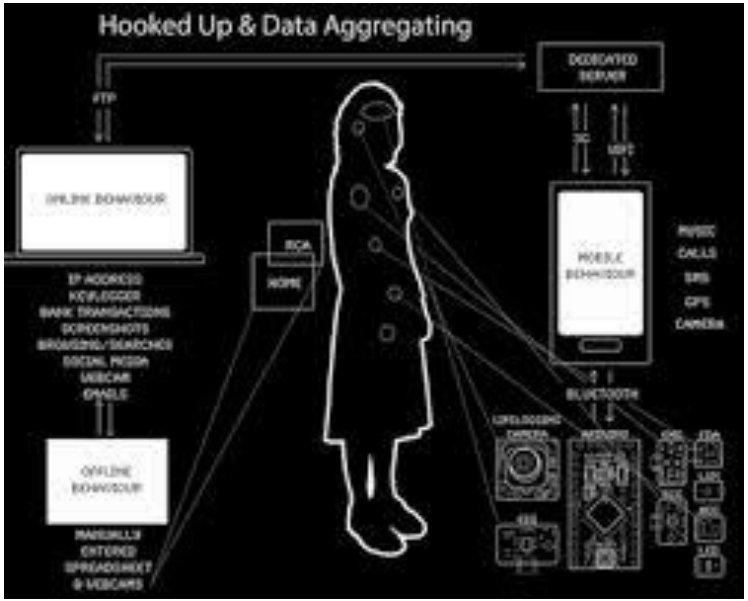
Acknowledging Value of Personal Information:a Privacy Aware Data Market for Health and Social Research

The context

- Increasingly more data is available about individuals
- Data has lot of potential value but, but it’s exploited by few actors
- Value produced with individuals data is not redistributed
- Data gathering relies on users unawareness, and lack of privacy consciousness
- Is there a way to empower data producers, acknowledging value?

Data as Labour

- Jennifer Lyn Morone, a “data artist”, offered collections of her data for sale
- Posner and Weyl, in “Radical Markets”, imagine that unions could/should arise to negotiate users data value



A personal data market

In these perspectives, creators “own” their own data. What does this means?

- They control access of the data they produce
- They can disclosure their data, in change for something

- There are actors that form a data “demand”, and that are willing to get users data
- They can extract value from the data (for instance, to seek correlations from big amounts of data)
- A primary (why not a secondary?) market can be established

Problems/obstacles

- How can users
  - control the access to their data?
  - disclose partial information in a reliable way?
  - ensure quality of data?
  - aggregate data from different sources?

- How can demand actors credibly offer rewards for data?

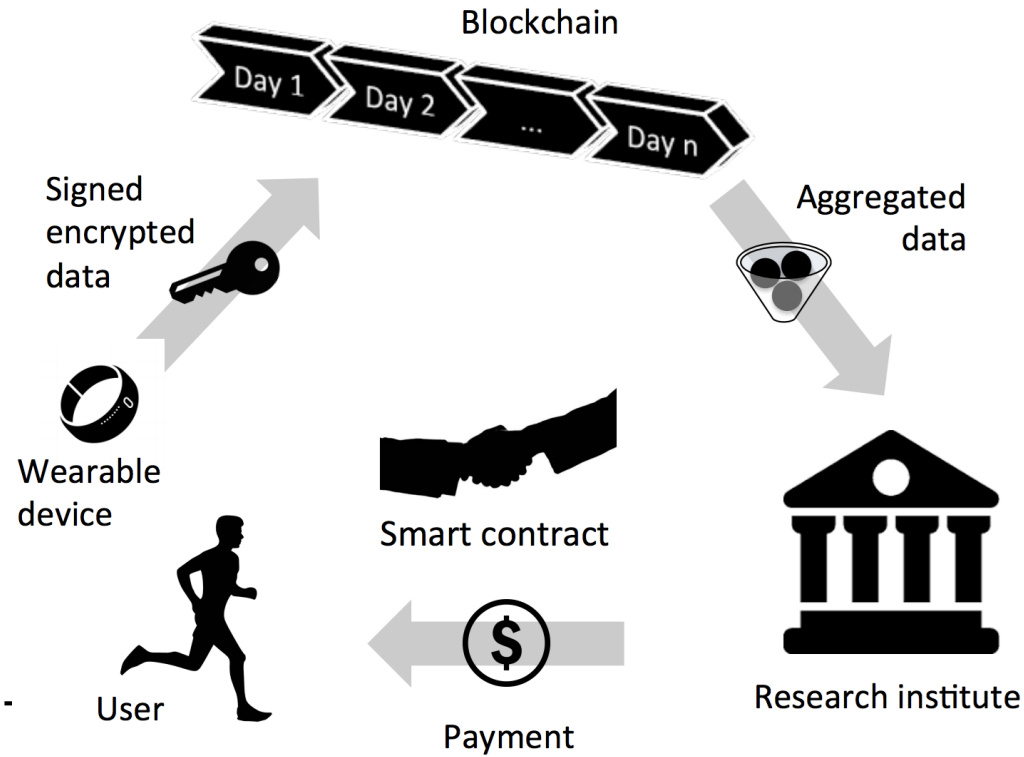
Data ownership

- Data is signed by the devices, and stored encrypted by the user
- Data can be timestamped on a blockchain
- The devices are trusted, meaning that they cannot be tampered with
- Example: a fitbit
- Problems: how can user
  - prove data source?
  - disclose data partially?

(non interactive) Zero Knowledge Proofs

- cryptographic devices that allow to convince a verifier (Bob) that Alice has possess some information (e.g. knows the preimage to a hash), without revealing any other information
- ZKP are very general
- In this case, the user could produce a proof that:
  - A certain number is indeed the result of the application of a function (say average) over a sequence of values signed by a certified device

The market


Blockchain
Day 1
Day 2
...
Day n
Signed encrypted data
Aggregated data
Wearable device
Smart contract
Research institute
User
Payment

State of the art

- Zcash uses it for confidential transactions
- Aztec is using zkproofs to enable anonimity in ethereum smart contracts
- There are libraries to generate proofs and provers:
  - ZoKrates
  - Pinocchio

Trusted computing (in embedded devices)

- **Endorsement Key**: key generated at production time, available only to cryptographic circuitry (cannot be extracted)
- **Memory curtaining**: memory with restricted access, even by software.
- **Remote attestation**: ability to prove that a device is running a given software on a given hardware

Current work

- Exploring useful functions over a set of signed data (e.g. average, or subsampling) to produce proofs
- A smart contract that checks the proof and acts accordingly
- Implement a prototype on a embedded device that exploits secure boot to guarantee data produced
- Relationship with digital identities models such as Self Sovereign identity