Increasing Trust for Data Spaces with Federated Learning

Susanna Bonura, Davide Dalle Carbonare, Roberto Díaz-Morales, Ángel Navia-Vázquez, Mark Purcell and Stephanie Rossello*

Abstract Despite the need for data in a time of general digitisation of organizations, many challenges are still hampering its shared use. Technical, organisational, legal and commercial issues remain to leverage data satisfactorily, specially when the data is distributed among different locations and confidentiality must be preserved. Data platforms can offer "ad hoc" solutions to tackle specific matters within a data space. MUSKETEER develops an Industrial Data Platform (IDP) including algorithms for federated and privacy-preserving machine learning techniques on a distributed setup, detection and mitigation of adversarial attacks and a rewarding model capable of monetizing datasets according to the real data value. The platform can offer an adequate response for organizations in demand of high security standards such as industrial companies with sensitive data or hospitals with personal data. From the

*All authors have contributed equally and they are listed in alphabetical order:
Susanna Bonura
Engineering Ingegneria Informatica SpA - Piazzale dell'Agricoltura, 24 00144 Rome Italy, e-mail: susanna.bonura@eng.it

Davide Dalle Carbonare
Engineering Ingegneria Informatica SpA - Piazzale dell'Agricoltura, 24 00144 Rome Italy, e-mail: davide.dallecarbonare@eng.it

Roberto Díaz- Morales
Tree Technology - Parque Tecnológico de Asturias, CL Faya 18 33428 Llanera, Asturias, Spain e-mail: roberto.diaz@treetk.com

Ángel Navia-Vázquez
University Carlos III of Madrid, Avda. Universidad, 30, 28911-Leganés, Madrid, Spain. e-mail: angel.navia@uc3m.es

Mark Purcell
IBM Research Europe, IBM Campus, Damastown Industrial Estate, Mulhuddart, Dublin 15, Ireland. e-mail: markpurcell@ie.ibm.com

Stephanie Rossello
KU Leuven Centre for IT & IP Law - imec, Sint-Michielstraat 6 bus 3443, B-3000 Leuven, Belgium e-mail: stephanie.rossello@kuleuven.be

architectural point of view, trust is enforced in such a way that data has never to leave out its provider's premises thanks to federated learning. This approach can help to better comply with the European regulation as confirmed from a legal perspective. Besides, MUSKETEER explores several rewarding models based on the availability of objective and quantitative data value estimations, which further increases the trust of the participants in the data space as a whole.

Keywords

Industrial Data Platform, Federated Learning, Data Spaces, Data Value Estimation, GDPR, Trust, MUSKETEER.

1 Introduction

Thanks to important advances in the recent years, machine learning has led to disruptive innovation in many sectors, for instance industry, finance, pharmaceutical, healthcare or self-driving cars, just to name a few. Since companies are facing increasingly complex tasks to solve, there is a huge demand for data in these areas. However, the task can be challenging also because it does not only depend on the companies themselves. For example, the healthcare sector has started to use machine learning to detect illnesses and support treatments. However the necessity to use appropriate datasets, composed of data from enough patients suffering a given illness and related treatments, can be hindered by the limited number of patients that can be found in the historical medical records of a single hospital. This issue could be solved if people and companies were given an adequate way to share data tackling the numerous concerns and fears of a large part of the population, that form barriers preventing the development of the data economy:

- Personal information leakage: the main concern of the population is the fear about possible information leakage. However companies, in order to run their analysis need digital information such as images or healthcare records containing very sensitive information.
- Confidentiality: A company can benefit from jointly created predictive models, but the possibility of leaking some business secrets in the process could lead to disadvantage this company vis-à-vis its competitors.
- Legal barriers: governments, in order to regulate the use of data, have defined legal constraints that impact the location of data storage or processing.
- Ownership fear: Some data could be very valuable. Some companies and people could benefit economically from providing access to these data. But digital information could be easily copied and redistributed.
- Data value: data owners could provide data with low quality, or even fake data, so that, effectively, there would only be limited value for other partners in using

this data. Hence, a key challenge is to provide mechanisms for monetising the real value of datasets and avoiding a situation where companies acquires a dataset without information about its usefulness.

In order to remove these barriers, several technologies have emerged to improve the trustworthiness of machine learning. Aligned with priorities of the Big Data Value Association Strategic Research, Innovation and Deployment Agenda such as identifying strong and robust privacy-preserving techniques, exploring and engage a broad range of stakeholder's perspectives or providing support in directing research efforts to identify a smart mix of technical, legal, ethical and business best practices and solutions[33], the MUSKETEER project developed an Industrial Data Platform including algorithms for federated and privacy-preserving machine learning techniques, detection and mitigation of adversarial attacks, and a rewarding model capable of monetizing datasets according to the real data value. We will show in this chapter how these challenges are tackled by the platform architecture but also how these techniques improve the compliance with certain principles of the EU regulation, and eventually the necessary data value estimation needed to balance the contributions of the platform stakeholders creating incentive models. Ultimately, the contributions from MUSKETEER help to increase the level of trust among participants engaged in federated machine learning.

2 Industrial Data Platform, an architecture perspective

The MUSKETEER platform is a client-server architecture, where the client is a software application that in general is installed on-premise and run at every end user site. This software application is named the client connector in the MUSKETEER taxonomy. On the server side of MUSKETEER, resides the central part of the platform that communicates with all the client connectors and acts as a coordinator for all operations. Users of the MUSKETEER Industrial Data Platform, interact with the client connector installed on their side and that client will communicate with the server to perform several actions on the platform. In Fig. 1 we show the topology of a MUSKETEER installation.

Often in client-server architectures, the means of communication between remote modules is direct, i.e. each module has a communications component that essentially presents an outward facing interface that allows remote modules to connect. This is usually accomplished by publishing details of an IP address and port number. For operations beyond the local area network, this IP address must be Internet-addressable. The actual implementation of the communications can vary: examples are direct socket communications, REST, grpc etc.

There are a number of security and privacy challenges to these traditional approaches, that the MUSKETEER architecture addresses. Allowing direct connections from the outside world is a potential security risk, both from a malicious actor perspective but it is also susceptible to man-in-the-middle attacks. These attacks often target known vulnerabilities in the host operating system or software stack. It

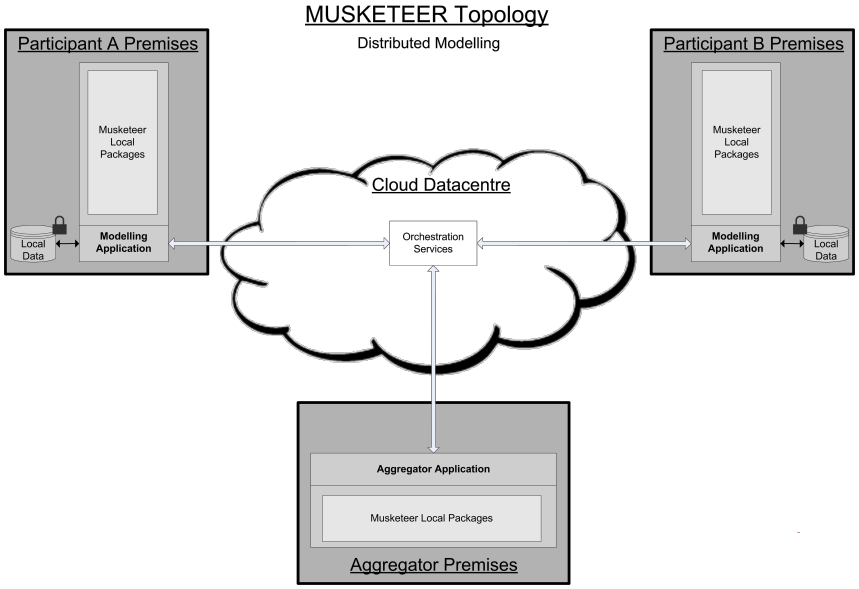

Fig. 1 MUSKETEER Topology.

is also possible for these attacks to operate bidirectionally, whereby a benign entity might be attacked, and potentially sensitive data may be at risk. Furthermore, firewall policies in different organisations may not permit Internet-based traffic, further restricting platform use.

In the MUSKETEER architecture, there are no direct connections between participants, or aggregators. All interactions occur indirectly through the MUSKETEER central platform, as depicted by *Orchestration Services* in Fig. 1. The central platform acts as a service broker, orchestrating and routing information between participants and aggregators. In this way, only the connection details for the broker is made available, with all other entities protected from direct attack. Such an architecture slightly differs from current reference models promoted by International Data Spaces Association (IDSA) and the Gaia-X initiative. Although largely aligned with most of the concepts included in these models (containerization, secured communication, etc.), there is an important difference with the privacy by design dimension included in the MUSKETEER architecture. Both IDSA and Gaia-X models rely on mutual trust between participants in the same ecosystem, while participants in MUSKETEER never have direct interactions.

2.1 Client connector

The Client Connector is a software component that is installed at the client site, as depicted by *Musketeer Local Packages* in Fig. 1. Within the Client Connector, two types of architectures have been designed: the first one implements a Cluster mode, the second one implements a Desktop mode.

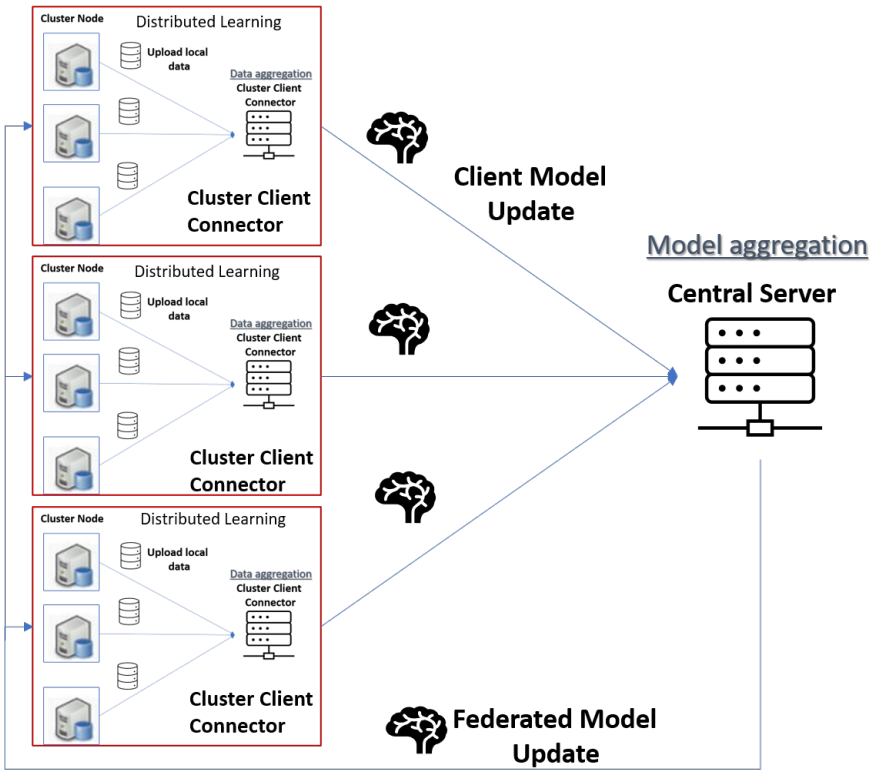

Fig. 2 Federated Machine Learning through the Cluster Client Connector.

The Cluster Client Connector (Fig. 2) supports the storage and the processing of large data sets before applying the machine learning federation, through horizontal scalability and workload distribution on multiple nodes of the cluster. Within a Cluster Client Connector, distributed machine learning algorithms have the potential to be efficient with respect to accuracy and computation: data is processed in parallel in a cluster or cloud by adopting any off-the-shelf efficient machine learning algorithm (e.g. Spark's MLlib). In this way we combine the benefits of distributed machine learning (inside the Client Connector) with the benefits of federated machine learning (outside the Client Connector).

The Desktop Client Connector (Fig. 3) is used when data is collected in a non-centralized way and there is no need to use a cluster to distribute the workload, both in terms of computing and big data storage. Anyway, the Desktop version could also leverage GPUs for the training process, enabling the processing of a large amount of data in terms of volume. Finally, the Desktop Client Connector can be easily deployed in any environment thanks to the use of Docker in order to containerize the Client Connector application. Docker containers ensure a lightweight, standalone and executable package of the software that includes everything needed to run the Desktop Client Connector: operating system, code, runtime, system tools, libraries and settings. The are also quite secure since it is possible to limit all capabilities except those explicitly required for any processes. (https://docs.docker.com/engine/security/).

Moreover extra layers of security can be added by enabling appropriate protection systems like AppArmor (https://packages.debian.org/stable/apparmor), SELinux (https://www.redhat.com/it/topics/linux/what-is-selinux), GRSEC (https://grsecurity.net/), so enforcing correct behavior and preventing both known and unknown application flaws are exploited.Finally, the Docker Engine can be configured to run only images signed using the Docker Content Trust (DCT) signature verification feature.

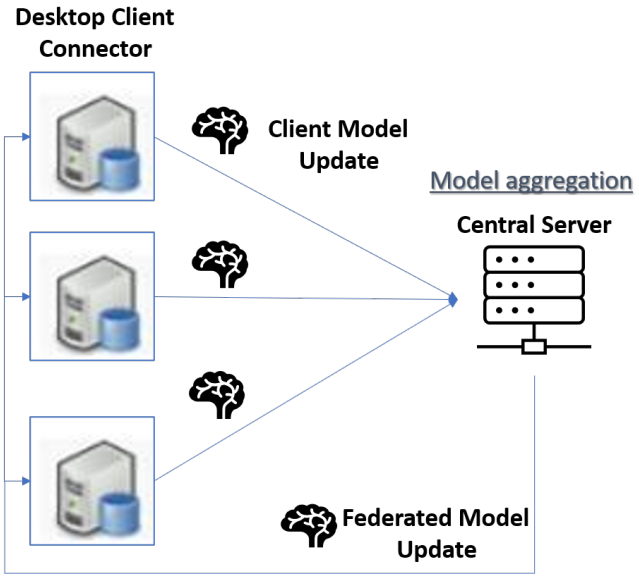

Fig. 3 Federated Machine Learning through the Desktop Client Connector.

In this way the whole Desktop Client Connector application can be easily deployed in a secure sandbox to run on the host operating system of the user.

2.2 Micro-services

For a viable federated learning platform, trust in the platform is an important requirement. This trust includes privacy protection for sensitive data, which remains on-premise, but also for platform user identities and communications. Ideally, no given user should be able to discover the identity or geographic location of any other user. Additionally, threats from traditional cyber-security attacks should be minimised.

The MUSKETEER server platform, depicted by *Orchestration Services* in Fig. 1, is a collection of cloud-native micro-services. These micro-services manage the life cycle of the federated learning process, using underlying cloud services such as a relational database, cloud object storage and a message broker.

By employing a brokered architecture, the MUSKETEER server platform enables outbound-only network connections from platform users. Users initiate connections to the platform, and do not need to accept connections. This ensures that users are not required to present Internet-facing services, having open ports readily accessible by external, potentially malicious actors. Additionally, all users must register with the platform, by creating a username/password combination account and all communications use at least TLS 1.2, with server platform certificate validation enabled.

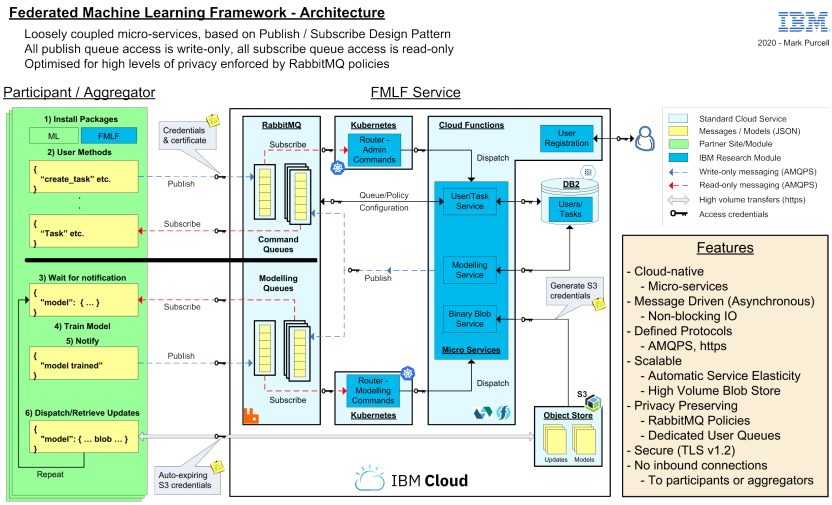

Fig. 4 MUSKETEER micro-services - from [26]

Once registered with the MUSKETEER server platform, each user is assigned a dedicated private message queue, which is read-only. This ensures that only the server platform itself can add messages to the queue but also, that only the assigned user has the appropriate privileges to view the contents of their queue. As the server platform

is broker based, the client connector simply invokes the appropriate procedure to subscribe to the assigned user queue.

As shown in Fig. 4 an important function of the server platform is the routing of messages between participants and aggregators, and how the micro-services interact to achieve this. For example, when an aggregator starts a round of training, an initial model may be uploaded to the platform's object storage. During this process, the aggregator obtains write-only privileges to a specific storage location for that model. Upon completion of the upload the aggregator publishes a message to initiate training, with an included checksum for the model. The platform receives this message and routes it to the queues of multiple users who are part of the federated learning task. Read-only privileges to download the aggregator's model are generated and appended to the message. Multiple participants receive these messages in parallel. They download the model, verify the checksum and start local training, all via the Client Connector. Upon completion each participant performs a similar operation to the aggregator and ultimately, all participant model updates are routed back to the aggregator for model fusion. This routing is deployed within a Kubernetes cluster, leveraging its high-availability features for an always-on, responsive system.

During the fusion process, the aggregator may employ a data contribution value estimation algorithm. Such an algorithm may identify high value contributions, and potentially assign a reward to the originating user, promoting a federated learning data economy. The server platform supports this by providing the capability to the aggregator to store information pertaining to the data value and potential reward. This is discussed in more detail in section 4.

By providing this capability, the server platform is in fact recording each step of the federated learning process. The combination of the recordings at each step, by the end of the federated learning process, enables a view of the complete model lineage for the final model. This lineage includes details such as updates provided per user, when, and of what value.

This architecture is instantiated for use in the MUSKETEER project. The server side (micro-services) is also integrated with *IBM Federated Learning* [22] and is available in the community edition [14]. The community edition supports multiple connection types, one of which is a HTTPS based connection, using REST, which requires IP addresses to be supplied to participants and aggregators. As previously discussed, there are a number of potential security issues with this approach, which the inclusion of the MUSKETEER option alleviates. Other federated learning platforms also exist, many of which display similar potential security issues due to the direct communication mechanisms employed.

So far we have described the technological means used to increase the trust of the user on the platform, basically focusing on data/communications security aspects and data confidentiality protection provided by the federated learning approach. In what follows we provide a legal perspective about the trust required in any data space by further explaining the regulatory data protection (compliance with GDPR principles). Finally, we will focus on the description of several data valuation mechanisms potentially leading to objective credit assignment and reward distribution schemes that further increase the end user trust on the data space operation.

3 Industrial Data Platform, a legal perspective

3.1 The broader policy context

Driven by the significant benefits that the use of Big Data analytics technologies (including machine learning) can have for our society, the European Union ("EU") has in the past decade taken several steps towards creating favorable conditions for what is calls a "thriving data-driven economy" [6] and a "common European dataspace" [7]. Key in these steps is the objective to foster access to and availability of large datasets for re-use for innovation purposes [9]. This is confirmed in the most recent Communication from the European Commission a "European Strategy for Data", where the Commission announces its intention to establish "EU-wide common interoperable data spaces in strategic sectors" ([10], p. 16). These spaces, the European Commission goes on, will include "data sharing tools and platforms" ([10], p. 17).

3.2 Data sharing platforms

Industrial data platforms were already mentioned by the Commission in its earlier guidance on the sharing of private sector data ([8], p. 5). In the aforementioned guidance, the Commission identifies industrial data platforms as one of the modes through which data can be shared among businesses and it describes these as "platforms dedicated to managing regular data interactions with third parties [and which] offer functionalities when it comes to data exchange [. . .] storage inside the platform and [. . .] additional services to be provided on top of the data (based on data analytics)."([8], p. 5).

In academic literature, ([28] p. 10) similarly describe data sharing platforms as entities providing "the technical infrastructure for the exchange of data between multiple parties". These scholars discuss several core functions of data sharing platforms and identify the "creation and maintenance of trust [among data users and data suppliers]" as one of their key functions ([28], p. 14). Indeed, they point out that, in order for the platform to achieve its main goal which is to match suppliers of data with users thereof, it is essential that suppliers trust that the data they supply will not be used illicitly and that users trust that the data supplied is fit for use ([28], pp. 13–14). As correctly remarked by these scholars, technology can be a key enabler for trust among users and suppliers of a data platform ([28], p. 17).

Aside from a possible lack of trust in the data, users and suppliers thereof, there may be legal reasons inhibiting the sharing of data among businesses. Crucially, when it comes to the sharing of personal data among businesses, the latter will often qualify as a processing of personal data falling under the scope of application of the General Data Protection Regulation ("GDPR"). Although the GDPR does not

prohibit the sharing of personal data among businesses as such, it does impose a number of conditions under which such sharing is allowed to take place.

3.3 Federated learning as a trust enabler: some data protection considerations

Federated learning has recently been emerging as one of the technologies aimed at overcoming some of the trust and, more specifically data protection concerns, related to the sharing of personal data. Indeed, federated learning differs from traditional centralized machine learning paradigms, since it does not require that the raw data used to train a machine learning model are transferred to a central server for the training to occur. Instead, under the federated learning paradigm, the machine learning model is trained locally, i.e. on the premises of the data suppliers, under the coordination of a central server. Therefore, under a basic federated learning process, only the local updates to the machine learning model leave the premises of the data suppliers and are sent to the central server for aggregation.

As implicitly recognized by several data protection authorities [2, 4] and the German Data Ethics Commission ([5], p. 120) federated learning can facilitate compliance with some principles of the GDPR. Indeed, as pointed out by the Norwegian Data Protection Authority, federated learning helps reducing the amounts of data needed for training a machine learning model ([4], p. 26). Therefore, if the training data qualifies as personal data, federated learning can help complying with the principle of data minimization set forth in article 5.1 (c) GDPR. This principle requires personal data to be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed". Moreover, since under the federated learning paradigm the training data is not transferred to a central server, the possibilities of such data being re-purposed by that server are also reduced. If the training data qualify as personal data, this means that federated learning could also facilitate compliance with the principle of purpose limitation set forth in article 5.1.(b) GDPR. This principle requires personal data to be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes [. . .]". Federated learning can hence be considered as a technique that helps implementing the principle of data protection by design, contained in article 25.1 GDPR. This principle requires controllers of personal data to "[. . .] implement appropriate technical and organizational measures [. . .] which are designed to implement data-protection principles, such as data minimization, in an effective manner [. . .]".

Despite the advantages that federated learning presents from a data protection perspective, it is not, as such, a silver bullet. We name some of the reasons for this. First, as also remarked by [2], the updates that data suppliers share with the central server could, in certain cases, leak information about the underlying (personal) training data to the central server or a third party ([23], para. 1.2). It is hence important to combine federated learning with other privacy preserving technologies, such as Multi-Party

Computation, differential privacy ([21], p. 11) and homomorphic encryption ([32], pp. 3–4). Second, "federated learning has by design no visibility into the participants local data and training" ([1], para. 1). This may render federated learning vulnerable to (data and model) poisoning attacks by training participants [17], which could, in turn, in some instances, impair the performance of the final machine learning model. Therefore, the use of federated learning may require an increased attention to not only technical but also organizational accountability measures. The latter may include a careful due diligence investigation into the training participants' compliance with the GDPR (and other relevant legislation) and envisioning contractually binding protocols specifying (among others requirements mentioned in the aforementioned EC Guidance on sharing of private sector data [9]).

Another key point to consider is about the quality requirements the training data should meet in light of the purpose of the final machine learning model and the population to which it will be applied. To this purpose, we will describe in the next section several data value estimation approaches that can be used to assess the quality of the data provided by each participant, so that the platform is ultimately able to reward every participant proportionally to the contribution to the final model. The availability of such data value estimations is key to the deployment of a true data economy.

4 Industrial Data Platform, objective Data Value Estimation for increased trust in Data Spaces

As already mentioned, another key requirement for a secure industrial data platform is to measure the impact of every data owner on the accuracy of the predictive models, thus allowing to monetize their contributions as a function of their real data value.

Today data has become the new gold, as it serves to power up advanced artificial intelligence (AI) models that form the core of an unlimited number of highly profitable processes, ultimately generating a potentially enormous business value. The importance of collecting large amounts of data as a way to obtain increasingly complex (and therefore accurate) AI models without the problem of overfitting (that is, complex models that perform well in the presence of input patterns never seen before) is out of the question.

For example, everywhere we are witnessing a struggle to capture as much information as possible from users in the context of mobile applications, to be used or resold for different purposes without any reward for data producers. In this well-known example the users give their consent (very often inadvertently) for their data to be used by third parties when they install and accept the terms and conditions of a certain application. A fairer scenario would be the one where users[1] are aware

[1] In what follow we will refer as "user" to any entity, either person or organization, that has some data of potential interest to a given process.

of their potential valuable data and agree to share it hoping to receive some compensation in return. It is currently debated that users should be paid for their data in a fairly direct way to foster the data exchange and ultimately improve many AI models. Many economists and politicians believe that data should be treated as an asset, with the possibility of protecting its specific use by third parties and the right of the data owner to sell it for different purposes, like any other “physical" good [31]. In economic terms, data is "non-rival" in the sense that it can be unlimitedly used multiple times for different purposes, unlike other physical goods, which can only be used once [16]. The current situation tends to be the opposite of the desired one, since in most cases large companies accumulate and have the rights over an increasing amount of data, to the detriment of the users who generated them.

The concept of an ideal data market has been studied in [16] where different alternatives (companies own data, people own data, data sharing is not allowed) have been compared against an optimal economic model administered by a benevolent ruler. As a conclusion of this research, it appears that the situation closest to the ideal reference model is the one in which users handle their own data. On the other hand, the case (more common today) in which companies own the data, the privacy of the users is not respected and the data is not shared efficiently with other companies. Finally, when data is not shared at all, economic growth tends to come to an end. Therefore a reasonable approach would be to allow users to retain the ownership and control over their data, and get a revenue whenever they contribute to any machine learning or AI model. The question still to be answered is how to adequately estimate that reward.

As discussed in [15] there are several families of pricing (rewarding) strategies, such as "query-based pricing", which sets the price according to the number of data views [19], ‘data attribute-based pricing‘" which fixes prices according to data age or credibility [13], and “auction-based pricing" which set prices based on bids among sellers and buyers [20]. The aforementioned methods, although potentially useful in certain contexts, have a significant drawback, in the sense that prices (rewards) are set independently of the task to be solved or of the actual utility of the data for the model to be trained. In what follows we will restrict ourselves to the data value concept that is linked to a real value for a given task, usually the training of a Machine Learning or AI model.

This data value estimation process is of great interest in a wide range of scenarios with different data granularity. On the one hand we may have situations where every user provides a unique training pattern (for example, a person offers data from the clinical record) and a potentially very large number of participants is needed to train a model (millions of people?). On the other side, we have scenarios where a reduced number of entities (organizations, companies, groups) offer a relatively large amount of data (e.g., several companies try to combine their efforts to improve a given process by joining their respective accumulated experience). The first type of scenarios can be associated with the concept of a Personal Data Platform (PDP), where users are individuals who offer their own data for commerce. This is the kind of scenario illustrated in the pioneering work by Google [25] and other in the context of mobile phones [18][27]. The latter example is associated with the concept of Industrial Data

Platform (IDP), where the number of participants is not that high (context also known as enterprise Federated Learning [32]), but each provides a good amount of training samples. The MUSKETEER platform is oriented towards the latter, and it aims at becoming an IDP offering a variety of possible confidentiality/privacy scenarios, named as Privacy Operation Modes (POMs).

If we assume an scenario where a total amount of reward is to be distributed among the participants (data providers), according to the actual contribution of their respective data to the final model quality/performance, then it is possible to formulate the task as a "profit allocation problem". This type of situation has been studied extensively in the context of cooperative game theory, and the most popular solution is provided by the Shapley value estimation scheme [29][11]. This approach offers some attractive features: it is task-dependant, the data is valued only if it allows to improve the performance of the model, the reward is fully distributed among the participants, equal data contribution means equal reward, and the addition of several contributions gets a reward equal to the sum of the individual rewards. The calculation of Shapley values is quite simple. If we consider N participants and S is a subset of players and $U(S)$ is the utility function that measures the performance of the model produced with the data from users in the set S. Then, the Shapley value s_i for user i is defined as:

$$s_i = \sum_{S \subseteq I \setminus \{i\}} \frac{1}{N\binom{N-1}{|S|}} [U(S \cup \{i\}) - U(S)] \tag{1}$$

According to the expression in 1, the Shapley value is computed as the average utility gain obtained when player i is added to any other[2] group of participants. Despite the relatively simple definition of the Shapley's values, their computation requires an exponential number of different utility computations (each one of them usually requiring to train a brand new model). Therefore, Shapley's approach poses some computational challenges if we opt to use a brute force approach. Some works indicate that it is possible to reduce the exponential computational cost to a linear or logarithmic scale by benefiting from a knowledge transfer between trained models, exploiting some peculiarities of a given machine learning model [15] or using Monte Carlo estimations of the utility values [24].

All the above mentioned optimized methods assume we have an unlimited access to the training data and that we can run the training procedures an unlimited number of times, a situation which is rarely found in real world situations. Even so, gathering large amounts of data in the same place faces many barriers, such as the growing number of regulations that limit the access/sharing of the information, with the ultimate intention of protecting the privacy and property rights of users (e.g. GDPR [3] or HIPAA [12]).

As already presented in the previous sections, various architectures have emerged in an attempt to circumvent these data exchange restrictions and ultimately facilitate the training of models with increasing amounts of data while preserving the data privacy/confidentiality. For many years the field of Privacy Preserving Machine

[2] All possible combinations must be considered.

Learning (a.k.a. Privacy Preserving Data Mining) has produced solutions relying on different security mechanisms (Secure Multiparty Computation or Cryptography, among others). It is obvious that the data value estimation in these scenarios has an additional degree of complexity, sometimes unaffordable. Lately, the Federated Learning paradigm has emerged as a less complex approach to the problem of training models while preserving data confidentiality. In a Federated Learning context we face many restrictions on accessing training data, and the training process is typically only run once. Therefore, the traditional data value estimation methods cannot be used directly in this context.

An interesting approach is the one presented in [30], where the interchanged values (models, gradients) during the federated learning process are used to reconstruct the variety of models needed to estimate Shapley values using 1. In this way we can calculate estimates of the different models that would be obtained if different combinations of data sets were used, without the need to train them from scratch. Obviously, an exact reconstruction of all models is not possible and we only get estimates, but it is shown in [30] that good approximations are possible.

The procedure is as follows. It is assumed that there is a validation set available in the aggregator, so that for each possible model trained with a subset S of the training data it is possible to calculate the corresponding utility $U(S)$ needed to estimate the Shapley values. We also assume that the aggregator has access to the following information:

- The initial global (epoch 0) model weights $M^{(0)}$
- The global model weights at epoch n, $M_{all}^{(n)}$
- The model increments[3] contribution from participant m at epoch n, $\Delta_m^{(n)}$

Taking into account all this information, in [30] two approaches are proposed for Data Shapley value estimation. The first one estimates at epoch n the model trained with the datasets from the set of users in set R[4], M_R^n, as the cumulative update from the initial model, i.e.:

$$M_R^{(n)} = \sum_{i=0}^{n} M_g^{(0)} + \sum_{m \in R} \Delta_m^{(n)} \tag{2}$$

and using these model estimates, the corresponding utilities and Data Shapley values in 1 can be calculated, averaging the estimates across all epochs. This approach is prone to divergences from the real model, since the accumulation takes place with respect to the initial (random) model.

The second approach is based on updating the global model $M_{all}^{(n-1)}$ obtained at every step $n-1$ with the contributions from all participants, so the different submodels are estimated using updates with partial data. For example, the model trained with the datasets from the set of users R at epoch n, M_R^n, is estimated as:

[3] If model weights are exchanged instead of gradient updates, the increments can be obtained as a difference between models.

[4] R can be set to S or $S \subseteq I \setminus \{i\}$, as needed.

$$M_R^{(n)} = M_R^{(n-1)} + \sum_{m \in R} \Delta_m^{(n)} \tag{3}$$

such that more accurate submodel estimates are obtained, but they are influenced by the contributions from other participants, since $M_R^{(n)}$ is calculated using information from all contributors.

Notwithstanding the restrictions mentioned above, both methods appear to provide reasonable Data Value estimates in a Federated Learning environment, as evaluated in [30]. Note that under the approaches described above, the Shapley values are calculated exactly but are based on model estimates. Therefore, the quality of those estimates will determine the precision of data value estimates according to Shapley principles.

Various MUSKETEER privacy modes of operation (POM) do not exactly follow Federated Learning principles and use other security/privacy mechanisms (Secure Multi-Party Computing, Homomorphic Encryption), and it remains to be analyzed how to extend the procedures described above to adapt them to the new scenarios.

The above described approach is perfectly valid under "honest but curious" security assumptions, where the participants are assumed not to act outside of the defined protocols (which is the case of the MUSKETEER platform), and therefore they can fully trust the aggregator in the sense that they are confident in that it will always declare the correct (estimated) credit allocation values.

However, in some other situations, the aggregator could act maliciously and, after using participant data for a given task, could declare a lower value than actually estimated. In this different security scenario, a different approach would be needed. Also, it would be of great interest to be able to estimate the Data Shapley values *before* actually training any model, so that preliminary data negotiation can be established before actually participating in the training process.

We are exploring the extent to which the Data Value can be estimated using a collection of statistics calculated on each participant, but which do not contain enough information to train the global model. In the MUSKETEER context we are interested in answering the following questions (and hence we are investigating in that direction):

- To what extent is it possible to estimate the Data Values before actually training the model, based on locally pre-calculated statistical values.
- To what extent can the incremental approach proposed in [30] be extended to scenarios other than Federated Learning, where other privacy mechanisms are used (Two-party computation, Homomorphic encryption, etc.)

5 Conclusion

In this chapter we described an Industrial Data Platform (IDP) for federated learning offering high standards of security and other privacy preserving techniques (MUS-

KETEER). Our approach shows how trust respectful of privacy can be enforced from an architecture point of view but also how the techniques used can support the compliance with certain GDPR principles from a legal perspective. Besides, leveraging more data on such data platforms requires incentives that fairly reward shared data, thereby we also discuss different strategies of data value estimation and reward allocation in a Federated Learning scenario.

Acknowledgements This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 824988.

References

1. Bagdasaryan, E., Veit, A., Hua, Y., Estrin, D., Shmatikov, V.: How to backdoor federated learning. In: International Conference on Artificial Intelligence and Statistics, pp. 2938–2948. PMLR (2020)
2. Binns, R., Gallo, V.: Data minimisation and privacy-preserving techniques in ai systems [www document] (2020). *https://ico.org.uk/about-the-ico/news-and-events/ai-blog-data-minimisation-and-privacy-preserving-techniques-in-ai-systems/*
3. Council of European Union: Council regulation (EU) no 269/2014 (2016). *http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1416170084502&uri=CELEX: 32014R0269*
4. Datatilsynet: Artificial intelligence and privacy (2018)
5. Daten Ethik Kommission: Opinion of the data ethics commission (2019)
6. European Commission: Towards a thriving data-driven economy (2014)
7. European Commission: Guidance on sharing private sector data in the european data economy (2018)
8. European Commission: Guidance on sharing private sector data in the european data economy (no. swd(2018) 125 final) (2018)
9. European Commission: Towards a common european data space (2018)
10. European Commission: A european strategy for data (2020)
11. Ghorbani, A., Zou, J.: Data shapley: Equitable valuation of data for machine learning. In: Proc. 36th International Conference on Machine Learning, PMLR, vol. 97, p. 2242–2251 (2019)
12. Gunter, K.: The hipaa privacy rule: practical advice for academic and research institutions. Healthcare financial management : journal of the Healthcare Financial Management Association **56**, 50–4 (2002)
13. Heckman, J., Boehmer, E., Peters, E., Davaloo, M., Kurup, N.: A Pricing Model for Data Markets. In: Proc. iConference'15 (2015)
14. IBM: IBM Federated Learning. https://github.com/IBM/federated-learning-lib (2020)
15. Jia, R., Dao, D., Wang, B., Hubis, F., Gurel, N., Li, B., Zhang, C., Spanos, C., Song, D.: Efficient Task-Specific Data Valuation for Nearest Neighbor Algorithms. Arxiv http://arxiv.org/abs/1908.08619 (2019)
16. Jones Charles & Tonetti, C.: Nonrivalry and the Economics of Data. American Economic Review **110**, 2819–2858 (2020)
17. Joseph, A.D., Laskov, P., Roli, F., Tygar, J.D., Nelson, B.: Machine learning methods for computer security (dagstuhl perspectives workshop 12371). In: Dagstuhl Manifestos, vol. 3. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2013)
18. Konecny, J., McMahan, H.B., Ramage, D., Richtarik, P.: Federated optimization: Distributed machine learning for on-device intelligence. Arxiv http://arxiv.org/abs/1610.02527 (2016)
19. Koutris, P., Upadhyaya, P., Balazinska, M., Howe, B., Suciu, D.: Query-Based Data Pricing. In: Proc. of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems., vol. 62 (2012)

20. Lee, J., Hoh, B.: Sell your experiences: a market mechanism based incentive for participatory sensing. In: Proc. IEEE International Conference on Pervasive Computing and Communications (PerCom) (2010)
21. Li, T., Sahu, A.K., Talwalkar, A., Smith, V.: Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine **37**(3), 50–60 (2020)
22. Ludwig, H., Baracaldo, N., Thomas, G., Zhou, Y., Anwar, A., Rajamoni, S., Ong, Y., Radhakrishnan, J., Verma, A., Sinn, M., et al.: Ibm federated learning: an enterprise framework white paper v0. 1. arXiv preprint arXiv:2007.10987 (2020)
23. Lyu, L., Yu, H., Yang, Q.: Threats to federated learning: A survey. arXiv preprint arXiv:2003.02133 (2020)
24. Maleki, S., Tran-Thanh, L., Hines, G., Rahwan, T., Rogers, A.: Bounding the estimation error of sampling-based shapley value approximation with/without stratifying. Arxiv http://arxiv.org/abs/1306.4265 (2013)
25. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-Efficient Learning of Deep Networks from Decentralized Data. In: Procs. of AISTATS, pp. 1273–1282 (2017)
26. Purcell, M., Sinn, M., Simioni, M., Braghin, S., Tran, M.N.: D3.2 - Architecture Design – Final Version. https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5cf9bc07f&appId=PPGMS (2020)
27. Ramaswamy, S., Mathews, R., Rao, K., Beaufays, F.: Federated learning for emoji prediction in a mobile keyboard. Arxiv http://arxiv.org/abs/1906.04329 (2016)
28. Richter, H., Slowinski, P.R.: The data sharing economy: on the emergence of new intermediaries. IIC-International Review of Intellectual Property and Competition Law **50**(1), 4–29 (2019)
29. Shapley., L.S.: A Value for n-person Games. In: Annals of Mathematical Studies: contributions to the Theory of Games, vol. 28, p. 307–317. Princeton University Press (1953)
30. Song, T., Tong, Y., Wei, S.: Profit Allocation for Federated Learning. In: Proc. 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA (2019)
31. Walsh, D.: How Much Is Your Private Data Worth - and Who Should Own It?. Insights by Stanford Business. https://www.gsb.stanford.edu/insights/how-much-your-private-data-worth-who-should-own-it (2019)
32. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated Machine Learning: Concept and Applications. ACM Transactions on Intelligent Systems and Technology (TIST) **10**(2), 1–19 (2019)
33. Zillner, S., Bisset, D., Milano, M., Curry, E., García Robles, A., Hahn, T., Irgens, M., Lafrenz, R., Liepert, B., O'Sullivan, B., Smeulders, A., (eds.): Strategic Research, Innovation and Deployment Agenda - AI, Data and Robotics Partnership. Third Release. BDVA, euRobotics, ELLIS, EurAI and CLAIRE (2020)