Sharing Personal Data Through Semantic Containers

Christoph Fabianek – OwnYourData
Elmar Kiesling – TU Wien, Linked Data Lab
Fajar Juang Ekaputra – TU Wien, Linked Data Lab
Personal Data Ecosystems

- **Services**
  - Healthcare providers
  - Financial Institutions
  - Mobile providers
  - Educational Institutions

- **Personal Data Store**
  - **Government** (obtain)
  - **Users** (collect)
  - **Researchers** (donate)
  - **Businesses** (provide/sell)
  - **Users** (share)

- **Data Source**
- **Data Using Service**

**Phases**
- request/subscribe
- provide/sell
Interoperability Challenges in Personal Data Ecosystems

- Policies and consent
- Transparency
- Privacy
- Provenance and reproducibility
- Data quality
OwnYourData is a Personal Data Store created in 2015 based on

- Data Sources: connect to data available in your life
- Records: E-2-E encrypted and immutable (blockchain) data vault
- Analysis: weekly reports and interactive visualizations

Extensible through plugins and fine-grained permissions
Functionality (data sources, analysis) encapsulated in docker containers

Exchange data with others
Currently:

data export → email / online storage → data import

Goals:

– Authentication → sign w/ private key
– Data Integrity → store hash in blockchain
– Clean Data → define and verify data quality w/ semantics
– Restrictable Usage → distribute usage policy with data
– Provenance → full history of data sources and processing
– Standardized Access → use REST-API

..and all that fully automated for data provider and consumer
Semantic Containers: Key ideas

- **Encapsulation** (data + processing)
- **Compartmentalization**
- **Self-descriptive** (data and services)
- **Shared vocabularies**
- **Graph-based data model**
- **Higher-level interoperability**
- **Legal Provenance**
Semantic Containers and Interoperability

L4+ Higher-level interoperability (enabled by Semantics):
- Semantic API description
- Approved usage policy/consent management
- Privacy enforcement
- Provenance and integrity tracking
- Data request specifications

L3 Semantic interoperability:
- Common meaning through shared vocabularies
- Foundation for semantic validation, integration, discovery

L2 Syntactic: JSON via RESTful web API

L1 Technical: use of standard internet protocols
Internal data storage and processing technologies are not prescribed.

External interface:
- Machine-readable data and metadata
- Developer-friendly: RESTful, JSON-LD
- Standard (semantic) web technologies

Semantic technologies:
- **Representation:**
  - RDF provides flexible graph-based data model
  - JSON-LD imposes semantic structure on JSON with @context annotations
- **Vocabularies:** RDF Schema
- **API description:** Hydra
- **Validation:** SHACL Shapes Constraint Language
- **Transformation:** RML mapping language to convert structured data into RDF
- Apache Jena for Java-based processing containers (but: flexible)
- in the beginning: configure / make / make install
- now: docker run

- Apply the same principle to data sharing
  - Standardized API
  - encapsulated logic for read/write data, sign & verify data, write to blockchain, store & verify semantics, check usage policy, update provenance
Step 1: Setup container
- Base Container is initialized with semantic description

Result:
- *Publish the container as template for a specific scenario*
Step 2: Fill Container with data
- Upload data yourself or accept data from others

Result:
- Any data written to Semantic Container is validated automatically
- Metadata (provenance, data integrity) is generated automatically
Step 3: Process data

- Apply processing steps

Result:

- Automatic verification of usage policy
- New semantic container with updated metadata
Illustrative UC 1: Earthquake data

Source: ZAMG earthquake data [http://geoweb.zamg.ac.at/static/event/lastday.json](http://geoweb.zamg.ac.at/static/event/lastday.json)

Service: [https://vownyourdata.zamg.ac.at/api/seismic](https://vownyourdata.zamg.ac.at/api/seismic)

- **Authenticity**
  - data signed with ZAMG email

- **Integrity**
  - hash value is store in Ethereum Blockchain

- **Access**
  - REST-API with optional query parameters

Example: earthquakes in the last 7 days in a 100km radius around Vienna
[https://vownyourdata.zamg.ac.at/api/seismic?query=%7B%22lat%22%3A48.21%2C%22long%22%3A16.37%2C%22radius%22%3A100%2C%22duration%22%3A7%7D](https://vownyourdata.zamg.ac.at/api/seismic?query=%7B%22lat%22%3A48.21%2C%22long%22%3A16.37%2C%22radius%22%3A100%2C%22duration%22%3A7%7D)
ZAMG Integration into OwnYourData

Record 56764

<table>
<thead>
<tr>
<th>Field</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>&quot;value&quot;:&quot;e76445456d258a9ef4... show more decrypt data</td>
</tr>
<tr>
<td>Hash Value</td>
<td>4aa0dc2a240d9b1c31c8435bc1e43668f5c97eaddacd2f3ad83e68fd03fb2084</td>
</tr>
<tr>
<td>Merkle Root Hash</td>
<td>6e5310ca29180b0e50dad293076d5d1a91d9fdaf5760a0a804d1b406b6c67c3c8 (show audit proof)</td>
</tr>
<tr>
<td>Transaction</td>
<td>5cb6aa335390a950a16c029b5d9b6f3f3f50226598edcdabc45a8064a14e374 (show on Etherscan.io)</td>
</tr>
<tr>
<td>Blockchain Timestamp</td>
<td>2018-08-13T02:01:56Z</td>
</tr>
<tr>
<td>Verified Source</td>
<td>signed by <a href="mailto:a.krimbacher@zamg.ac.at">a.krimbacher@zamg.ac.at</a></td>
</tr>
</tbody>
</table>

OwnYourData.eu © 2018 Data security Imprint

FAKULTÄT FÜR INFORMATIK
Illustrative UC 2: Data Donation Scenario

- Goal: donate data in a secure and privacy-respecting manner

- Use approach from ZAMG
  - semantically describe data
  - define allowed usage policy (from SPECIAL project)
  - delineate provenance of data

- 2 Challenges:
  - Semantic description
  - Technical infrastructure
Use Case 2: Data Donation

https://donate.ownyourdata.eu
Use Case 2: Data Donation
(Simplified) Example

```xml
<?xml version="1.0" encoding="UTF-8"?>

<Record type="HKQuantityTypeIdentifierStepCount" sourceName="iPhone" sourceVersion="9.0.2" device="&lt;HKDevice: 0x1c0485af0&gt;,
name:iPhone, manufacturer:Apple, model:iPhone, hardware:iPhone8,1,
software:9.0.2" unit="count" creationDate="2018-06-17 07:08:33 +0200"
startDate="2018-06-17 07:00:17 +0200" endDate="2018-06-17 07:01:44 +0200"
value="91"/>

...

<Record type="HKQuantityTypeIdentifierStepCount" sourceName="iPhone" sourceVersion="9.0.2" device="&lt;HKDevice: 0x1c0485af0&gt;,
name:iPhone, manufacturer:Apple, model:iPhone, hardware:iPhone8,1,
software:9.0.2" unit="count" creationDate="2018-06-17 07:08:33 +0200"
startDate="2018-06-17 07:00:17 +0200" endDate="2018-06-17 07:01:44 +0200"
value="91"/>

...

[ ]

[ { "date": "2018-06-17", "value": "3450" },
{ "date": "2018-06-18", "value": "6010" }
]
```
Use Case 2: Data Donation (Simplified) Example

Wrap the data contribution and impose meaning through external vocabulary:

```json
{
    "@context": {
        "sc": "http://semantics.id/ns/semcon#"
    },
    "@graph": [
        {
            "@type": "sc:Container",
            "sc:data": [
                {
                    "@type": "sc:Contribution",
                    "sc:content": {
                        "@type": "sc:ContributionContent",
                        "sc:contributionData": [
                            {
                                "date": "2018-06-17",
                                "value": "3450",
                                "@type": "sc:StepCount"
                            },
                            {
                                "date": "2018-06-18",
                                "value": "6010",
                                "@type": "sc:StepCount"
                            }
                        ]
                    }
                }
            ]
        }
    ]
}
```
Use Case 2: Data Donation (Simplified) Example

What about consent, provenance, and integrity?

```json
{
    "@context": {
        "sc": "http://semantics.id/ns/semcon#",
        "spl": "http://www.specialprivacy.eu/langs/usage-policy#",
        "svl": "http://www.specialprivacy.eu/vocabs/locations#",
        "svn": "http://www.specialprivacy.eu/vocabs/recipient#",
        "svpr": "http://www.specialprivacy.eu/vocabs/processing#",
        "svdu": "http://www.specialprivacy.eu/vocabs/duration#",
        "prov": "http://www.w3.org/ns/prov#",
        "foaf": "http://xmlns.com/foaf/0.1/"
    },
    "@graph": {
        "@type": "sc:Container",
        "sc:data": [
            {
                "@type": "sc:Contribution",
                "sc:content": {
                    "@type": "sc:ContributionContent",
                    "sc:contributionData": [
                        {
                            "date": "2018-06-17",
                            "value": "2345",
                            "@type": "sc:StepCount"
                        },
                        {
                            "date": "2018-06-18",
                            "value": "101",
                            "@type": "sc:StepCount"
                        }
                    ],
                    "sc:contributionConsent": {
                        "spl:hasData": "spl:anyData",
                        "spl:hasDuration": "svdu:StatedPurpose",
                        "spl:hasLocation": [
                            "svl:OurServers",
                            "svl:EU"
                        ],
                        "spl:hasProcessing": [
                            "svpr:Aggregate",
                            "svpr:Analyze",
                            "svpr:Anonymize"
                        ],
                        "spl:hasPurpose": [
                            "svpu:Current",
                            "svpu:Admin",
                            "svpu:Develop",
                            "svpu:Historical"
                        ],
                        "spl:hasRecipient": [
                            "svr:Ours",
                            "svr:Same"
                        ],
                        "spl:hasStorage": "spl:AnyStorage"
                    },
                    "sc:contributionProvenance": {
                        "@type": "prov:Activity",
                        "dc:description": "web site upload",
                        "prov:wasAttributedTo": {
                            "@type": "foaf:Person",
                            "foaf:name": "Christoph Fabianek",
                            "foaf:mbox": "christoph.fabianek@gmail.com"
                        },
                        "prov:wasGeneratedBy": "http://donate.ownyourdata.eu"
                    }
                }
            },
            {
                "@type": "sc:DataContributionProof",
                "sc:contentHash": "7024dabd93cfcc3eab0c0d7265c94a6a7129d6c4effc8b1fc3253e910",
                "sc:contentHashVerification": "https://seal.ownyourdata.eu/?hash=7024dab........e910"
            }
        }
    }
}
```
Conclusions

- **Interoperability**
  - is key for the adoption of personal data stores (ingress and sharing..)
  - is a prerequisite for the emergence of personal data ecosystems
  - Not just on a technical level – also trust, consent, provenance, billing..

- **Semantic containers**
  - can support a wide range of interaction and transaction models
    (data sharing, donation, selling, ..)
  - could become a platform for personal data ecosystems

- **Use cases:**
  - Data Donation as a first use case
  - Various ideas in the works
What’s next?

- Survey and develop vocabularies for various categories of personal data
- Ingress:
  - Data collection: mobile app, standard importers
  - GDPR right of access requests → standard importers (Expedite project)
- Mashups and sharing
- Data market participation (SEMCON project)
- Developer infrastructure
- Incorporate payment

<table>
<thead>
<tr>
<th>What’s next for you..</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Try OwnYourData: <a href="https://ownyourdata.eu">https://ownyourdata.eu</a></td>
</tr>
<tr>
<td>• Propose Data Donation Scenarios: <a href="https://donate.ownyourdata.eu">https://donate.ownyourdata.eu</a></td>
</tr>
<tr>
<td>• Join the W3C Data Privacy Vocabularies and Controls Community Group: <a href="https://www.w3.org/community/dpvCG/">https://www.w3.org/community/dpvCG/</a></td>
</tr>
</tbody>
</table>
Questions for the Audience

- What criteria are necessary to proof data ownership and right to use?
- How do you like the approach from a developer perspective?
- Is Docker a good platform choice?
- What other use cases for semantic containers come to your mind?
- Do you see any conflict between the encapsulation and self-containment of semantic containers and the Linked Open Data idea?
- What technical and non-technical factors could drive/hinder participation in personal data ecosystems?