

Grant Agreement Number: 825225

Safe-DEED

www.safe-deed.eu

D8.6 Safe-DEED Volume

Deliverable number	<i>D8.6</i>
Dissemination level	<i>Public</i>
Delivery date	<i>November 30th 2021</i>
Status	<i>Final version</i>
Author(s)	<i>Gert Breitfuss, Stefan Gindl, Lukas Helminger, Petr Knoth, Yiannis Markopoulos, Mark de Reuver, Abdel Aziz Taha, Mihnea Tufis</i>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 825225.

Changes Summary

Date	Author	Summary	Version
06.10.2021	Stefan Gindl	Deliverable skeleton	0.1
29.10.2021	Gert Breidfuss, Stefan Gindl, Lukas Helminger, Petr Knoth, Yiannis Markopoulos, Mark de Reuver, Abdel Aziz Taha, Mihnea Tufis	Initial draft	0.2
22.11.2021	Gert Breidfuss	Complementary input	0.3
23.11.2021	Stefan Gindl	Revision	0.4
28.10.2021	Mihnea Tufis	Complementary input	0.5
29.10.2021	Stefan Gindl	Revision	0.6
30.11.2021	Mark de Reuver	Internal review I	0.7
30.11.2021	Gert Breidfuss	Internal review II	0.8
30.11.2021	Stefan Gindl	Final version	1.0

Executive summary

The present deliverable D8.6 Safe-DEED Volume is a report summarizing the major outputs and activities of Safe-DEED. Safe-DEED focuses on the investigation of privacy-preserving technologies and methods from three directions, (i) the business perspective, (ii) the legal perspective, and (iii) the technological perspective. The project has produced artefacts for each of these three perspectives. This report follows this triangle of perspectives and is segmented accordingly.

We start by elaborating business aspects in Section 2. The section covers the business-related Safe-DEED demonstrator and the data-driven business model toolkit. Section 3 gives details about the technological artefacts and approached investigated and developed in Safe-DEED. This covers the four components for data valuation, private set intersection, deanonymization, and lead-time based pricing. For the latter, Safe-DEED developed another demonstrator, which the respective sub-section describes. In Section 4 we detail legal aspects relevant for Safe-DEED. This covers legal and ethical requirements, the bridging between impact and value assessment, building trust in data markets, as well as a teaching module.

Table of Contents

1	Introduction	7
2	Business Aspects	7
2.1	The Safe-DEED Demonstrator.....	7
2.1.1	Objectives.....	7
2.1.2	Achievements	8
2.1.3	Implications and Recommendations.....	9
2.2	The Data-driven Business Model Toolkit.....	9
2.2.1	Objectives.....	9
2.2.2	Achievements	10
2.2.3	Implications and Recommendations.....	11
3	Technology	11
3.1	The Data Valuation Component	11
3.1.1	Objectives.....	11
3.1.2	Activities	11
3.1.3	Key findings	12
3.1.4	Implications / Recommendations	14
3.2	PSI Component.....	14
3.2.1	Objectives.....	14
3.2.2	Achievements	14
3.2.3	Implications and Recommendations.....	15
3.3	Deanonimization Component	15
3.3.1	Objectives.....	15
3.3.2	Achievements	15
3.3.3	Implications and Recommendations.....	17
3.4	Lead-time Based Pricing.....	18
3.4.1	Objectives.....	18
3.4.2	Achievements	18
3.4.3	Implications and recommendations.....	19
4	Legal Aspects	20
4.1	Legal and Ethical Requirements.....	20
4.1.1	Objectives.....	20
4.1.2	Achievements	20
4.1.3	Implications and Recommendations.....	21
4.2	From Impact Assessment to Value Assessment.....	22
4.2.1	Objectives.....	22

4.2.2	Achievements	22
4.2.3	Implications and Recommendations.....	22
4.3	Fostering Trust in Data Markets	23
4.3.1	Objectives.....	23
4.3.2	Achievements	23
4.3.3	Implications and recommendations	23
4.4	Syllabus for Teaching Module.....	23
4.4.1	Objectives.....	23
4.4.2	Achievements	24
4.4.3	Implications and Recommendations.....	25
5	Conclusion.....	25

List of Figures

Figure 1: The welcome screen of the Safe-DEED WP6 demonstrator.	8
Figure 2: Risk analysis for invoice data.	9
Figure 3: The business model toolkit consists of the data map, the business canvas, and the data service cards.	10
Figure 4: A photo of the data service cards.	11
Figure 5: Private set intersection of two datasets using Psittacus.	14
Figure 6: Tabular data de-anonymisation risk analysis	16
Figure 7: Invoices data risk analysis	16
Figure 8: Demonstrator Supplier View.	19
Figure 9: Demonstrator Customer View.	19
Figure 10: Defining the pricing algorithms.	20

1 Introduction

The establishment of a lively data-driven economy rests on at least two major pillars: a notion of value for data and the trust in methods to ensure that data is not misused. Without these two pillars, any initiative in the area of data exchange and data trade is at risk of failure. In case stakeholders do not see any value in the exchange of data, or if they fear their data might be misused, they are likely to withdraw from respective initiatives, e.g. data markets. Safe-DEED tackles these two challenges. As mentioned on the its website, Safe-DEED works towards

“A competitive Europe where individuals and companies are fully aware of the value of the data they possess and can feel safe to use it.”¹

The project works towards this goal by covering the three perspectives “business aspects”, “technological aspects”, and “legal aspects”. For each perspective, Safe-DEED created artifacts for stakeholders in the area. For instance, the business perspective produced the data-driven business model toolkit, providing tools for organizations, companies, startups, etc., to identify innovative business ideas and tackle data-driven innovations.

We start our report with the business aspects, continue with technological aspects followed by legal aspects, and close the document with concluding remarks.

2 Business Aspects

2.1 The Safe-DEED Demonstrator

2.1.1 Objectives

The demonstrator applications aim at providing beyond the state-of-the-art solutions to the following business needs of an evolving data driven entrepreneurship:

Joint Data usage within corporate environments: It addresses the need of secure joint data usage between different units in a corporate environment. Private information should not be disseminated in the company. Nevertheless, the need to share department data e.g. between the marketing and strategy departments is impacted by the lack of technology to securely share the data thus bringing barriers to efficient business decisions. Taking into account the danger of de-anonymization due to the availability of relevant department data corpus, we use technology components to perform business analysis on the data in a way that is meaningful to the recipient department while hiding private information.

Joint Data usage between different enterprises in the same domain: It addresses the secure data exchange need between different enterprises who find it useful to analyse each other’s data. GDPR compliance, lack of trust, and potential leak of trade secrets are some of the challenges enterprises face while trying to capitalise on each other’s data knowledge. Nevertheless, should such challenges are overcome; the potential of such a data based knowledge exchange will be immense.

Joint Data usage between different enterprises in different domains: Companies that own a large amount of data and provide services to millions of customers, often do not have the ability to analyse them in order to obtain insights that will help to improve or define business strategies. For this reason, it is of great importance for enterprises to be able to share their data with specialised external entities to perform the analysis. Although certain agreements are signed about the use of these data, private

¹ <https://safe-deed.eu/>, last accessed Nov 30, 2021.

information should be kept secure. An indicative threat is that anonymized data can be combined with various freely available datasets so as private information can be revealed.

Data valuation: Even among large companies, many have no data valuation process in place resulting to lack of knowledge of the value that such data assets can bring to the enterprise. Safe-DEED aims at providing tools to facilitate the assessment of data value, thus incentivizing data owners to make use of the cryptographic protocols to create value for their companies and their clients.

2.1.2 Achievements

A strategic decision was taken in the project to implement **two versions of the demonstrator:**

- The first version will be fully functional aiming at being used within the professional community building actions. This demonstrator version will be characterised confidential according to the project DoA.
- The second version will be used to address the wide audience through the project web site. This demonstrator version will be public and will be implemented beyond the DoA mandates. It will expose the complete functionality with predefined scenarios and datasets to the wide community, but users will not be able to upload their own data due to security reasons. Safe-DEED project considers this as a very useful channel to increase awareness and disseminate the project achievements. Such a strategic decision to implement a public version was taken in the project since the research results are considered to be sound and can easily be adopted by the professional community.

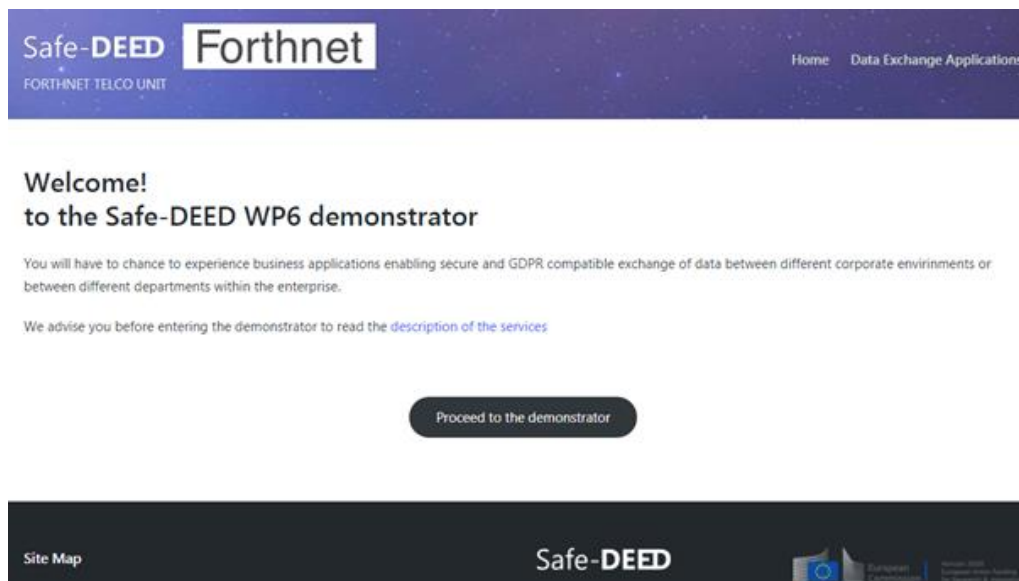
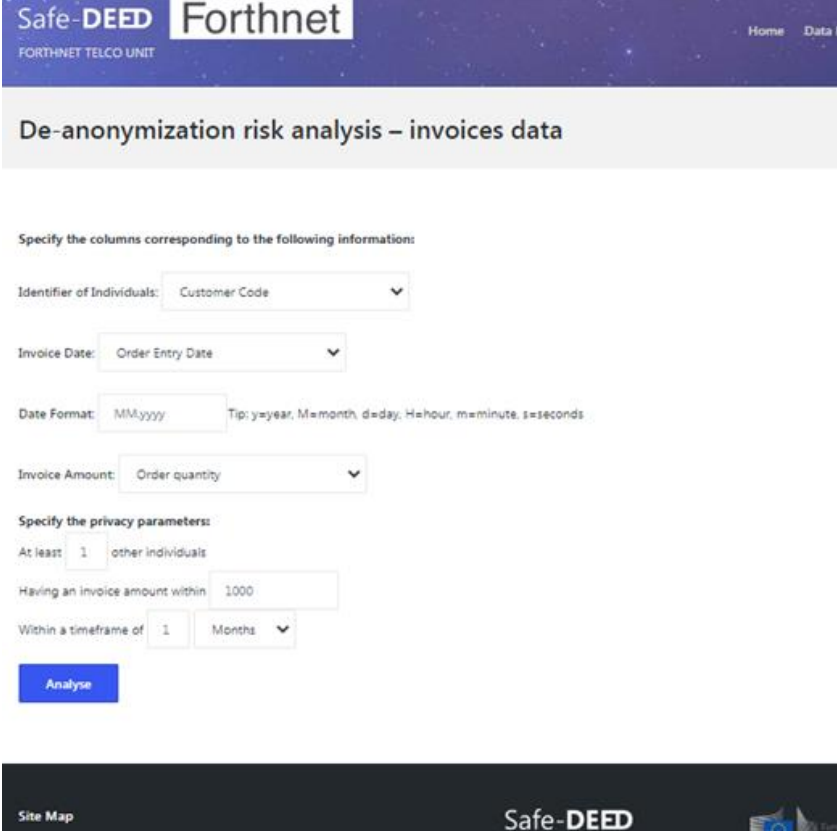


Figure 1: The welcome screen of the Safe-DEED WP6 demonstrator.



The screenshot shows the 'De-anonymization risk analysis – invoices data' page on the Safe-DEED website. The page header includes 'Safe-DEED Forthnet' and 'FORTHNET TELCO UNIT'. The main content area contains a form with the following fields and options:

- Specify the columns corresponding to the following information:**
 - Identifier of Individuals: Customer Code (dropdown)
 - Invoice Date: Order Entry Date (dropdown)
 - Date Format: MM/yyyy (text input) with a tip: Tip: y=year, M=month, d=day, H=hour, m=minute, s=seconds
 - Invoice Amount: Order quantity (dropdown)
- Specify the privacy parameters:**
 - At least: 1 other individuals (text input)
 - Having an invoice amount within: 1000 (text input)
 - Within a timeframe of: 1 Months (dropdown)
- Analyse** (blue button)

The footer of the page includes 'Site Map' and 'Safe-DEED'.

Figure 2: Risk analysis for invoice data.

2.1.3 Implications and Recommendations

Based on the demonstrator's evaluation feedback from various workshops within NOVA, presentations to companies and users via the demo link on our Safe-DEED website we can draw the following conclusions.

- The demonstrator is considered appealing to the professional community
- The demonstrator is seen as an interactive tool to supplement digital transformation presentations to large enterprises
- It is recommended to be able to accommodate additional scenarios, customised per case, via a wizard
- It can be provided as a white label service to consultants to augment their presentations to clients
- It can be used for training purposes.

2.2 The Data-driven Business Model Toolkit

2.2.1 Objectives

The main objectives of WP2 (Economic Aspects and Business Models) were to demonstrate the economic value and the development of new multi-actor business models for privacy enhancing and data valuation technologies. The project results regarding this task settings are reported in the two deliverables D2.2 (Business models for use cases and generic business models) and D2.5 (Quantification of the economic impact). To assist the business model development process a toolkit of three data-driven business supporting tools have been designed and evaluated.

2.2.2 Achievements

To develop suitable BM supporting tools for Safe-DEED technologies we examined 1) data-driven BMs in general, 2) the advances of Safe-DEED technologies for data-driven BM and 3) existing types of tools and methods for developing BMs.

- Since data is the central resource in a data-driven Business Model we designed a visual collaboration tool (Safe-DEED Data Map) to identify possible data sources for developing new data-driven services. The main aim of the Safe-DEED Data Map is to establish a common understanding of available data in a company concerning Safe-DEED technologies.
- The Safe-DEED Data Service Cards (DSC) aim to enhance or develop new data-based services through the systematic combination of data sources, analysis methods, customer benefits and revenue opportunities. The present 50 Data Service Cards can be used as inspiration in the development process of data-driven innovations.
- The Safe-DEED Data-Driven Business Canvas particularly considers the development process of a structured value proposition, the identification of customer benefits and the assessment of financial aspects. Furthermore, the canvas aims to support the necessary interdisciplinary communication in the innovation process of data-driven services.

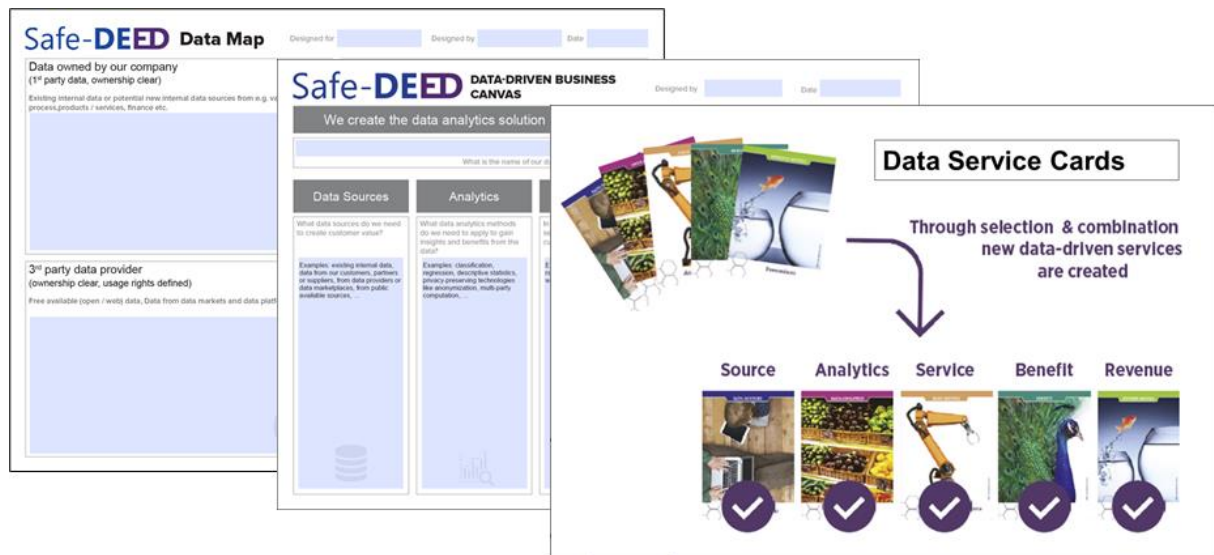


Figure 3: The business model toolkit consists of the data map, the business canvas, and the data service cards.

To make the developed tools best possible available to a large audience the Safe-DEED toolset is available on the tool section of the platform <https://businessmakeover.eu/tools>. The Safe-DEED tools can be reached directly via the following links:

- **Data Map:** <https://businessmakeover.eu/tools/safe-deed-data-map>
- **Data Service Cards:** <https://businessmakeover.eu/tools/safe-deed-data-service-cards>
- **DDB Canvas:** <https://businessmakeover.eu/tools/safe-deed-data-driven-business-canvas>

Besides the available online versions, the Safe-DEED Data Services Cards are also available as a printed version (in English and German) for use offline in the form of innovation workshops or lectures.



Figure 4: A photo of the data service cards.

2.2.3 Implications and Recommendations

So far more than 20 workshops (online and offline) have been conducted and the gathered feedback in general was very positive, especially regarding usefulness and ease of use e.g. “Basically a useful tool, the cards offer ideas about the most important services - especially useful if you are completely ‘planless’”; “I find the Data Service to be an interesting and useful tool.”; “The tool speeds up the ideation process”; “The cards are a good source of inspiration”. The evaluation results also show that people who used the Data Service Cards in the workshop would use the cards for the actual development of data-driven services (intention to use). In summary, it can be stated that the Data Service Cards are perceived as a tool that facilitates the development of data-driven services. Next steps are:

- Further development of all three tools based on the gathered evaluation results.
- Extension of the toolset along the data service innovation process.
- Translation of the tools into other European languages e.g. French, Italian, Spanish
- Strengthening the partnership with Innovalor (who hosts businessmakeover.eu) as an exploitation partner of the toolset.
- Establish the toolset as an education standard for data driven business

3 Technology

3.1 The Data Valuation Component

3.1.1 Objectives

The main objective was to design and implement a Data Valuation Component (DVC) used to assess the value of structured data, considering the context in which it is used as well as in a context-free setting. A secondary objective was to create awareness about the importance of quantifying the value of data and to create a community of practitioners around the DVC.

3.1.2 Activities

Our activities focused around three main areas: requirements and design, research and development.

Requirements and Design

In this task we conducted an industry survey in order to gauge the attitude of various data-centred businesses towards data production and consumption, as well as their interest in a data valuation platform. The findings of this survey helped us define initial requirements for the data valuation

component, which quickly materialised in a high-level architecture design and in the initial baseline prototype of the platform.

Research

These activities were dedicated to understanding the state of the art of data valuation. This has proven more challenging than initially thought, because while everybody has an intuition that data is valuable, there is no clear definition of what the value of data is. Most of the initial studies and platforms that we reviewed tried to answer the question “how much is my/your data worth?”, therefore assuming an equivalence between the value of data and the price tag attached to data. In parallel our research was looking into previous attempts at data valuation, investigating attempts to define the value of data, the properties that contribute to the difficulty of establishing its value, methods and methodologies for quantifying it. As a result, we were able to articulate our own data valuation process, and focus our research on the specific pillars that support it:

- formalising data valuation contexts.
- state of the art review of data quality frameworks, measures and metrics.
- data usability in context.
- metrics aggregation.
- economic value of data.

Development and Testing

The results of our research activities have materialised in the development of the Data Valuation Component, a web application that receives an input data set, together with a context and outputs a report on the value of data, comprising various scores and data profiles. The main modules and features of the DVC are:

- module for gathering and assessing context. In the latest version the collection and scoring of contexts can be performed in a secure manner, by using PSI protocols.
- data quality assessment module, comprising 6 data quality metrics.
- data utility assessment module, allowing for the usability of the data set with out-of-the-box algorithms for clustering, classification and regression.
- chance estimators and de-anonymisation risk analysis
- score aggregation and reporting module.

The testing of the platform was realised as part of the two versions of the demonstrator and the personal data trials as reported in Section 2.1.

3.1.3 Key findings

The initial survey revealed interesting attitudes of data-driven companies with respect to data and data value. While unsurprisingly, it showed that companies are aware that data is valuable both to themselves and to other stakeholders, it also revealed a clear reluctance to engage in business transactions involving data sets, amongst the main concerns cited being: legal and security uncertainties, lack of maturity of data markets, a lack of alignment between their current business model and engaging in transactions with data. The survey also revealed a clear necessity for a data valuation platform, and it was able to distill some expected requirements around the capacity to quantify data quality, data provenance, data compliance to standards, as well as a clear interest in ethics and compliance to data protection frameworks (e.g., GDPR).

Second, our research on data production and consumption ecosystems showed that the advent of data processing technologies and the development of data collection capabilities has triggered two paradigm shifts: i) a change in the production-consumption paradigm, according to which there is a clearer distinction between producers and consumers, their expectations and their views towards the same data asset, and ii) a change in the data exchange ecosystem, in which new roles are appearing (data brokers, platforms for data monetisation) and old stakeholders are still trying to find their role (governments) or need to be more fairly represented (private citizens, in the case of personal data). The difficulty of

assessing the value of data stems from these two shifts, both centered around the ubiquity of data and is best explained by what Shoshana Zuboff calls “surplus data”².

At the same time, our research on previous attempts to data valuation revealed other key attitudes towards data, many coming from reports about personal data monetisation platforms:

- pricing data is difficult and there is no consensus on how to do it well.
- there is a large discrepancy between how individuals self-valuate their data and how much they release it for, and such differences can be stark even between demographic categories.
- there is a discrepancy between the perceived risk related to sharing data and the willingness to do so. These last two points may also be illustrative of a lack of data culture in an increasingly data-centric society.

Other insights referred to different data properties and data types and their influence on data value:

- processed data tends to be more valuable than raw data.
- aggregation decreases value, but may help in dealing with privacy concerns.
- there seems to be an ordering in terms of the difficulty of acquisition: financial data is the most difficult to obtain, followed by behavioural data and finally demographic data.
- dynamic data tends to be more valuable than static data.
- infinitely shareable data leads to the loss of its value; this in fact is one of the key points related to valuating data as an intangible asset using the same type of economic models.

All these observations motivated the need for introducing the notion of contexts in which data is valued. Despite recognising the role of contexts, the literature of how to formalise and integrate them is even scarcer than that on data valuation. Our first version of the context formalisation combines ideas from research on mapping data properties to data value³, datasheets for data sets⁴ and nutrition labels for data⁵.

With respect to economic methods for data valuation, our research outlined the use of three general approaches: cost of investment, market (equivalent) pricing and impact assessment. Each of them has their advantages and limitations, which makes them context-dependent; unfortunately, their application was reported only in very specific use cases and they seem to be too generic to adapt to the many particularities arising from the contextual nature of data valuation.

One of the most important pillars supporting data value is data quality and a large portion of our research was dedicated to reviewing data quality frameworks, dimensions and metrics, their application in different contexts and how do they translate to data value. As opposed to the other research areas that we covered, data quality assessment benefits from decades of research, resulting in a plethora of mature metrics that can be adjusted to various assessment contexts. However, with no consensus on what are the key data quality dimensions, we followed recommendations from literature⁶ and implemented the six most common ones: accuracy, completeness, domain validity, format validity, uniqueness, and timeliness.

² Zuboff, S. (2019). *The Age of Surveillance Capitalism*. Profile Books Ltd.

³ Kannan, K., Ananthanarayanan, R., and Mehta, S. (2018). What is my data worth? From data properties to data value. <http://arxiv.org/abs/1811.04665>, last accessed Nov 30, 2021.

⁴ Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H., and Crawford, K. (2020). Datasheets for Datasets. <http://arxiv.org/abs/1803.09010>, last accessed Nov 30, 2021.

⁵ Holland, S., Hosny, A., Newman, S., Joseph, J., and Chmielinski, K. (2020). The Dataset Nutrition Label: A Framework to Drive Higher Data Quality Standards. In D. Hallinan, R. Leenes, S. Gutwirth, and P. De Hert (Eds.), *Data Protection and Privacy: Data Protection and Democracy* (pp. 1–26). Oxford: Hart Publishing.

⁶ Askham, N., Cook, D., Doyle, M., Fereday, H., Gibson, M., Landbeck, U., Lee, R., Maynard, C., Palmer, G., and Schwarzenbach, J. (2013). *The Six Primary Dimensions for Data Quality Assessment—Defining data quality dimensions*. DAMA UK.

Concerning aggregation for the purpose of reporting, we also followed recommendations from literature, which advise to take into account the complex, multi-dimensional nature of the concept and report it as such, as opposed to forcing a unique quantitative result.

3.1.4 Implications / Recommendations

1. There is a strong need for data value assessment and a lack of tools supporting it. With our DVC, data value assessment is here and will continue to improve in the years ahead.
2. Companies and private citizens can profit from data value assessment. If implemented and deployed correctly, this can lead to secure, trusted and fair data markets.
3. With data markets in place, DVC will increasingly help to automatize valuation of data from automatically extracted properties and defined contexts.
4. Data value assessment shouldn't be limited to an economic or financial assessment. In the case of personal data, its value assessment has the potential to educate citizens of our data-centric reality about the real implications of sharing their data. Moreover, this should give an impulse to legal experts and politicians to create adequate legal frameworks to deal with the pressure from data driven companies as well as legal and ethical challenges raised by data ownership, personal data transactions etc.

3.2 PSI Component

3.2.1 Objectives

The objective of this task was to develop a PSI protocol suitable for enterprise-scale data sets. This PSI protocol should enable business partners to capitalize on their data sets by analysing the combined data without sharing any confidential data. Further, the goal was to focus on concrete instantiations and implementations of the protocol. In the end, there should be a software library implementing the personal data use-cases for WP6.

3.2.2 Achievements

We have developed Psittacus (D5.12), which is a secure PSI software library. It allows two parties to find the intersection in their data privately (Figure 5). Thereby, neither party learns information from the protocol execution except for the elements in the intersection. The privacy of Psittacus is based on advanced cryptographic methods. There is no need for a trusted third party. Psittacus is a prime example of privacy by design (PbD). PbD is currently heavily encouraged by regulators in the EU. In particular, PbD is necessary to meet the General Data Protection Regulation (GDPR) requirements. To sum up, Psittacus is privacy-enhancing technology for data sharing complying with the GDPR.

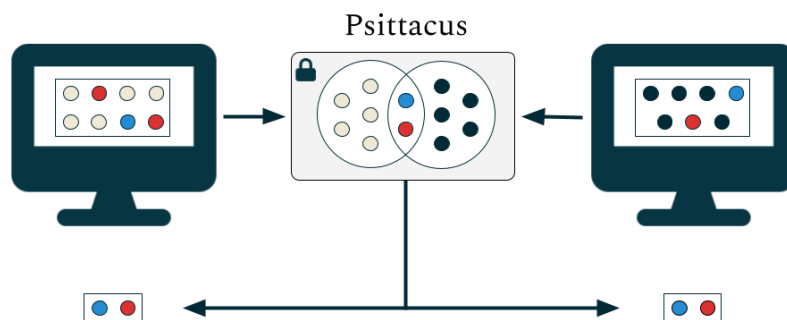


Figure 5: Private set intersection of two datasets using Psittacus.

3.2.3 Implications and Recommendations

Psittacus, our PSI solution, is suitable for enterprise-scale data sets. Although research on more efficient PSI solutions continues, we show that scalability is already possible in real-world experiments (one server in Frankfurt and one in Paris). More concretely, the intersection with Psittacus is practically instantly for datasets with up to 10k entries. Even datasets with up to one million entries are being processed in 1-2 minutes (details in D5.12).

In addition, we found that scalability is necessary but not sufficient. If we want to see widespread adoption from industry, it is equally important to have a business-friendly solution, i.e., one that integrates well into existing systems and workflows.

- PSI solutions should be more heavily encouraged by regulators, e.g., through the GDPR codes of conduct requiring PSI in specific settings.
- The community should not only focus on efficiency but also on developing more business-friendly PSI solutions with a high TRL.

3.3 Deanonymization Component

3.3.1 Objectives

De-anonymisation is the process of identifying individuals in a dataset that does not contain any personally identifiable information (PII). Full name, social security number, and address are PIIs because they directly identify individuals. However, there are attributes that do not directly identify individuals and that, when combined, can be used to identify individuals called quasi-identifiers (QIs). It has been shown that 87% of the U.S. population is uniquely identifiable by their combination of gender, date of birth, and ZIP code. It is also demonstrated how this information can be used to breach the privacy of individuals. Such privacy breaches can occur with different kinds of data. It has been shown that four spatio-temporal data points are enough to identify 95% of the individuals in a human mobility dataset consisting of one and a half million individuals.

There are only two tools exist that provide de-anonymisation risk analysis: ARX and X²R². ARX is the current state-of-the-art in anonymisation and de-anonymisation risk analysis, having been launched in 2012 and still being actively developed. However, the main limitation of these tools is that they only deal with (what we define as) tabular data, in which one record corresponds to one individual, and are not able to deal with complex, high-dimensional data, such as location data. Additionally, the visualisation of the de-anonymisation risks is limited. Note that, at the time of writing, the X²R² tool or its source code could not be found online, but its publication does not feature dealing with complex, high-dimensional data, or advanced visualisation techniques.

The objectives of Task 5.4 *De-anonymisation* are to provide de-anonymisation risk analysis tools that extend the capabilities of existing tools and cover their limitations above. In particular, the objective of this Task is to provide de-anonymization risk analysis tools that are capable to deal with data with complex structures than tabular data and to provide visualization of the data risk to raise awareness of data controllers about the de-anonymization risk in their data.

3.3.2 Achievements

Besed on a study of literature on (de-)anonymisation and the GDPR, we defined a 3-step procedure which data controllers should follow before sharing their datasets. We applied this procedure to the project's datasets provided by FNET. We showed that the defined procedure makes the data controllers aware of the de-anonymisation risks in their datasets and helps them in deciding the appropriate anonymisation measures.

For supporting data controllers, we developed three de-anonymisation risk analysis tools corresponding to the different types of FNET’s data: tabular, invoices, and aggregated:

Tabular data: We refer to a tabular dataset as a dataset where each line corresponds to one individual. Such is the case of the *Assets* table. The privacy threat in tabular datasets is that of an individual being de-anonymised through their QIs, which could lead to an adversary gaining knowledge about sensitive information of individuals (if such information exists in a dataset). We have developed a tool that reveals the likelihood of de-anonymisation by an adversary that already possesses information on the QIs of individuals, can designate the QIs critical for a de-anonymisation, and reveals the extent to which a dataset is de-anonymisable. Figure 6 depicts a snapshot of the tool we have developed. Each point in the interactive plot represents a unique combination of QIs, with the x-axis referring to the number of QIs in a combination and the y-axis referring to the probability of de-anonymisation, if an adversary possesses the information of those QIs.

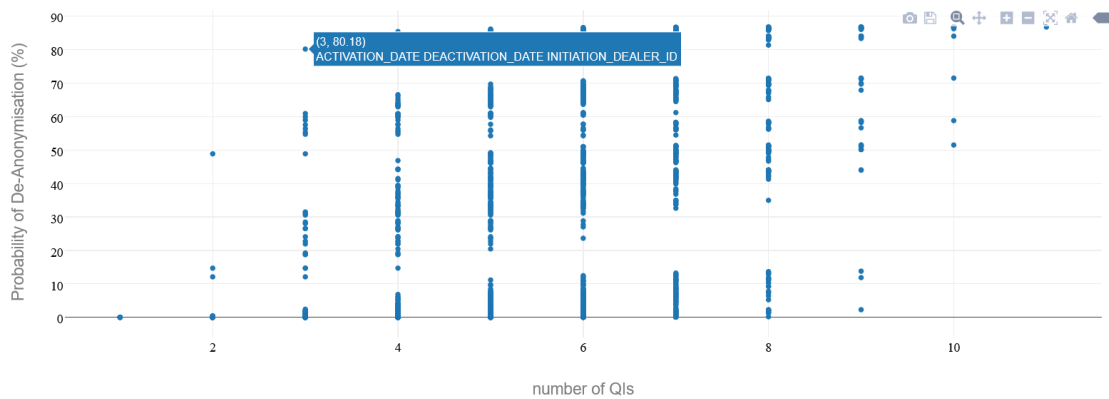


Figure 6: Tabular data de-anonymisation risk analysis

Invoices: The *Invoices* table is of different nature than *Assets*, since every individual has more than one invoice in the table and has different privacy threats. De-anonymising an individual from the exact invoice amount is highly unlikely, since this information is hard to be acquired by an adversary and it would rather be the sensitive information an adversary aims to find out. The privacy concern in this case is whether the information contained in the dataset in the *Invoices* table is sensitive. *SRs* is of similar nature; however, it does not contain any sensitive information.

To that purpose, we have developed a tool that visualises the risks of any given invoices dataset, taking as input 3 privacy parameters: **a.** the number of individuals, **b.** an invoice amount, **c.** a timeframe. Figure 7 depicts a snapshot of this tool. Each point in the plot represents a unique invoice amount (y-axis) at a specific time (x-axis). In Figure 7, a point is colored green if there are at least 2 distinct individuals (parameter **a**) having an invoice of ± 5000 (parameter **b**) within ± 1 month (parameter **c**); colored red, otherwise. This output provides two insights:

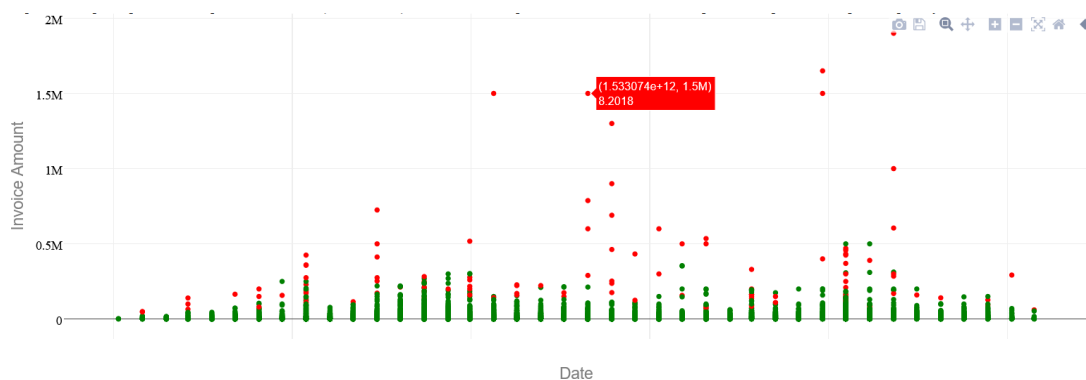


Figure 7: Invoices data risk analysis

Date aggregates: An aggregated dataset is a dataset that contains aggregate values referring to individuals. Such is the case of the viewership dataset, where certain columns correspond to a grouping of individuals (e.g. shares on Facebook). While the individuals cannot be identified through those aggregated values, the values of potentially sensitive attributes may be inferred when the aggregate values corresponding to individuals are low. In the case of the viewership dataset, there are no sensitive attributes, but if there were such, for example an attribute "sum of income in November of those who shared the video", then there would be a privacy breach if $shares = 1$.

To that purpose, we have developed a tool that visualises the risks of any given aggregated dataset, taking as input one privacy parameter: the number of individuals in an aggregation – k .

The **de-anonymisation risk analysis tool** is a web application with three modules, each corresponding to a type of data: tabular, invoices, and aggregated. It was developed in Java and uses Spring Boot as a web server. A prototypical user interface (UI) is found in its code repository, but in the WP6 demonstrator it comes with a proper UI.

We identified the following challenges of (de-)anonymisation for data sharing that the industrial and scientific community has to deal with.

- **Lack of awareness on (de-)anonymisation:** In general, laypeople that do not have a scientific or engineering background are not aware of the risks of de-anonymisation, and anonymity is viewed as simply removing the direct identifiers. The same situation, unfortunately, exists in the industry as well, as reported by popular news media.
- **Lack of detailed guidelines:** Privacy in data sharing is a complex issue and it should be studied in more detail and have a broader coverage in the regulations and guidelines. Even though WP29 and other authorities provide guidelines that help data controllers comply with the regulations, more details should be provided on anonymising datasets case-by-case, especially, on the balance between privacy and utility, and cases where even a low privacy guarantee results in a tremendous loss of information and value.
- **Lack of open-source tools from complex data:** We identify the need of reproducing the most important methods in the anonymisation literature and packaging them as open-source, easy-to-use tools. This challenge may be faced by researchers and developers reproducing existing and developing new anonymisation methods and making them open source.

3.3.3 Implications and Recommendations

The work on Task 5.4 *De-anonymisation* resulted in the following:

- **Raising awareness:** Laypeople that do not have prior knowledge on anonymisation view it as removing the PII's only and are not aware of the de-anonymisation risks in their datasets. De-anonymisation risk analysis tools can reveal and highlight the de-anonymisation risks in datasets and the extent to which they are de-anonymisable.
- **Compliance to GDPR:** As described previously, GDPR does not specify which privacy models are suitable in which cases and implies that the data controller should become aware of the de-anonymisation risks in their datasets. De-anonymisation risk analysis tools help in the compliance to GDPR since they report on how much datasets conform to privacy models and raise the awareness of the de-anonymisation risks.
- **Aiding in the anonymisation measures and their extent:** Deciding the anonymisation measures and their extent is the core challenge in the anonymisation process. De-anonymisation risk analysis tools, if designed properly, can help in this decision and its extent since they can reveal the distortion that is required for a dataset to comply to privacy models. Examples are described in the next section.

3.4 Lead-time Based Pricing

3.4.1 Objectives

Concerning Lead-Time Based Pricing (LTBP) in the Safe-DEED project, different perspectives on LTBP were analyzed. LTBP is a concept of a dynamic pricing strategy, which is aimed to lead to several benefits for both customer and supplier. A famous example for it is the aviation industry where tickets are getting more expensive the closer they get bought to the departure. In our hypothetical implementation, late orders (specifically orders that have been placed after the standard delivery time, which usually is around several weeks), will get priced using a price adder. During the project, different implementations of the LTBP use case were analyzed.

The overall goal of this project is the exploration of the concepts of Lead-Time Based Pricing (LTBP) and multi-party computing (MPC).

LTBP, a method of revenue management, gained major interest across several industries including the airline, service and process industry. The state of the art for LTBP in the semiconductor domain contains prices that are set with predefined lead times. However, this state of affairs disregards the performance reserves that actually exist today. Suppliers would often be able to stay below the contractual lead times if this is linked to a lead time dependent price in a systematic way. However, due to the pre-negotiated purchase prices and the lack of a platform that enables the dynamic exchange of pricing data, system-to-system order processes are missing out on these potentials for being fair to suppliers and customers.

3.4.2 Achievements

On the one hand, the application of Multi Party Computing on LTBP was developed and evaluated, on the other hand, multiple possible pricing options were simulated. A main focus for that was the exploration of different pricing algorithms for the LTBP implementation. Due to the lack of data that occurs because dynamic pricing is not being used yet, first of all, qualified synthetic data (QSD) needed to be developed, which simulated customer data, using the rare data that is existent. Then, several pricing algorithms ran on the different QSD sets to find out if there is one pricing algorithm that suits most assumptions about customer behavior best.

The LTBP Demonstrator. The data required for LTBP is very sensitive and when implementing this method of Revenue Management, it is essential that the privacy of this data is secured. With the use of secure multi-party computing, it is possible to provide both parties involved with the required security for Lead-Time Based Pricing decisions. We integrated new multi-party computing technology into a demonstrator of an order management platform. The demonstrator offers an easy and secure platform for exchanging sensitive data needed for LTBP while keeping the sensitive data of the supplier (see Figure 8) and the customer (see Figure 9) private and secure.

The usage of secure MPC technology can increase the trust in LTBP and thereby the acceptance of this method of revenue management. The usage of MPC is valuable for ensuring the privacy of the order data creating additional trust in secure LTBP.

The acceptance of the order platform and the different security measures will be determined in future projects. These security measures can be valuable for ensuring the privacy of the order data creating additional trust in secure LTBP.

Figure 8: Demonstrator Supplier View.

Figure 9: Demonstrator Customer View.

3.4.3 Implications and recommendations

The key findings of the project are that besides the quantitative aspect on evaluating different pricing algorithms, which was mainly the development of financial figures, qualitative aspects needed to be evaluated either. First of all, there was the ease of application and understanding for the customer. Since the price would always get newly calculated for every late order, it would not be understandable for the client to have these price changes. Therefore, a continuous pricing function is not useful for the customer but can be taken as an orientation of e.g. a pricing bucket function, which gives more transparency to the customer (see Figure 10).

Once hypothetically implemented in a suitable pricing algorithm, LTBP has a potential to not only generate more revenue for the supplier, it also stabilizes the complex supply chain. Due to long lead times, late orders occur often and lead to an unstable production since they are not planned from the beginning and therefore lead to shortages after the production. An incentive to order early enough would reduce these late orders since the customers need to place them early enough to get the best price in the end.

- The approach on the pricing algorithm is a key factor in success or failure of dynamic pricing and should therefore be respected in pricing policies.

- Not only the quantitative aspect is important so further research on customer behavior in presence of dynamic pricing is interesting.
- During the project we explored multiple algorithms that can be used for LTBP and analyzed their compatibility with MPC. During the trialing of the demonstrator interviewees stated that their understanding of the technologies highly increased. The Demonstrator can be used in the future to introduce experts within Infineon to LTBP and MPC.
- The privacy ensuring aspect of MPC technology can also be used in a different context within Infineon. During interviews it became clear that within Infineon there is an interest in using MPC to increase privacy also outside the field of LTBP.
- MPC technology could be used to reduce the bullwhip effect. The bullwhip effect occurs due to the high complexity of the supply chains in the semiconductor industry. Here MPC technology can be used in order to enable a privacy preserving and secure exchange of sensitive information that can increase the stability of the supply chain.

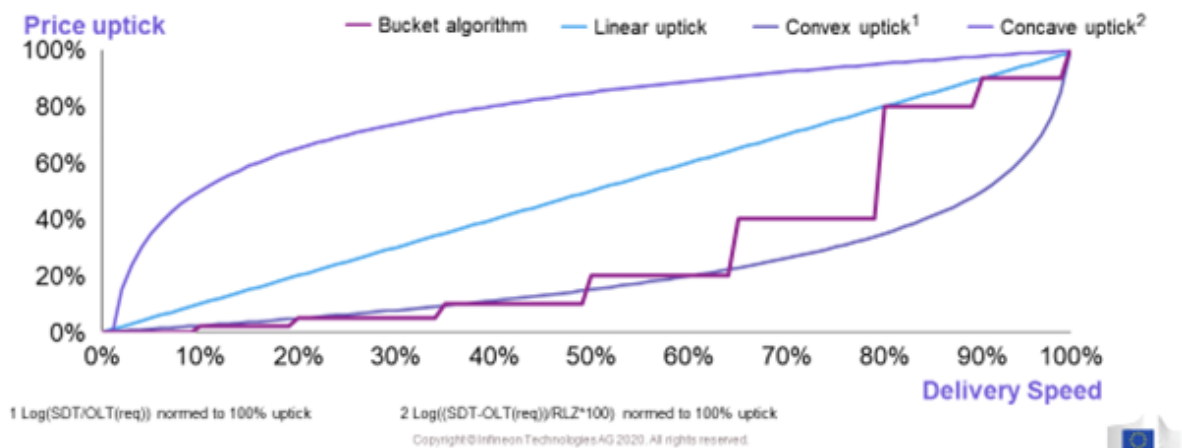


Figure 10: Defining the pricing algorithms.

4 Legal Aspects

4.1 Legal and Ethical Requirements

4.1.1 Objectives

The first task aimed at providing a high-level description of the EU legal frameworks applicable to the Safe-DEED project. The analysis took into account the legislative initiatives already implemented and those still in the approval stage at the time of writing. This task specifically considered the main requirements that should be taken into account by Safe-DEED partners in the design and deployment phase of the project. This task aimed at laying out the general legal/ethical frameworks applicable to the project in a first deliverable, after which the two use cases (processing of personal data and processing of non-personal data, respectively) were more closely scrutinized.

4.1.2 Achievements

This task has set out the main EU legal/ethical framework applicable to the purposes of the project. Firstly, the legal/ethical rules relevant for the exploitation of PETs were outlined. Specific focus was devoted to the GDPR and the E-Privacy Directive, as well as to ethical aspects involving the management of personal data and to the latest initiatives regarding the regulation of AI. Secondly, legal/ethical rules relevant for the development of new MPC methods were drawn up. We have specifically delved into the Free Flow of Non-personal Data Regulation and the Platform to Business Regulation. Additional emphasis was put on security and competition law aspects which may stem from

the current EU framework. Thirdly, an outline was made of relevant consumer protection laws. Furthermore, in two concrete deliverables, the legal/ethical requirements intended to support the activities of the Safe-DEED partners were outlined, providing partners with a clear and understandable overview of the measures they must implement to ensure compliance of their activities with all legal/ethical frameworks. As compliance with legal requirements is an on-going process, an enduring dialogue between KU Leuven and other Safe-DEED partners has been established and reinforced. The aim was to provide guidance in the identification of the technical and legal aspects that need to be implemented with appropriate compliance measures. This has resulted in the “Legal Q&A”, available on the Safe-DEED website.

Table 1: Illustrative excerpts of some data processing requirements, as outlined by WP3 as part of this task.

NON- PERSONAL DATA	Make data available for competent authorities	<ul style="list-style-type: none"> • Art 5 FFNPR • Art 6 FFNPR 	Make data available to competent authorities upon request	• DEPLOYMENT
	Draft of a Code of Conduct	<ul style="list-style-type: none"> • Art 6 FFNPDR 	To develop a self-regulatory code of conduct that each party has to comply with	• DEPLOYMENT
PERSONAL AND NOT- PERSONAL DATA	Preparation of data usage agreement	<ul style="list-style-type: none"> • Commission Staff Working Document – Guidance on sharing the private sector data in the European data economy. (COM (2018) 232 final) 	Prepare the draft of the agreement that complies with guiding principles and consider competition law restrictions	• DEPLOYMENT

4.1.3 Implications and Recommendations

For privacy and data protection law, the GDPR’s general principles, data controllers’ obligations and data subjects’ rights were outlined. Moreover, the E-Privacy Directive’s security aspects were touched on, as well as the general principles of the proposed E-Privacy Regulation. The various considerations relevant for the exploitation of PETs developed in the Safe-DEED project were presented. For the ethical guidelines, the EDPS’ Ethics Advisory Group 2018 Report (‘Towards a digital ethics’) was highly relevant, as were the General Ethics Guidelines for trustworthy AI. For the use of MPC within the project, it was found that the Free Flow on Non-Personal Data Regulation and the Platform-to-Business Regulation were mostly relevant. For security aspects, the project’s success hinges on compliance with the NIS Directive, the Cybersecurity Act, and the EU Encryption Framework. Concerning the latter, it was found that the ENISA Opinion Paper on Encryption, the Progress Reports on ‘moving towards an effective and genuine Security Union’ and the European Electronic Communications Code formed the main framework for the project. Lastly, this task found that for competition law and consumer law, Safe-DEED technical partners should at all times take notice of the TFEU and the Merger Regulation, and the Digital Content Directive. This task has also provided a detailed list of actions to be carried out by Safe-DEED partners in the deployment phase of the project, according to their role, to comply with the identified instruments. The task has also shown that the legal framework regarding this use case is

developing fast, which has translated in a conscientious follow-up by WP3. This has *inter alia* resulted in the aforementioned Safe-DEED Legal Q&A.

This task aimed to make sure that the other work packages were aware – and respected – the applicable legal and ethical framework in the project, and thus may serve as a legal/ethical guidance for future research projects of similar nature.

4.2 From Impact Assessment to Value Assessment

4.2.1 Objectives

This task has explored the emerging body of literature on impact assessments in the digital realm. Moving forward from the structure of the Data Protection Impact Assessment, (DPIA), this task has explored the potential benefits that might arise from an assessment that takes into account not only privacy and data protection normative aspects, but also ethical values and social norms that are usually not considered in a DPIA. As a result, the deliverable aims to move forward from a Data Protection Impact Assessment to a more comprehensive Data Valuation Impact Assessment. It was believed that such a “DVIA”, considering ethical, normative and social instances, should allow for a closer alignment between company’s procedures and those factors considered by judicial and administrative authorities when assessing a certain data processing activity. Second, knowing that an entity exchanging data in a data-market had to carry out a DVIA might represent an attractive factor for entities interested in joining the data-market, enhancing economic positive outcomes as a result of their overall compliance with legal and ethical norms and principles. This task therefore scrutinized the scope of such a DVIA, and its relevance for the Safe-DEED project as a whole.

4.2.2 Achievements

WP3 has assessed the new notion of Data Valuation Impact Assessment. Fundamental research was conducted to assess how to move from DPIAs to an assessment that also considers the value of data. In doing so, the non-economic considerations of data valuation were outlined, after which the main problems in assessing value in a data marketplace context were described. WP3 has scrutinized five common methods to assign economic value to personal data. Moreover, the data ownership debate and property law issues (including IP) were assessed. Based on these (non-)economic insights, a Safe-DEED Data Protection Value Impact Assessment was put forward. In a first place, the primary legislative, ethical and economic considerations were put forward. Additional attention was devoted to fundamental rights and values, risks and mitigation measures. Finally, the Safe-DEED Data Protection Value Impact Assessment was demonstrated with a practical ‘Covid-19’ use case.

4.2.3 Implications and Recommendations

Based upon the developed Safe-DEED Data Protection Value Impact Assessment, it was found that the use of MPC or differential privacy as cryptographic measures will, first of all, guarantee a substantial reduction of risks linked to the processing of personal data as requested by the GDPR. To achieve such purpose, compliance of such cryptographic measures with the EU framework will be ensured following the measurable criteria listed by the Court. Multi-party computation and other cryptographic measures should ensure the impossibility for third parties to identify data subjects from the available information. Concretely, the identification process should require third parties a disproportionate effort in terms of time, cost and workforce applicable to the whole data process. Besides, regardless of the exchange of personal data occurring between the two or more entities, this task found that the used cryptographic procedures applied to such data make the risk of identification insignificant. To conclude, the use of multi-party computation and other specific cryptographic measures will support activities of parties involved in data processing activities in complying with the EU law, respect fundamental rights and individual ethical values, with a consequent overall positive outcome for society.

- Making use of a DVIA instead of a mere DPIA allows for a more comprehensive legal/ethical a priori assessment of data processing.
- The Safe-DEED Data Protection Value Impact Assessment may serve as an example for further research projects.
- The particular use of MPC encryption has passed the DVIA check.

4.3 Fostering Trust in Data Markets

4.3.1 Objectives

This task has built on the insights gained in the previous one, with the broader aim of engendering more trust (individual and societal) in data markets. A selection of concrete issues that are particularly problematic in this context have been investigated. The research done within this task has combined traditional legal desk-research (focusing in particular on hard law, official policy documents, and reputable scholarship) with surveys and semi-structured interviews with project stakeholders, in order to formulate a more grounded perspective on the key challenges (and solutions) for generating trust.

4.3.2 Achievements

Within this task, WP3 has formulated a definition of trust that can be used as aim within the project, based on a literature review of the relevant legal/ethical literature. This concept of “e-trust” has been scrutinized extensively in both D3.6 and D3.7. Besides, we have analyzed the ethical and normative tensions brought about by the commodification of (personal) data, particularly in light of the values that underlie the European identity (i.e. the respect of human dignity, freedom, democracy, equality, security, the rule of law). The outcome has resulted in a better articulation of these tensions and identifying lines that should not be crossed. In addition, we have identified and investigated how the developed technology can foster rather than undermine normative and ethical values. In particular, the trade-off has been investigated between on the one MPC encryption and on the other hand the (in)ability to exercise data subject rights. Finally, we have assessed in what ways the use of codes of conduct and/or MPC encryption may foster e-trust. As part of this task, we have also developed a teaching module of 90 minutes, aimed at students in all relevant disciplines. This specific accomplishment has been more extensively outlined in the next part.

4.3.3 Implications and recommendations

A definition of e-trust has been put forward, which goes beyond the constraints of the Safe-DEED project. The main antecedents and consequences of trust in the data market context have been outlined, and two means to foster trust (MPC encryption and codes of conduct) have been provided.

- Develop an autonomous European definition of e-trust, for which the Safe-DEED research may provide a solid basis.
- Promote the use of MPC encryption and codes of conducts to foster trust in data marketplaces, for which the Safe-DEED research may equally be a solid foundation.

4.4 Syllabus for Teaching Module

4.4.1 Objectives

This task focused on the development of a teaching module for students from relevant disciplines, to familiarize them with Safe-DEED’s research objectives. The syllabus targeted two audiences: (I) researchers within and external to the consortium; and (II) students from all relevant disciplines. Given these two groups’ different levels of expertise, it has been opted to write an accessible syllabus, ideally understandable for those with no prior knowledge of various subject matters

4.4.2 Achievements

WP3 has created an extensive 154-page syllabus on the Safe-DEED project. Concretely, the syllabus consisted of three main parts. The first part provided an introduction to the various EU initiatives regarding the advancement towards a European data-driven economy. A second part dealt with the ethical and legal considerations relevant in the Safe-DEED project. A final part discussed a number of challenges to the project. The first part consisted of two chapters, respectively serving as an introduction to the “value of data” and to “data marketplaces”. The second part comprised three chapters, correspondingly dealing with ethical guidelines, the legal framework on the processing of personal data and the legal framework on the processing of non-personal data. These chapters thus mainly hinged on the insights from the previous tasks in WP3. The final part consisted of four chapters, discussing the issues of data valuation, organizational trust, and the technical and legal challenges to the use of MPC encryption. The syllabus has been distributed internally and may be used as a support tool in future occasions. In addition, each chapter has been turned into a video lecture as well. Besides the syllabus, WP3 has thus gone a bit further, and has created nine video lectures, recorded in a professional studio, and edited accordingly. Nine research experts from KU Leuven CiTiP have each presented a lecture. The combination of video lectures and a written syllabus should allow for the most engaging teaching module, familiarizing students with the Safe-DEED project and the general research areas the project relates to.

Table 2 : The contents of the teaching module.

Part I	
The advancement toward a European data-driven economy: Introduction	
Chapter 1	The Value of Data
Chapter 2	Data Marketplaces
Part II	
The advancement toward a European data-driven economy: Interdisciplinary Considerations	
Chapter 3	Ethical Guidelines
Chapter 4	The Protection of Personal Data
Chapter 5	The Protection of Non-Personal Data
Part III	
The advancement toward a European data-driven economy: Challenges & Opportunities	
Chapter 6	The Valuation of Data
Chapter 7	Organizational Trust
Chapter 8	Secure Multi-Party Computation Encryption (MPC)
Chapter 9	MPC: Legal Questions and Answers
Conclusion	

Table 3: Description and goals of the teaching module.

Lectures series: overview & goals

<p>Why is this subject matter relevant?</p> <ol style="list-style-type: none"> 1. The value of data 2. Data marketplaces 	}	Introduction
<p>What is the current legal/ethical framework?</p> <ol style="list-style-type: none"> 3. Ethical Guidelines 4. Legal Framework I: personal data 5. Legal Framework II: non-personal data 		State of affairs
<p>How to overcome impediments?</p> <ol style="list-style-type: none"> 6. Data valuation 7. Organizational Trust 8. MPC: introduction to encryption 9. MPC: legal issues 	}	Safe-DEED research goals:
		<ol style="list-style-type: none"> 1. Data valuation protocol: empowering all data owners 2. Fostering trust in data markets, encompassing a sound legal/ethical framework 3. Enabling large-scale secure multi-party computation

4.4.3 Implications and Recommendations

The syllabus has *inter alia* been used during a presentation for students from interdisciplinary backgrounds at the KU Leuven. As hoped, this teaching module helped translate abstract concepts into tangible, easily understandable subject matters, as acknowledged by the students.

As a recommendation for future research, it is advised to incorporate a teaching module in further projects. Such a module helps break down the project in its key component and increases visibility amongst external actors.

5 Conclusion

The work in Safe-DEED has researched and developed significant contributions for a safe and commercially beneficial exchange of data. Safe-DEED achieved this by tackling the business, technological, and legal perspective related to this area. Safe-DEED developed artifacts and tools for each of the three perspectives, which are detailed in this report. The combination of these perspective lays a solid basis for the successful establishment of a data-driven economy within Europe. Safe-DEED reached both a scientific and a business-related audience, which both benefit from the creations of the project.