

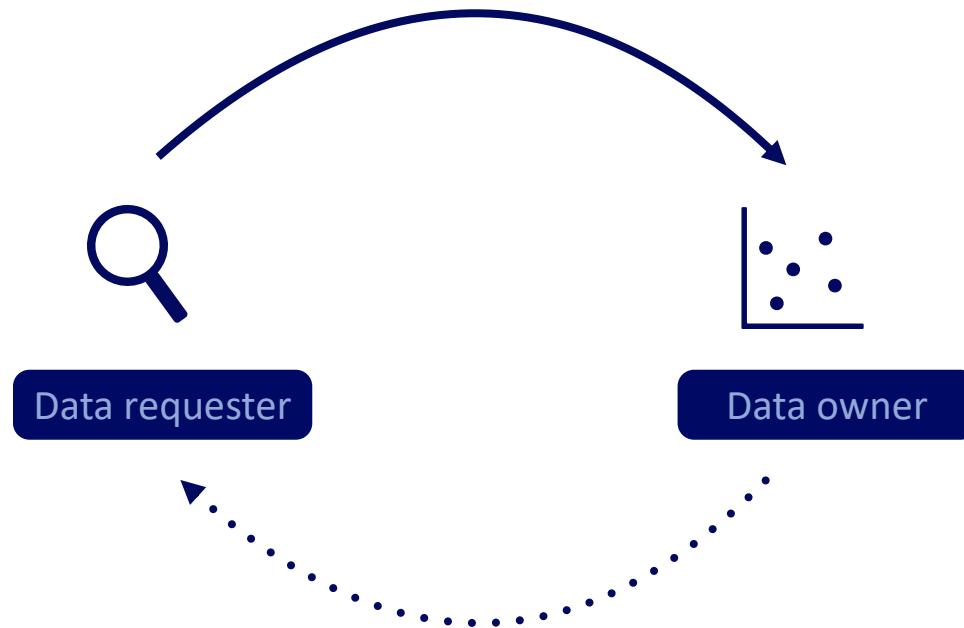
DATA EXCHANGE DEMO

Share data while retaining control and confidentiality of your data

SURF

Gains and difficulties of sharing confidential data

- + Access to non-public data.
- + Potential new research and collaborations.
- More work to manage confidential data.



- ~ Possible to gain new insights.
- Risks on privacy and security.
- Additional work without direct return on investments (ROI).

! Gain is usually with the data requester, burden is with the data provider

Willingness to share data

ROI



Return on Investment (ROI) is determined by the balance between effort it takes to share data, and the gain received by sharing data



+

Trust



Trust is determined by the balance between the risks (due to privacy or competition), and the control (due to verification and security) of sharing and usage of data



Type of Data Owners



Data aggregators

Health care (Palga, NZa)
Social-economic (CBS, municipalities)



Hospitals + medical institutions

Hospital (AMC, vuMC, St. Antonius)
Insurance companies (Zilveren Kruis)



Onderzoekers + universiteiten

Universities (Twente, Wageningen, Groningen)
Researchers



Bedrijven

Friesland-Campina, Elsevier

Privacy sensitive

Competitive data

Methods to Ease Data Sharing

Agreements

- Stipulation of what can/cannot be done
- Signing of contract or NDA
- Dispute resolution process

Registration

- Authentication
- Verification of credential
- Reputation score
- Policy framework
- Audit trails

Pseudonymization

- Filtering (on records)
- Pruning (on properties)
- Aggregation (combining records)
- Make coarse grained buckets
- Slight alteration of data
- One-way hashing
- One-time identifiers

Data Vault

- Data source retains control
- Delegate permissions
- No central data lake
- Data marketplace

Secure Containers

- Bring algorithm to data
- Trusted third party
- Share output instead of data

Secure Computing

- Secure multi-party computation
- Homomorphic encryption
- Garbled Circuits
- Zero-knowledge proof

Example: Find the average income



Run #1

- 21 people
- Algorithm verified
- Outcome guaranteed not to be traceable to individual people

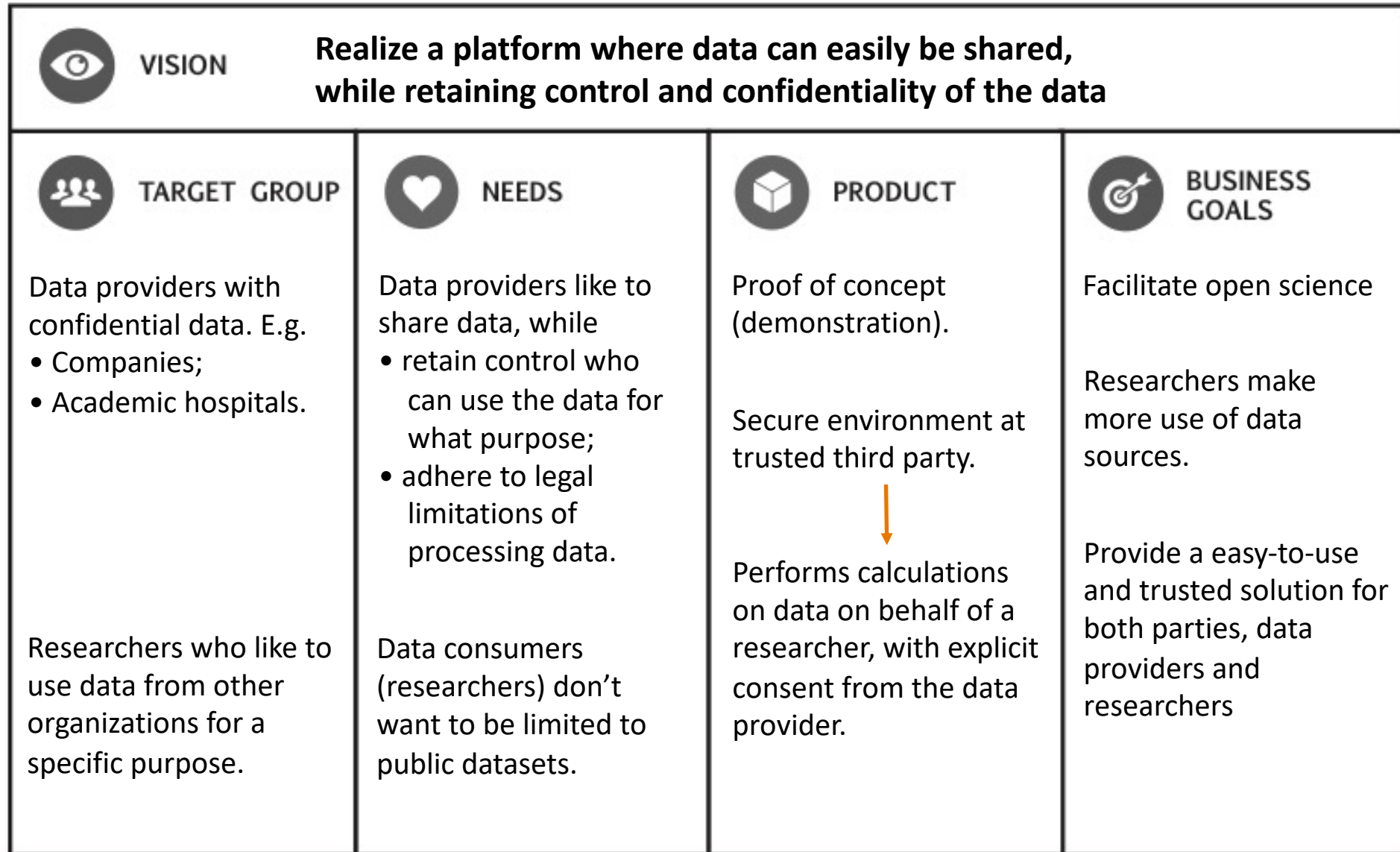


Run #2

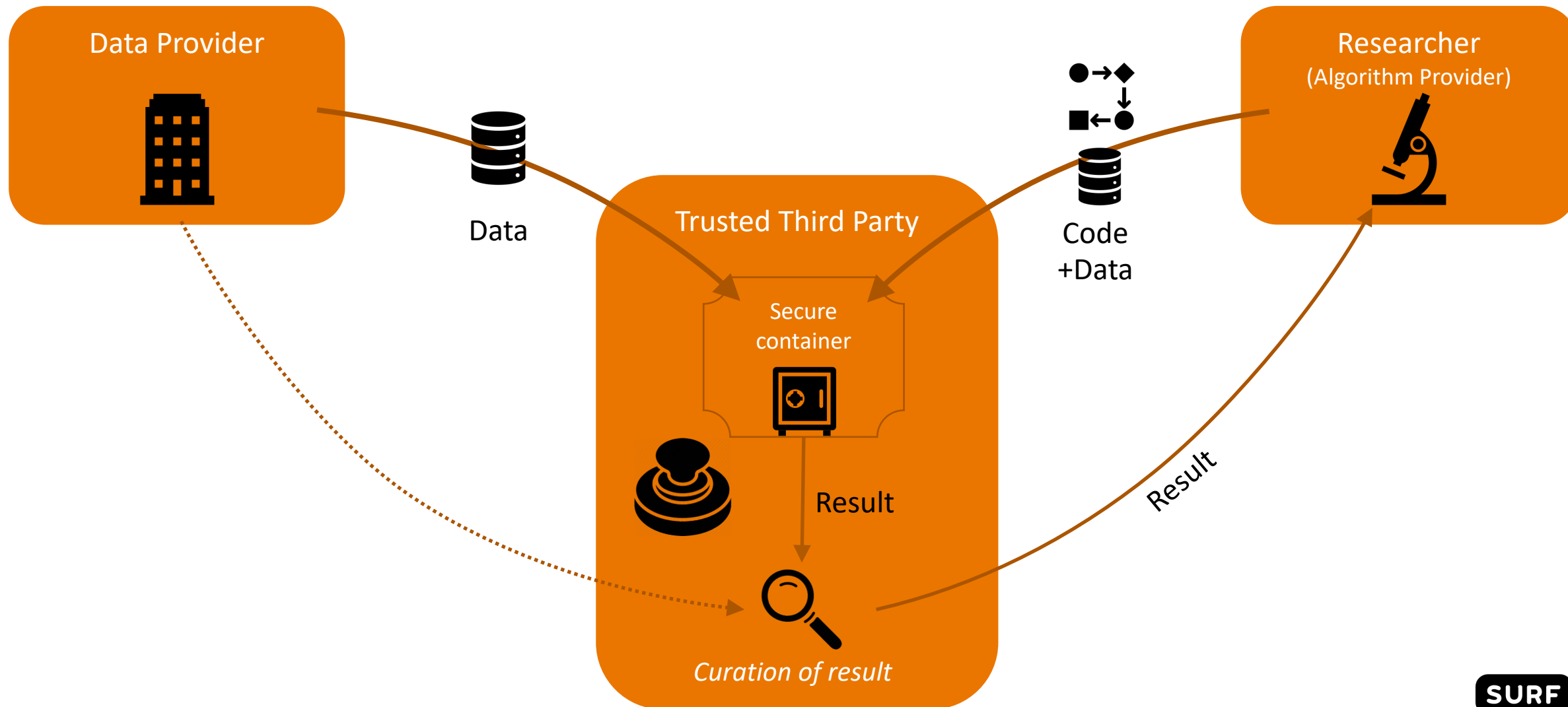
- 22 people (same 21 and 1 other)
- Algorithm verified
- Outcome guaranteed not to be traceable to individual people

Even if individual runs are fine, combining two runs may reveal confidential data

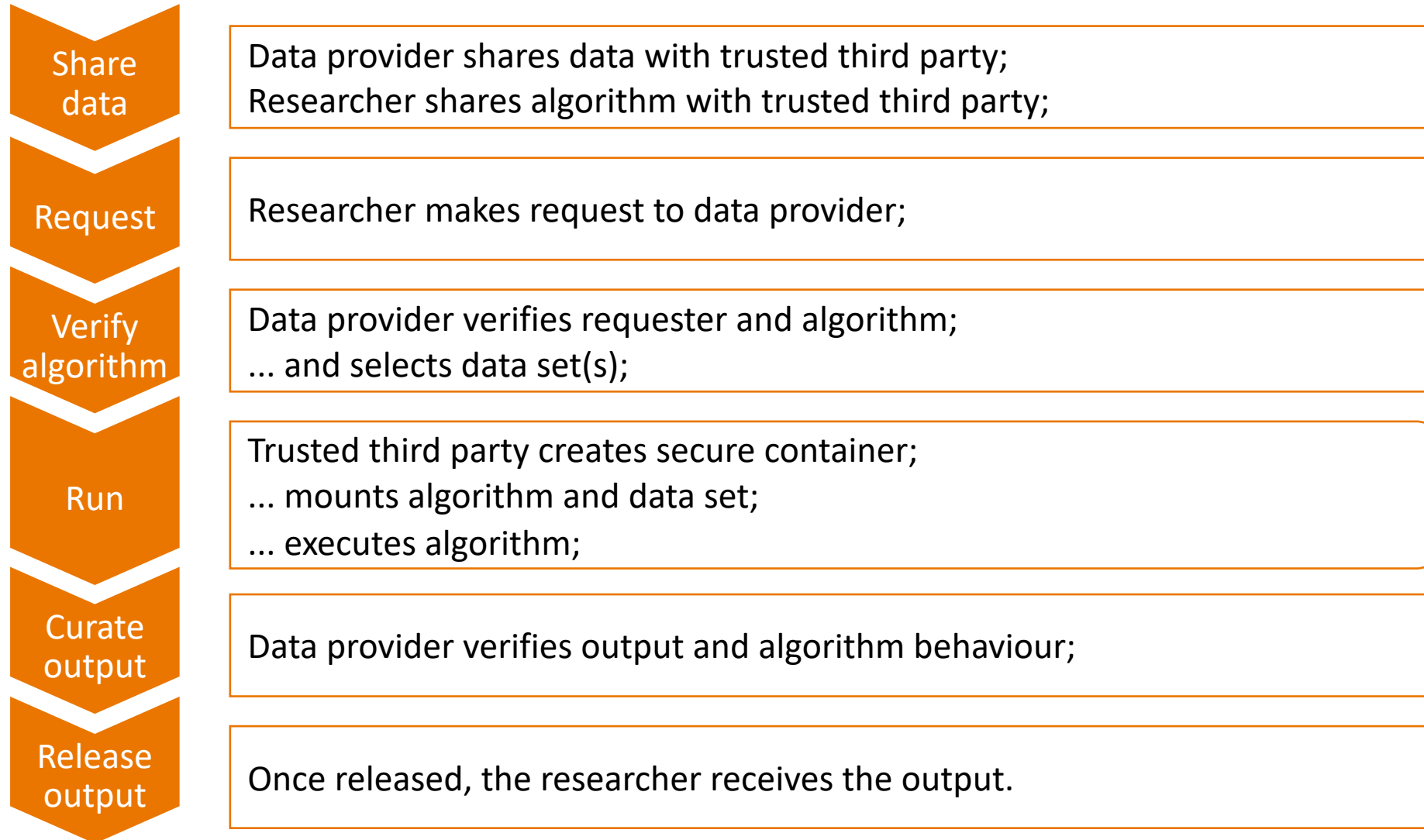
Data Exchange



Collaborating without direct Sharing Data



Workflow

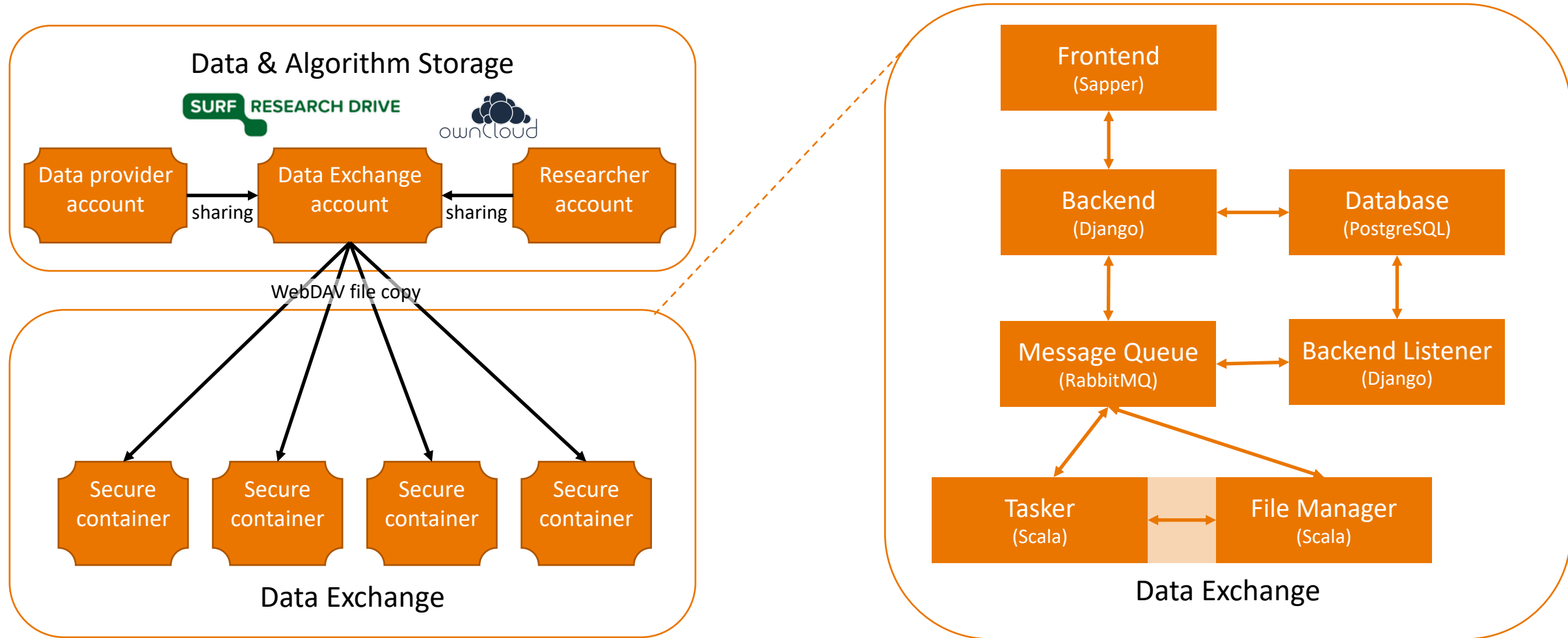


Permission Models

One-off permission	Trust a researcher	Run on a data stream
<p>The data provider permits a researcher to run a specific algorithm once on a specific dataset.</p>	<p>The data provider permits a researcher to run any algorithm on a specific dataset.</p> <p>The permission can be revoked at any time.</p> <p>Example use cases:</p> <ul style="list-style-type: none">• the data provider trust the researcher to always write benevolent code• the researchers wants to tweak the algorithm, and run it on a sample dataset every time.	<p>The data provider permits a researcher to run a specific algorithm on any data set in a selected folder. Every time a new dataset is added to the folder, the algorithm is automatically run.</p> <p>The permission can be revoked at any time, but is also automatically revoked as soon as a change to the shared algorithm is detected.</p>

Currently supported permission models

Technical Implementation of the prototype






External integrations

Internal Components



The screenshot shows a web browser window displaying the SURF Research Drive interface. The browser's address bar shows the URL: `https://researchdrive.surfsara.nl/index.php/apps/files/?dir=/&view=sharing`. The SURF Research Drive logo is visible in the top navigation bar. On the left, a sidebar menu includes options like 'All files', 'Favorites', 'Shared with you', 'Shared with others', 'Shared by link', 'External storage', 'Deleted files', and 'Settings'. The main content area displays a list of files and folders. The files listed are:

Name	Share time
 ugly_cats_and_dogs	Shared with Data Exchange ... 2 days ago
 cat-looks-like-dog-1.jpg	Shared with Data Exchange ... 13 hours ago
 random_numbers.txt	Shared with Data Exchange ... 2 days ago

Below the file list, it indicates '1 folder and 2 files'. The interface also shows a search bar and a user profile 'freek@macfreek.nl' in the top right corner.

Files - SURF - Research Drive My Files

https://researchdrive.surfsara.nl/index.php/apps/files/?dir=/&fileid=81846599

SURF RESEARCH DRIVE Freek Dijkstra

Menu Files

- All files
- Favourites
- Shared with you
- Shared with others
- Shared by link
- External storage
- Deleted files
- Settings

Name	Size	Modified
calculate_sum.py	< 1 KB	a month ago
calculate_average.py	< 1 KB	a month ago
demo1_data	< 1 KB	14 days ago
demo1_code	< 1 KB	a month ago
Data Exchange	0 KB	2 months ago
ComputeWizards	0 KB	3 months ago
catordog	121.8 MB	13 hours ago
catdog	121.8 MB	a day ago

calculate_sum.py < 1 KB, a month ago

Activities Comments **Sharing**

User and Groups Public Links

Data Exchange

The screenshot shows a web browser window with the URL <https://dataexchange.surfsara.nl/tasks/request>. The page title is "DataExchange (Demo)". The navigation bar includes "Home", "My permissions", "Create Request" (highlighted in blue), and "Log out (freek.dijkstra@surfsara.nl) — Algorithm owner (Toggle)".

Create request

Request Permission for a dataset

Type of permission: Run once

The selected algorithm will be ran on the selected dataset of the data owner exactly once.

Select algorithm: calculate_sum.py

Data owner email: catordog

Dataset description: Can I run the sum on your numbers?

Run an algorithm with continuous permission

Select algorithm: Select algorithm

Select dataset: Select algorithm first.

Run

Step 1. Accept algorithm Step 2. Run algorithm Step 3. Release output

Algorithm Owner
freek.dijkstra@surfsara.nl

Permission Type
one time permission

Permission Information
The selected algorithm will be ran on your selected dataset once.

Algorithm Name
calculate_sum.py

Algorithm Dependencies
sys

Algorithm Length
Lines: 22, Words: 44, Characters: 522

Choose dataset
random_numbers.txt

```
calculate_sum.py
#!/usr/bin/env python
import sys

def main():
    try:
        filename = sys.argv[1]
    except IndexError as err:
        sys.stderr.write("Please spec
        return -1

    sum = 0
    with open(filename, 'r') as numbe
        for number in numbersfile:
            try:
                sum += int(number)
            except ValueError as err:
```

Run algorithm on data to see output and go to step 2 Reject request

← Back

Step 1. Accept algorithm **Step 2. Run algorithm** Step 3. Release output

Run algorithm

Completed: Creating container

Completed: Installing dependencies

Completed: Downloading data and algorithm to container

Completed: Blocking all outside access to container

- . Verifying algorithm
- . Running algorithm on data
- . Saving output
- . Deleting container including data and algorithm
- . Wrapping up..

The screenshot shows a web browser window with the address bar displaying `https://dataexchange.surfsara.nl/tasks/80`. The page has a progress bar at the top with three steps: "Step 1. Accept algorithm", "Step 2. Run algorithm", and "Step 3. Release output", with the third step being the active one. The main content area is divided into three columns:

- Algorithm Owner:** freek.dijkstra@surfsara.nl
- Algorithm Name:** calculate_sum.py
- Output:** 30717
- Permission Type:** one time permission
- Algorithm Dependencies:** sys
- Permission Information:** The selected algorithm will be ran on your selected dataset once.
- Algorithm Length:** Lines: 22, Words: 44, Characters: 522
- Used Dataset:** random_numbers.txt

At the bottom of the page, there are two buttons: "Reject output" (red) and "Release output" (green).

My Files | Files - SURF - Research Drive | +

https://dataexchange.surfsara.nl/tasks/80

Execution finished

Data Owner	Algorithm Name	Output
freek.dijkstra@surfsara.nl	calculate_sum.py	30717
Permission Type one time permission	Algorithm Dependencies sys	
Permission Information The selected algorithm will be ran on the selected dataset of the data owner exactly once.	Algorithm Length Lines: 22, Words: 44, Characters: 522	
	Choose dataset random_numbers.txt	

Shared with others - SURF - Re X Manage Data

https://dataexchange.surfsara.nl/manage_data 80% Search

DataExchange (Demo) Home Manage Data Review Requests Log out (freek@macfreek.nl) — Data owner (Toggle)

Manage shared Files and Folders

random_numbers.txt **Withdraw Data**

Permissions

With	Algorithm	Type	
freek.dijkstra@surfsara.nl	calculate_sum.py	one time permission	Reject Permission

Runs

Algorithm Owner	Passed	When	Action
freek.dijkstra@surfsara.nl	Passed	20-11-2019	See log
freek.dijkstra@surfsara.nl	Rejected	18-11-2019	See log

ugly_cats_and_dogs **Withdraw Data**

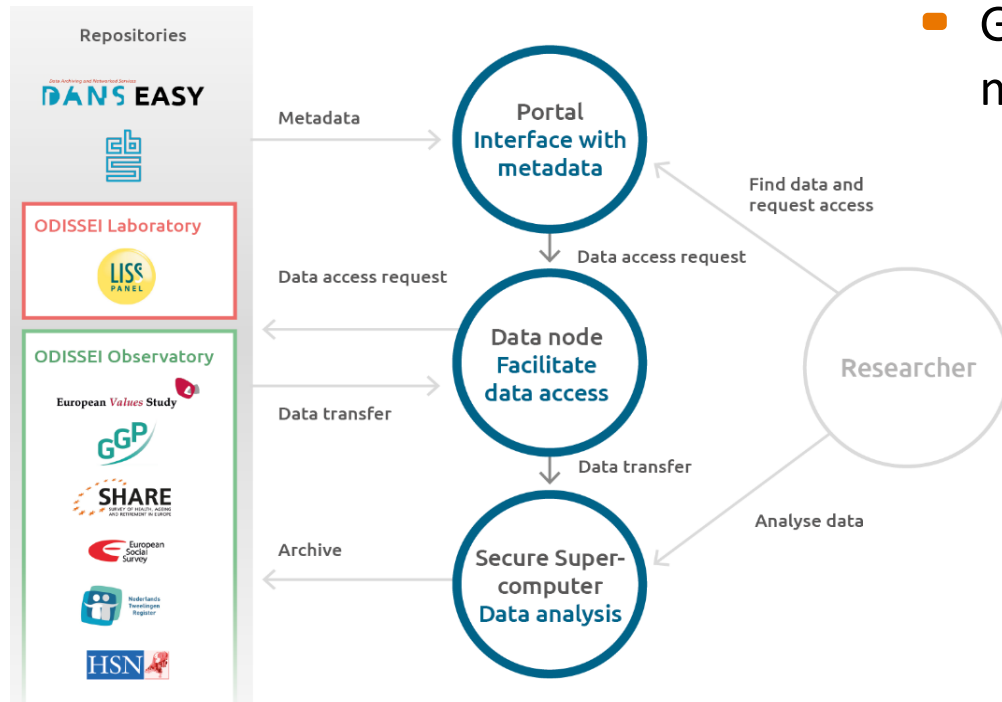
Related Projects

ODISSEI Secure Supercomputer (OSSC)

- In production
- Processes CBS micro-data on Cartesius
- Does pseudonymization as well

AMdEX

- Collaboration of interested parties
- Initiated by Amsterdam Economic Board
- Goal is to build an infrastructure for multiple Data Marketplaces



COLLABORATION WITHOUT SHARING DATA

 Freek Dijkstra

 Freek.Dijkstra@surfsara.nl

 www.surf.nl

 This presentation is available under the
creative commons attribution 4.0 license

Driving innovation together

 SURF