

# Analyze Decentralized Personal Health Data using Solid, Digital Consent, and Federated Learning

Citation for published version (APA):

Sun, C., van Soest, J., & Dumontier, M. (2023). Analyze Decentralized Personal Health Data using Solid, Digital Consent, and Federated Learning. In *Semantic Web Applications and Tools for Health Care and Life Sciences 2023* (Vol. 3415, pp. 169-170)  
[https://bbuelens.solidcommunity.net/public/html/SolidSymposium2023\\_ChangSun.pdf](https://bbuelens.solidcommunity.net/public/html/SolidSymposium2023_ChangSun.pdf)

## Document status and date:

Published: 01/01/2023

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

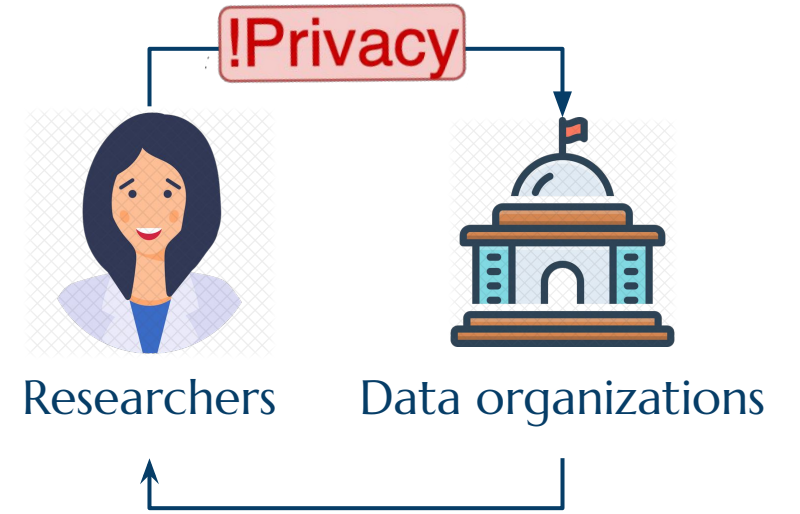
Download date: 29 Apr. 2024



# Analyze Decentralized Personal Health Data using Solid, Digital Consent, and Federated Learning

**Chang Sun, Ph.D.**  
Institute of Data Science, Maastricht University

Solid Symposium 2023 - Applications for Health Data Management with Solid  
31 March 2023





## Title and reference

**Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance)**

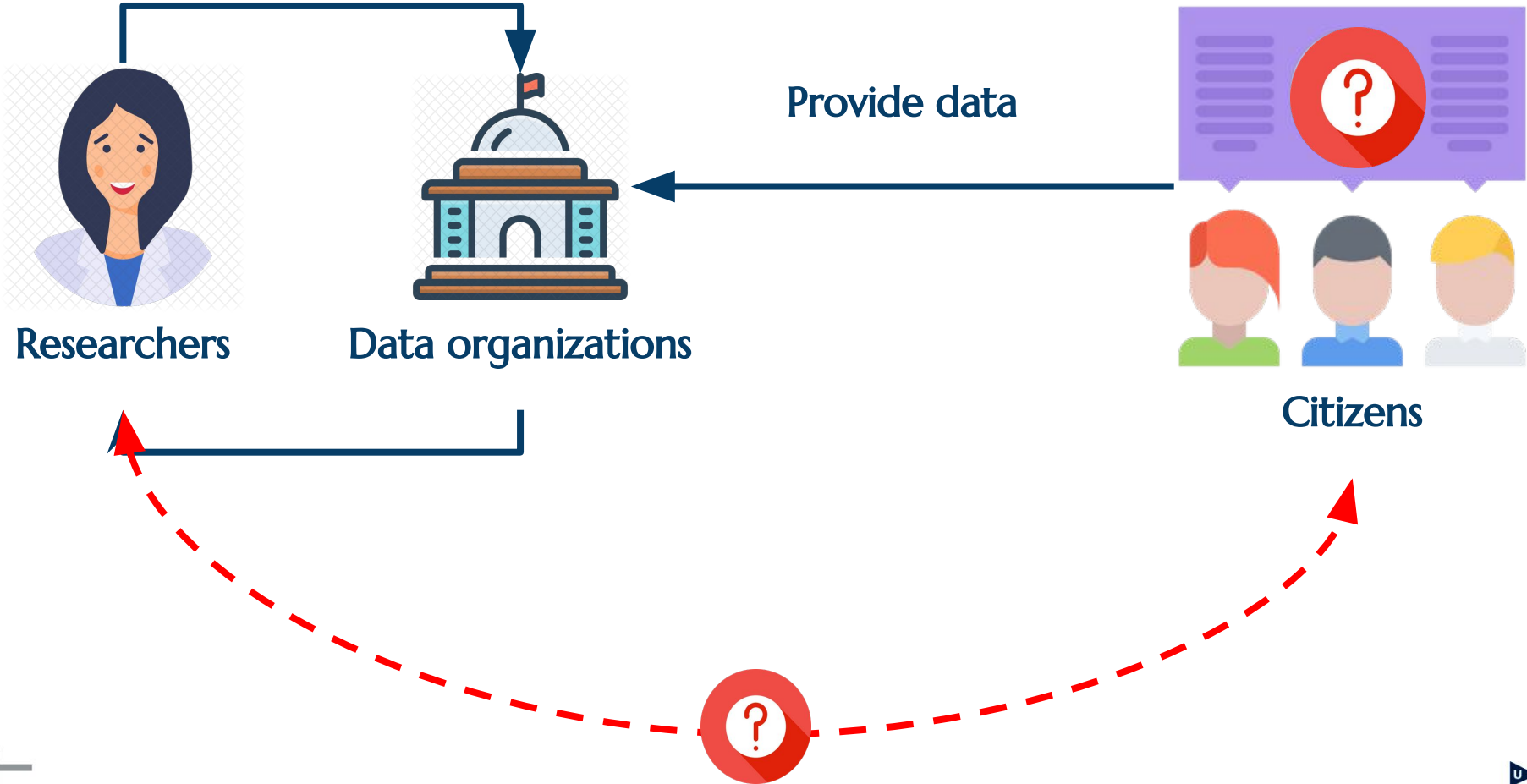
*OJ L 119, 4.5.2016, p. 1–88 (BG, ES, CS, DA, DE, ET, EL, EN, FR, GA, HR, IT, LV, LT, HU, MT, NL, PL, PT, RO, SK, SL, FI, SV)*

In force: This act has been changed. Current consolidated version: [04/05/2016](#)

ELI: <http://data.europa.eu/eli/reg/2016/679/oj>



A lot work focused on analysing the data controlled by organizations, but what about using data directly controlled by individuals?

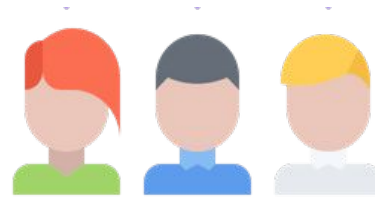


## Citizens would like to have more access and control over the data they generate.

5

*EU should make major investments in technologies and infrastructures that **enhance data access and use**, while giving us as well as public and private organisations **full control over the data they generate**.*

***Law and technology** should enable us to make available their data for the public interest without any direct reward.*



*We would be willing to **make such data available**, especially for health-related research and for aspects relating to the locality we live in (e.g. mobility, environment).*

*There are **no sufficient tools and mechanisms** for us to 'donate' our data.*

# A key challenge

6

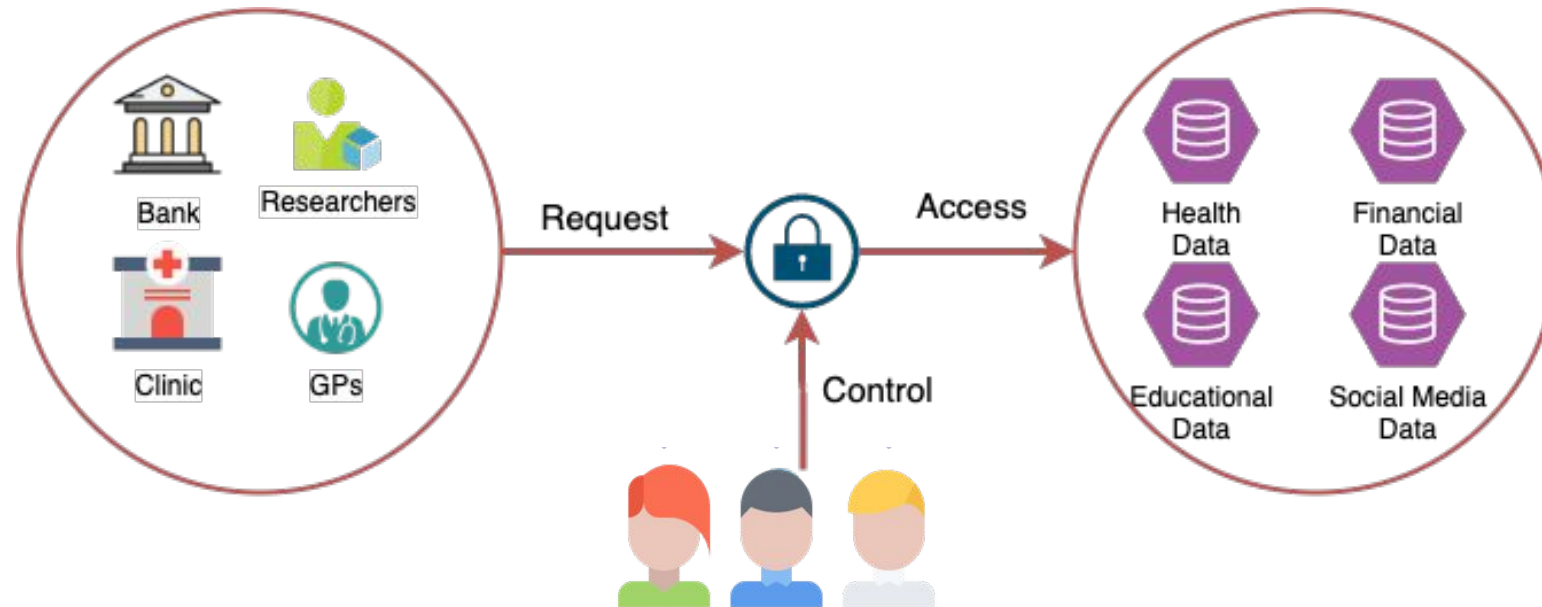
*“Individuals value the high level of protection granted by the GDPR and ePrivacy legislation. However, they suffer from the absence of technical tools and standards that make the exercise of their rights simple and not overly burdensome.”*

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



# Solutions:

provide individuals **the tools** to decide at a **granular level** what is done with their data using consent management, personal information management, decentralised networking, etc, in order to give greater oversight, transparency, control over their personal data.



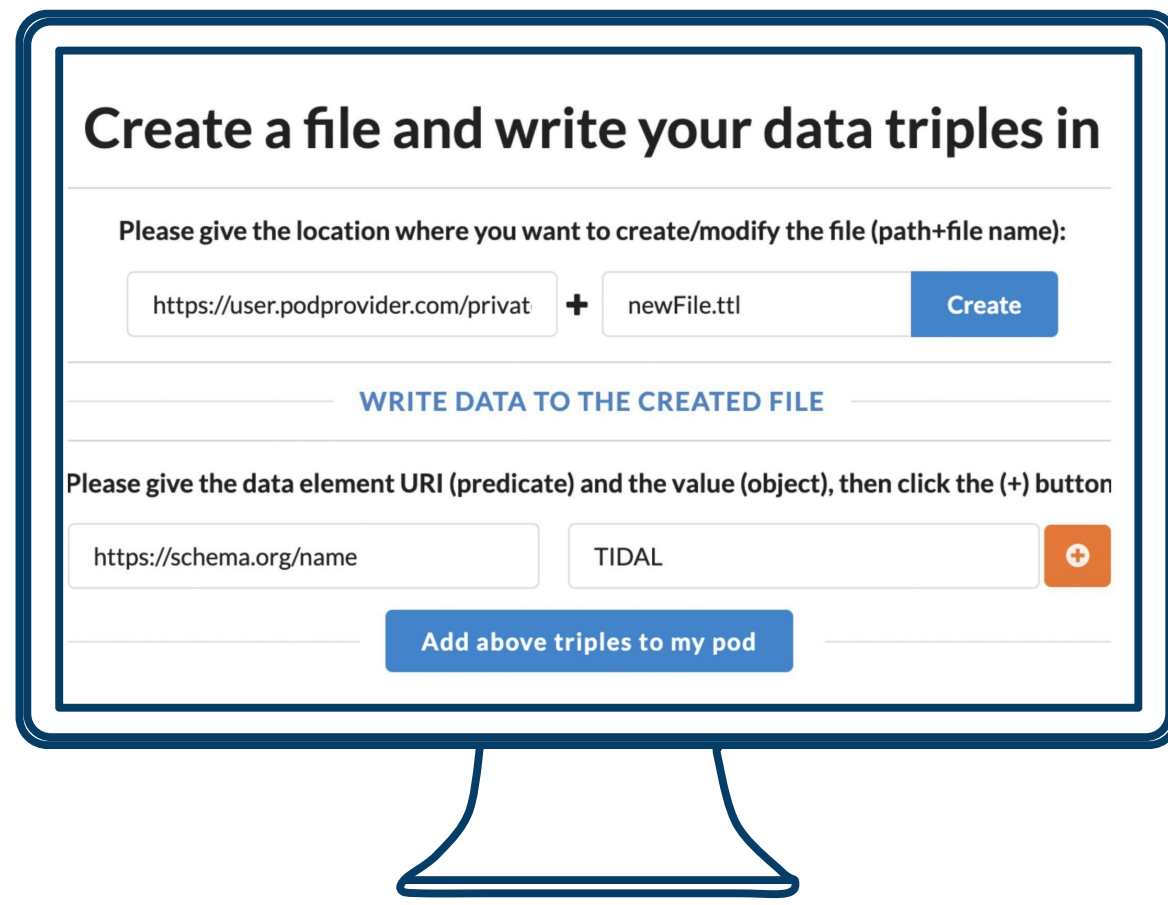
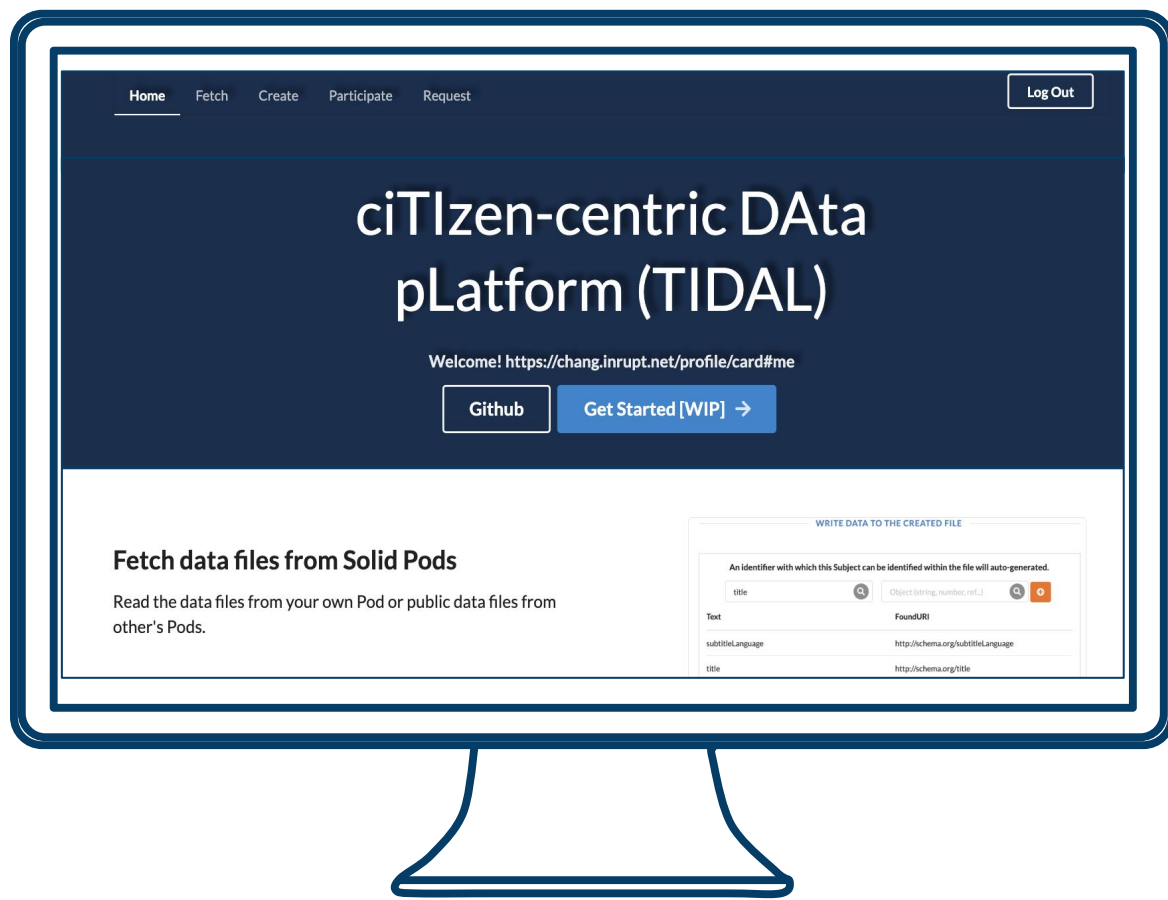
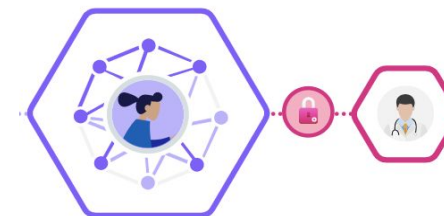


# ciTizen-centric DAta pLatform (TIDAL)

8



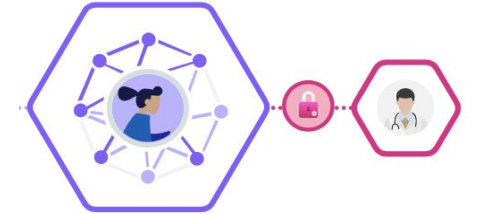
**Based on Solid (SOcial Linked Data) to store and structure personal data, and give individual control access to their fine-grained personal data**



# ciTizen-centric DAta pLatform (TIDAL)



Based on Solid (SOcial Linked Data) to store and structure personal data, and give individual control access to their fine-grained personal data



Integrating Data Privacy Vocabulary to structure personal data requests as digital consents

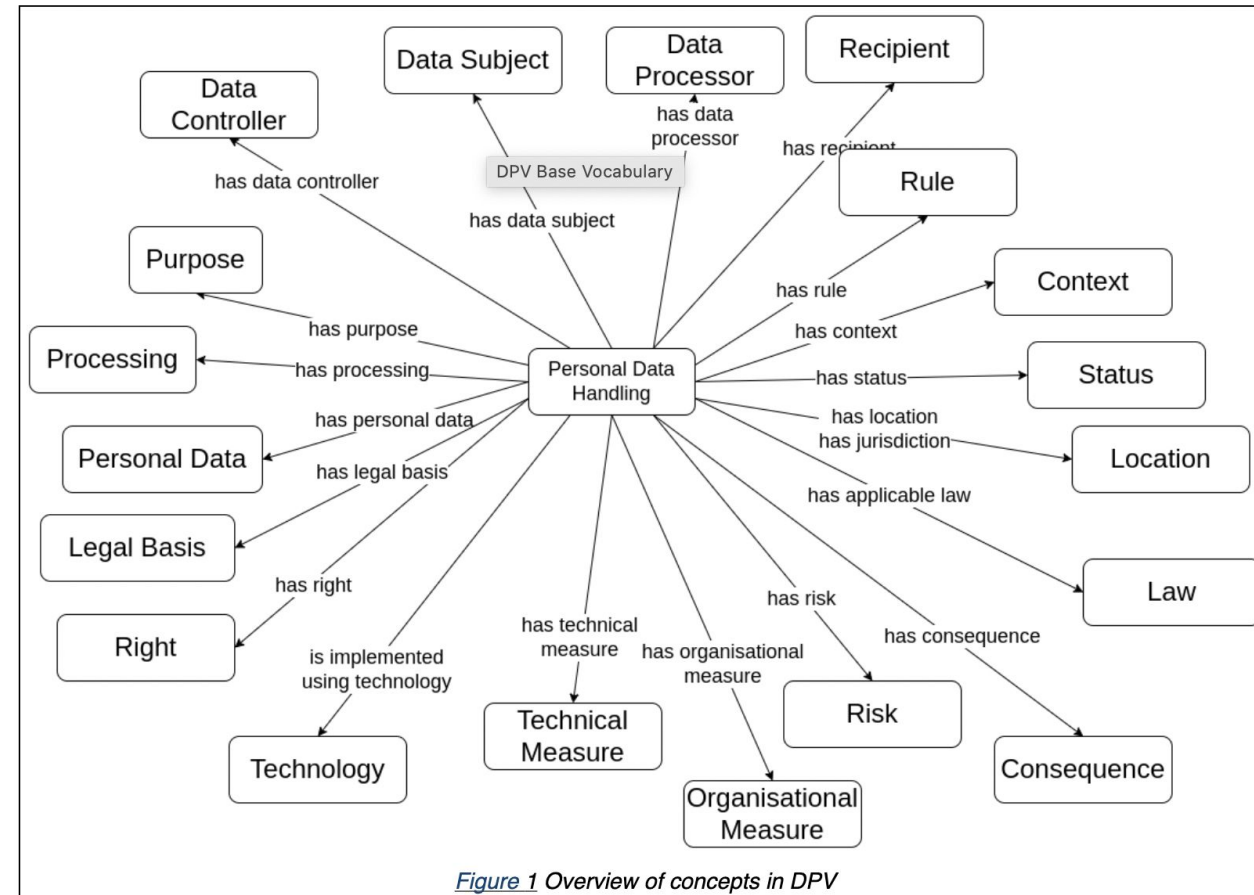


Figure 1 Overview of concepts in DPV

# ciTizen-centric DAta pLatform (TIDAL)

10

```
@prefix : <https://exampleparticipant.solidprovider.com/profile/card#>.
@prefix SNOMEDCT: <http://purl.bioontology.org/ontology/SNOMEDCT/>.

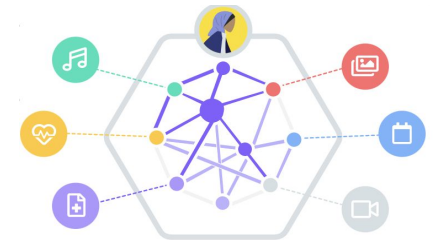
:me a SNOMEDCT:116154003; # Patient
    SNOMEDCT:397669002 "27"^^xsd:int; # Age
    SNOMEDCT:50373000 "165"^^xsd:int; # Height
    SNOMEDCT:726527001 "55"^^xsd:int; # Weight
    SNOMEDCT:263495000 SNOMEDCT:248152002; # Gender, Female
    SNOMEDCT:271649006 "110"; # Systolic blood pressure
    SNOMEDCT:271650006 "90"; # Diastolic blood pressure
    SNOMEDCT:405751000 SNOMEDCT:44054006. # Type 2 diabetes
```

try

Listing 4: An example of the RDF data file in a participant's Solid pod.



**Formulating personal data and data request into RDF format** with integration of vocabulary services and standards



# ciTizen-centric DAta pLatform (TIDAL)

11



**Based on Solid (SOcial Linked Data)** to store and structure personal data, and give individual control access to their fine-grained personal data



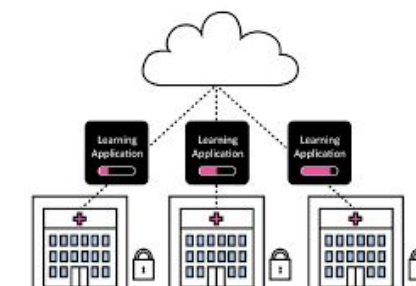
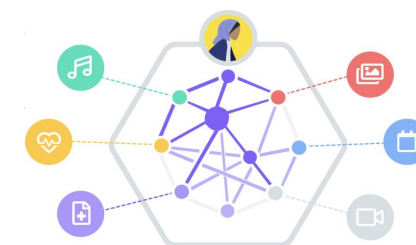
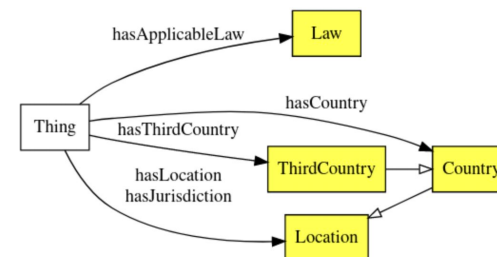
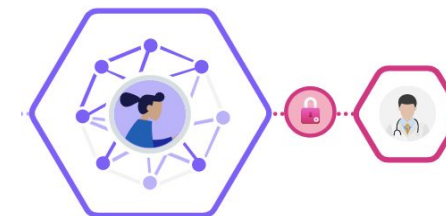
**Integrating Data Privacy Vocabulary** to structure personal data requests as digital consents to meet the requirements of GDPR



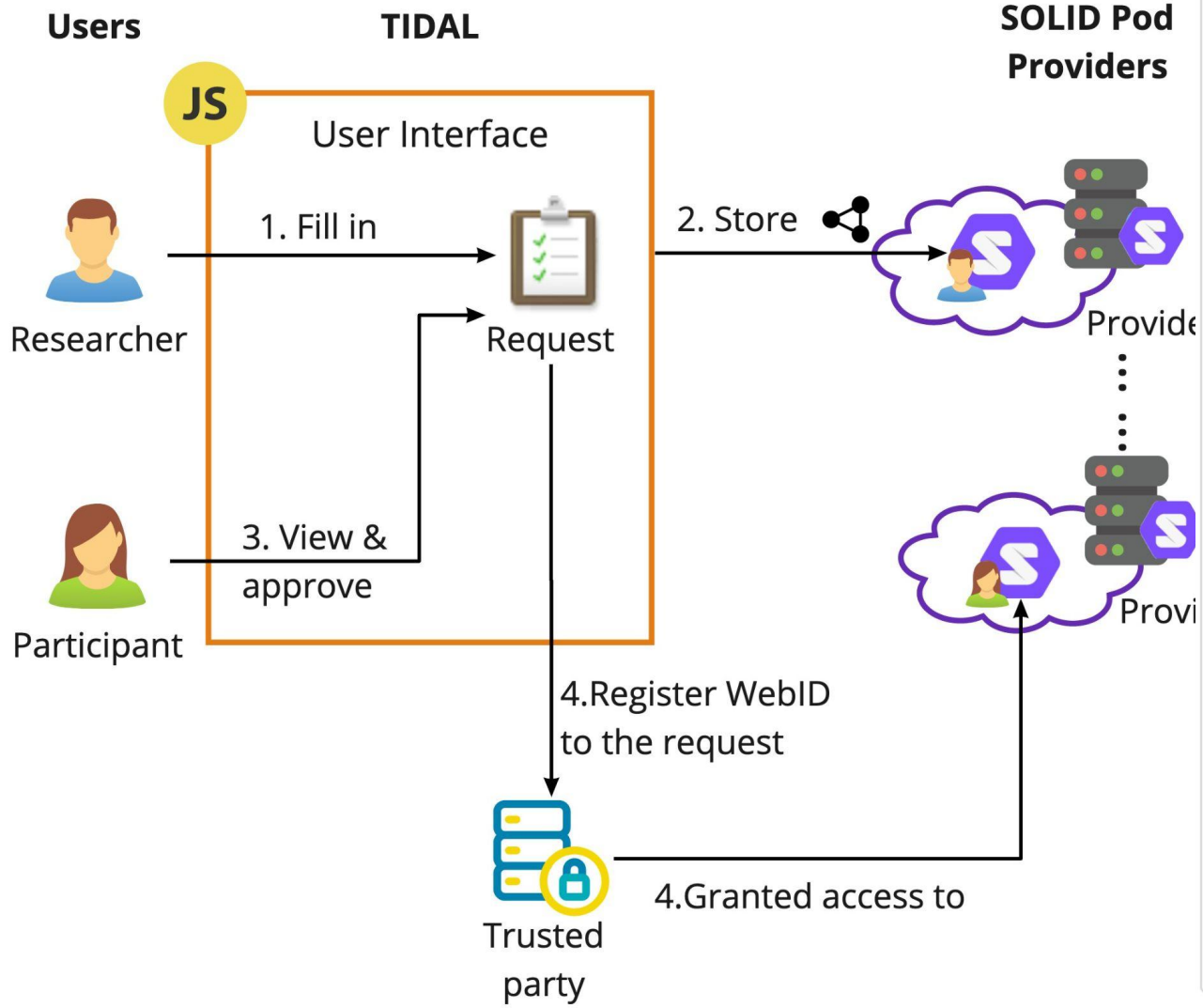
**Formulating personal data and data request into RDF format** with integration of vocabulary services and standards



**Integrating federated learning (Personal Health Train)** using the parameters promised in the data request and only the results are sent to the researchers



# TIDAL framework - Data request



Please Note: This request form is structured using the [Data Privacy Vocabulary \(DPV\)](#). DPV provides terms (classes and properties) to describe and represent information related to processing of personal data based on established requirements such as GDPR.

**Purpose of your research** ⓘ \*

Research and Development ✕

**Description of your purpose:** ⓘ

Learn association between diabetes status and dietary pattern Recommender

**Personal data categories** ⓘ \*

Medical Health [Special] (hysical Health, Mental Health, DNA Code, Disability, Health History) ✕

Demographic (Physical Trait, Income Bracket, Geographic) ✕

**Data elements (URI)** ⓘ \*

diagnosis +

**Searching terms from BioPortal ontologies**

90	NCIT	<b>Diagnosis</b> <a href="http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C15220">http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#C15220</a> The investigation, analysis and recognition of the presence and nature of disease, condition, or injury from expressed signs and symptoms; also, the scientific determination of any kind; the concise results of such an investigation.
100	PREMEDONTO	<b>Diagnosis</b> <a href="http://purl.obolibrary.org/obo/NCIT_C15220">http://purl.obolibrary.org/obo/NCIT_C15220</a> The investigation, analysis and recognition of the presence and nature of disease, condition, or injury from expressed signs and symptoms; also, the scientific determination of any kind; the concise results of such an investigation.
	CRISP	<b>diagnosis</b> <a href="http://purl.bioontology.org/ontology/CSP/4000-0159">http://purl.bioontology.org/ontology/CSP/4000-0159</a> general term for detecting and classifying diseases.
	IOBC	<b>Diagnosis</b> <a href="http://purl.jp/bio/4/id/200906001611549035">http://purl.jp/bio/4/id/200906001611549035</a>

**Consent Duration (Days)** \*

90

**Number of instances (minimal)** \*

100

**Data Processing Category** ⓘ \*

> Analyse ✕

**Analysis Model** \*

Linear Regression

**Consequences of data processing and impact of your research:**

Help diabetes patients understand the impact of their diet pattern

**Publish**

# TIDAL framework - Data request and consent

13

```
@prefix : <http://exampleresearcher.solidprovider.com/public/request.ttl#>.
@prefix schema: <https://schema.org/>.
@prefix exre: <http://exampleresearcher.solidprovider.com/profile/card#>.
@prefix dpv: <http://w3id.org/dpv#>.
@prefix dpvvpd: <http://w3id.org/dpv/dpv-pd#>.
@prefix dpvtech: <http://w3id.org/dpv/dpv-tech#>.
@prefix SNOMEDCT: <http://purl.bioontology.org/ontology/SNOMEDCT/>.

:161964062096710764675982245664
  a schema:AskAction, dpv:PersonalDataHandling;
  dpv:hasLegalBasis dpv:Consent;
  rdfs:label "Learn association between diabetes status and dietary pattern";
  schema:collectionSize 32;
  schema:creator exre:me;
  schema:DataFeedItem SNOMEDCT:10396001, SNOMEDCT:230125005, SNOMEDCT:56718006, SNOMEDCT:73211009;
  schema:dateCreated "2021-01-18T00:00:00Z"^^XML:dateTime;
  dpvtech:isImplementedUsingTechnology dpvtech:LinearRegression;
  dpv:hasImpact "Help diabetes patients understand the impact of their diet pattern";
  dpv:hasContext SNOMEDCT;
  dpv:hasDataController exre:me;
  dpv:hasExpiryTime "2021-12-31T00:00:00Z"^^XML:dateTime;
  dpvvpd:hasPersonalData dpvvpd:Health;
  dpv:hasProcessing dpv:Analyse;
  dpv:hasPurpose dpv:ResearchAndDevelopment.
```

# TIDAL framework - Data request and consent

Home Fetch Create **Participate** Request

Log Out

## All ongoing data requests



### Viewing

View the published data requests from researchers.



### Period

Enter the specific end date for using your data.



### Respond

Click decline or approve to the data request.

### CommercialInterest

### ResearchAndDevelopment

### Security

#### Researcher A

Institute A



Class of purpose: [CommercialInterest](#)  
Purpose: How XXX health app can help patient with obesity to have healthier lifestyle.  
Personal Data Category: [HealthRecord](#)  
Data Processing Category: [Analyse](#)  
Requested data:  
[Weight change](#)  
[Running](#)  
[Body mass index](#)  
End date: 2021-09-15T00:00:00Z  
Instances: 300  
Analysis: Linear Regression

Withdrawal Date: dd/mm/yyyy

Data Recipient:

Decline

Approve

#### Researcher B

Institute B



Class of purpose: [ResearchAndDevelopment](#)  
Purpose: Learn association between diabetes status and dietary pattern  
Personal Data Category: [Health](#)  
Data Processing Category: [Analyse](#)  
Requested data:  
[Hyperlipoproteinemia diet](#)  
[Diabetes mellitus](#)  
End date: 2021-08-01T00:00:00Z  
Instances: 32  
Analysis: Linear Regression  
Consequence of data process: Help diabetes patients understand the impact of their diet pattern

Requested data is not detected in your pod!

Send a message to the researcher.

#### Researcher C

Institute B



Class of purpose: [Security](#)  
Purpose: Improve the security of health data storage and sharing  
Personal Data Category: [HealthRecord](#)  
Data Processing Category: [Profiling](#)  
[PseudoAnonymise](#)  
Requested data:  
[Blood pressure](#)  
[Heart rate](#)  
[Plasma glucose](#)  
End date: 2022-01-01T00:00:00Z  
Instances: 265  
Analysis: Linear Regression

Requested data is not detected in your pod!

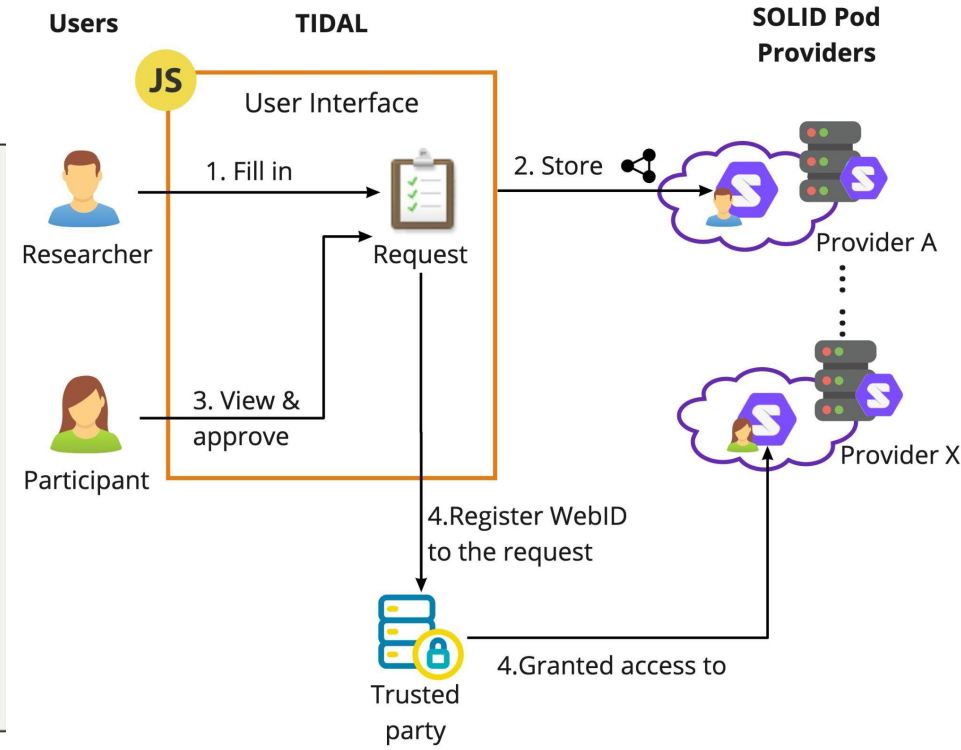
Send a message to the researcher.

# TIDAL framework - Data request and consent

```
@prefix : <http://exampleParticipant.solidProvider.com/private/participation#>.
@prefix part: <https://exampleparticipant.solidprovider.com/profile/card#>.
@prefix req: <https://exampleresearcher.solidprovider.com/public/request.ttl#>.
@prefix extp: <http://exampltrustedparty.solidprovider.net/profile/card#>
@prefix app: <https://examplesolidapp.com/

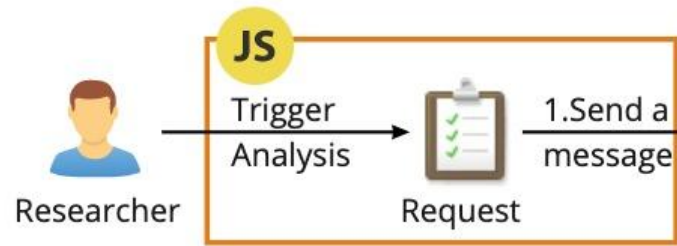
:16197041266295299657542155198
  a schema:JoinAction, dpv:Consent;
  schema:dateCreated "2021-02-18T00:00:00Z"^^XML:dateTime;
  dpv:DataSubject part:me;
  dpv:hasConsentNotice req:161964062096710764675982245664;
  dpv:hasExpiry schema:false;
  dpv:hasExpiryTime "2021-12-31T00:00:00Z"^^XML:dateTime;
  dpv:hasProvisionMethod app:participate;
  dpv:hasProvisionTime "2021-02-18T00:00:00Z"^^XML:dateTime;
  dpv:hasWithdrawalTime "2021-09-18T00:00:00Z"^^XML:dateTime;
  dpv:hasRecipient extp:me.
```

Listing 2: An example of generated participation statements in a RDF format in the participant's Solid pod.





# TIDAL framework - Data Analysis



**Chang Sun**  
Institute of Data Science

Purpose: InruptUser64VarTest  
Requested data:  
<http://dbpedia.org/ontology/income>  
<http://xmlns.com/foaf/0.1/age>  
<http://schema.org/height>  
<http://schema.org/weight>  
<http://purl.bioontology.org/ontology/SNOMEDCT/4085780>  
... .. 59 more data elements  
End date: 2021-11-24T00:00:00Z  
Instances: 128  
Analysis: Linear Regression

Data collection progress - 0%

Stop collection    Trigger analysis

**Chang Sun**  
Institute of Data Science

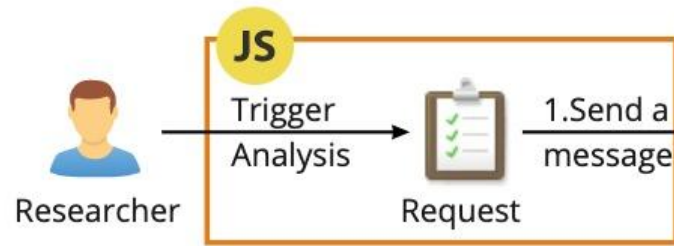
Purpose: TestUser4VarTest  
Requested data:  
<http://solidtest.org/variable/100001>  
<http://solidtest.org/variable/100002>  
<http://solidtest.org/variable/100003>  
<http://solidtest.org/variable/100004>  
End date: 2022-01-01T00:00:00Z  
Instances: 128  
Analysis: Linear Regression

Data collection progress - 100%

Stop collection    Trigger analysis

# TIDAL framework - Data Analysis

17

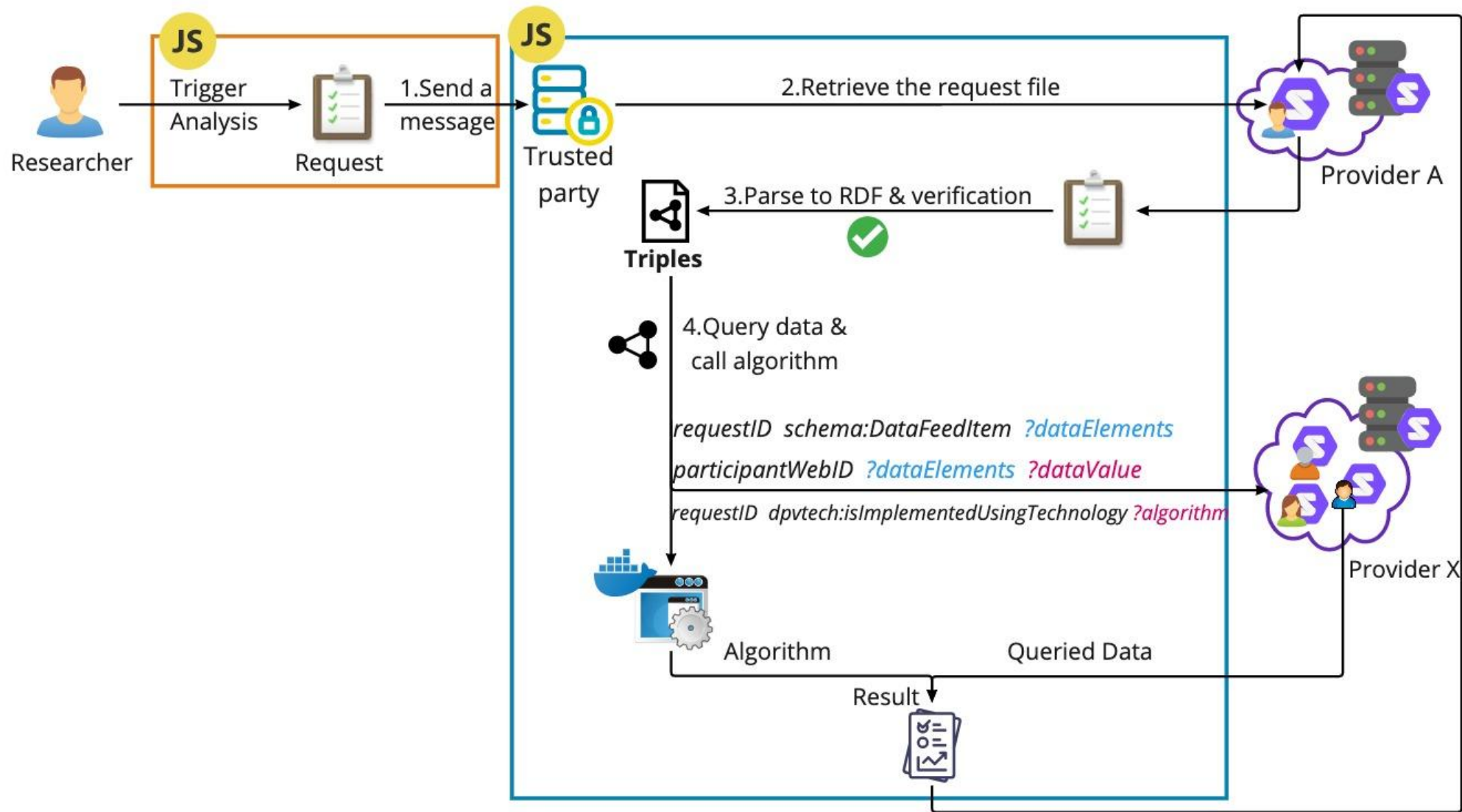


```
@prefix : <http://examplertrustedparty.solidprovider.net/inbox/triggermessage#>.
@prefix req: <http://examplerresearcher.solidprovider.com/public/request.ttl#>.
@prefix exre: <http://examplerresearcher.solidprovider.com/profile/card#>.

:160622932739325095672093710975
  schema:actionStatus schema:ActivateAction;
  schema:creator exre:me;
  schema:dateCreated "2021-04-20T09:37:57.499Z"^^XML:dateTime;
  schema:target req:161964062096710764675982245664.
```

Listing 3: An example of generated trigger message (Activate Action) sent by the researcher.

# TIDAL framework - Data Analysis



- Individuals data will be read only when the analysis gets executed.
- The analysis can only be triggered by the data requester under conditions (have enough participants, in the valid period, for the indicated purpose, etc)
- Individuals can revoke the permission anytime.
- Individuals can change the data values at any moment (even after giving researchers permission, when the experiments are being conducted)
- (Future work) Individuals will be able to indicate their preference to which data requests they want to receive.
- and more...

**Paper:** ciTizen-centric DATA pLatform (TIDAL): Sharing Distributed Personal Data in a Privacy-Preserving Manner for Health Research

**Open Source tool and code:** <https://github.com/sunchang0124/TIDAL>

**Web application is live** →

Demo Youtube Video:



Contact:

- [chang.sun@maastrichtuniversity.nl](mailto:chang.sun@maastrichtuniversity.nl)
- [www.linkedin.com/in/chang-sun-maastricht/](https://www.linkedin.com/in/chang-sun-maastricht/)

