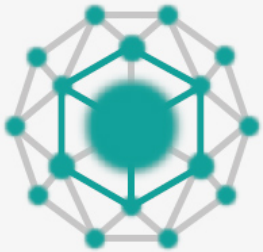


MY HEALTH
MY DATA

A blockchain-based project providing privacy-enhancing solutions for data transactions between people, clinical institutions, and industries

Edwin Morley-Fletcher

LYNKEUS.

Atelier Blockchain pour la Santé – 21 Octobre 2020

emf@lynkeus.com

Blockchain's basics

- Blockchain is a disruptive innovation regarding uncertainties that traditionally have implied the need of relying on some amount of trust for coping with them.
- It is a technology providing transparent and secure storage and transfer of data without having recourse to a central authority. With blockchain all data transfers become traceable and auditable by participants to the ledger.
- Blockchain allows to envisage a distributed rather than a hierarchical foundation of trust. It allows "trust-less certainty".
- If internet has dramatically reduced transaction costs on information, blockchain can do the same regarding the exchange of data incorporating value.
- Blockchain allows to digitize value transfers and use self-enforcing "smart contracts" for automating the enactment of contractual rules (*code is law*) .
- Blockchain allows to have recourse to issuing digital tokens for crowdsourcing (*tokenomics*).

Blockchains change the balance between hierarchies and markets

- What Internet did to transaction costs regarding information, blockchain can do regarding trust.
- Reliable data-rich information systems become possible without being paralysed by excessive transaction costs.
- The limitation of decentralised solutions was in the past the all-pervasive (and excessively reductive) role played in market-based solutions by just one synthetic information: the price.
- The advantage of hierarchical solutions was based on the capacity to tackle uncertainty by centrally ordering a multiplicity of information about which actions to follow.
- The blockchain methodology, by allowing solutions based on decentralised protocols,
  - removes the friction and costs of current intermediaries
  - makes it possible to develop distributed and transparent systems
    - where empowerment can be shared
    - asymmetries can be balanced
    - qualitative aspects can be taken into account.

What is the potential of blockchains as social technologies?

- Blockchains are technical solutions that facilitate the smooth functioning of an ecosystem by:
  - managing and implementing decision-making through automated consensus
  - creating incentives that nudge participants into behaving constructively
  - generating trust and transparency
- By enacting decentralised information systems with inherent data integrity, blockchains constitute strong anti-corruption tools and feature as relational software
  - enabling new trust mechanisms capable of transforming social relations
  - reducing transaction costs
- There are close similarities between blockchains and bureaucracies, though bureaucracies are centralised and blockchains are distributed:
  - both are defined by the rules and execute predetermined rules
  - both work as information processing machines
  - both work as trust machines

Bureaucracies are thus natural candidates to have centralisation being replaced by federated blockchain systems.

- Sustainable growth implies doing more with existing resources and attracting more resources to expand the scale of operations: blockchains reduce costs and increase the flow of funds, helping social innovation organisations to scale up, by enabling marketplaces and the issuance of alternative currencies and tokens.

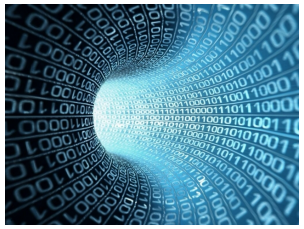
How can blockchains translate into innovative social policies?

- Social protection systems are a fundamental element of pride for EU countries, though differing between bismarckian and beveridgean models (including also Mediterranean varieties) and experiments combining both approaches.
- Traditionally the welfare states are characterized as mainly exerting a “piggy bank” or a “Robin Hood” role:
  - the first helping people to insure against social risks and redistribute resources over the life cycle
  - the second entailing measures to reduce social exclusion by redistributing income and wealth
- This varying mix is translated into a combination of four key functions: regulatory, redistributive, insurance, production
- Historically, these functions have been largely centralised, but could be reorganised in different manners, for instance as Personal Welfare Accounts, operating in blockchain-regulated social markets, where beneficiaries could have incentives to:
  - check that the universal (but not unlimited) coverage is responsibly guaranteed by public agencies and private organisations
  - play a role in changing the market through informed decision and collective actions

Can blockchains translate into opportunities for healthcare?

- Blockchains reside at the nexus of several disciplines which are key for providing healthcare solutions: cryptography, game theory, tokenomics, network theory.
- There is a huge potential for:
  - employing cryptographic and algorithmic methods to record and synchronise health data transactions across distributed networks in an immutable manner.
  - using smart contracts as coded instructions which execute on the occurrence of an event and extend the functionality of blockchains from storing transactions to performing computations.
  - developing multi-sided platforms where data providers (being both clinical institutions and individuals), researchers and industries can all rely on data integrity and security and mutually reinforce network effects.
  - allowing to manage data flows and usage, based on individual free choice and self-determination, making dynamic data portability in real time possible for individuals and companies, along with various compensation models.
  - applying Health Big Data to Artificial Intelligence and Machine Learning for medical knowledge discovery.

Hurdles and constraints



- There is an inherent tension between the rationale of the blockchain technology and some structural elements of the General Data Protection Regulation (GDPR).
- Data minimisation, the right to amendment, and the right to be forgotten, are deeply at odds with blockchain immutability and require that personal data be stored off-chain in order to make them modifiable and deletable.
- The off-chain health data storage solution is also advisable on a technological ground with regard to present blockchain scalability limitations.
- There is a discrepancy in the way anonymisation is referred to in the GDPR, on a risk-based approach (as “personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable by all the means reasonably likely to be used”), and how it is defined by the European Data Protection Board (where “anonymisation results [only] from processing personal data in order to irreversibly prevent identification”).
- The subsequent regulatory uncertainty makes it extremely difficult to obtain anonymised data from clinical institutions (also because of the new heavy sanctions falling on non-compliant Data Controllers).
- Whereas pseudonymised data require on principle a specific legal ground, such as an explicit personal consent, for being shared with third parties.
- Notwithstanding all the promises they entail, Big Data and AI are therefore difficult to apply at scale in medicine, given that effective data sharing is still the exception in healthcare.

What have been MHMD basic architectural features?

- A private permissioned blockchain recording all transactions related to Off-chain data stored by multiple hospital repositories and by individuals
- A Metadata Catalogue allowing to safely inspect what health-data are available
- The possibility of making use of Smart Contracts for automatically checking the needed consent.
- Privacy-enhancing technologies for assuring GDPR compliance and advanced ways of handling data.
- An overall Privacy-by-Design and Compliance Assessment

What about data handling? Two key modalities

1. “Visiting mode”: bringing the algorithms to the data

Secure computation, which permits running AI without disclosing neither data nor algorithms, is performed through three tools:

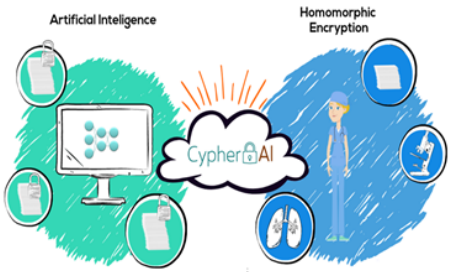
1. Homomorphic Encryption
2. Secure Multiparty Computation
3. Federated Deep Learning with an untrusted Black Box

2. The sharing mode

- Sharing health data may be risky even in a blockchain:
  - What happens after data download is not under control of the MHMD blockchain.
  - The risk of data breaches increases with the number of copies shared
  - MHMD has investigated what contribution can come from sharing Synthetic Data
- According to various scenarios of trust, and privacy-preserving needs, MHMD health data can be published as pseudonymous or anonymous data.
- AMNESIA is used to guarantee k-anonymity.
- Synthetic data, nevertheless, have proven to be a powerful solution to scale up data sharing in privacy-constrained environments such as healthcare, especially if operated in conjunction with Differential Privacy (DP).

Homomorphic Encryption

- Homomorphism is the property of an encryption scheme that allows to perform operations on encrypted data.
- Once the results are available, they are sent back and decrypted at the source. The computing service has access only to the encrypted data, and since the decryption key is not available to the service, no personal or useful information can be extracted.
- The solution enacted by the Transylvania University of Brasov within MHMD is based on Fully Homomorphic Encryption (FHE) making use of MORE (Matrix Operation for Randomization or Encryption) as encryption scheme, enabling the computations within a neural network model to be directly performed at a relatively small computational overhead, with the addition of an obfuscation layer.
- This solution was Awarded the Innovation Radar Prize 2019 in the category Industrial & Enabling Technologies.

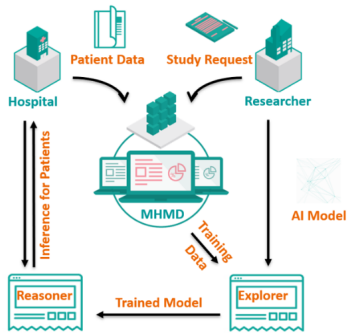


Secure Multiparty Computation

- SMPC is a cryptography modality aiming to create methods for parties to jointly compute a function over their inputs, keeping these inputs private.
- SMPC allows a set of distrustful parties to perform the computations in a distributed manner, while each of them alone remains oblivious to the input data and to the intermediate results.
- The computation is considered secure if, at the end, no party knows anything except its own input and the results.

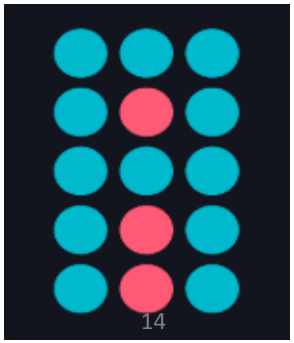
Federated Deep Learning with an untrusted Black Box

- Jointly developed by Siemens Healthineers and Athena RC, using SMPC and Differential Privacy.
- A federated learning platform allows training of complex deep-learning models without ever centralizing or exposing the underlying raw data, while providing formal privacy guarantees through DP mechanisms and SMPC cryptographic techniques.
- The “black-box” ML modules can be provided by third parties and executed locally at the data providers to support model training.
- As such, this solution can be an important building block of a privacy-by-design data sharing system for healthcare research. It supports advanced analytics tools like DeepExplorer (for training models) and DeepReasoner (for deploying trained models at clinical and research facilities on incoming local data)
- Privacy Preserving Machine Learning (PPML) is an emerging field in data science. As yet, its foundation on SMPC still implies a large communication overhead which makes it hard to use where very large amounts of data are required, since communication and computation costs are greatly affected by the increase of the number of involved parties or of the model’s complexity.



Synthetic Data

- Synthetic data are fully artificial data, automatically generated by making use of machine learning algorithms, based on recursive conditional parameter aggregation, operating within global statistical models.
- High-quality synthetic data closely resemble the real data and are a suitable substitute for processing and analysis.
- They typify the case of “personal data [which are] rendered anonymous in such a manner that the data subject is not or no longer identifiable” (Recital 26 GDPR).



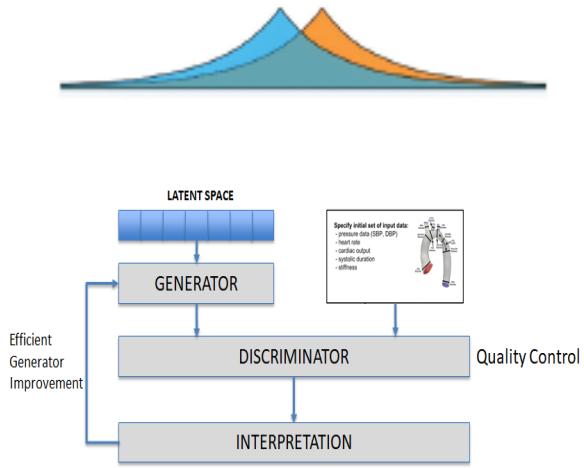
GANs

- Generative Adversarial Networks (GANs) have been one of the most important innovations in deep learning.
- GANs are based on two models playing recursively against each other.
- The Generator learns to capture and recreate the data distribution.
- The Discriminator estimates again and again whether the data created by the Generator is fake or not.
- This allows (with important interpretability issues) to assess the overall statistical resemblance of two sets of data.

How to bridge GDPR privacy and health data

- Synthetic data provide realistic data while not exposing identifiable information, in support of both medical-AI technologies and traditional biomedical products development.
- They achieve anonymity breaking the link between private information and data's information content.
- Values in the original database are algorithmically substituted with those taken from the database's statistical distributions, to create entirely new records.
- In MHMD they have been successfully used to publish health data and health imaging data, to train machine learning tools, and to test clinical decision support applications.

Generating differentially-private synthetic data



- Differential privacy provides an until-now lacking mathematical foundation to privacy definition:
- "Differentially Private Synthetic Data Generation is a mathematical theory, and set of computational techniques, that provide a method of de-identifying data sets—under the restriction of a quantifiable level of privacy loss. It is a rapidly growing field in computer science"

[National Institute of Standards and Technology Differential Privacy Synthetic Data Challenge 2019: Propose an algorithm to develop differentially private synthetic datasets to enable the protection of personally identifiable information while maintaining a dataset's utility for analysis]

Further projects & developments

EuCanImage

- EuCanShare is a cross-jurisdiction (Canada-EU) large scale research cardiac data repository enabled by a blockchain
- Kraken is a self-sovereign identity system leveraging blockchain technologies for establishing a self-governing biomedical data market-place for individual and institutional data in healthcare and education
- EuCanImage tackles the combined usage of blockchain and AI in the healthcare domain, and particularly in oncology, with the aim of attaining better results in terms of data management and sharing, and more trustworthy and reliable AI models and training processes.
- In managing data used for AI training in a distributed learning environment, a blockchain can become particularly useful by creating a tamper proof record of how an AI model has been implemented, storing it in a secure decentralised manner.
- An AI passport will provide critical information about any given AI model implemented within the EuCanImage framework, including details about the data used, the training pipeline, changes and updates to the model, as well as the outcome of the testing and validation process, allowing to audit AI algorithms throughout their lifecycle.