Manifesto v0.1

Substra Foundation

Objective of this document

This document is Substra Foundation's manifesto. It describes its environment, the challenges it tackles, its vision, approach, origin and ambitions, the community of users and contributors it aims to federate and stimulate.
The purpose of this manifesto is to provide enough information for anyone to easily get a good sense of what are Substra Foundation and this trustless, traceable, privacy-preserving learning and open source adventure.
It is an early version (v0.1) of the manifesto. Remarks, questions and all kind of feedbacks, more than welcome, will contribute to complement and improve future versions: this is a collaborative effort.

Content

TL; DR 2

AI is promising, but... 2

Trustless, traceable, privacy-preserving ML orchestration 2

Substra Foundation advances trustless privacy-preserving learning 3

- A non-profit to host and drive the open source Substra project 3
- Core partnership with Owkin to prepare the open source collaborations 3
- Ambitions for the years to come 4
- Community, contributors 4
- Independence 4

Annex - Definitions and additional information 6

TL; DR

Substra Foundation is an independent non-profit organization aiming to promote privacy-preserving machine learning and foster scientific and economic collaborations.

It coordinates an open source project named Substra, enabling by-design trustless, traceable, privacy-preserving ML orchestration.

It aims at federating a vibrant community of users and contributors around the Substra software framework. The first contributor and a core partner is Owkin, a fast-growing health data AI startup that dedicates a complete technical team to the development of the framework.

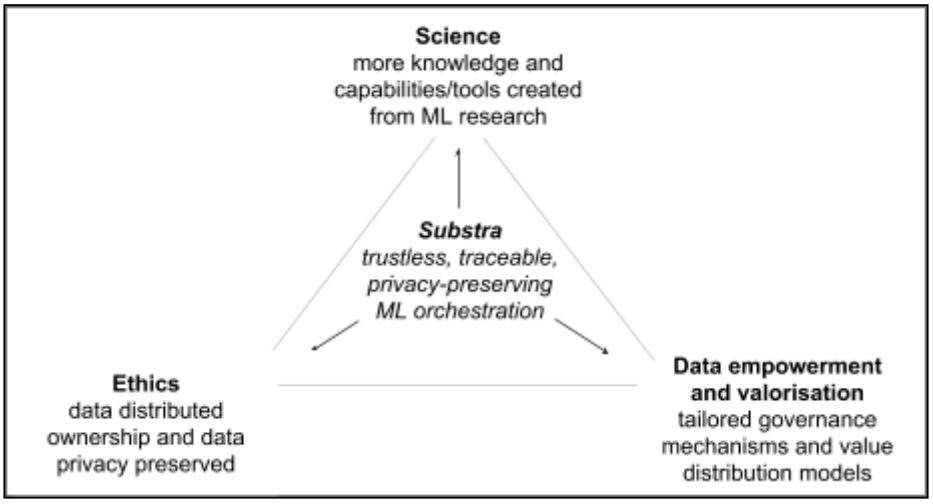
The first public release of the framework (H2 2019), developed by Owkin, will mark the start of the open source collaborations with the broader community.

AI is promising, but...

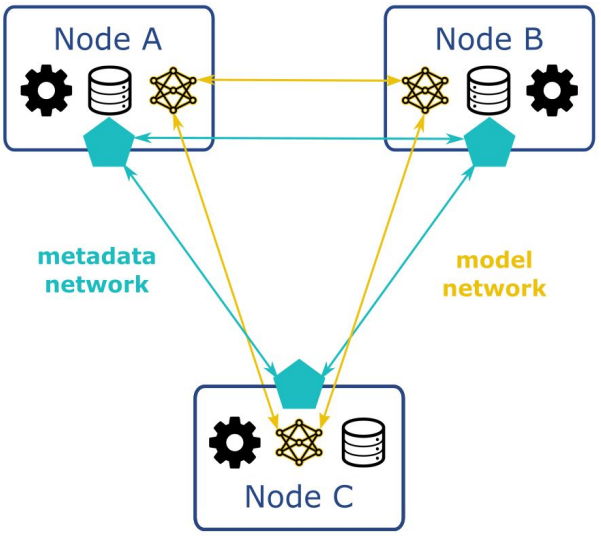
AI is making tremendous progresses, creating vast transformation and innovation opportunities in all sectors. One prerequisite of AI technologies is the collection of data and its processing into exploitable datasets. Storage, processing and sharing of huge amounts of information may constitute a considerable threat to data privacy and governance, and public and private awareness is raising as illustrated by the growing fear of value appropriation by private tech giants, or by the recent enforcement of GDPR in Europe. These threats are even more acute when the underlying data in a machine learning ecosystem is sensitive, either by its personal or confidential nature.

Trustless, traceable, privacy-preserving ML orchestration

At Substra Foundation we believe in and advocate for a model where data providers retain full control of their data, while these data are unlocked, in privacy-preserving conditions, for concrete and impactful ML projects.



Substra is a software framework building upon the leading distributed ledger technology Hyperledger Fabric. It offers by-design trustless, traceable, privacy-preserving machine learning orchestration among multiple parties.
Data analysis algorithms *travel* to distributed training dataset nodes, and computations are performed on local, disposable, secure computing containers at each node. The trustless nature of the distributed ledger framework enforces the preservation of data privacy and provides an incorruptible traceability of all operations.



It enables multi-party data analysis and machine learning collaborations, it facilitates a large variety of strategies for privacy-preserving data sharing.

Substra Foundation ambition is to facilitate multi-partner collaborative ML projects

An independent non-profit to host and drive the open source Substra project

Substra Foundation is an independent non-profit created early 2018 by passionate individuals.

Its mission is to coordinate the open source project Substra, foster the emergence of a community of users and contributors, research value repartition models, and promote trustless privacy-preserving learning approaches by-design.

It is important for Substra Foundation to maintain its independence from other organizations and its agility as a small and innovative organization. Thus, to welcome new members in its governance over time, it will only accept individuals (no organizations or representatives of organizations), accomplished contributors, privacy-preserving learning enthusiasts, wishing to participate in the administrative management of the non-profit.

Core partnership with Owkin to prepare the open source collaborations

Substra Foundation works in close partnership with Owkin, a fast-growing health data AI startup that dedicates a complete technical team to the development of the framework and is the first contributor to it. In the course of H2 2019, the first public release of the framework, developed by Owkin, will mark the start of the open source collaborative approach on which Substra Foundation is working.

Ambitions for the years to come

As of 2019 Substra Foundation is engaged in large collaborative research projects in France and in Europe in the health sector. It aims at growing its reach and developing the adoption of trustless privacy-preserving learning and of the Substra Framework in multiple industries and geographies. But its most challenging ambition is certainly to federate a vibrant community of contributors and users (see next section).
Over the long run, its ambitions include:

- Demonstrating the security of the framework and its compliance with data protection regulations (e.g. GDPR), have it seen as a piece of solution to privacy preservation requirements;
- Leveraging the framework mechanisms to foster scientific rigour in ML model development and evaluation;
- Integrating value repartition bricks/features in the framework;

- Cultivating its independence and be considered a third-party of trust;
- Developing a sound revenue model to ensure its autonomy and sustainability.

Community, contributors

As mentioned in the above section a key ambition is to federate a vibrant community, to make the whole more than the sum of the parts. It is a challenge essential to the success of the project. All this is work in progress:

- contributors will constitute the *College of Contributors*;
- enthusiasts about using the Substra framework are invited to participate in the work of the *College of Users*.

If it proves relevant in the future, these colleges could be subdivided into thematic or sectoral colleges (e.g. a health college). Other ideas open for reflection include the animation of a bug bounty program, the deployment of a devnet, a support program for researchers and startups building tech components and services on top of Substra, etc.
Substra Foundation is currently working on how to facilitate and stimulate the emergence of these communities, organize and animate the colleges and their respective work, ensure a good coordination between all parties. The objective is to create the conditions for an open, fruitful, enthusiastic collaboration.

SUBSTRA
FOUNDATION

Get in touch


substra.org


/SubstraFoundation


@Substra_org


hello@substra.org

Annex - Definitions and additional information

Hyperledger Fabric: the world leading private and permissioned blockchain framework. Hyperledger Fabric is one of the Hyperledger open source projects hosted by the Linux Foundation. It has been widely adopted as a reference framework for implementing blockchain-based services in business ecosystems. Substra Framework is built upon Hyperledger Fabric and its core components (distributed ledger, identities and membership mechanisms, smart contracts, consensus mechanisms, etc.). Link

Distributed ledger: a distributed ledger is a consensus of replicated, shared, and synchronized digital data geographically spread across multiple sites, countries, or institutions. There is no central administrator or centralized data storage. A peer-to-peer network is required as well as consensus algorithms to ensure replication across nodes is undertaken (source: Wikipedia).

Trustless: Substra Framework is a 'trustless' ML orchestration framework. The word 'trustless' might be ambiguous in certain circumstances. We believe it should be used as 'doesn't require trust *a priori* between parties': the code implementation of the software enables parties to collaborate without trusting each other, it technically guarantees that actions and transactions will be performed as defined in the rules agreed upon. What is required is to 'trust the code': it might not be straightforward and even require some audit effort, but in many cases it is easier than trusting a number of other independent organisations.

Privacy-preserving: Substra Framework is a 'privacy-preserving' ML orchestration framework. It enables data analysis and machine learning computations on data without transferring the data to anyone and without giving data scientists read access to these data. In that sense we call it privacy-preserving, with the word 'privacy' referring to both the privacy of the dataset for the organisation managing it, or the privacy of personal data for the individuals these data refer to.

Machine learning orchestration: in contexts where multiple parties collaborate for elaborating machine learning models, the different operations (e.g. algorithms transfers, training computations, model evaluations, predictions…) need to be orchestrated in time and space. Such an orchestration is done over a network connecting the parties, and requires complete traceability of all operations, identities certifications, security (among others). Substra Framework enables the implementation of applications or services requiring trustless, traceable, privacy-preserving machine learning orchestration.