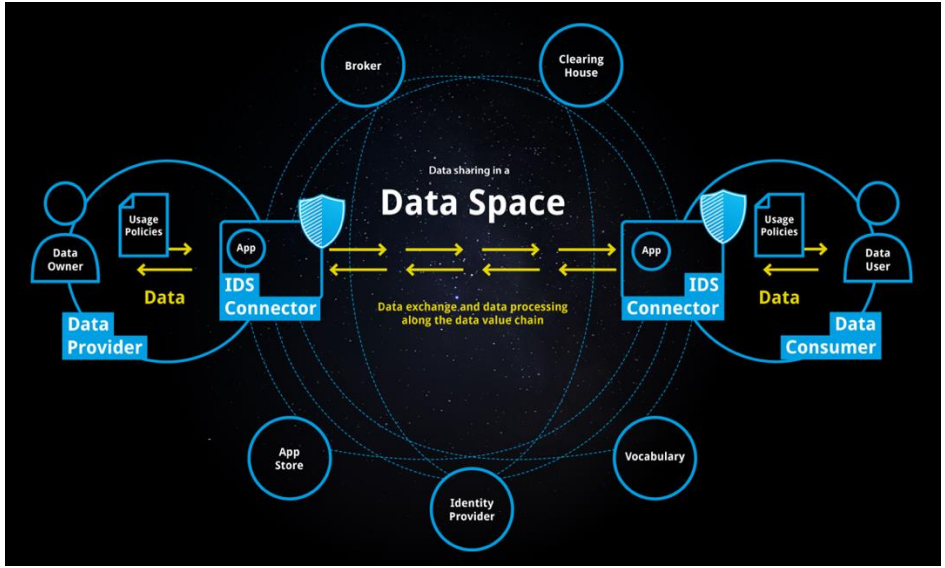
eSAAM 2024

Privacy-friendly sharing of health data using a reference architecture

A.Shayan Ahmadian
Postdoctoral researcher
universität koblenz
weiter:denken

Health Data Spaces – Motivation and challenges (1/2)



https://internationaldataspaces.org/why/data-spaces/

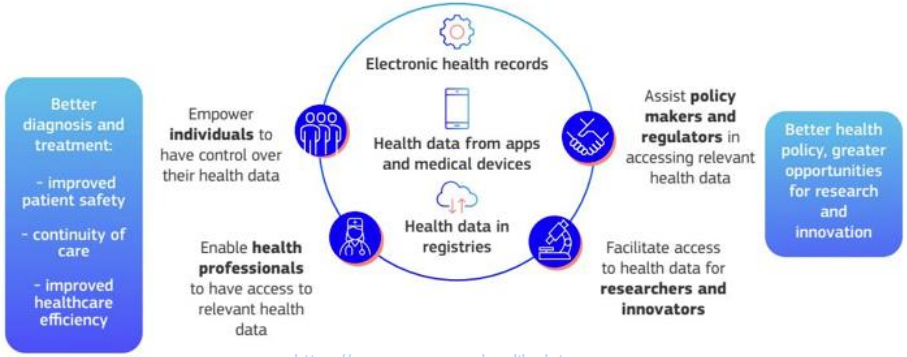
Strategic requirements of International Data Spaces (IDS):

Trust
Security and data sovereignty
Ecosystem of data
Standardized interoperability
Value adding apps
Data market

Data spaces for different domains:

Automotive industry (Catena-X)
Mobility, Transport and Tourism (EONA-X)
Skills analytics and matching (Prometheus-X)

Health Data Spaces – Motivation and challenges (2/2)



https://www.european-health-data-space.com

- Empowering **individuals** through increased digital **access** to and **control** of their electronic personal health data.
- Fostering a **single market** for electronic health record systems, relevant medical devices and high risk AI systems.
- Providing a trustworthy and efficient set-up for the use of health data for **research**, **innovation**, **policy-making** and **regulatory** activities (secondary use of data).

Health Data Spaces – Motivation and challenges (2/2)

- Empowering individuals through increased digital access to and control of their electronic personal health data.
- Fostering a single market for electronic health record systems, relevant medical devices and high risk AI systems.
- Providing a trustworthy and efficient set-up for the use of health data for research, innovation, policy-making and regulatory activities (secondary use of data).

Health Data Spaces – Motivation and challenges (2/2)

Health data spaces face critical challenges in addressing privacy concerns… [1]

Maintain the **privacy** and **confidentiality** of sensitive health data.

Public embarrassment

RQ1: How can a data sharing contract be designed for health data spaces to specify privacy and security requirements?

RQ2: How can the IDS RAM be adapted to realize a privacy friendly sensitive data sharing in health data spaces?

[1] J. S. Marcus et. al., The European Health Data Space. Policy Department for Economic, Scientific and Quality of Life Policies Directorate-General for Internal Policies. December 2022.

There is no formal or widely adopted structure for contract offers when it comes to health data spaces

A structured contract:

- Outlines the scop, purpose, responsibilities and terms of data sharing.
- Establishes a framework that reduces ambiguity, mitigates disputes and facilitates communication.
- Fosters trust and compliance.

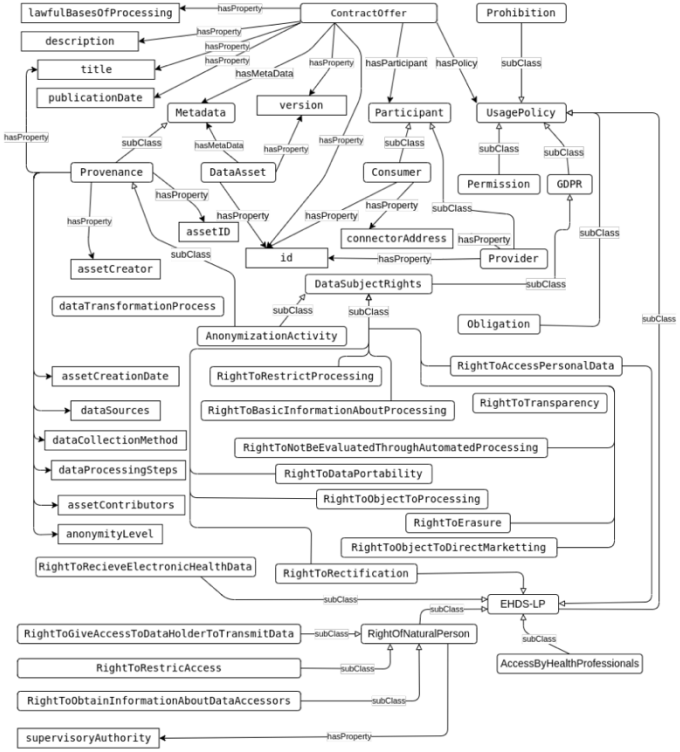
First an **ontology** that represents the **domain** of **contracts** in **heath** data spaces is proposed.

The ontology defines the foundational terminologies to establish a clear and unambiguous understanding of key terms.

Initially GDPR and EHDS are used to establish an ontology

The **state of the art** is analyzed and we leverage the terminologies presented in:

- Usage Contract Language (IDS RAM),
- Policy Expression technique,
- ODRL,
- GDPRtEXT,
- PROV-O,



An excerpt structure of the contract offers

```
{
    "@context": {
        "co": "http://contract.contology#",
        "consumer": {
            "@type": "co:Consumer" },
        "contract_id": {
            "@type": "co:contractID" },
        "contract_version": {
            "@type": "co:contractVersion" },
        "data_asset": {
            "@type": "co:DataAsset" },
        "description": {
            "@type":  "co:description" },
        "lawful_basis_for_sharing" {
            "@type": "xsd:string" },
        "meta_data": {
            "@type": "co:Metadata" },
        "policies": {
            "@type": "co:Policy",
            "@container": "@set" },
        "provider": {
            "@type": "co:Provider" },
        "publication_date": {
            "@type": "co:publicationDate" },
        "title": {
            "@type": "xsd:string" },
        "version": {
            "@type": "co:version"} } }
```

Unique identifier, title, description, metadata, parties involved (provider, consumer), policies.

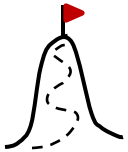
An excerpt structure of the contract offers - Example

```
{
    "@context": {
        "co": "http://contract.contology#",
        "consumer": {
            "@type": "co:Consumer" },
        "contract_id": {
            "@type": "co:contractID" },
        "contract_version": {
            "@type": "co:contractVersion" },
        "data_asset": {
            "@type": "co:DataAsset" },
        "description": {
            "@type": "co:description" },
        "lawful_basis_for_sharing" {
            "@type": "xsd:string" },
        "meta_data": {
            "@type": "co:Metadata" },
        "policies": {
            "@type": "co:Policy",
            "@container": "@set" },
        "provider": {
            "@type": "co:Provider" },
        "publication_date": {
            "@type": "co:publicationDate" },
        "title": {
            "@type": "xsd:string" },
        "version": {
            "@type": "co:version"} } }
```

The data provider does not want **the shared data** to be **retained** for more than 3 months (Article 5(1)c of the GDPR known as **data minimization**).

```
{ "policy-class": "dataMinimization",
  "policy-properties": {
      "action": "delete",
      "constraint": [{
          "leftOperand": "duration",
          "operator": "lt",
          "rightOperand": {
              "@value": "P90D",
              "@type": "xsd:date" } }] } }
```

An excerpt structure of the contract offers - Example

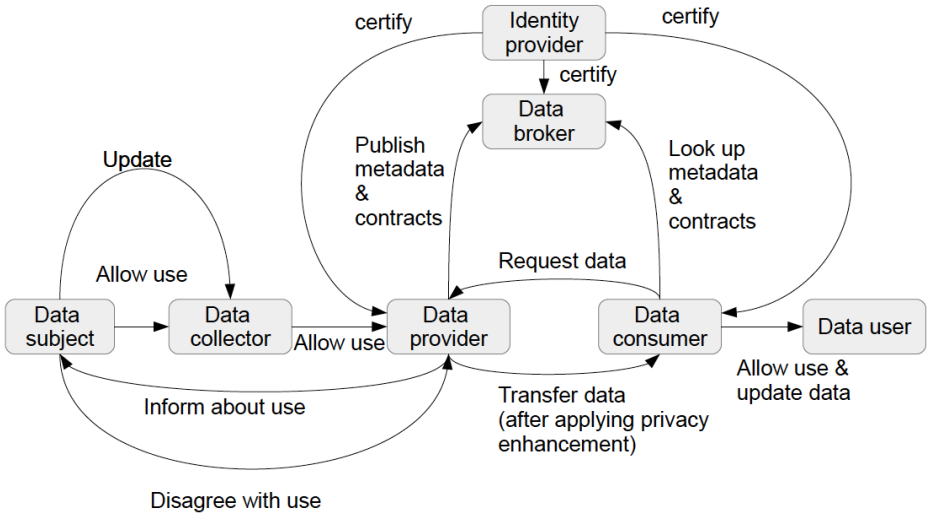


RQ1: How can a data sharing contract be designed for health data spaces to specify privacy and security requirements?

The data provider does not want **the shared data** to be **retained** for more than 3 months (Article 5(1)c of the GDPR known as **data minimization**).

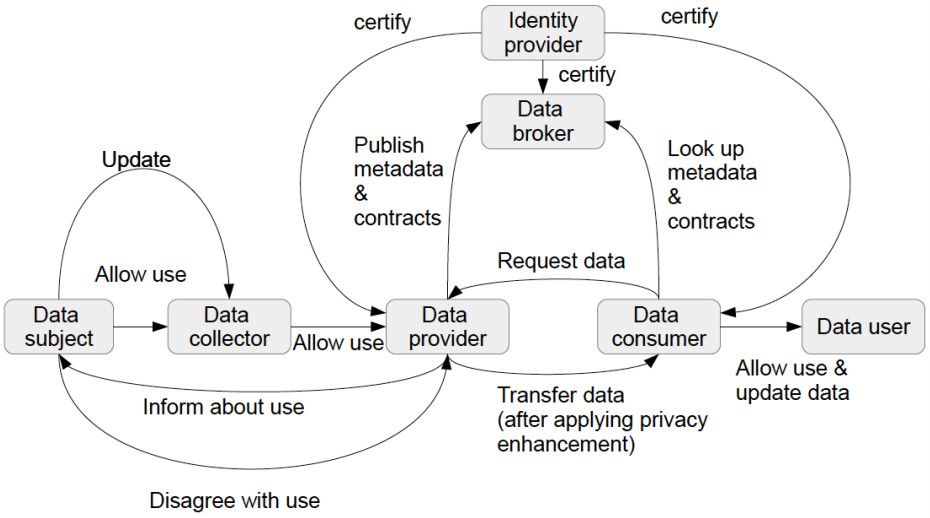
```
{ "policy-class": "dataMinimization",
  "policy-properties": {
      "action": "delete",
      "constraint": [{
          "leftOperand": "duration",
          "operator": "lt",
          "rightOperand":  {
              "@value": "P90D",
              "@type": "xsd:date" } }] } }
```

Our reference architecture is an adaption of the IDS RAM



The **data subject** (patient) is introduced as an active role that has **control** over the.

Our reference architecture is an adaption of the IDS RAM

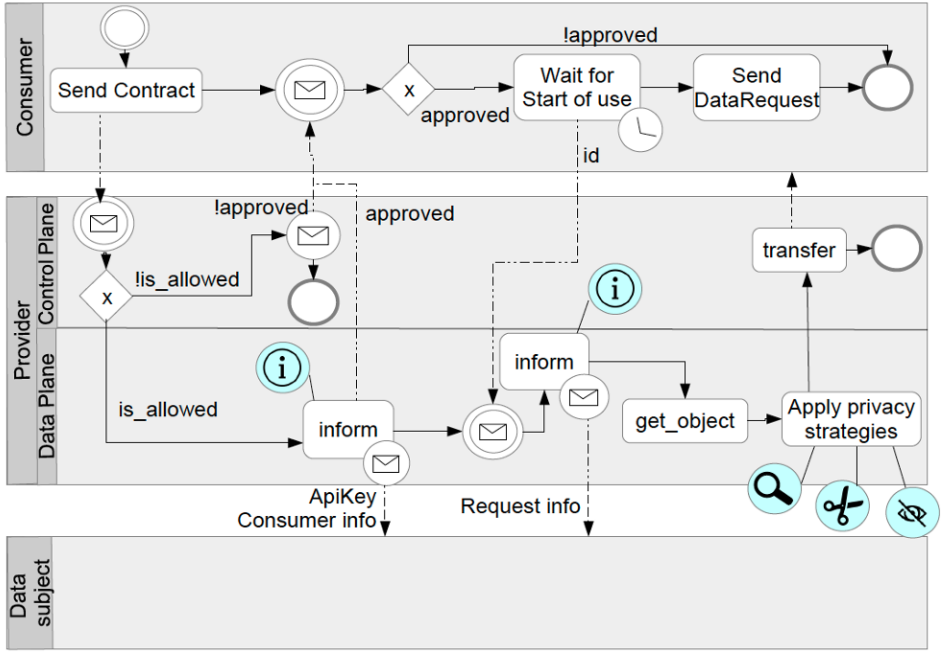


- **Minimise**: Sensitive data that is not needed should be deleted.
- **Abstract**: Sensitive data should be only processed in the highest required granularity.
- **Hide**: The access to sensitive data should only be granted to users with the right clearance.
- **Separate**: The processing or storing sensitive data should occur on different locations to avoid profiling.
- **Inform**: The person to whom the data belongs, is informed about the data processing.
- **control**: The person to whom the data belongs has the ability to update or delete their data.
- **Enforce**: All parties should enforce appropriate rules in their entire structure.
- **Demonstrate**: All parties should demonstrate that they obey the rules and how they do it.

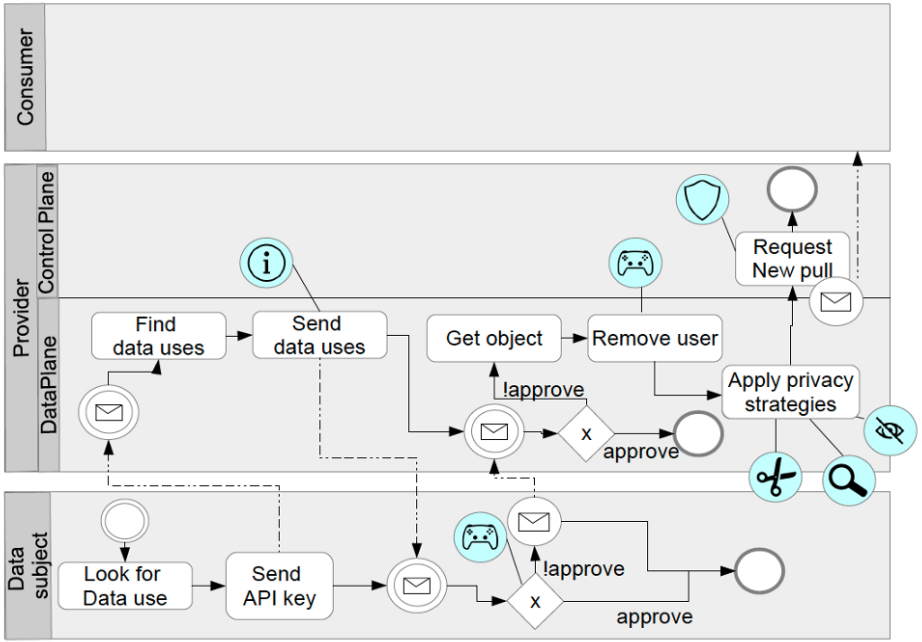
The **data subject** (patient) is introduced as an active role that has **control** over the.

Well-established **privacy design strategies** are used to realize privacy by design principle (**GDPR - Article 25**).

The interactions between the components are adapted to particularly give data subjects control over their data

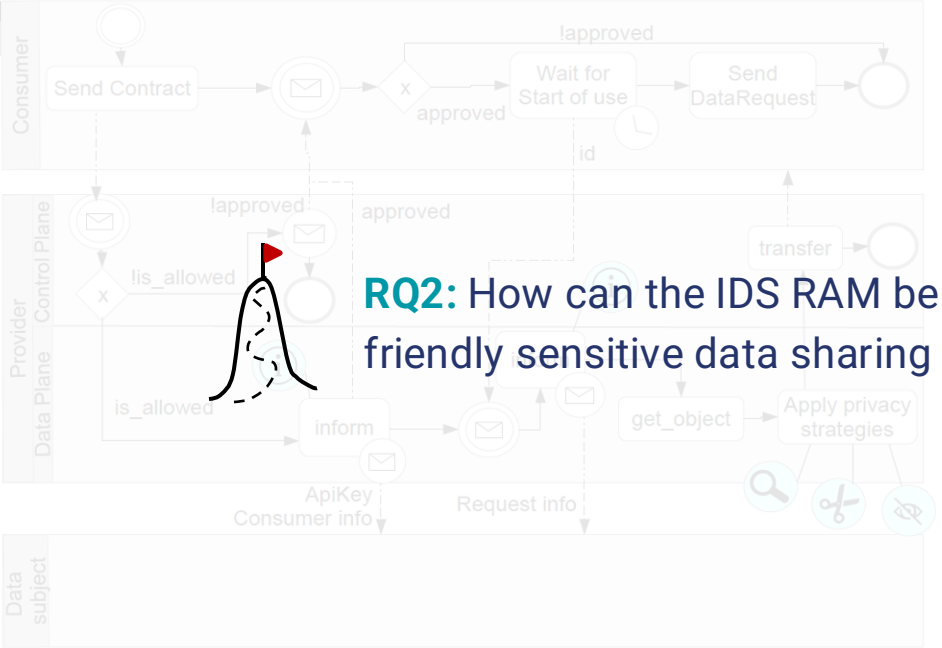


The data flow of a **data transfer**

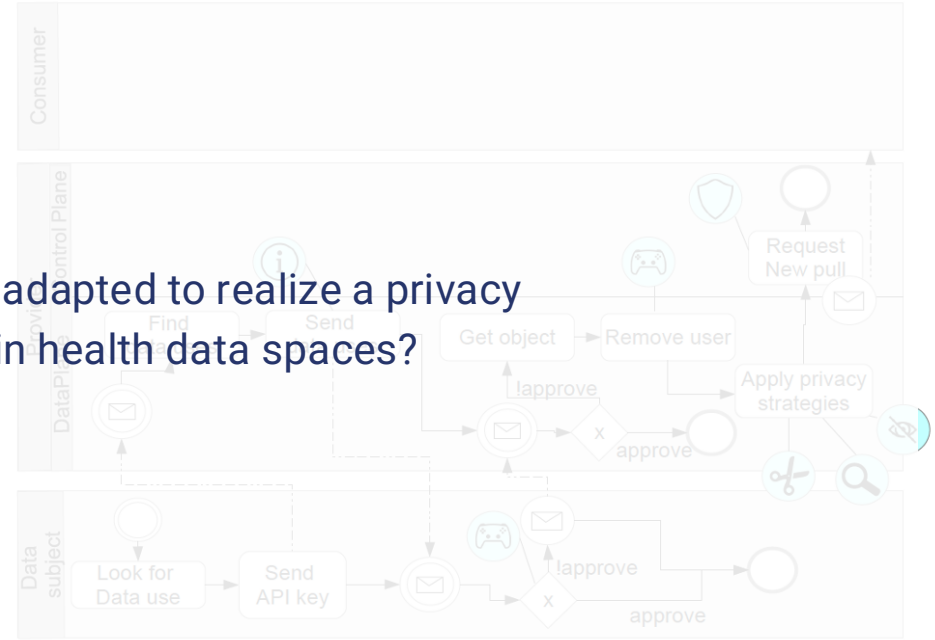


The data flow of an **update operation**

The interactions between the components are adapted to particularly give data subjects control over their data



The data flow of a data transfer



The data flow of an update operation

RQ2: How can the IDS RAM be adapted to realize a privacy friendly sensitive data sharing in health data spaces?

Evaluation – An ontology should accurately capture the concepts with the seeking domain (1/2)

Our focus will be accuracy and coverage. The evaluation metrics are defined as follows:

$$Precision = TP/(TP + FP)$$

$$Recall = TP/(TP + FN)$$

$$F1 - Score = 2 * (Precision * Recall)/(Precision + Recall)$$

$$Coverage = \left(\frac{\#OC \in corpus}{|ontology|}\right)$$

TP is the correctly identified concepts from the ontology in the reference corpus
FP is incorrectly identified concepts
FN are concepts not identified
$\#OC \in corpus$ specifies the count of concepts of ontology in the corpus
$|corpus|$ and $|ontology|$ signify the total number of concepts in the corpus and the ontology

Evaluation – An ontology should accurately capture the concepts with the seeking domain (2/2)

Precision, **Recall** and **F1-Score** are widely used in information retrieval and natural language processing to evaluate the accuracy of ontology-based information extracting.

Table 1: Contract ontology evaluation properties.

	TP	FP	FN	$\#OC \in corpus$	$\lvert corpus \rvert$
GDPR	178	0	221	178	3826
EHDS	136	0	263	136	1344

Table 2: Contract ontology evaluation results.

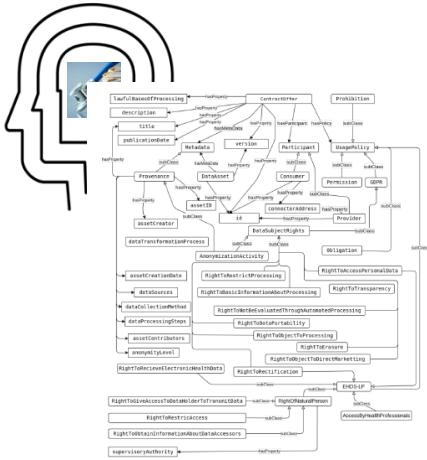
	Precision	Recall	F1-Score	Coverage
GDPR	1.0	0.45	0.62	0.09
EHDS	1.0	0.34	0.51	0.07

Refine the **ontology** and the **contract** structure definition by considering various **regulations** such as cyber security act, resilience act, and AI Act.

Refine the **methodology** of identifying key terms in the regulations **coding** (for instance Hypothesis coding).

How **statically** (design time) and **dynamically** (run time) the principle of **data minimization** can be realized?

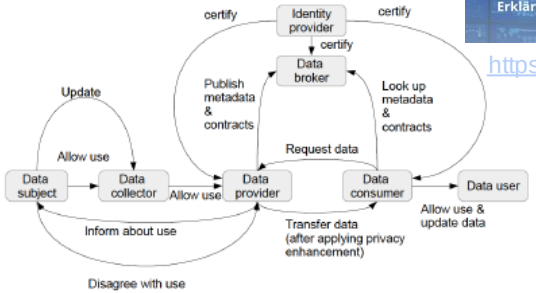
Privacy-friendly sharing of health data using a reference architecture for health data spaces

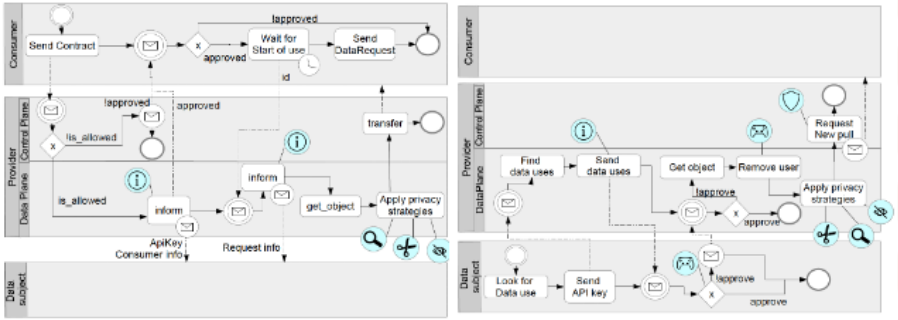


RQ1: How can a data sharing contract be designed for health data spaces to specify privacy and security requirements?


```
{ "policy-class": "dataMinimization",
  "policy-properties": {
      "action": "delete",
      "constraint": [{
          "leftOperand": "duration",
          "operator": "lt",
          "rightOperand": {
              "@value": "P90D",
              "@type": "xsd:date" } }] } }
```


https://covid-ai.uni-koblenz.de





RQ2: How can the IDS RAM be adapted to realize a privacy friendly sensitive data sharing in health data spaces?

Thank you