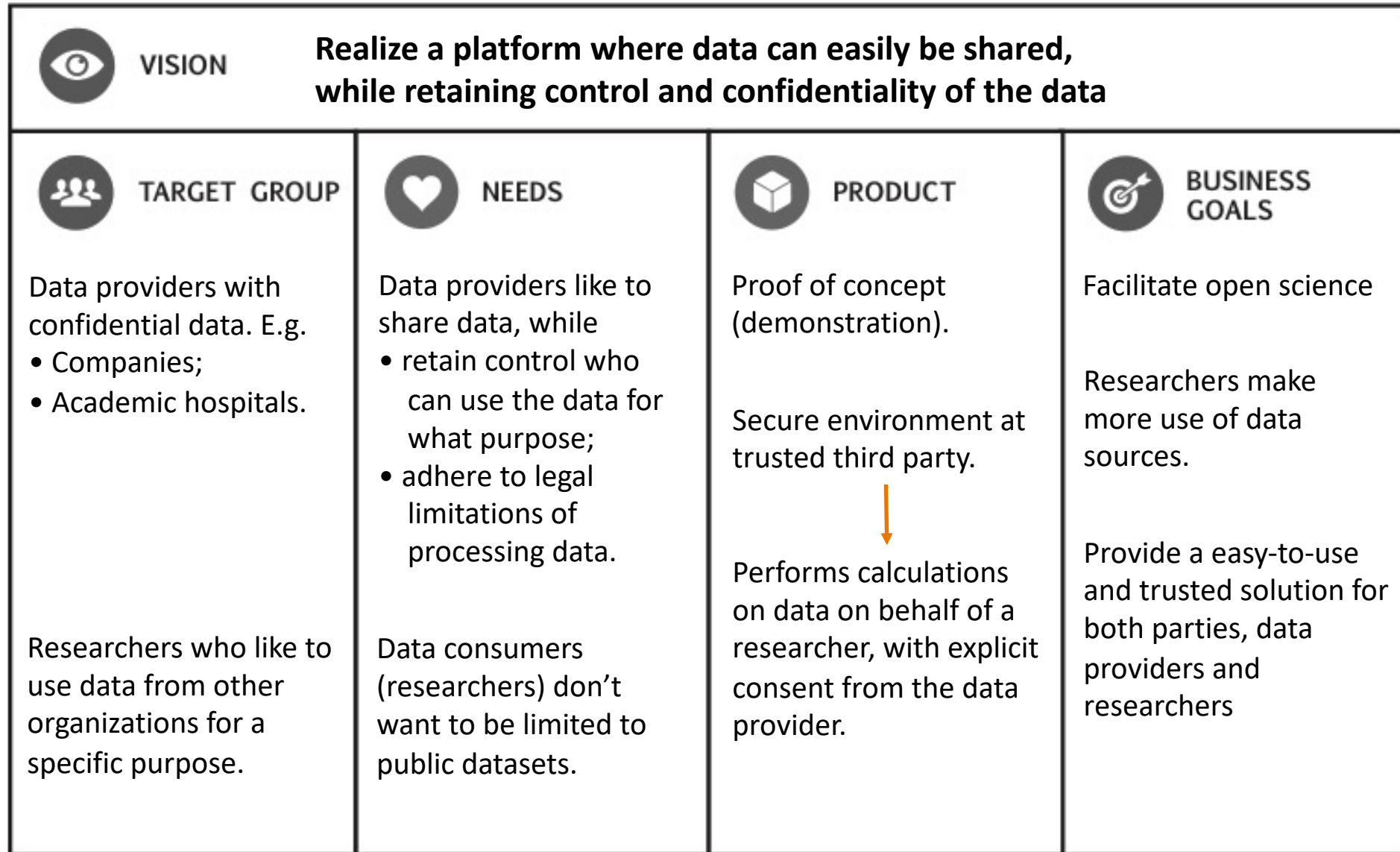


DATA EXCHANGE DEMO

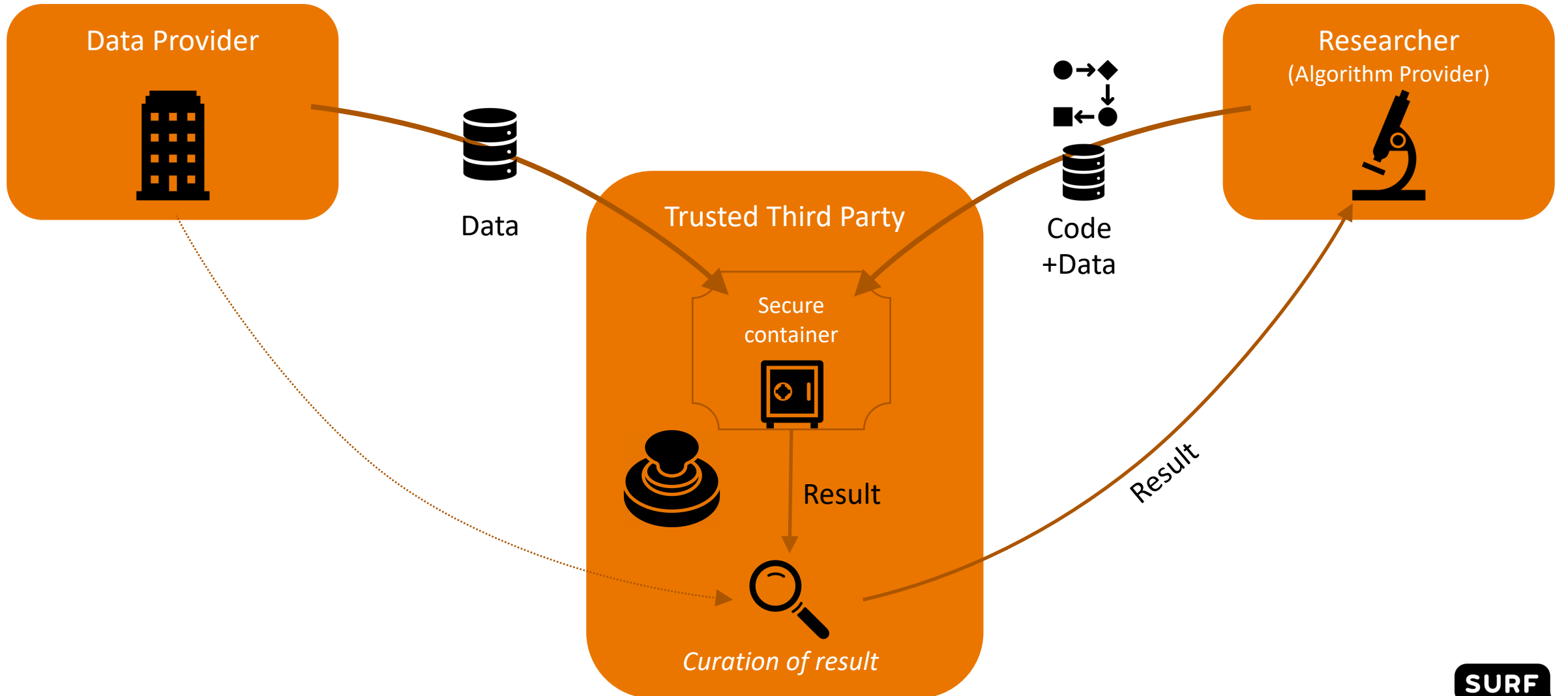
Share data while retaining control and confidentiality of your data

SURF

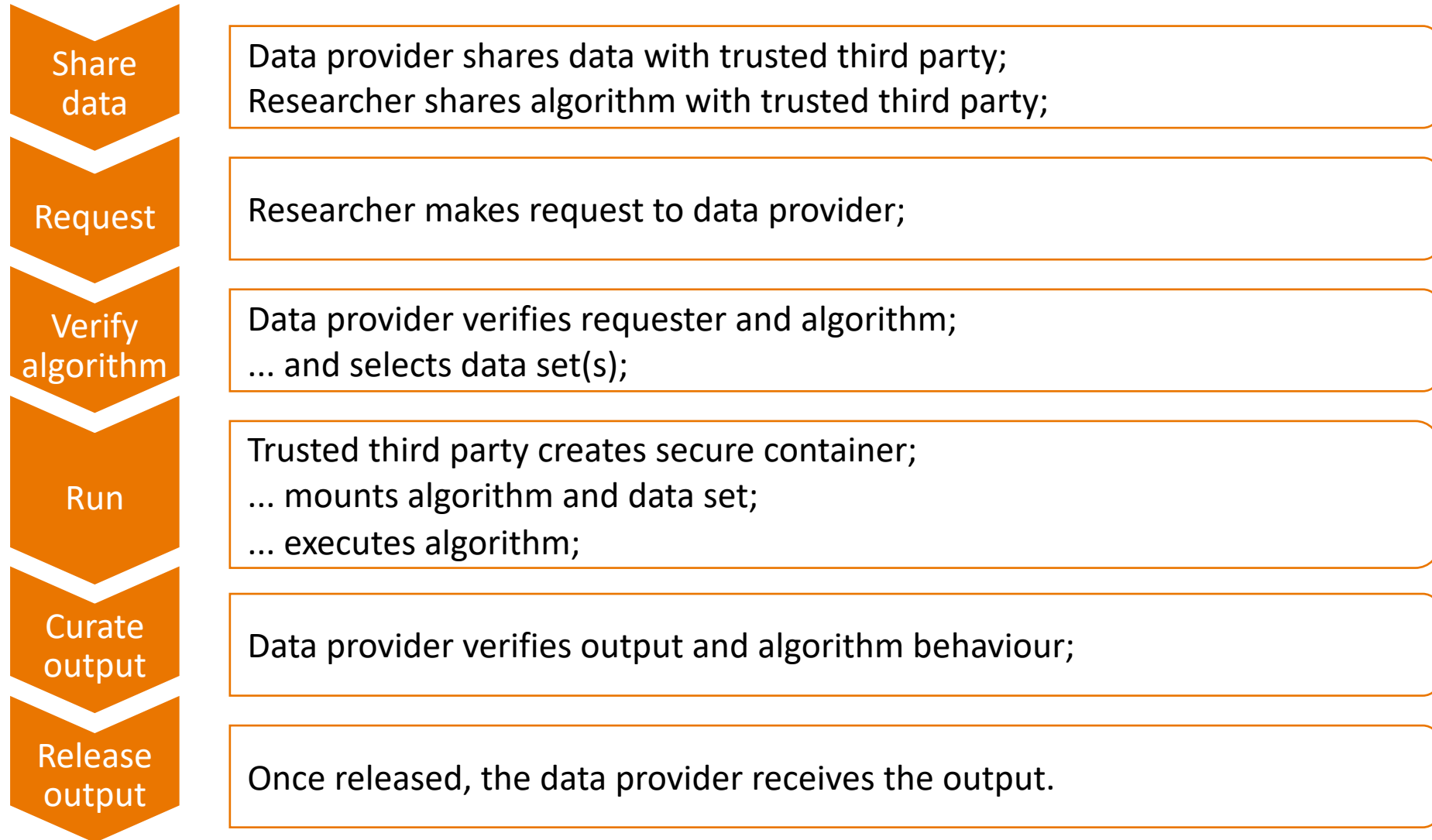
Data Exchange



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

Alternative methods

Methods to collaborate without sharing confidential data:

- Trusted third party
 - Special case: Run algorithm at data provider's site
- Only sharing aggregated results
- Masking data (anonymization, or pseudonymization)
- Database matching techniques (bloom filters, or secure set intersection)
- Privacy preserving techniques (differential privacy, or mixing synthetic data)
- Secure multiparty computation (garbled circuits)
- Calculations on encrypted data (homomorphic encryption)

Implementation (Proof of Concept)

- Working prototype
- Non-production (not scalable nor fast, not rigorously tested)
- Data stored at ResearchDrive (OwnCloud implementation at SURF for researchers)
- Data sharing: <https://dataexchange.surfsara.nl/>
(simple password to emphasis it is a demonstration only: demo / dex)
- **Goal is to understand user requirements**



Mike Kotsur



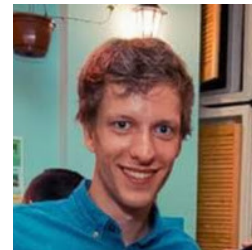
Rienk Koenders



Sijmen Schoon



Tijs Teulings



Sander
van Wickeren



Axel Berg



Hylke Koers

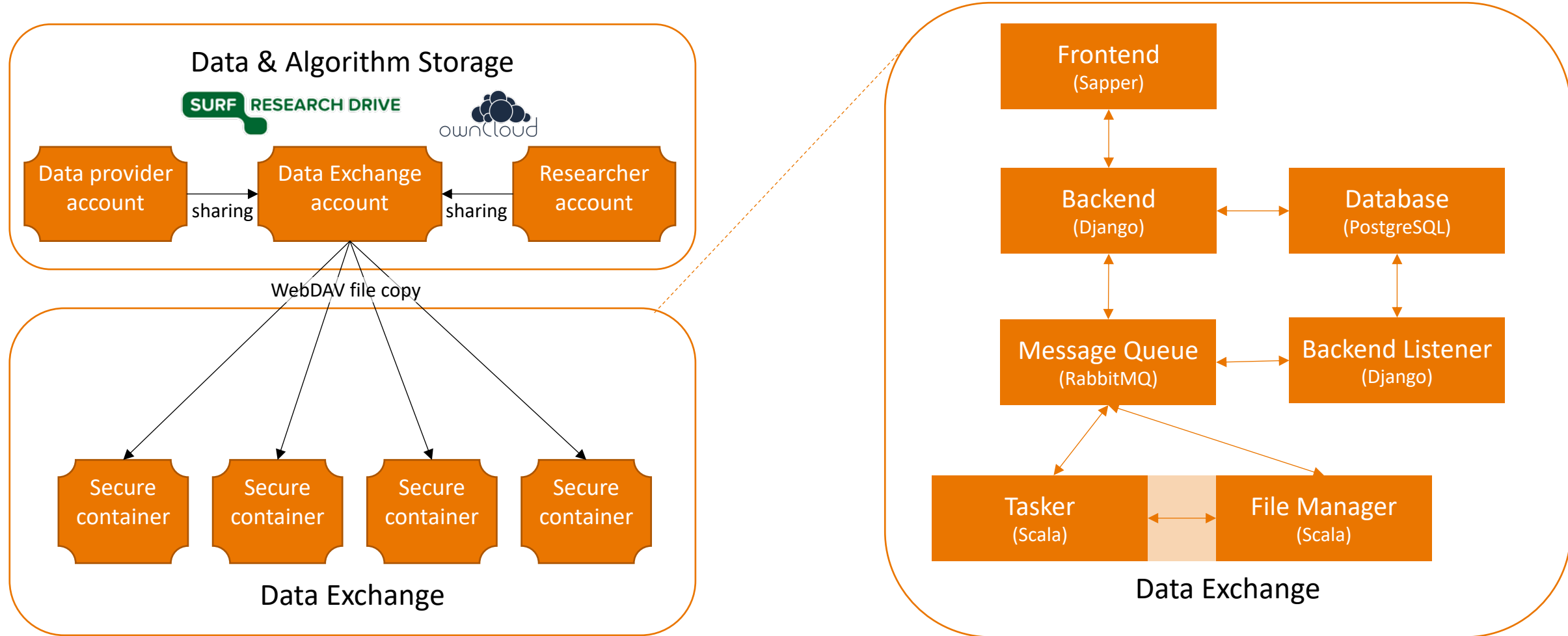


Gerben
van Malenstein



Freek Dijkstra

Technical Implementation of the prototype



External integrations

Internal Components






Risks and Mitigations

Risk	Mitigation
Data is leaked to outside world	Researcher can never view the raw data, only the result
Data is used in other ways than intended	Data provider can review algorithm
Algorithm is leaked to outside world	Algorithm is not reviewed by data provider, researcher is trusted to write benevolent code only *
Output contains confidential information	Data provider curates output before releasing it to researcher
Malicious algorithm tries to copy data to remote server	No network access is allowed in secure container
Malicious algorithm tries to embed data in output	Data provider can review algorithm
Algorithm is altered after it is shared	Permissions involving this algorithm are automatic revoked
Researcher can no longer be trusted	Permission can be revoked by data provider at any time
Trusted third party can no longer be trusted	Sharing of data to trusted third party can be revoked at any time
Data is corrupt or data provider can no longer be trusted	Researcher should look for other data sources
Data can't leave premises, not even to a trusted third party	Secure container can be run at premises of data provider *

* Not yet implemented in the prototype

The screenshot shows a web browser window with the address bar displaying `https://researchdrive.surfsara.nl/index.php/apps/files/?dir=/&view=sharing`. The page header is green with the SURF Research Drive logo and the user email `freek@macfreak.nl`. A left sidebar contains navigation options: All files, Favorites, Shared with you, Shared with others, Shared by link, External storage, Deleted files, and Settings. The main content area shows a file list with columns for Name and Share time. 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".

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

My Files Files - SURF - Research Drive +

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

DataExchange (Demo) Home My permissions **Create Request** 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:

Data owner email:

Dataset description:

Run an algorithm with continuous permission

Select algorithm:

Select dataset: Select algorithm first.

The screenshot shows a web browser window with the URL `https://dataexchange.surfsara.nl/tasks/80`. The browser tabs include "Shared with others - SURF - Re" and "My Files". The page content is organized into three steps: "Step 1. Accept algorithm" (highlighted in blue), "Step 2. Run algorithm", and "Step 3. Release output".

Algorithm Owner: freek.dijkstra@surfsara.nl

Algorithm Name: calculate_sum.py

Algorithm Code:

```
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:
```

Permission Type: one time permission

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

Algorithm Dependencies: sys

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

Choose dataset: random_numbers.txt

At the bottom, there are two buttons: "Run algorithm on data to see output and go to step 2" (green) and "Reject request" (red).

The screenshot shows a web browser window with the address bar displaying `https://dataexchange.surfsara.nl/tasks/80`. The page features a progress bar with three steps: "Step 1. Accept algorithm", "Step 2. Run algorithm" (which is currently active and highlighted in blue), and "Step 3. Release output". Below the progress bar, a blue header reads "Run algorithm". Underneath, a list of completed tasks is shown, each preceded by a green "Completed:" label and a small green checkmark:

- Completed: Creating container
- Completed: Installing dependencies
- Completed: Downloading data and algorithm to container
- Completed: Blocking all outside access to container

Below these, a list of tasks is shown with a small grey dot indicating they are in progress:

- 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 browser tabs include "Shared with others - SURF - Re" and "My Files". The page content is divided into three steps: "Step 1. Accept algorithm", "Step 2. Run algorithm", and "Step 3. Release output", with the third step being the active one.

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

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

Shared with others - SURF - Re X My Files

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

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

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	

Reject output Release output

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	

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**

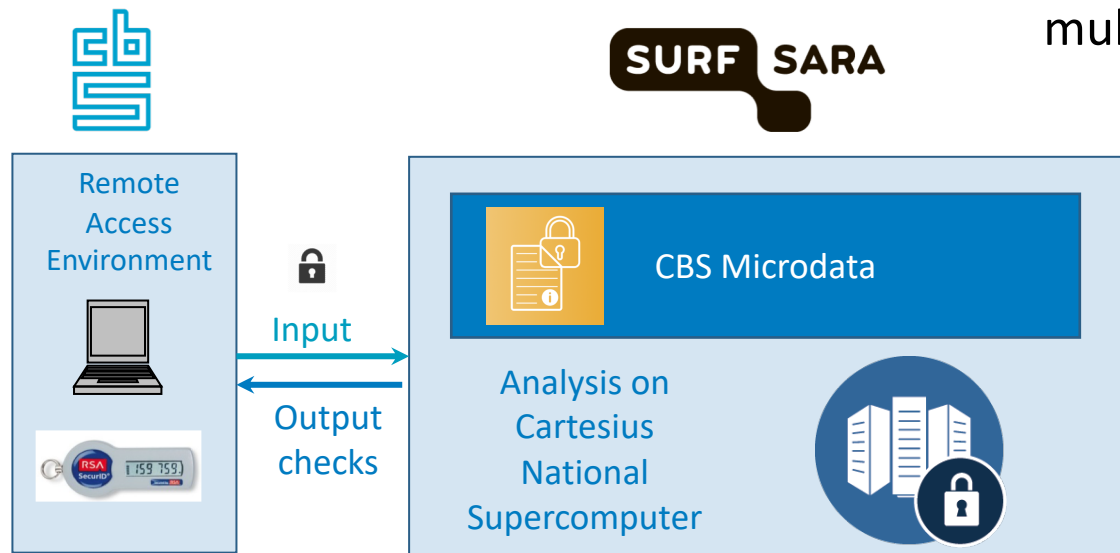
Related Projects

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



Partnership Questions

- Who may benefit from a data exchange?
 - Are there researchers that want to use confidential data?
 - Who are the data providers in this case?
 - Under what conditions would these data providers release their data?
- What should the role of SURF?
 - Service provider; software developer; community manager; ...
- Should SURF turn this prototype into a pilot?
- Are there other projects we should collaborate with?

Technical Questions

- Is a trusted third party the right approach?
- What is the trust relation?
 - Does the data provider trust the researcher?
 - Does the data provider trust the algorithm?
- More advanced user scenarios (e.g. with 3 parties):
 - Patient trust a hospital with their data
 - Hospital trust a researcher with the patient data
 - What are the implications for the current demo with 2-part user-scenario?
 - Who gives what permissions, and is that a continuous permission? How to withdraw permissions?

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