel". It is better to use pre-existing and tested systems and integrate them in the FDS, thus certifying them as being secure. Initiatives that have attempted to create all the technology and protocols from scratch have been delayed enormously and have encountered multiple difficulties. No FDS initiative that has created the majority of their systems on their own has been successful. On the other hand, if the stakeholders and the creator of the FDS are willing then making everything from scratch is a possibility, it does not mean the FDS will surely fail. It will however take longer and there is potential for more problems to appear. Considering all this, the attributed impact score for this success factor is two.

During the comparison of each initiative and literature research the following factors have been observed in all initiatives, including those postponed or failed. Therefore, they were excluded from the "success factors list".

Excluded factors are:
(i) **The FDS ecosystem structure**.
The structures are all unique and differ based on the use case

(ii) The industry in which the FDS operates.
The industry mainly affects complexity and development speed but not the success rate of FDS.

(iii) The employed technology.
While certifications are needed to have a successful implementation which one is used does not impact the success rate. All the technology used should be catered for that the specific use case.

Sub question 5: What implication do the found success factors have for research and industry?

The above defined factors have several implications for industry, research and for society in general that need to be considered when aiming for a successful FDS.
With respect `to research, studies should not solely focus on the technical aspects of FDS but also on the stakeholders' requirements and how to implement throughout the system the concepts of Federated Values. This means, for example the stakeholder's needs to: (1) ensure data sovereignty and data quality; (2) create trust among platforms and participants; (3) provide a framework to enable interoperability; (4) be open and

neutral to any participating party; (5) ensure rapid deployment of new IT services to support business processes.

Moreover, research in the field of FDS should include not only technical aspects but also social studies to determine the interaction between companies in a FDS.

With regards to the industry, designing a FDS around existing technologies speeds up the process (no need to "re-invent the wheel"). Due to an abundance of new data creating new ways to share and handle the data is paramount, hence the FDS. These new systems must include the proper values to bring people, companies and technology together in an efficient ecosystem centered around well-defined values embraced by all.

Automating every aspect that does not require human values and is a repetitive process is an efficient way to streamline processes and obtain accurate and reproducible results. The speed of data sharing has increased, the access to data has widened but in the same time the volume of data produced grows exponentially. Hence, to keep the pace, faster, more efficient solutions for data handling must be implemented by the companies and this presumes standardization, smart algorithms and a high level of automation. Consequently, standardization of protocols and techniques needs to pick up.

More research into the values at the core of FDS is required.

In addition, research in the areas where implementations take a longer time is needed to identify ways to speed up the process so those industries can be competitive.

5. Conclusion

The field of FDS is relatively young with ambitious initiatives that emerged in the past years, covering various industries from healthcare to space exploration. Due to the novelty of the FDS a systematic overview of these initiatives is missing. This study fills a knowledge gap by identifying the main traits of a successful FDS, i.e., one where sovereign data exchange is taking place, the technology behind it is fully functional, participants can fully access it as a service, that has improved its goal process and is fully deployed.

In the aim to identify the success factors, the available literature, including a set of 10 successful initiatives was examined by considering use cases put forward by companies and different organizations. This enabled to identify 13 factors that may be associated with successful FDS. To give a quantitative measure of each factor's impact on successful new implementations, an arbitrary score from 1 to 5 was assigned. The scores proposed were reviewed and validated by an expert. Still, due to the small pool of data available it is premature to claim absolute weights of each of these traits on the success of an FDS. Nonetheless, the analysis indicates that the most important are: (i) the set of values defined in a framework which ensure the fair and efficient use of the data in the benefit of all participants; these which are translated further into clear FDS structure and technical specifications; (ii) focusing on a narrow, well defined objective; (iii) ensuring the clear understanding and close collaboration of stakeholders from design to implementation; (iv) taking a stepwise approach to implementation, preferably by including pilot tests; (v) standardization of protocols and processes, automation and the ability to integrate various existing software solutions. In the same time, the success of an FDS does not depend on its structure, the target industry or activity field or the type of technology used, as long as that technology is adequate for the intended use. The success factors identified in this study need to be validated in future works by analyzing larger sets of data, once more information will become available. The limited amount of information available definitely affects the identified success factors and may be biased towards the mobility sector since the majority of the successful initiatives come from that domain. Further validation should be done with more interviews with experts in the field.

The findings have a series of implications for both industry and research. In a nutshell, FDS bring a shift in social paradigm, therefore social aspects and participant requirements are equally important to consider as are technical matters.

FDS are about bringing together people, companies and technology in an efficient ecosystem centered around well-defined values embraced by all.

6. Future Work

Most of FDS analyzed initiatives analyzed in this study are scheduled to reach deployment stage at the end of 2023. In this context, as the information from these initiatives will become available in the future, the analysis of potential success factors should be revised by including the new data. More interviews with experts in the FDS area are required and finally more research is needed to identify the interactions between stakeholders and the FDS.

In Europe, the vision for a common data space [20], digitizing the industry [5] and the efforts to regalement the use of AI [25] go in parallel. FDS are at the intersection of these pathways. As it is anticipated that AI will radically transform industry and society, FDS will be crucial for the AI training to allow safe access to training data while ensuring data sovereignty [26]. Thus, artificial intelligence powered systems may also take the role of a traditional stakeholder in the ecosystem such as a route optimizing agency.

Acknowledgements.

The author thanks Mr. Simon Dalmolen, lead architect for data sharing at TNO, Groningen, Netherlands for kindly providing his view as an expert in FDS and for the help in revising the success factors and their impact. Special thanks to my supervisor, dr. IR. Hans Moonen for his supervision and guidance throughout the thesis.

.

References

[1] https://www.statista.com/statistics/871513/worldwide-data-created/

[2] Bernier A., Molnár-Gábor F., Knoppers B.M. et al (2023). Reconciling the biomedical data commons and the GDPR: three lessons from the EUCAN ELSI collaboratory. Eur J Hum Genet. https://doi.org/10.1038/s41431-023-01403-y

[3] Mertens C., Alonso J., Lázaro O, Palansuriya C, Böge G., Nizamis A., et al., A Framework for Big Data Sovereignty: The European Industrial Data Space (EIDS), in Edward Curry, Simon Scerri, Tuomo Tuikka (Editors), Data Spaces Design Deployment and Future Directions, Springer , 2022, p. 201-226, ISBN 978-3-030-98636-0, https://doi.org/10.1007/978-3-030-98636-0

[4] Dutch Institute for Advanced Logistics, TKI Dinalog, The Logistics Data Sharing Infrastructure, Whitepaper, August 2020. Available at https://www.dinalog.nl/wpcontent/uploads/2020/08/Dinalog_Whitepaper-Data-Infrastructure_D. Accessed June 23, 2023.

[5] European Commission, A European strategy for data, Brussels 2020, available at: https://digital-strategy.ec.europa.eu/en/policies/strategy-data. Accessed on May 5, 2023

[6] Gaia-X, https://www.gaia-x.eu/

[7] Hallock H, Marshall S. E., 't Hoen P. A. C., Nygård J.F., Hoorne B., Fox C., Alagaratnam S. (2021) Federated Networks for Distributed Analysis of Health Data. Frontiers in Public Health (2021) 9712568, https://doi.org/10.3389/fpubh.2021.712569

[8] International Data Spaces Association, Data Spaces Overview, Data Spaces Brochure 2022, available at https://internationaldataspaces.org/publications/about-idsa/. Accessed May 5 2023

[9] Jarke M., Otto B., Ram S. (2019). Data sovereignty and data space ecosystems. Business & Information Systems Engineering, 61, 5, 549–550; https://doi.org/10.1007/s12599-019-00614-2.

[10] Jurmu M., Niskanen I., Kinnula A., Kääriäinen J., Ylikerälä M., Räsänen P., Tuikka T. (2023). Exploring the Role of Federated Data Spaces in Implementing Twin Transition within Manufacturing Ecosystems, Sensors, 23, 4315. https://doi.org/10.3390/s23094315

[11] Jussen I., Schweihoff J. Dahm V., Möller F., Otto B. (2023). Data Sharing Fundamentals: Definition and Characteristics. In Proceedings of the 56th Hawaii International Conference on System Sciences, Hawaii: USA, Available at https://hdl.handle.net/10125/103083. Accessed June 23, 2023.

[12] Li S., Ngai E. C., Voigt T. (2023). An Experimental Study of Byzantine-Robust Aggregation Schemes in Federated Learning, IEEE Transactions on Big Data, 1-13. https://doi.org/10.1109/tbdata.2023.3237397

[13] Mertens C. (2023): The Data Spaces Radar. International Data Spaces Association. https://doi.org/10.5281/zenodo.7624017. Available at https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/Data_Spaces_Radar_Web_230307.pdf. Accessed May 5, 2023

[14] Otto B. (2022). A federated infrastructure for European data spaces. Communications of the ACM. 65. 44-45. 10.1145/351234128.

[15]Otto, B. (2022). The evolution of data spaces. In B. Otto, M. ten Hompel, S. Wrobel (Eds.), Designing dataspaces, Springer, https://doi.org/10.1007/978-3-030-93975-5_1

[16] Otto B., Hompel M. T., Wrobel S. (2022). Designing Data Spaces: The Ecosystem Approach to Competitive Advantage. Springer.

[17] Otto B, Jarke M. (2019). Designing a multi-sided data platform: findings from the International Data Spaces case. Electronic Markets, 29(4), 561–580. https://doi.org/10.1007/s12525-019-00362-x

[18] Pilots LivingLab Scoping, FEReRATED Milestone 4 , 29 October 2020, available at http://federatedplatforms.eu/index.php/library/item/federated-milestone-4-report-pilots-livinglab-scoping. Accessed May 5, 2023.

[19] Piest J.P.S., de Alencar Silva P., Bukhsh F.A., Aligning Dutch Logistics Data Spaces Initiatives to The International Data Spaces: Discussing The State of Development, Proceedings of the Workshop of 11th International Conference on Interoperability for Enterprise Systems and Applications Workshops , March 23–24 2022, Valencia Spain. Available at https://ceur-ws.org/Vol-3214/WS6Paper1.pdf. Accessed June 23, 2023

[20] Publications Office of the European Union, Corcho, O., Simperl, E., Data.europa.eu and the European common data spaces – A report on challenges and opportunities, 2022, https://data.europa.eu/doi/10.2830/91050

[21] Schleimer A. M. Fraunhofer I. Jahnke N. and Otto B (2023), Architecture Design Options for Federated Data Spaces, in Proceedings of the 56th Hawaii International Conference on System Sciences. January 2023, 3643-3652, available at https://hdl.handle.net/10125/103078, Accessed June 23, 2023

[22] Van der Keijl N., Moonen H. (2022). The what, how and why of data sharing - Port logistics more efficient reliable and sustainable. SmartPort Whitepaper, available at https://smartport.nl/wp-content/uploads/2023/03/SmartPort_White-paper_Data-Sharing_EN_final.pdf. Accesed May 5, 2023

[23] World Economic Forum, Federated Data Systems: Balancing Innovation and Trust in the Use of Sensitive Data, White paper 2019, available at https://www3.weforum.org/docs/WEF_Federated_Data_Systems_2019.pdf . Accessed June 23, 2023

[24] World Economic Forum, Sharing Sensitive Health Data in a Federated Data Consortium Model- An Eight-Step Guide, Insight Report, July 2020, available at https://www3.weforum.org/docs/WEF_Sharing_Sensitive_Health_Data_2020.pdf . Accessed June 23, 2023.

[25]European Commission (2021). "A European approach to Artificial intelligence," European Union, 2021. URL: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence.

[26] Netherlands AI Coalition towards a federation of AI data spaces, NL AIC reference guide to federated and interoperable AI data spaces, available at https://nlaic.com/wp-content/uploads/2022/02/Towards-a-Federation-of-AI-Data-Spaces.pdf. Accessed July 1, 2023.

[27]European Commision, H2020 Programme Guidelines on FAIR Data Management in Horizon 2020, version 3.0, 26 July 2016, available at https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf, acceded July 1, 2023.

[28] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) , Official Journal of the European Union, L152, vol 65, 3 June 2022, p 1-44, available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2022:152:TOC. Accesed July 1, 2023.

[29] Schlueter Langdon, C., Schweichhart, K. (2022). Data Spaces: First Applications in Mobility and Industry. In: Otto, B., ten Hompel, M., Wrobel, S. (eds) Designing Data Spaces . Springer, Cham. https://doi.org/10.1007/978-3-030-93975-5_30

[30] Nagel L., Lycklama D. (2021): Design Principles for Data Spaces. Position Paper. Version 1.0. Berlin, DOI: http://doi.org/10.5281/zenodo.5105744

[31] Vision document, Final version FEDeRATED Milestone 1, 17 December 2019, available at https://federatedplatforms.eu/images/Library/Vision/MILESTONE_1_FEDeRATED_VISION_17122019.pdf. Accessed June 23, 2023

[32] https://www.palantir.com/

[33] DL4LD, www.dl4ld.net

[34]https:// www.dinalog.nl/project/industry-4-0-drivensupply-chain-coordination-for-small-mediumsized-enterprises-iccos/

[35]https://www.nwo.nl/en/research-andresults/research-projects/i/42/35042.html

[36] https://www.dinalog.nl/project/daslogis-development-and-valorization-of-a-dutch-data-space-for-logistics/

[37] https://topsectorlogistiek.nl/wp-content/uploads/2022/07/20220614-BDI-Intro-FAQ-ENG.pdf

[38] https://mobility-dataspace.eu/

[39]https://internationaldataspaces.org/the-beauty-of-data-spaces-is-that-they-are-a-joint-effort/

[40] https://www.federatedplatforms.eu/index.php/the-federated-proposition

[41] https://internationaldataspaces.org/advaneo-launches-data-marketplace-to-guarantee-data-sovereignty-according-to-the-ids-standard/

[42]https://market40.eu/3dform/index.html

[43] https://gatewise.ecisolutions.com/

[44] https://mobilithek.info/

[45] https://internationaldataspaces.org/the-smart-connected-supplier-network-by-tno/

[46]https://www.federatedplatforms.eu/images/Library/Activity3/Final_MILESTONE_8_REPORT_11032022.pdf

[47] https://dih.telekom.com/en

[48] https://www.truzzt.com/en/

[49]https://internationaldataspaces.org/offers/certification/

[50]https://www.hsj.co.uk

Appendix

1.1. List of questions asked in the interview with the expert in FDS

Please note: These questions have been asked based on the original success factor list (Found in Appendix 1.2 A)

Q1: Do you believe that “Clear and Narrow Focus” is a success factor for FDS?
Q2: Do you agree with the impact score given to the success factor?
Q3: Do you believe that “Well defined application” is a success factor for FDS?
Q4: Do you agree with the impact score given to the success factor?
Q5: Do you believe that “Iterative approach to the FDS implementation” is a success factor for FDS?
Q6: Do you agree with the impact score given to the success factor?
Q7: Do you believe that “Seamless technical agreements between stakeholders” is a success factor for FDS?
Q8: Do you agree with the impact score given to the success factor?
Q9: Do you believe that “Implement Values into the FDS” is a success factor for FDS?
Q10: Do you agree with the impact score given to the success factor?
Q11: Do you believe that “Clear understanding of stakeholders and close collaboration with them from design to implementation” is a success factor for FDS?
Q12: Do you agree with the impact score given to the success factor?
Q13: Do you believe that “Invested stakeholders and pilot tests” is a success factor for FDS?
Q14: Do you agree with the impact score given to the success factor?
Q15: Do you believe that “Medium and big sized companies/organizations” is a success factor for FDS?
Q16: Do you agree with the impact score given to the success factor?
Q17: Do you believe that “Standardization of language” is a success factor for FDS?
Q18: Do you agree with the impact score given to the success factor?
Q19: Do you believe that “Certification for security and technology components” is a success factor for FDS?
Q20: Do you agree with the impact score given to the success factor?
Q21: Do you believe that “Certified data connectors” is a success factor for FDS?
Q22: Do you agree with the impact score given to the success factor?
Q23: Do you believe that “Use of IoT devices and multiple data streams” is a success factor for FDS?
Q24: Do you agree with the impact score given to the success factor?
Q25: Do you believe that “automation of processes” is a success factor for FDS?
Q26: Do you agree with the impact score given to the success factor?
Q27: Do you believe that “Data driven tools” is a success factor for FDS?
Q28: Do you agree with the impact score given to the success factor?
Q29: Do you believe that “Integration of available software solutions for automation, data gathering and analysis” is a success factor for FDS?
Q30: Do you agree with the impact score given to the success factor?
Q31: Do you believe that “Mobile Apps or platforms for easy data visualization” is a success factor for FDS?
Q32: Do you agree with the impact score given to the success factor?
Q33: What do you believe is the biggest problem that FDS solve?

1.2 Changes in the list of success factors following the discussion with the expert in FDS

A. The original list of success factors

No	*Success factor (SF)*	*Impact (From 1=low to 5=high)*	*FDS initiatives containing the SF*
	Governance		
1	Clear and narrow focus	4	2,4,5,7
2	Well defined application	4	ALL
3	Iterative approach to the FDS implementation	2	1,2,3,4,5,7
4	Seamless technical agreements between stakeholders	4	1,6
	Business Value		
5	Implement Values into the FDS	5	ALL
6	Clear understanding of stakeholders and close collaboration with them from design to implementation.	4	ALL
7	Invested stakeholders and pilot tests	5	ALL
8	Medium and big sized companies/ organizations.	1	7
	Architecture		
9	Standardization of language	5	ALL
10	Certification for security and technology components	5	ALL
11	Certified data connectors	5	ALL
	Technology		
12	Use of IoT devices and multiple data streams	3	3,5,6
13	Automation of processes	3	ALL
14	Data Driven Tools	2	2,6
15	Integration of available software solutions for automation, data gathering and analysis	2	ALL
16	Mobile Apps or platforms for easy data visualization	2	3,4,5

B. The revised list based on the feedback from the expert in FDS

No	*Success factor (SF)*	*Impact (From 1=low to 5=high)*	*FDS initiatives containing the SF*
	Governance		
1	Narrow focus	4	2,4,5,7
2	Clear Governance Model	4	ALL
3	Iterative approach towards the FDS implementation	2	1,2,3,4,5,7
4	Seamless technical agreements between stakeholders	4	1,6
	Business Value		
5	Shared values & common goals	5	ALL
6	Close collaboration with stakeholders	2	ALL
7	Invested stakeholders and pilot tests	4	ALL
8	Medium and big sized companies/ organizations.	2	7
	Architecture		
9	F.A.I.R compliant	4	ALL
10	Certification of technology	5	ALL
	Technology		
11	Use of multiple data streams.	3	3,5,6
12	Automation of processes	3	ALL
13	Integration and adaptability	2	ALL