Law, Science and Technology
MSCA ITN EJD n. 814177

The use of Decentralized and Semantic Web Technologies for Personal Data Protection and Interoperability

Mirko Zichichi

Víctor Rodríguez Doncel
Universidad Politécnica de Madrid
Stefano Ferretti
University of Bologna

11/12/2019

Outline

- Introduction
  - Personal Data
  - Problem
  - GDPR
- State of the Art
  - Semantic Web
  - Solid *by Tim Berners Lee*
  - Distributed Ledger Technologies (DLTs)
- Moving Data Sovereignty Towards Users
- Scenario
- Model Architecture
- Vision

+ Personal Data

- Any piece of information that can **identify** or be identifiable to a natural person

- Generated by the interaction of a user with a software or a hardware in form of:

  numbers, characters, symbols, images, sounds, electromagnetic waves, bits, etc. [1]

- Collected to improve the **safety and security** in citizens surveillance

- But also for a "not so new" **data-driven economy**

+

Problem

- **Smart services** transform data into meaningful information needed by the liveness of the ecosystem they generate.

- These are becoming more and more targeted towards individuals **recommending** them opportunities and making their life easier

Good or Bad?

- Many businesses (**Data Controllers**) rely on data collected about their users, usually storing this personal information in corporate databases (**data silos**)

- Transactions between these businesses happen with **no transparency** for individuals that are not capable of determining the **fate** of their personal data

- Abuse of personal information (Cambridge Analytica 2018, Google's Nightingale 2019)

General Data Protection Regulation (GDPR)

GDPR [2] has empowered data privacy of citizens by radically changing operations carried out by data controllers

Requires data controllers to **release** to their users the complete dataset they collected on them, when requested.

- **No standards** for this requests
- There is the tendency to **hinder the progress** of these

GDPR **data portability** provides the right to have data directly transferred from one data provider to another, making a step towards user-centric platforms of interrelated services

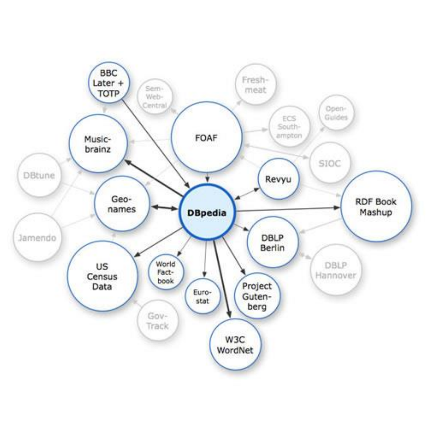
- **Interoperability** [3]

Semantic Web

Extension of the World Wide Web through standards provided by the World Wide Web Consortium (W3C)

Semantic Web brings structure to the meaningful contents of the Web by promoting **common data formats and exchange protocols** [4] e.g.:

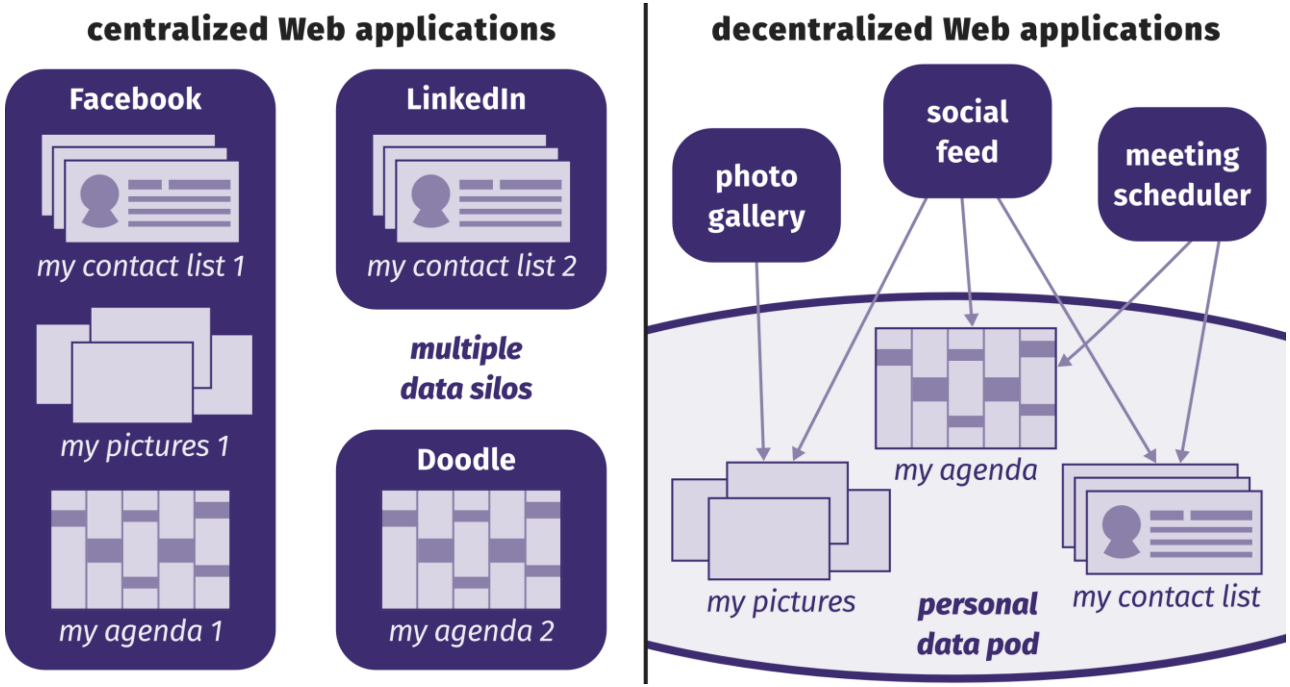
- **RDF** (Resource Description Framework) [5]
- **OWL** (Web Ontology Language) [6]



+ **Linked Data**: data published in a structured manner, in such a way that information can be found, gathered, classified, and enriched using annotation and query languages.

SOLID (Tim Berners Lee's project)

Involves the use of distributed technologies and Semantic Web integration in social networks. Born with the purpose of giving users their data sovereignty, letting them choose where their data resides and who is allowed to access and reuse it [7]



Distributed Ledger Technologies

ethereum

IOTA

- A software infrastructure maintained by a p2p network, where the network participants must reach a **consensus** on the states of transactions submitted to the distributed ledger

- A DLT brings trust when there are several parties that concur in handling some data in a **trustless** manner

- Ethereum **Smart Contract** [8] is a new paradigm of contract that does not completely embodies the same features of a legal contract, but can act as a self-managed structure able to execute code that forces agreements between two or more parts

- SCs remove the technology bond with finance and provide a new paradigm where **unmodifiable instructions** are executed in an **unambiguous manner** during a transaction between two parts

Moving Data Sovereignty Towards Users

Designing methods and systems to support the right of individuals to the **protection** of personal data, at the same favoring its **portability** and economic exploitation and fostering the social good

1. Designing methods and systems that store and transfer personal data in a **controlled, transparent and non-centralized** manner

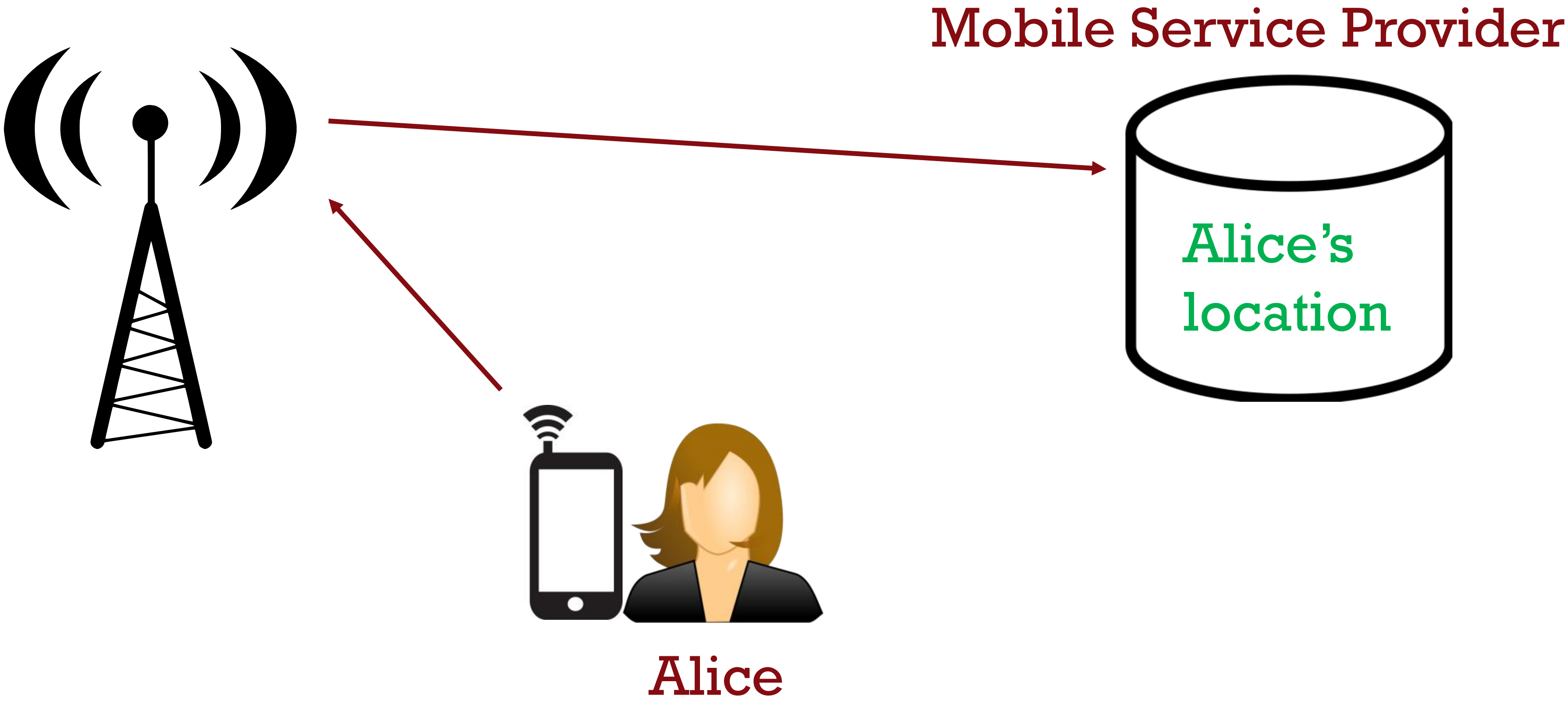
2. Understanding possible actors and manners to **infer** data analyzing social networks

3. Specifying languages and protocols that favor personal data **interoperability**

4. Represent and reason with **policies in smart contracts** to govern the access to personal data

Scenario (1/2)

Individual’s location data generated by a provider



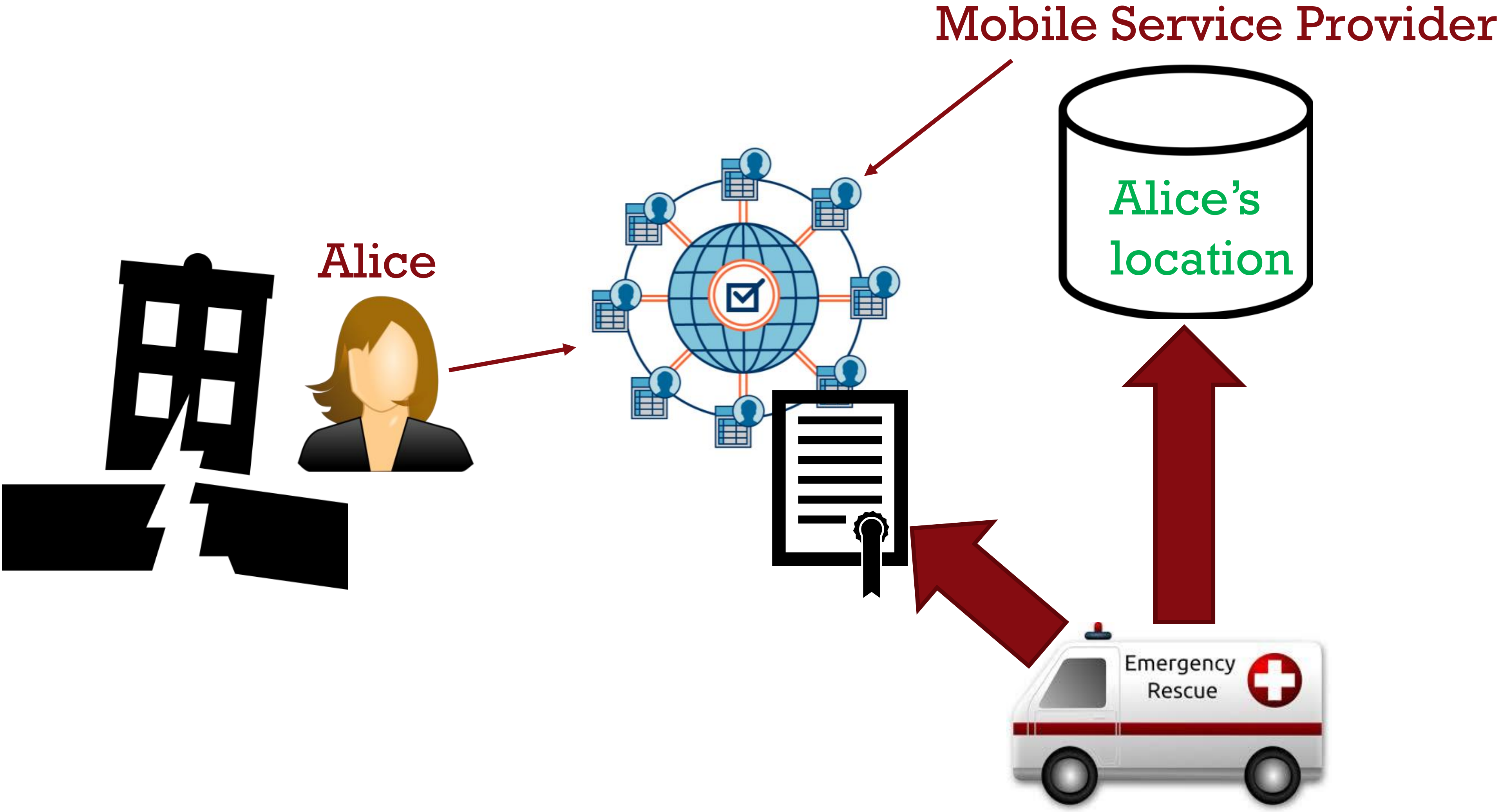
Mobile Service Provider

Alice’s location

Alice

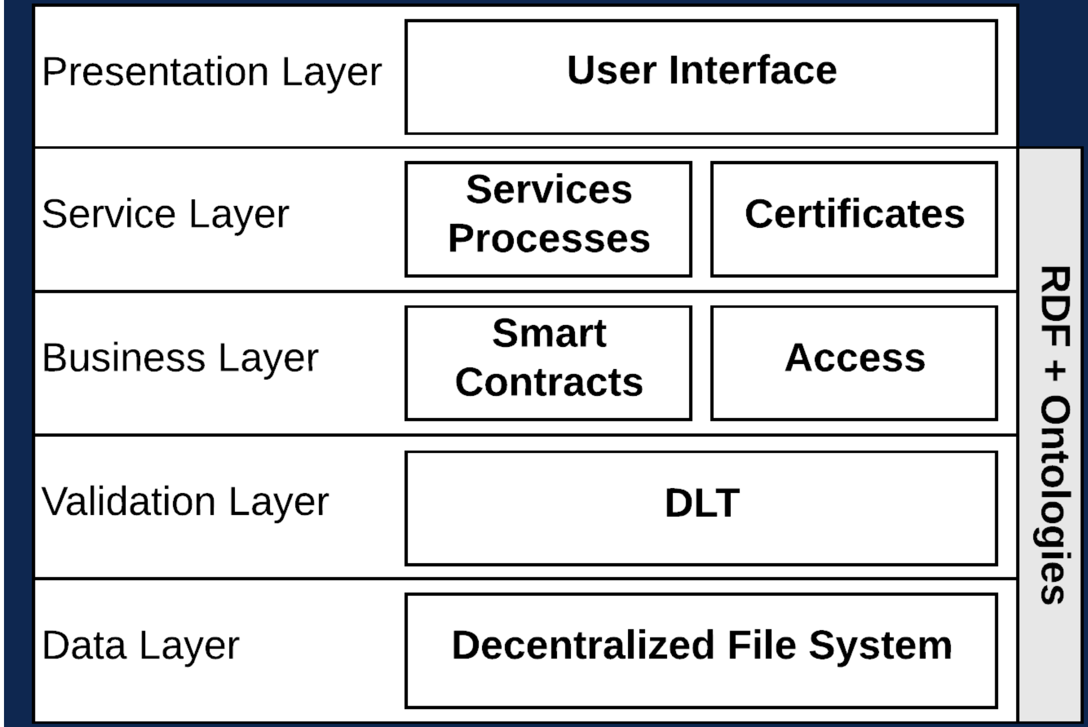
Scenario (2/2)

Individual's location data generated by a provider



Model Layered Architecture

- A unique **databox** for each data subject where **data flow** is ruled and data providers and consumers can meet to transact



Presentation Layer — User Interface

Service Layer — Services Processes — Certificates

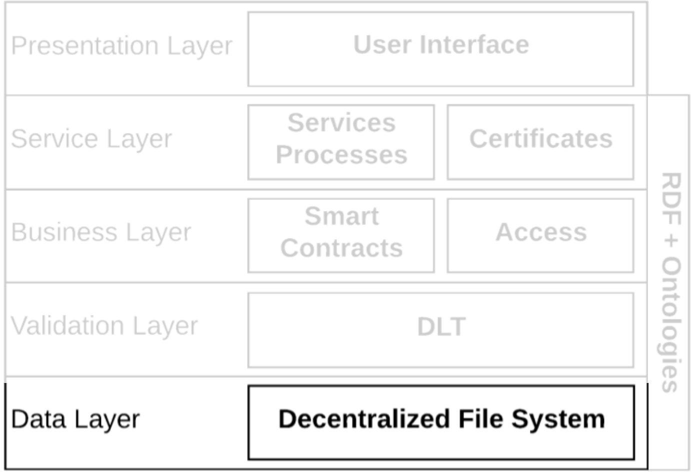
Business Layer — Smart Contracts — Access

Validation Layer — DLT

Data Layer — Decentralized File System

RDF + Ontologies

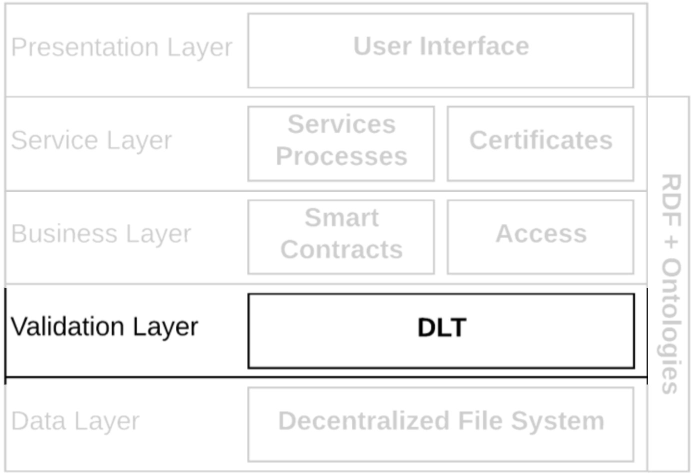
Model Layered Architecture



Presentation Layer	User Interface		
Service Layer	Services Processes	Certificates	RDF + Ontologies
Business Layer	Smart Contracts	Access	
Validation Layer	DLT		
Data Layer	**Decentralized File System**		

- **Decentralized File System**
  e.g. IPFS [9]
  allows storage and continuous data availability

Model Layered Architecture



Presentation Layer — User Interface

Service Layer — Services Processes — Certificates

Business Layer — Smart Contracts — Access

Validation Layer — DLT

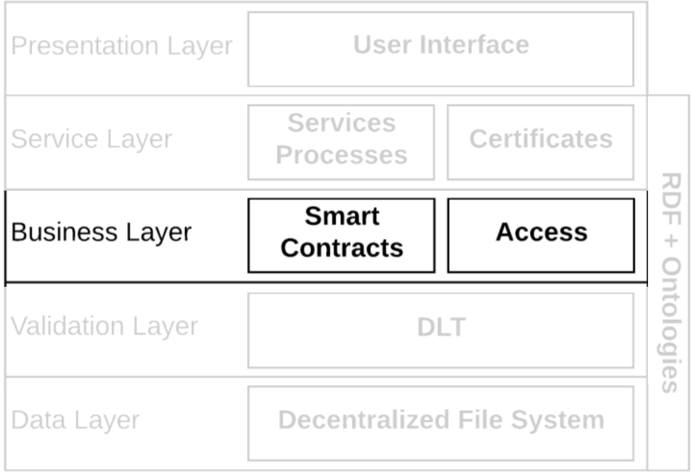
Data Layer — Decentralized File System

RDF + Ontologies

- **Decentralized File System**
  e.g. IPFS [9]
  allows storage and continuous data availability

- **Distributed Ledger Technology**
  e.g. IOTA [10]
  for data validation, no central point of failure, references immutability and most importantly traceability

Model Layered Architecture

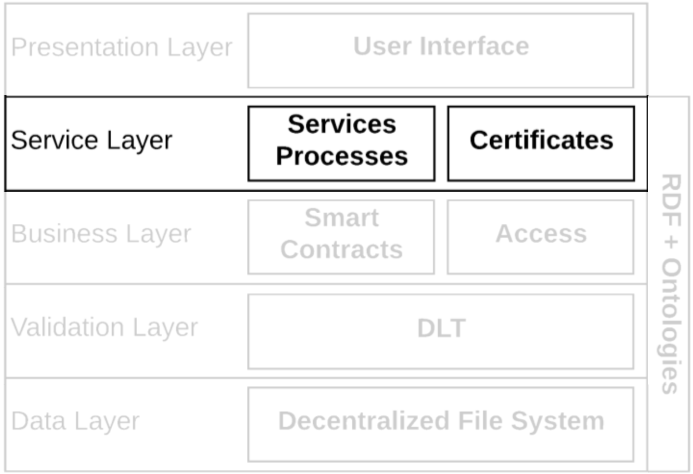


- **Decentralized File System**
  e.g. IPFS [9]
  allows storage and continuous data availability

- **Distributed Ledger Technology**
  e.g. IOTA [10]
  for data validation, no central point of failure, references immutability and most importantly traceability

- **Smart Contracts**
  e.g. Ethereum
  let users completely control the access to their personal data, expressing legal requirements and privacy preferences

Model
Layered
Architecture



Presentation Layer — User Interface

Service Layer — Services Processes; Certificates

Business Layer — Smart Contracts; Access

Validation Layer — DLT

Data Layer — Decentralized File System

RDF + Ontologies

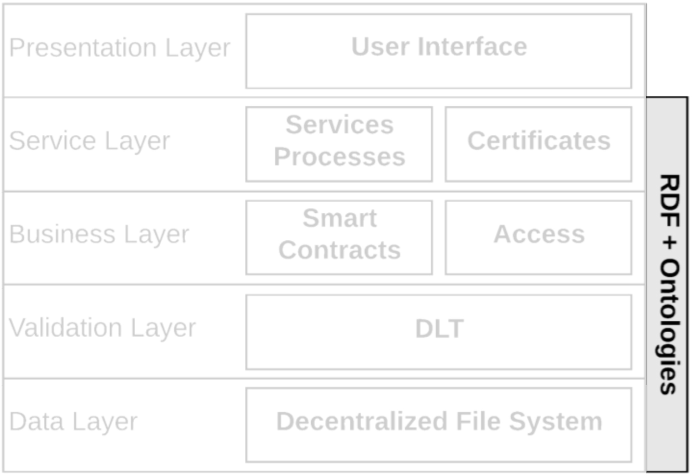
- **Decentralized File System**
  e.g. IPFS [9]
  allows storage and continuous data availability

- **Distributed Ledger Technology**
  e.g. IOTA [10]
  for data validation, no central point of failure, references immutability and most importantly traceability

- **Smart Contracts**
  e.g. Ethereum
  let users completely control the access to their personal data, expressing legal requirements and privacy preferences

- **Services and Certificate for granting Privacy**
  e.g. Zero-Knowledge Proof [11]
  The use of “suitable” data protection techniques allow to prove that an individual possesses a certain property without revealing his data.

Model Layered Architecture



Presentation Layer	User Interface	
Service Layer	Services Processes	Certificates
Business Layer	Smart Contracts	Access
Validation Layer	DLT	
Data Layer	Decentralized File System	

RDF + Ontologies

Semantic web based policies

- Through the use of ontologies it is possible to convey the meaning of data, hence to facilitate cross-domain applications and services
- New ontologies can be created whenever necessary but there is a set of *de facto* standard ontologies which should be reused whenever possible.
- The two advantages of 'interoperability' and 'reasoning' are:
  - Standard ontologies are recommended by the W3C and thus universally understood
  - Reasoning with the information represented using these data models is easy because they are mapped in a formal language

Vision

The main idea is that this model can lead personal data flow towards a "safe" place where the individual can enforce his rights.

- **Individual** are obviously favored because they assumes full control over such databox
- All the actors behind the decentralized structure are incentivized by the use of the **technology specification** itself, e.g. monetary retribution
- **Data providers and consumers** must be incentivized using common standards such as the ones provided by Semantic web, in addition to the GDPR requirements
- The data market generated behind the databox creates a **social system** that is matter of investigation to understand incentives and patterns

References

1. R. Kitchin, The data revolution: Big data, open data, data infrastructures and their consequences. Sage, 2014.
2. Council of European Union, “Regulation (eu) 2016/679 - directive 95/46,” pp. 1–88
3. P. De Hert, V. Papakonstantinou, G. Malgieri, L. Beslay, and I. Sanchez, “The right to data portability inthe gdpr: Towards user-centric interoperability of digital services,”Computer Law & Security Review,vol. 34, no. 2, pp. 193–203, 2018
4. T. Berners-Lee, J. Hendler, O. Lassilaet al., “The semantic web,”Scientific american, vol. 284, no. 5,pp. 28–37, 2001
5. https://www.w3.org/TR/rdf-syntax-grammar/
6. https://www.w3.org/TR/owl-features/
7. A. V. Sambra, E. Mansour, S. Hawke, M. Zereba, N. Greco, A. Ghanem, D. Zagidulin, A. Aboulnaga,and T. Berners-Lee, “Solid : A platform for decentralized social applications based on linked data,”2016
8. V.Buterin et al.,“Ethereum whitepaper” 2013.[Online]. Available: https://github.com/ethereum/wiki/wiki/White-Paper
9. Benet, J.: Ipfs-content addressed, versioned, p2p file system. arXiv preprint arXiv:1407.3561 (2014)
10. Popov, S.: The tangle (2015)
11. Feige, U., Fiat, A., Shamir, A.: Zero-knowledge proofs of identity. Journal of cryptology 1(2) (1988)