Law, Science and Technology
MSCA ITN EJD n. 814177

RIDE
Rights of Internet of Everything

Mirko Zichichi[1,2] and Michele Contu[2],
Stefano Ferretti[2], Víctor Rodríguez-Doncel[1]

[1]Ontology Engineering Group,
Universidad Politécnica de Madrid
[2]Department of Computer Science and Engineering,
University of Bologna

Ensuring Personal Data Anonymity in Data Marketplaces through Sensing-as-a-Service and DLTs

Overview

1. Personal Data
2. Anonymization by Aggregation
3. Service Design
4. Conclusion

Personal Data

Personal Data

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

Personal Data

- Any piece of information that can **identify** or **be identifiable** to a natural person
- Many businesses (**Data Controllers**) rely on data collected about their users, usually storing this personal information in corporate databases (**data silos**)

Personal Data

- Any piece of information that can **identify** or **be identifiable** to a natural person
- 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 provide a new class of **Smart Services** targeted towards individuals, that **recommend** them opportunities in order to make their life easier

Personal Data

- Any piece of information that can **identify** or **be identifiable** to a natural person
- 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 provide a new class of **Smart Services** targeted towards individuals, that **recommend** them opportunities in order to make their life easier

Good or Bad?
Cambridge Analytica 2018
Google's Nightingale 2019

Personal Data

- Any piece of information that can **identify** or **be identifiable** to a natural person
- 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 provide a new class of **Smart Services** targeted towards individuals, that **recommend** them opportunities in order to make their life easier

Good or Bad?
Cambridge Analytica 2018
Google's Nightingale 2019

- These data transactions happen with **no transparency** for individuals, that are not capable of determining the **fate** of their personal data

Databox: Personal Data Management and Interoperability

To ensure **sovereignty** of personal data and its **interoperability** we use the **databox model**: a data store model that acts as a **virtual boundary**, where individuals can **control** *how*, *when* and *what* data is shared with external parties.

Databox: Personal Data Management and Interoperability

To ensure **sovereignty** of personal data and its **interoperability** we use the **databox model**: a data store model that acts as a **virtual boundary**, where individuals can **control** *how*, *when* and *what* data is shared with external parties.

- provides means for individuals to **reflect** on their online presence

Databox: Personal Data Management and Interoperability

To ensure **sovereignty** of personal data and its **interoperability** we use the **databox model**: a data store model that acts as a **virtual boundary**, where individuals can **control** *how*, *when* and *what* data is shared with external parties.

- provides means for individuals to **reflect** on their online presence
- restores **agency** over their data

Databox: Personal Data Management and Interoperability

To ensure **sovereignty** of personal data and its **interoperability** we use the **databox model**: a data store model that acts as a **virtual boundary**, where individuals can **control** *how*, *when* and *what* data is shared with external parties.

- provides means for individuals to **reflect** on their online presence
- restores **agency** over their data
- enables a process of **negotiation** with other parties

Databox: Personal Data Management and Interoperability

To ensure **sovereignty** of personal data and its **interoperability** we use the **databox model**: a data store model that acts as a **virtual boundary**, where individuals can **control** *how*, *when* and *what* data is shared with external parties.

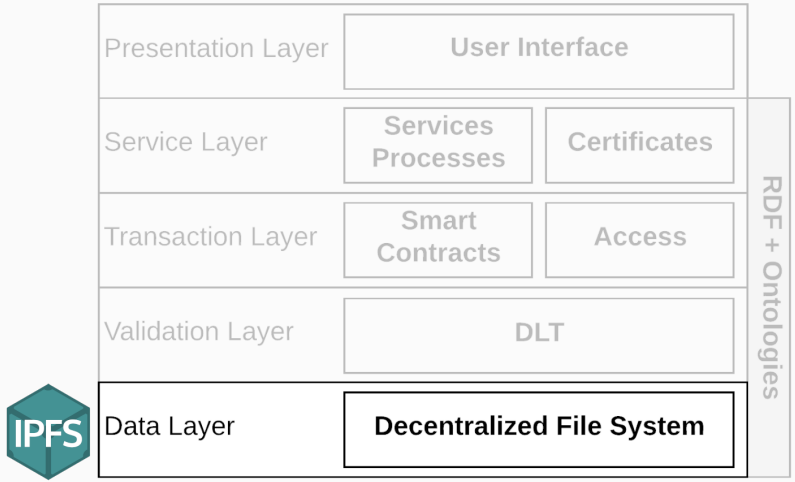
- provides means for individuals to **reflect** on their online presence
- restores **agency** over their data
- enables a process of **negotiation** with other parties
- moves towards the use of personal data for **data markets** and **social good**

Databox: Personal Data Management and Interoperability

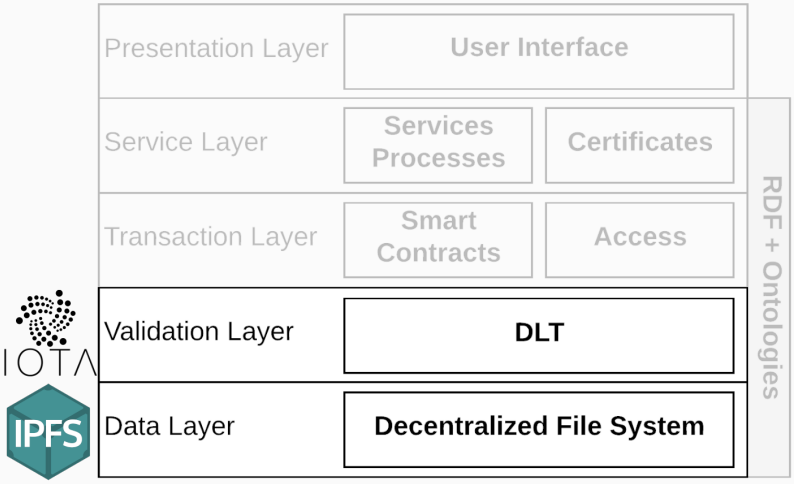
To ensure **sovereignty** of personal data and its **interoperability** we use the **databox model**: a data store model that acts as a **virtual boundary**, where individuals can **control** *how*, *when* and *what* data is shared with external parties.

- provides means for individuals to **reflect** on their online presence
- restores **agency** over their data
- enables a process of **negotiation** with other parties
- moves towards the use of personal data for **data markets** and **social good**
- user shares his data by defining some **policies** and preferences in compliance with GDPR

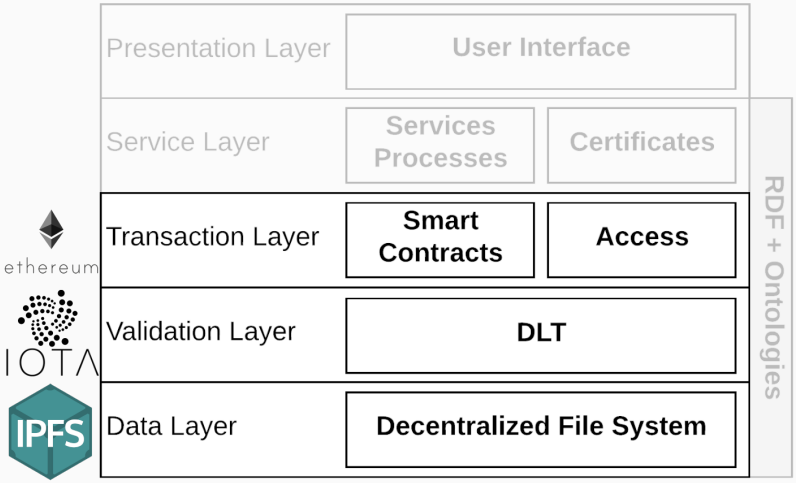
Databox Model Layered Architecture



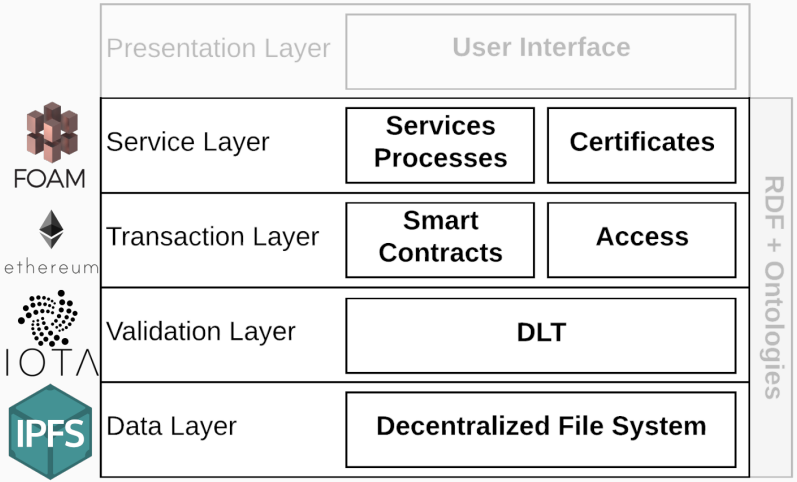
Databox Model Layered Architecture



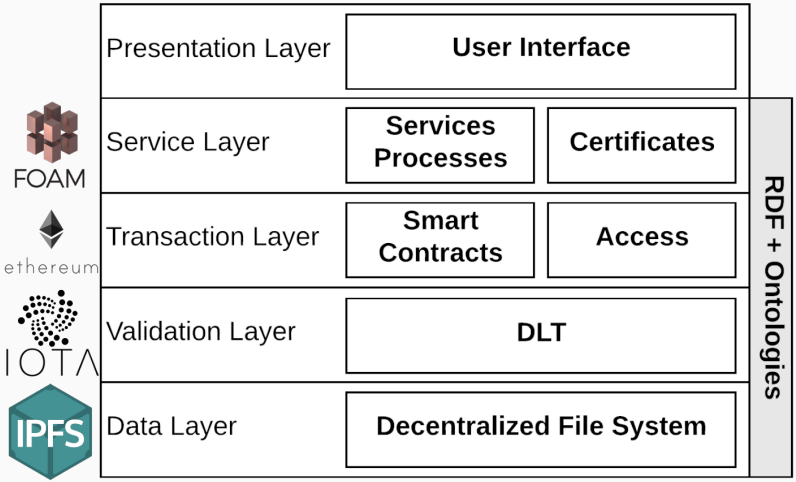
Databox Model Layered Architecture



Databox Model Layered Architecture



Databox Model Layered Architecture



Anonymization by Aggregation

Data Protection Techniques

Removing the association between an individual identifier (e.g. name) and a dataset is not enough

Data Protection Techniques

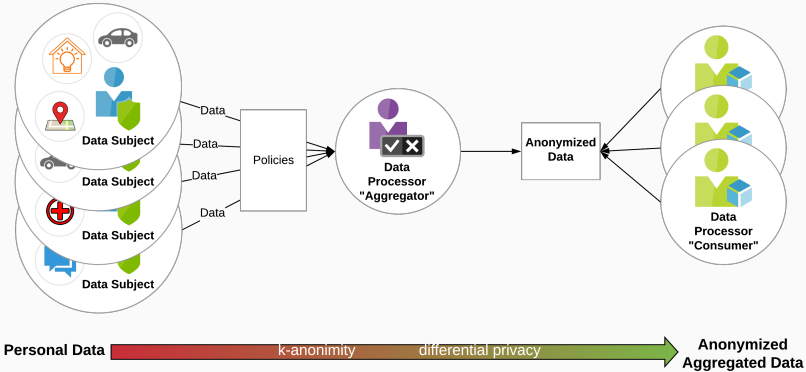
- ***k*-anonymity**
  A data release is said to have this property if the information for each individual contained in the release cannot be distinguished from at least $k - 1$ individuals, whose information also appear in the release

Data Protection Techniques

- ***k*-anonymity**
  A data release is said to have this property if the information for each individual contained in the release cannot be distinguished from at least $k - 1$ individuals, whose information also appear in the release

- **Differential Privacy**
  It is traditionally enforced by adding noise to the data released (typically using the Laplace distribution). Knowing the noise distribution allows to compensate the error when analyzing release.

Sensing-as-a-Service



SaaS has been introduced as a solution based on **IoT** infrastructures, usually implemented as a middleware that **aggregates** data coming from multiple sources, following specific **policies**

Service Design

”Feature” Smart Contract

By combining two kind of contracts data privacy can be incremented, leaving traces of the data flow and giving incentives to all the actors that correctly behave.

”Feature” Smart Contract

By combining two kind of contracts data privacy can be incremented, leaving traces of the data flow and giving incentives to all the actors that correctly behave.

- A **Feature Contract** is a smart contract owned by the data subject that points to a particular kind of data (i.e. the feature) that it is possible to invoke to get **access permissions**

"Feature" Smart Contract

By combining two kind of contracts data privacy can be incremented, leaving traces of the data flow and giving incentives to all the actors that correctly behave.

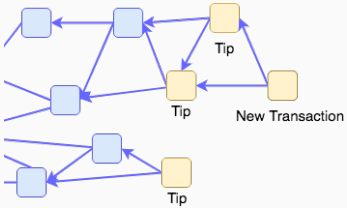
- A **Feature Contract** is a smart contract owned by the data subject that points to a particular kind of data (i.e. the feature) that it is possible to invoke to get **access permissions**
- An **Access Control List (ACL)** that associates an Ethereum address to a bundle of data

”Feature” Smart Contract

By combining two kind of contracts data privacy can be incremented, leaving traces of the data flow and giving incentives to all the actors that correctly behave.

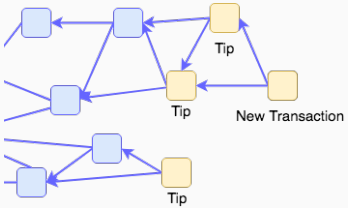
- A **Feature Contract** is a smart contract owned by the data subject that points to a particular kind of data (i.e. the feature) that it is possible to invoke to get **access permissions**
- An **Access Control List (ACL)** that associates an Ethereum address to a bundle of data
- Each feature’s data is maintained in different **MAM channels** →

IOTA MAM Channels



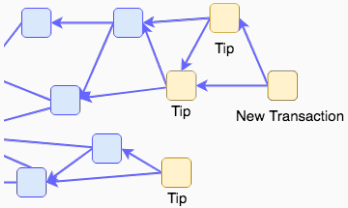
- The **IOTA** ledger is structured as a Direct Acyclical Graph (DAG) called the **Tangle**

IOTA MAM Channels



- The **IOTA** ledger is structured as a Direct Acyclical Graph (DAG) called the **Tangle**
- Vertices represent transactions and edges represent approvals: to issue a new transaction it is necessary to **approve two previous tip transactions**.

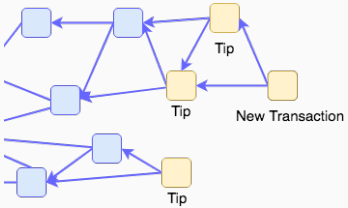
IOTA MAM Channels



- The **IOTA** ledger is structured as a Direct Acyclical Graph (DAG) called the **Tangle**
- Vertices represent transactions and edges represent approvals: to issue a new transaction it is necessary to **approve two previous tip transactions**.

- **Masked Authenticated Messaging (MAM)** is a second layer data communication protocol used to emit and access an **encrypted data stream** over the Tangle

IOTA MAM Channels



- The **IOTA** ledger is structured as a Direct Acyclical Graph (DAG) called the **Tangle**
- Vertices represent transactions and edges represent approvals: to issue a new transaction it is necessary to **approve two previous tip transactions**.

- **Masked Authenticated Messaging (MAM)** is a second layer data communication protocol used to emit and access an **encrypted data stream** over the Tangle
- MAM channels take the form of a **linked list** of transactions ordered in chronological order

”Aggregation” Smart Contract [1/2]

1. **Call for Data** - aggregator indicates the kind of data he is interested in (can be requested by someone else).

"Aggregation" Smart Contract [1/2]

1. **Call for Data** - aggregator indicates the kind of data he is interested in (can be requested by someone else).
   - Subjects interested in participating must provide a ***Proof of Sensing*** (PoSen).

"Aggregation" Smart Contract [1/2]

1. **Call for Data** - aggregator indicates the kind of data he is interested in (can be requested by someone else).
   - Subjects interested in participating must provide a ***Proof of Sensing*** (PoSen).
   - Call duration can depend on a prefixed time of closing or on the number of participants

"Aggregation" Smart Contract [1/2]

1. **Call for Data** - aggregator indicates the kind of data he is interested in (can be requested by someone else).
   - Subjects interested in participating must provide a ***Proof of Sensing*** (PoSen).
   - Call duration can depend on a prefixed time of closing or on the number of participants
2. k-DAO (Decentralized Autonomous Organization) formation

"Aggregation" Smart Contract [1/2]

1. **Call for Data** - aggregator indicates the kind of data he is interested in (can be requested by someone else).
   - Subjects interested in participating must provide a ***Proof of Sensing*** (PoSen).
   - Call duration can depend on a prefixed time of closing or on the number of participants
2. k-**DAO (Decentralized Autonomous Organization) formation**
   - The call is successfully closed only when $k \geq m$ data subjects that took part had been selected by the aggregator.

"Aggregation" Smart Contract [1/2]

1. **Call for Data** - aggregator indicates the kind of data he is interested in (can be requested by someone else).
   - Subjects interested in participating must provide a ***Proof of Sensing*** **(PoSen)**.
   - Call duration can depend on a prefixed time of closing or on the number of participants
2. ***k*-DAO (Decentralized Autonomous Organization) formation**
   - The call is successfully closed only when $k \geq m$ data subjects that took part had been selected by the aggregator.
   - The aggregator stakes a **safety deposit**, used to limit his malicious behavior.

"Aggregation" Smart Contract [1/2]

1. **Call for Data** - aggregator indicates the kind of data he is interested in (can be requested by someone else).
   - Subjects interested in participating must provide a ***Proof of Sensing*** (PoSen).
   - Call duration can depend on a prefixed time of closing or on the number of participants
2. k-**DAO (Decentralized Autonomous Organization) formation**
   - The call is successfully closed only when $k \geq m$ data subjects that took part had been selected by the aggregator.
   - The aggregator stakes a **safety deposit**, used to limit his malicious behavior.
   - The k-DAO indeed, in every moment can decide to **vote** to redeem this deposit if the aggregator **misbehaves**.

"Aggregation" Smart Contract [2/2]

3. **Aggregated data production** - The work of the aggregator is then to produce new data in form of anonymized aggregated data, providing k-anonymity and differential privacy by design.

"Aggregation" Smart Contract [2/2]

3. **Aggregated data production** - The work of the aggregator is then to produce new data in form of anonymized aggregated data, providing k-anonymity and differential privacy by design.
   - The aggregator must provide in the smart contract the exact quantity of data used from each subject's dataset $\rightarrow$ **Merkle trees**.

"Aggregation" Smart Contract [2/2]

3. **Aggregated data production** - The work of the aggregator is then to produce new data in form of anonymized aggregated data, providing k-anonymity and differential privacy by design.
   - The aggregator must provide in the smart contract the exact quantity of data used from each subject's dataset $\rightarrow$ **Merkle trees**.
   - k-DAO members can **validate** data used requesting (off-chain) leaves to the aggregator.

"Aggregation" Smart Contract [2/2]

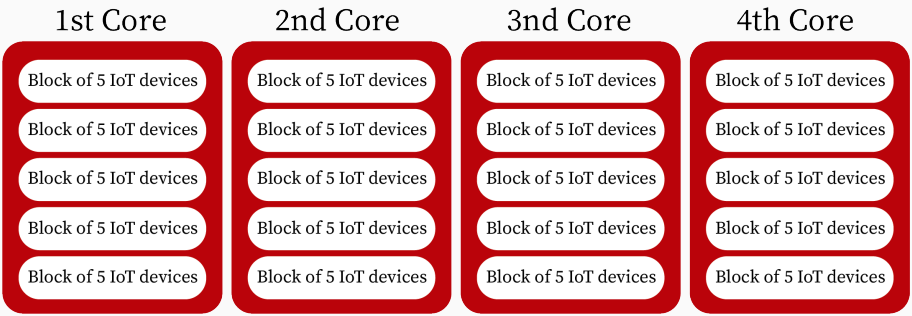
3. **Aggregated data production** - The work of the aggregator is then to produce new data in form of anonymized aggregated data, providing k-anonymity and differential privacy by design.
   - The aggregator must provide in the smart contract the exact quantity of data used from each subject's dataset $\rightarrow$ **Merkle trees**.
   - k-DAO members can **validate** data used requesting (off-chain) leaves to the aggregator.
4. **Aggregated data sale** - Data consumer can access to it through the contract and the payment is proportional to the contribute produced by each participant.

"Aggregation" Smart Contract [2/2]

3. **Aggregated data production** - The work of the aggregator is then to produce new data in form of anonymized aggregated data, providing k-anonymity and differential privacy by design.
   - The aggregator must provide in the smart contract the exact quantity of data used from each subject's dataset $\rightarrow$ **Merkle trees**.
   - k-DAO members can **validate** data used requesting (off-chain) leaves to the aggregator.
4. **Aggregated data sale** - Data consumer can access to it through the contract and the payment is proportional to the contribute produced by each participant.
   - Up to $n < k$ (with n predefined) can ask for a Proof of Sensing to the aggregator in order to check **quality of data**

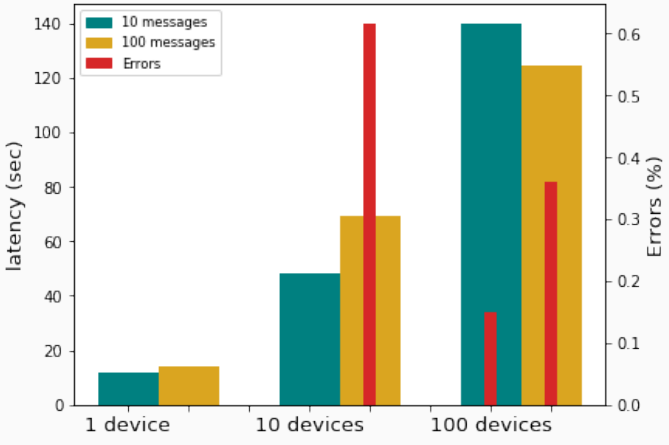
Evaluation: is IOTA network currently a suitable solution for IoT data?

We simulated up to **100 devices** that issue data to MAM Channels, querying randomly between **71 public IOTA full nodes** (remote PoW)



Evaluation: is IOTA network currently a suitable solution for IoT data?

We simulated up to **100 devices** that issue data to MAM Channels, querying randomly between **71 public IOTA full nodes** (remote PoW)



Conclusion

Conclusion

- **Service architecture** where different actors are incentivized to ensure the personal data **anonymity** when it is offered in a marketplace

Conclusion

- **Service architecture** where different actors are incentivized to ensure the personal data **anonymity** when it is offered in a marketplace
- **Personal data** managed through a **databox** with policies compliant to regulations (GDPR)

Conclusion

- **Service architecture** where different actors are incentivized to ensure the personal data **anonymity** when it is offered in a marketplace
- **Personal data** managed through a **databox** with policies compliant to regulations (GDPR)
- Future Work
  - Further experiments with IOTA and other scalable DLTs

Conclusion

- **Service architecture** where different actors are incentivized to ensure the personal data **anonymity** when it is offered in a marketplace
- **Personal data** managed through a **databox** with policies compliant to regulations (GDPR)
- **Future Work**
  - Further experiments with IOTA and other scalable DLTs
  - Complex queries on data stored in DLTs (e.g. Keyword search)