Maastricht University

Institute of Data Science

Creating and using your own personal data vault

Chang Sun

25 Nov 2020

For IDS Training Day 2020

Your expectations

- Have you heard/learnt personal data vault before?
- Why this is interesting for you?
- What do you expect to learn from this session?

Plan for the next one hour

- 20 minutes presentation about personal data vaults and dive into SOLID
- 20 minutes hands-on session
  - Create your own SOLID POD
  - Explore the file system
  - Understand how its access control works
  - Try existing applications
- 15 minutes discussion
- 5 minutes wrap-up

Objectives

- Learn what personal data vaults are and understand several existing technologies
- Create your own personal data vaults and understand/use the access control
- Get motivated to design your own Solid application to use data from personal data vaults

How we are using personal data for research now?

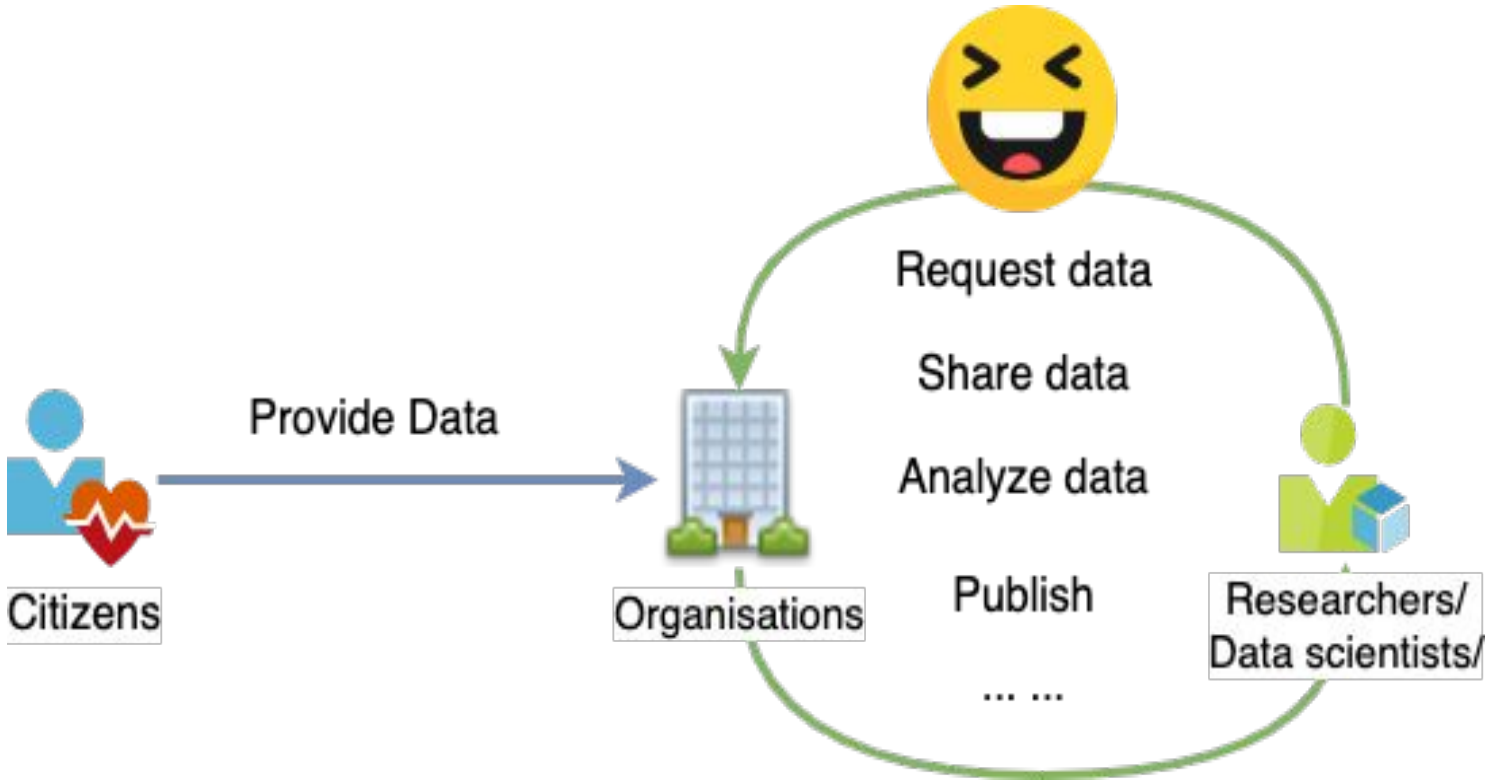
Citizens

Organisations

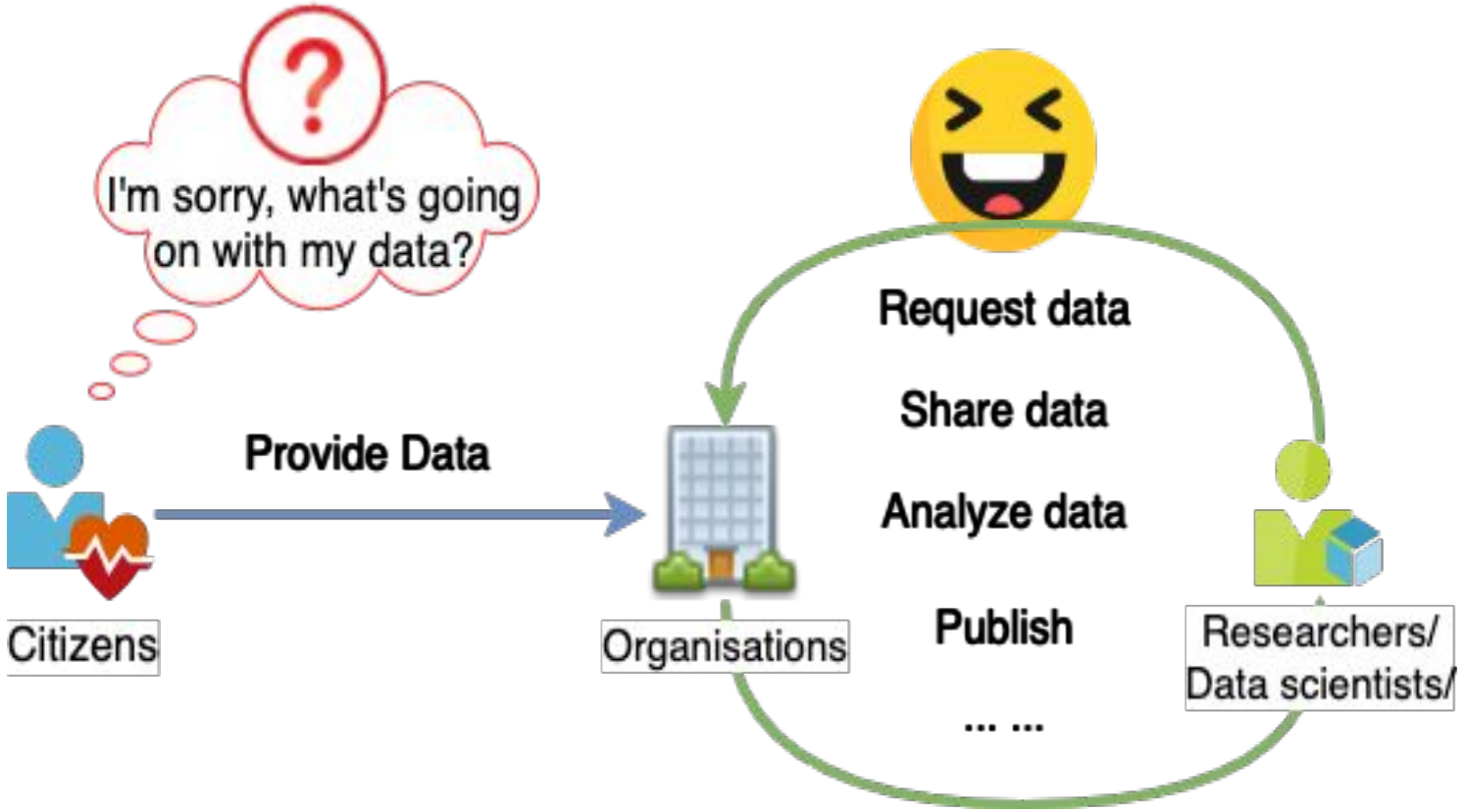
Researchers/
Data scientists/

How we are using personal data for research now?

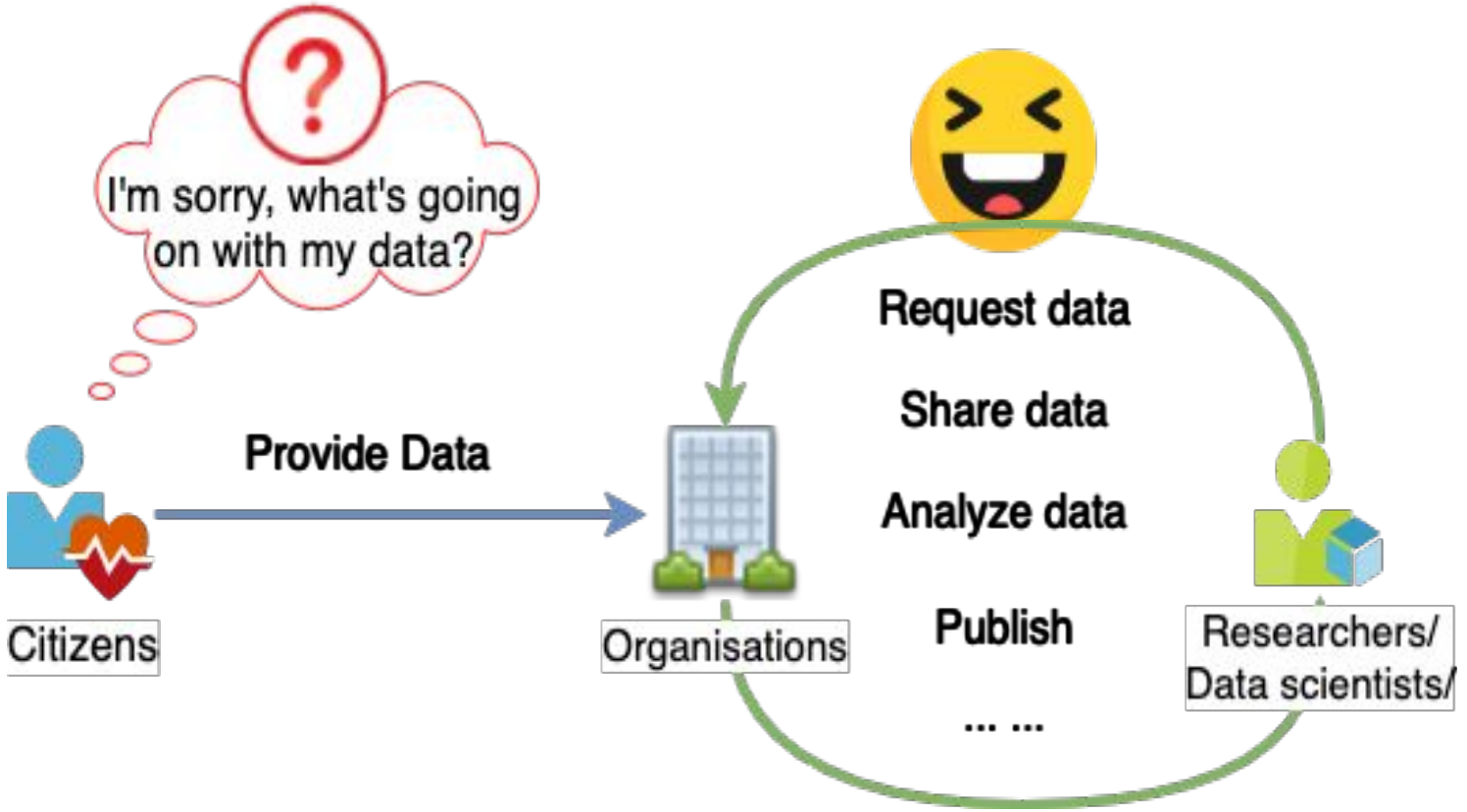
How we are using personal data for research now?



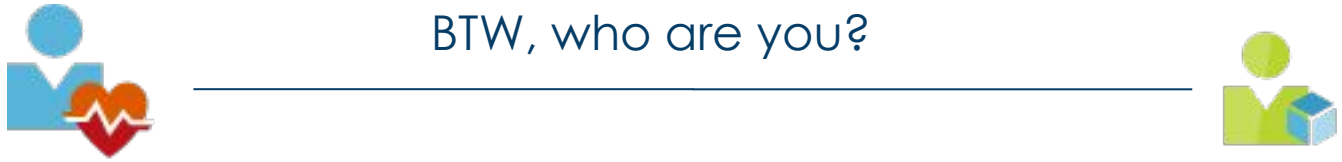
How we are using personal data for research now?



How we are using personal data for research now?



Why the current way using personal data for research is not good enough?



WHAT?
What data from me is being used?

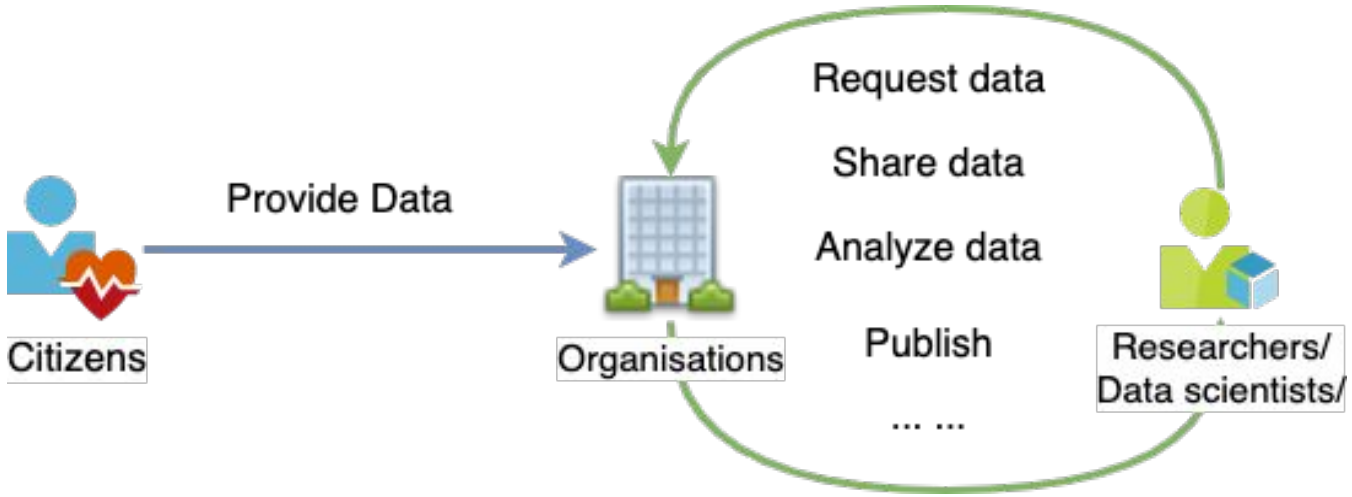
WHEN?
From when to when are you using it?

HOW?
How is my data being processed?

WHY?
Wait a minute, why do you need my data?

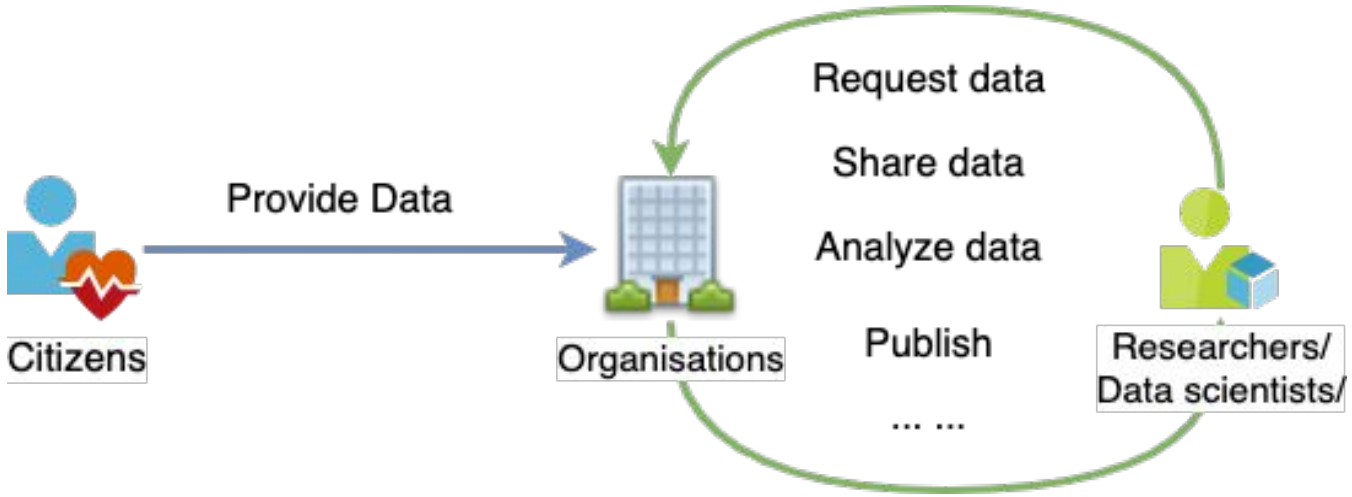
Organization-centric

Organization-centric



Organization-centric VS Citizen(Individual)-centric

Organization-centric



Citizen-centric

The key problem

"Individuals value the high level of protection granted by the GDPR and ePrivacy legislation. However, ***they suffer from the absence of technical tools and standards that make the exercise of their rights simple and not overly burdensome.***"

-- European Committee <An European Strategy for Data> 2020

https://ec.europa.eu/digital-single-market/en/european-strategy-data

The key problem

"Individuals value the high level of protection granted by the GDPR and ePrivacy legislation. However, ***they suffer from the absence of technical tools and standards that make the exercise of their rights simple and not overly burdensome****."*

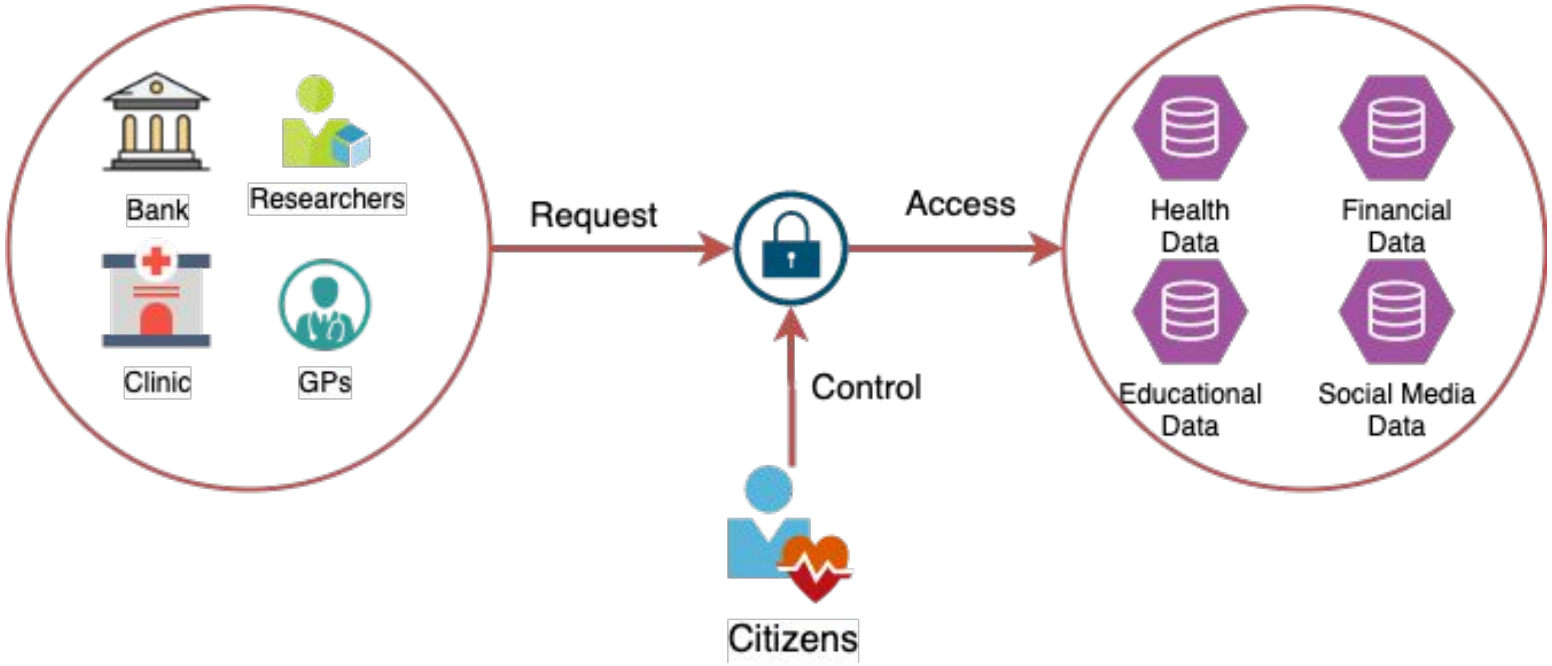
"In response to this, ***there are calls*** *to give individuals the tools and means to decide at a* ***granular level*** *what is done with their data.* ***Those tools and means*** *include consent management tools, personal information management apps, including fully decentralised solutions building on blockchain, as well as personal data cooperatives or trusts acting as novel neutral intermediaries in the personal data economy."*

-- European Committee <An European Strategy for Data> 2020

https://ec.europa.eu/digital-single-market/en/european-strategy-data

The key problem

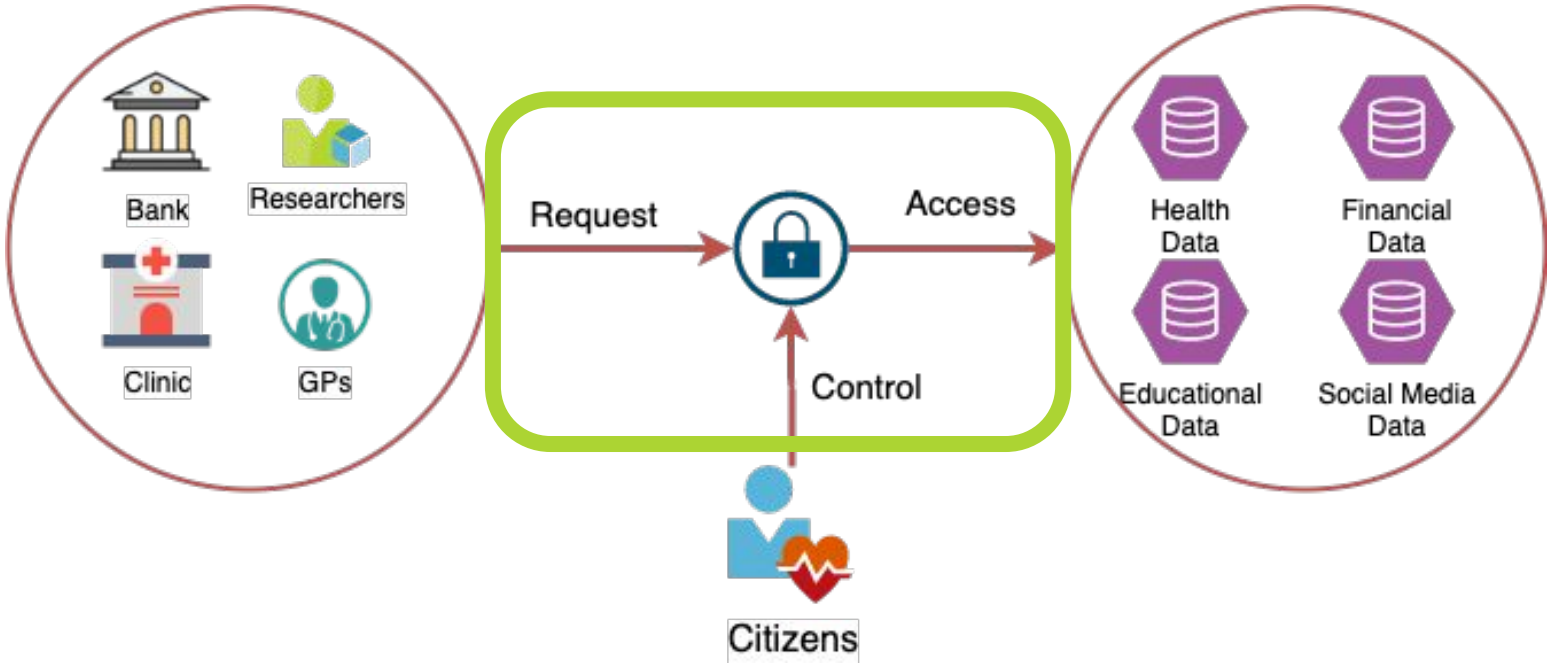
*"... **there are calls** to give individuals **the tools** and means to decide at a **granular level** what is done with their data. **Those tools and means** include consent management tools, personal information management apps...*



https://ec.europa.eu/digital-single-market/en/european-strategy-data

The key problem

*"... **there are calls** to give individuals **the tools** and means to decide at a **granular level** what is done with their data. **Those tools and means** include consent management tools, personal information management apps...*



https://ec.europa.eu/digital-single-market/en/european-strategy-data

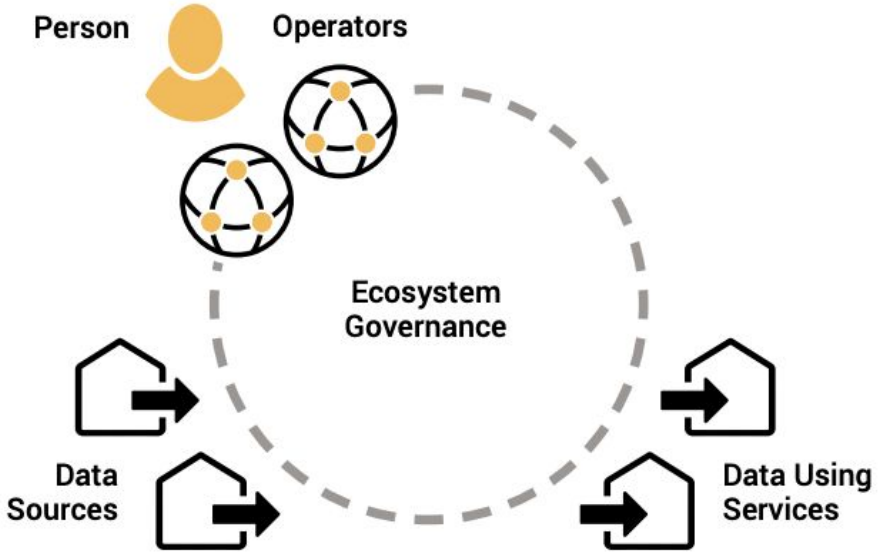
Personal Data Vault

- A Personal Data Vault (PDV) is a **privacy** architecture in which individuals retain **ownership** of their data. Data are routinely filtered before being shared with **content-service** providers, and users or data custodian services can participate in making **controlled** data-sharing decisions [1].

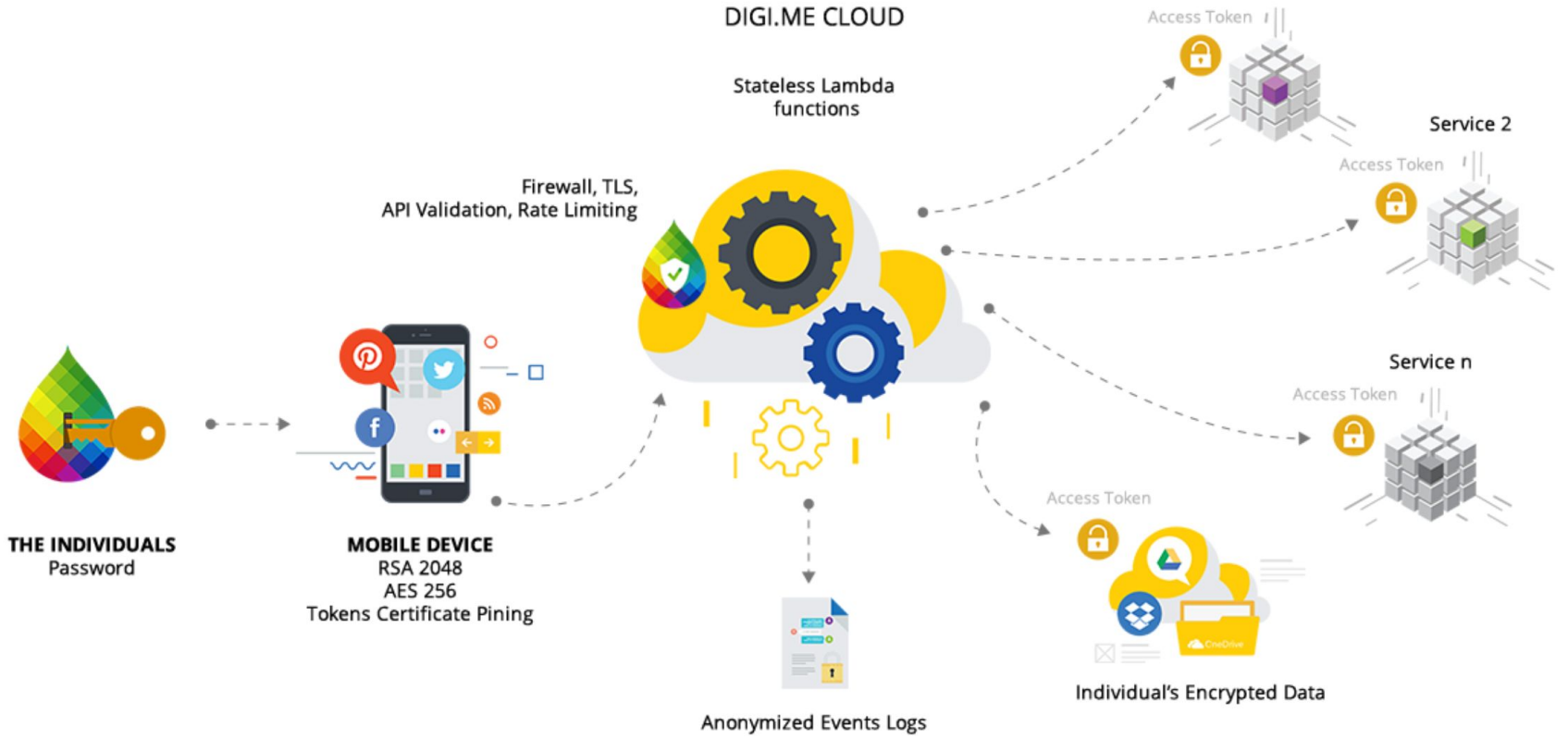
- Also named as
  - Personal Data or Information Management Systems concept
  - Personal Cloud,
  - Personal Data Server
  - Personal Data Store
  - Personal Data Locker
  - … …

[1] Min Y. Mun, Donnie H. Kim, Katie Shilton, Deborah Estrin, Mark Hansen, and Ramesh Govindan. 2014. PDVLoc: A Personal Data Vault for Controlled Location Data Sharing. ACM Trans. Sen. Netw. 10, 4, Article 58 (June 2014), 29 pages. DOI:https://doi.org/10.1145/2523820

Personal Data Vault Tools



- **Solid [2] (Social Linked Data)**
- MyData [3]
- Digi.me3 (Personal) [4]
- MyHealthMyData [5]
- DECODE [6]
- … many others …

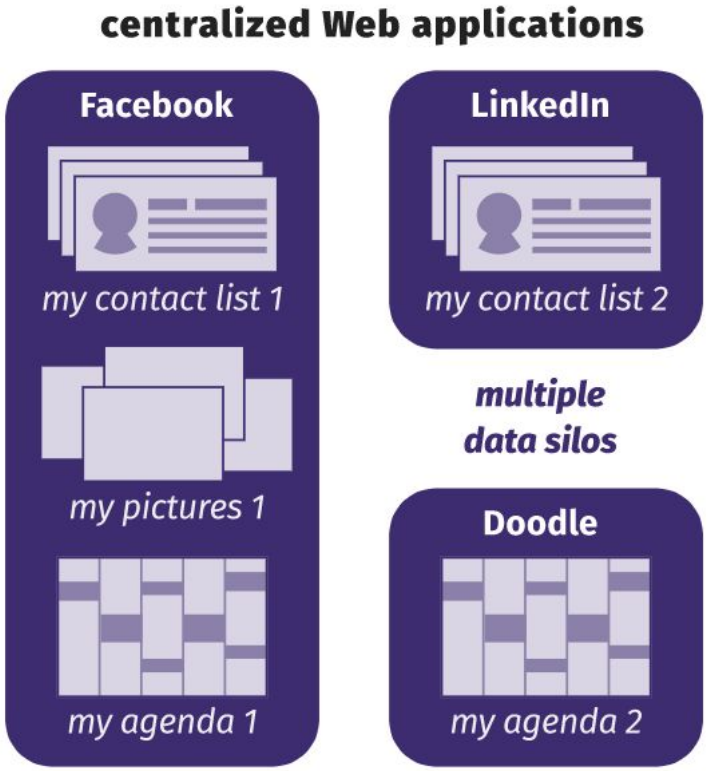


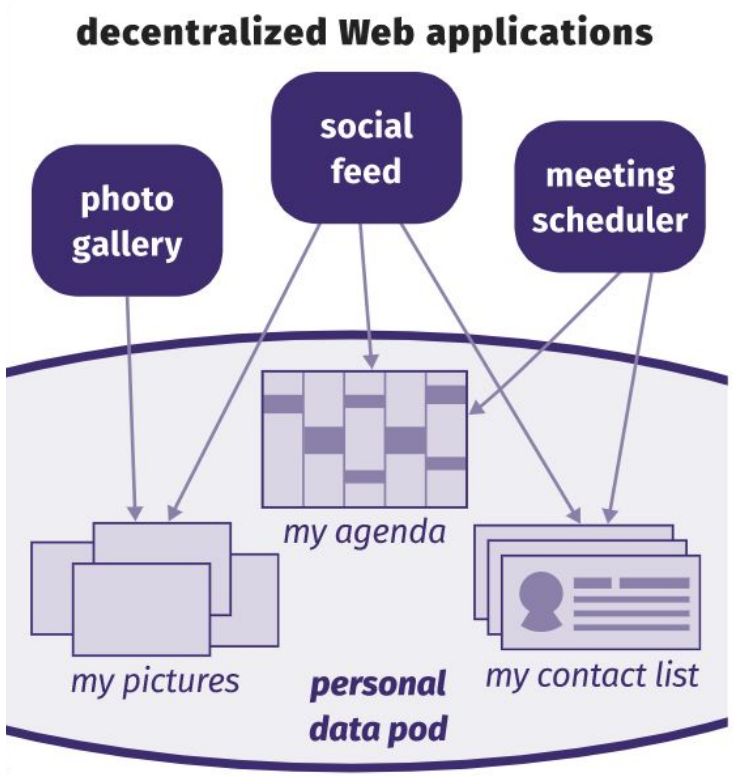
[2] https://solidproject.org/
[3] https://mydata.org/
[4] https://digi.me/
[5] http://www.myhealthmydata.eu/
[6] https://decodeproject.eu/

What is SOLID

Solid (Social Linked Data) is a web decentralization project led by Tim Berners-Lee.

Aims to radically change the way Web applications work today, resulting in true data ownership as well as improved privacy.





https://rubenverborgh.github.io/Solid-DeSemWeb-2018

Why SOLID?

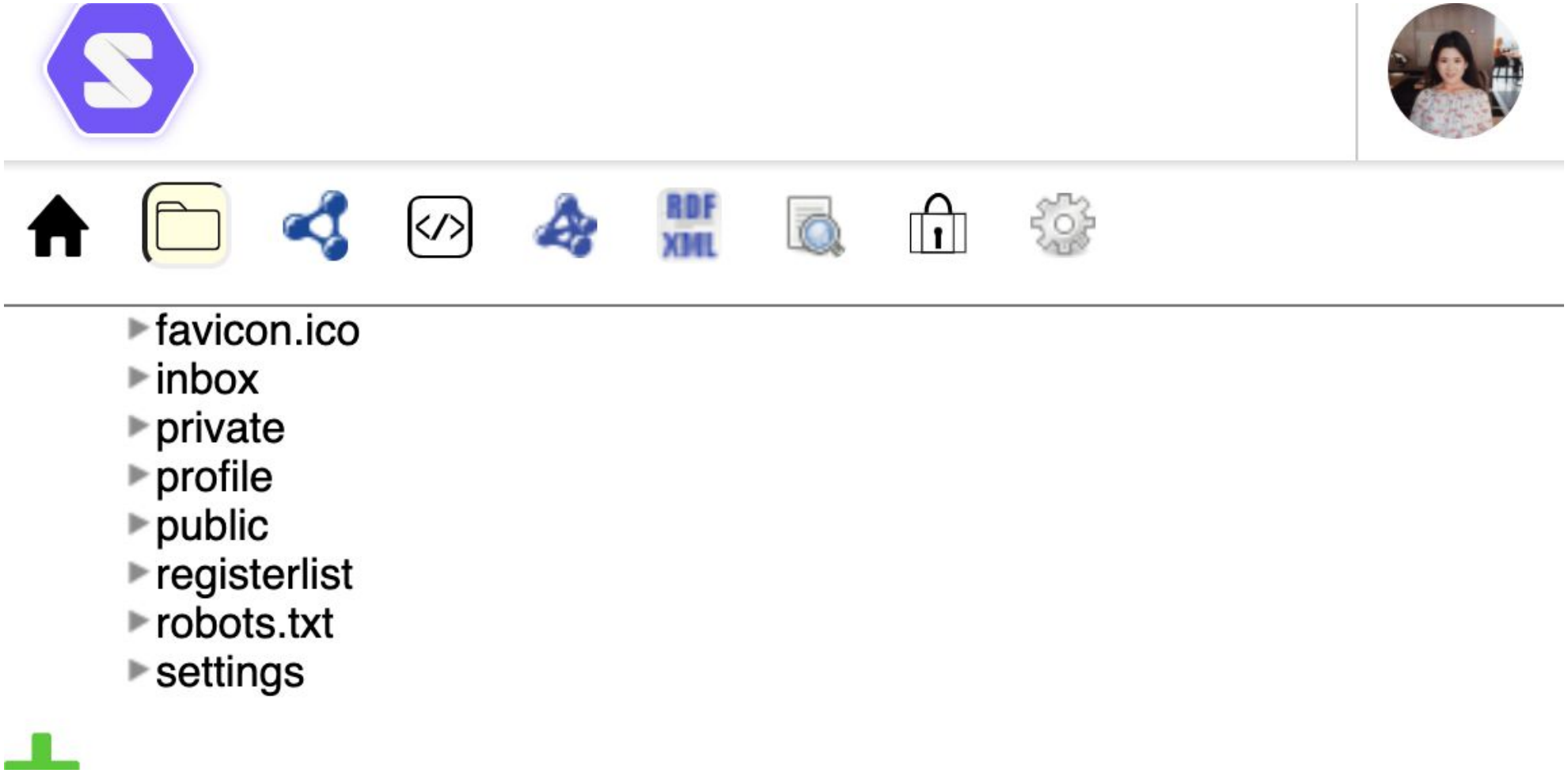
1. Use decentralized peer-to-peer networking
2. Based on semantic web tech and Linked data principle
3. Emphasize on interoperability
4. Rely on existing W3C standards and protocols
5. Being developed by open source community

Decentralized, Linked Data, FAIR, Open source… 🤓

Key components in SOLID

- WebID (e.g., https://chang.inrupt.net/profile/card#me)
- SOLID Pod
- Pod Provider
  - https://inrupt.net/ (USA)
  - https://solidcommunity.net/ (UK)
  - https://solidweb.org/ (Germany)
- SOLID Apps

SOLID - HOW IT LOOKS LIKE?



SOLID - Data file

healthrecord.ttl

```
@prefix : <#>.
@prefix schem: <http://schema.org/>.
@prefix c: </profile/card#>.
@prefix XML: <http://www.w3.org/2001/XMLSchema#>.
@prefix n0: <http://xmlns.com/foaf/0.1/>.

:159335416259554103796776206 23
    a schem:dataFeedElement;
    schem:creator c:me;
    schem:dateCreated "2020-06-28T14:22:42Z"^^XML:dateTime;
    n0:age 27;
    <http://dbpedia.org/ontology/income> 2000.
```


SOLID - ACCESS CONTROL

- Owners - read, write, and control sharing.
- Editors - read and change
- Posters - add new info and read but not change existing info
- Submitters - add new info but not read any
- Viewers - read but not change info
  - Based on file level
  - Give access to someone (everyone or none), some groups, some apps

SOLID - APPLICATIONS

- Many apps are available on https://solidproject.org/use-solid/apps with different functions: social, games, rating , movies...and more
- Tools and libraries to design your own SOLID App are open-source
  - Tutorial: https://solidproject.org/for-developers/apps/first-app
  - Tools list: https://solidproject.org/for-developers/apps/tools

Hands-on Time

PBL Style :)

- 10 minutes together (learn basis) +
- 10 minutes breakout groups (explore by yourself to answer some questions)

2nd 10mins - Breakout groups

Apple:

QA1: What are the differences among access control options (from both file and user level)?
QA2: What if you have two pods and WebID? Can you link them? How?

Banana:

QB2: How can we query data from pods? Can we use SPARQL?
BQ2:Can we query data from multiple pods at once? How? What are the limitations?

Cherry:

QC1: Can one WebID have/control several pods which are hosted by multiple providers? How?
QC2: Can we know if the user has the data we want before query the actual data? How can we have an overview of one data pod without revealing any actual data in the pod?