



# Introduction to SANE

Final EPI conference, 7 March 2024

Lucas van der Meer  
Ahmad Hesam  
Martin Brandt  
Annette Langedijk  
Freek Dijkstra



# | Agenda

**01**

SSH Data sharing challenges

**02**

What is SANE?

**03**

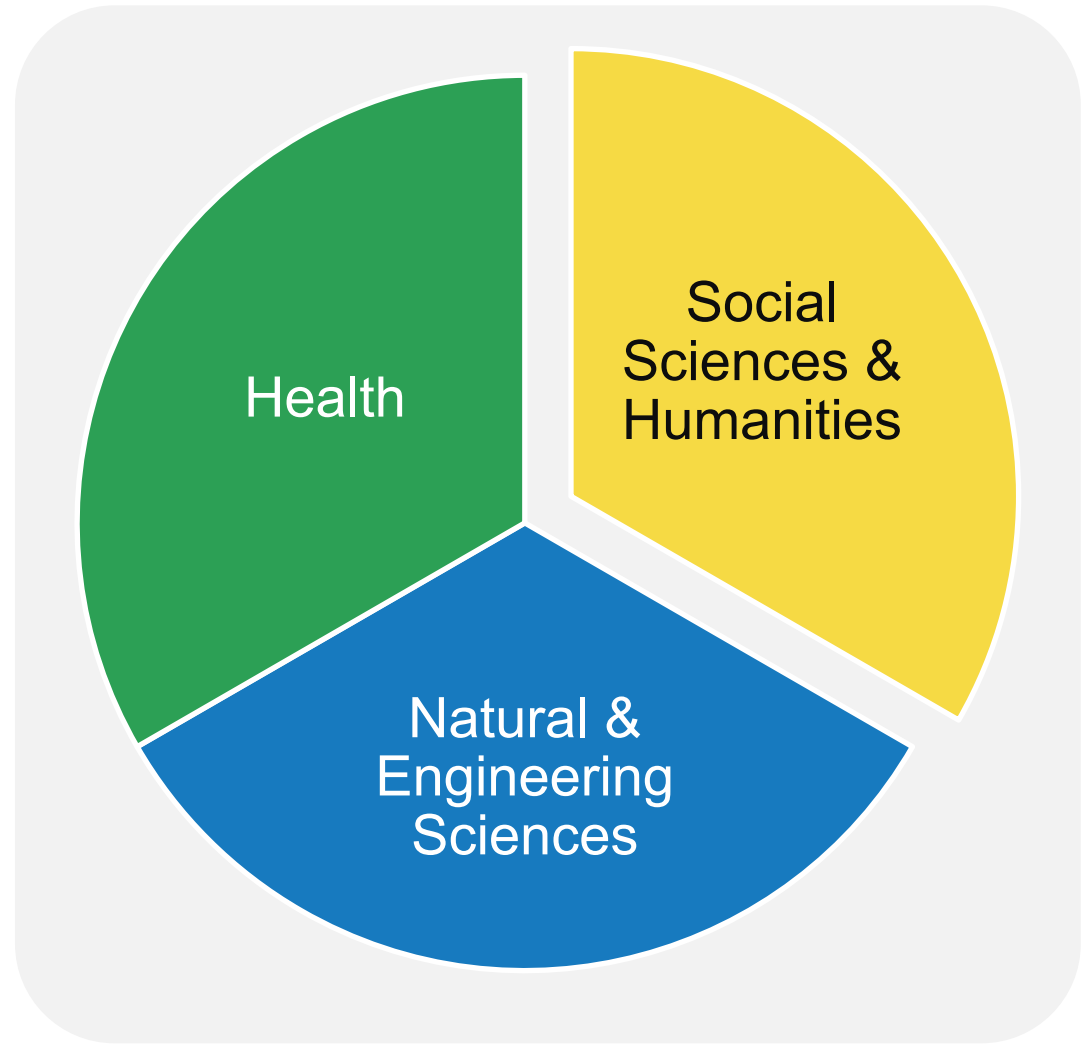
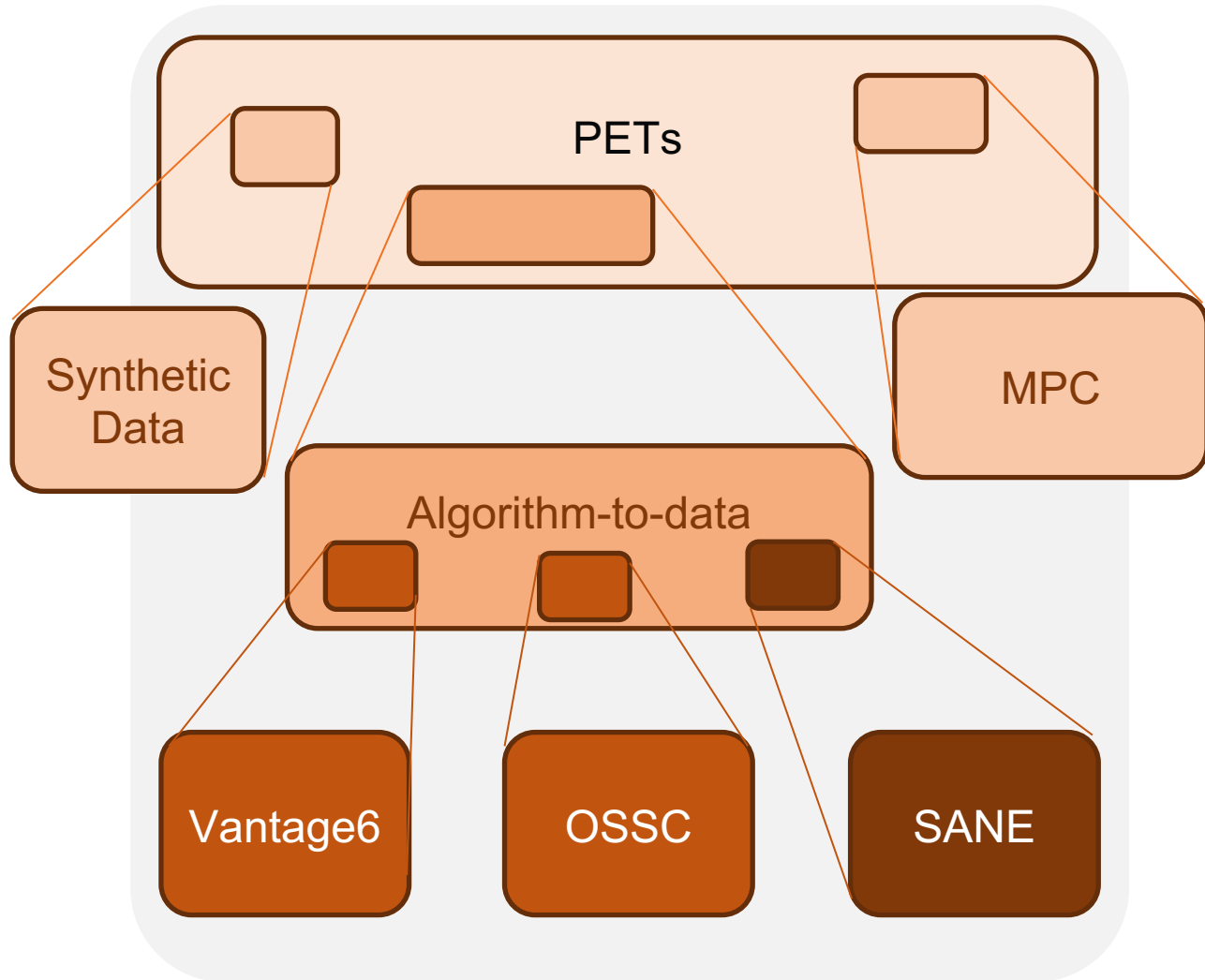
Relation to other projects

**04**

Q&A



# Scope



**data sharing challenges in  
social sciences and humanities**



# | Non-research data has great scientific value, but providers often struggle to share



**Privacy**



**Copyright**



**Competition**

# | Typical data provider would need (1)

## Complete control over the data:

- Data not sent to researcher
- “Non-consumptive use”
- Review any output

## Pseudonymised data

## Data upload not allowed

## Check on researcher & purpose

## Trust the research software





# | Typical data provider would need (2)

## Additional requirements:

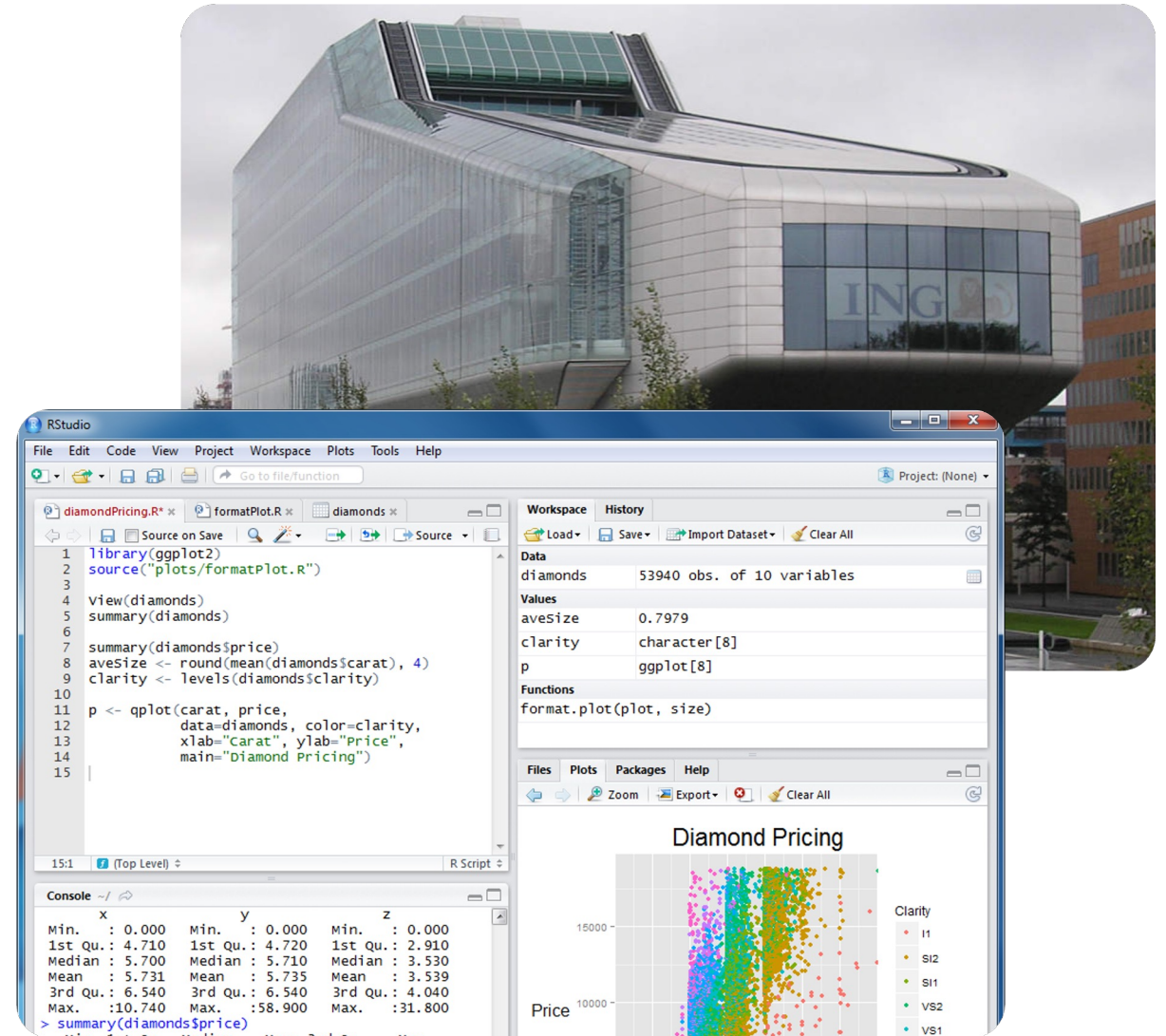
- Researcher cannot see the data
- Data & software upload allowed



# Researcher would need (1)

“To what extent does PIN transaction match official inflation statistics?”

- Play (tinker) with the data
- Specific characteristics of the combined data determine consequent analytical steps
- Combine it with my own data
- Extract and download aggregated non-sensitive results
- Use R, Python

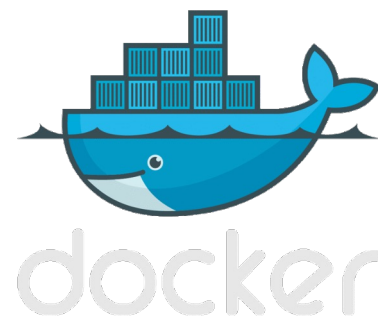
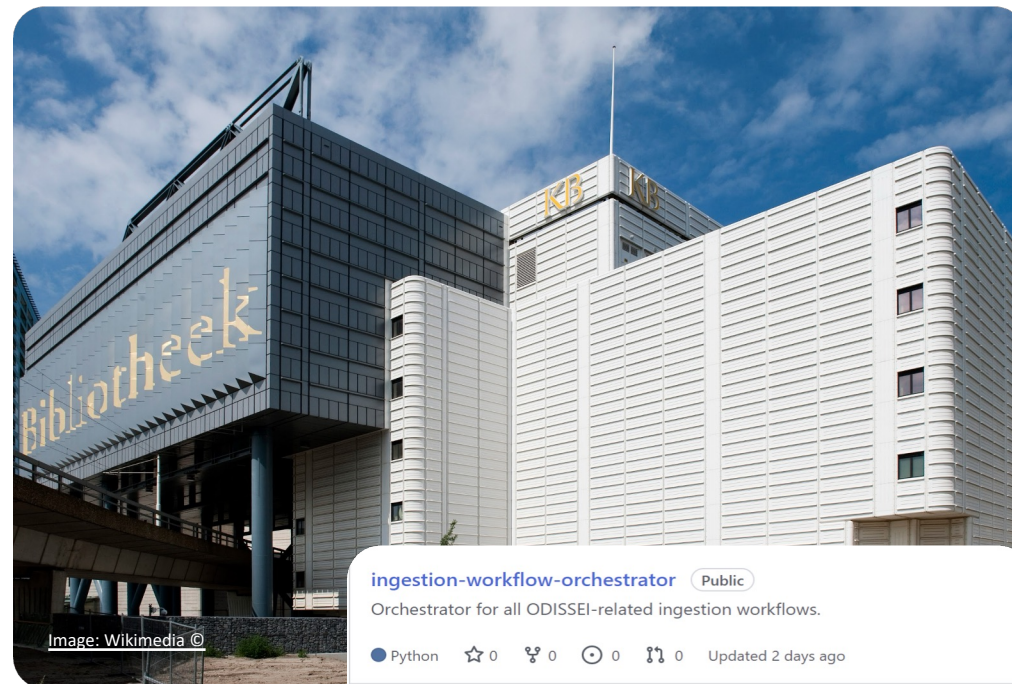




# Researcher would need (2)

“How can a publisher evaluate the impact of converting any backlist ebooks titles?”

- Mine the KB collections
- Use existing software and algorithms, e.g. NLP pipelines
- Use own software or algorithm
- Upload a container
- Extract and download non-sensitive results



**ingestion-workflow-orchestrator** Public  
Orchestrator for all ODISSEI-related ingestion workflows.  
● Python ☆ 0 🍷 0 🔄 0 🔗 0 Updated 2 days ago

**Custom-Metadata-Blocks** Public  
All the tsv-files for the custom metadata blocks will be stored here.  
☆ 0 📄 Apache-2.0 🍷 0 🔄 0 🔗 0 Updated 2 weeks ago

**version-tracker** Public  
Stores workflow versioning information.  
● Makefile ☆ 0 🍷 0 🔄 0 🔗 0 Updated 3 weeks ago

**publication-date-updater** Public  
Updates the publication date of an existing dataset in a dataverse.  
● Python ☆ 0 🍷 0 🔄 0 🔗 0 Updated on Dec 14, 2022

| **Best practice:**  
**Five safes**



Office for  
National Statistics







**SANE:**  
Secure  
ANalysis  
Environment

| **What is SANE?**



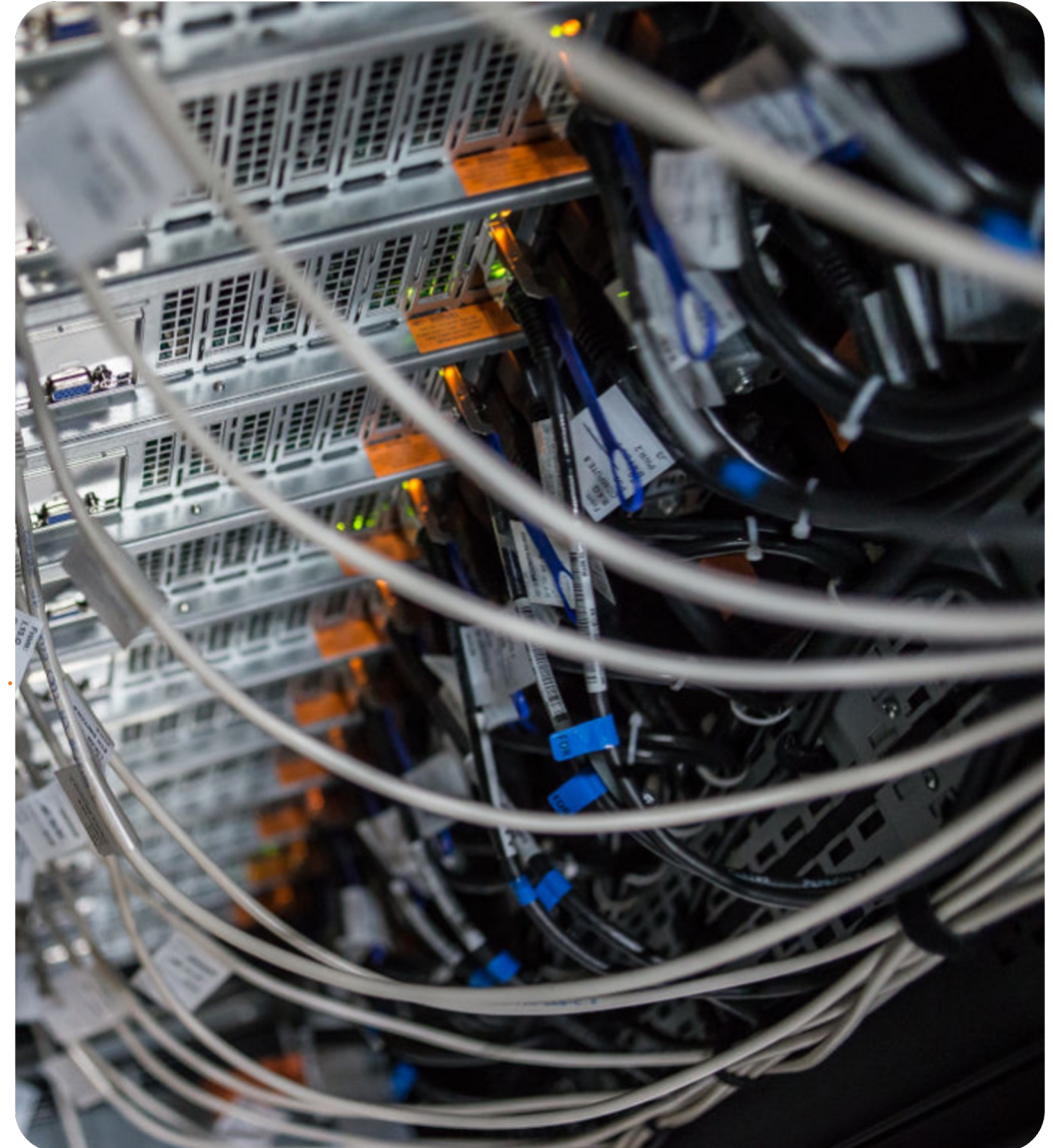
# SANE – Secure ANalysis Environment

- Virtual, fully shielded computing environment
- Enable access to the sensitive data
- Pre-approved analysis software
- Data provider maintain complete control
- Penetration tested
- Runs on ISO-27001 certified SURF Research Cloud

---

**Tinker SANE:** allows the researcher to see, manipulate and play with the data.

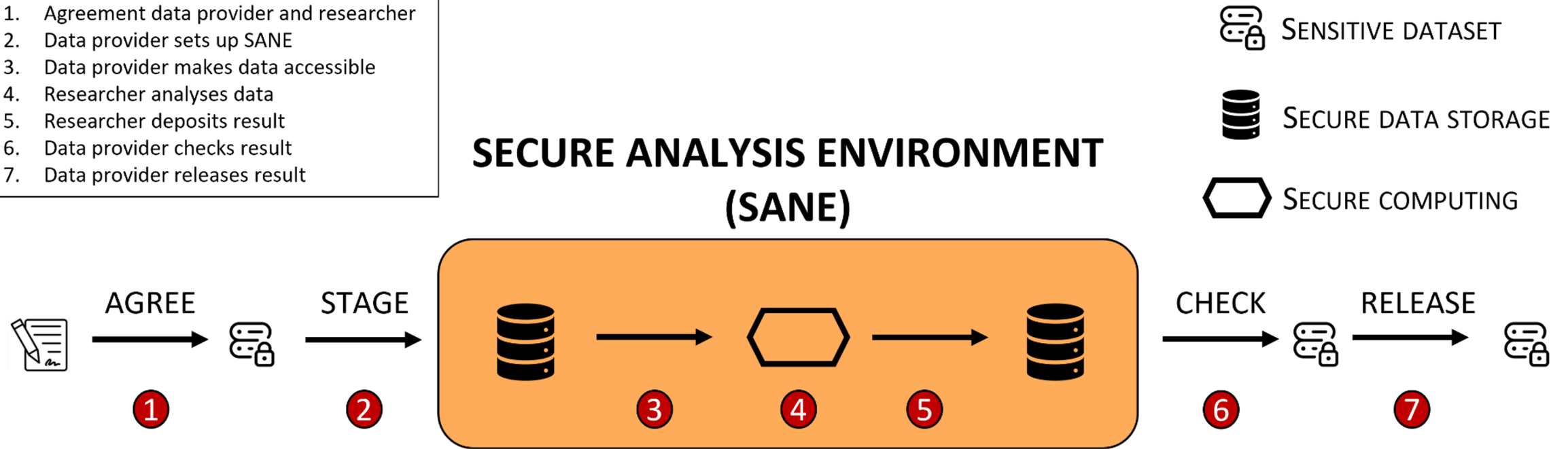
**Blind SANE:** the researcher submits an algorithm without being able to see the data and the data provider approves the algorithm and output.



# SANE – Workflow

## STEPS

1. Agreement data provider and researcher
2. Data provider sets up SANE
3. Data provider makes data accessible
4. Researcher analyses data
5. Researcher deposits result
6. Data provider checks result
7. Data provider releases result



| Screenshots ~~Demo~~

## | Roles

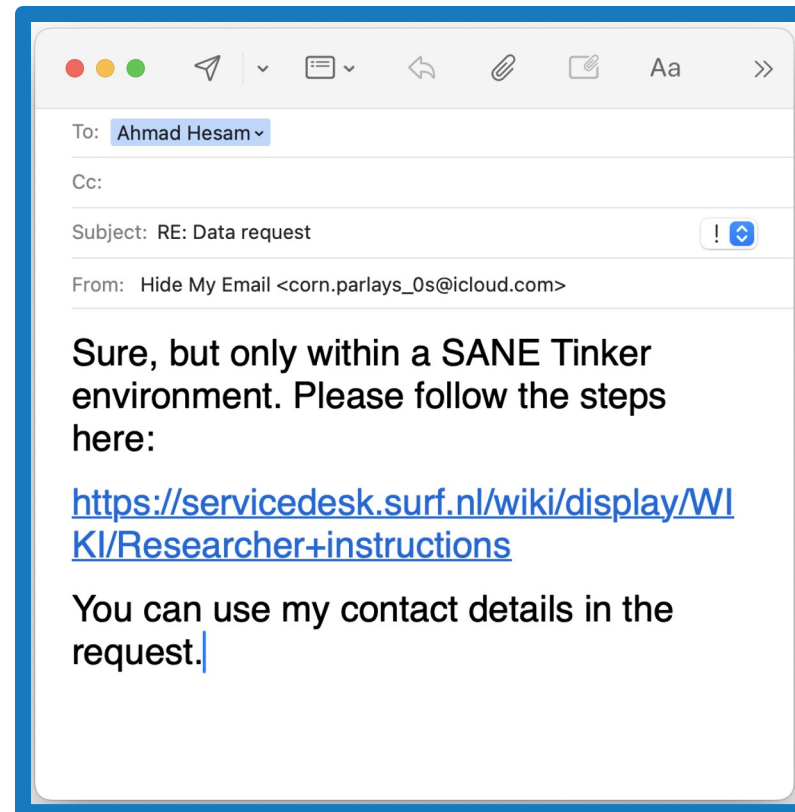
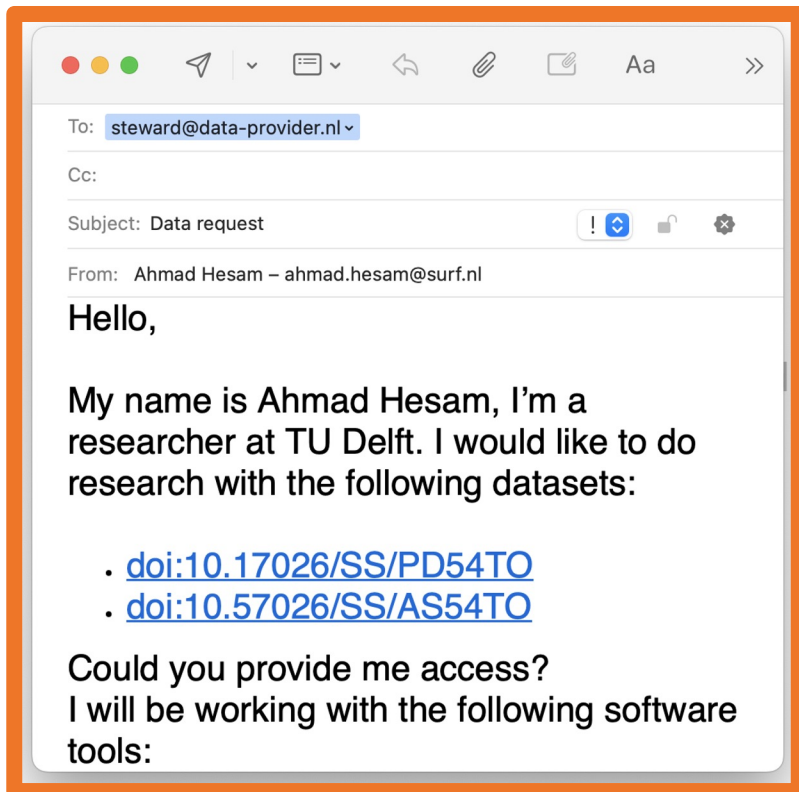
Data Provider

Researcher

SURF Advisor



# Data Agreement Step



# SANE Project Request

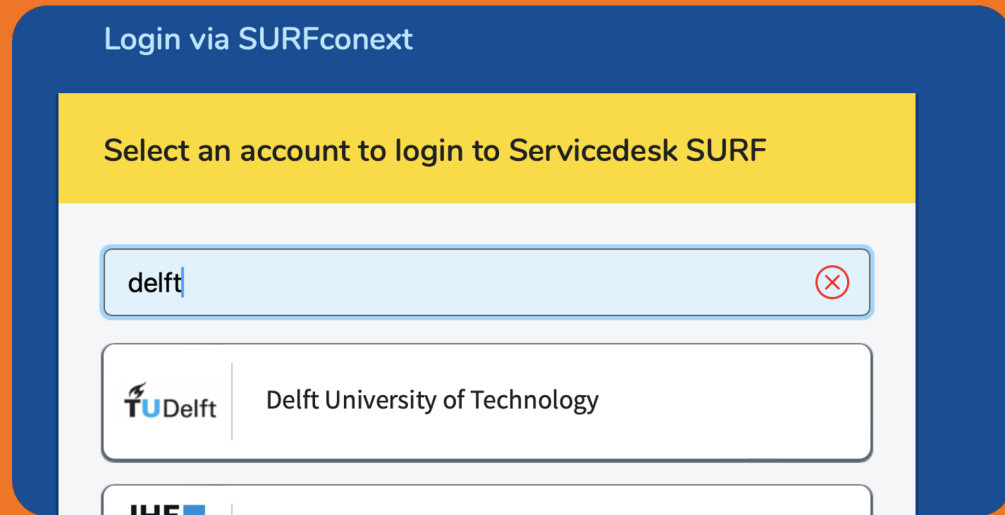
Login via SURFconext

Select an account to login to Servicedesk SURF

delft

TU Delft | Delft University of Technology

IHE

A screenshot of the SURFconext login interface. At the top, it says "Login via SURFconext". Below that is a yellow banner with the text "Select an account to login to Servicedesk SURF". There is a search bar containing the text "delft" with a clear button (an 'x' in a circle) to its right. Below the search bar, there are two account cards. The first card shows the TU Delft logo and the text "Delft University of Technology". The second card is partially visible and shows the IHE logo.

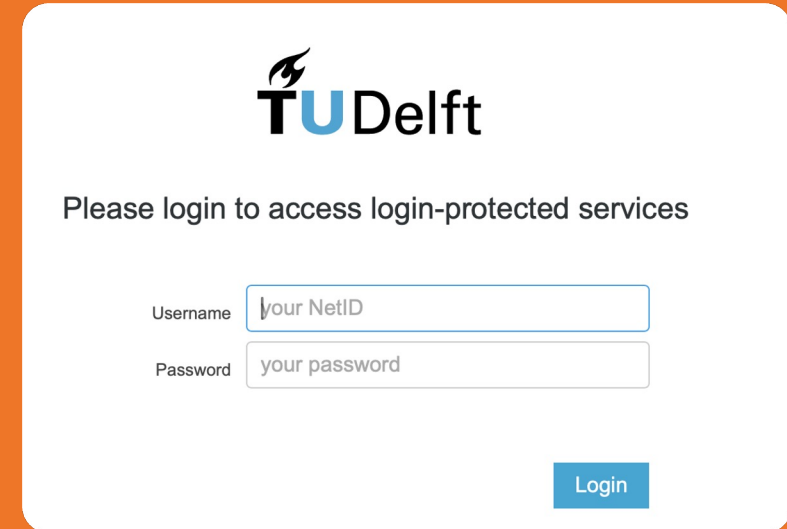
TU Delft

Please login to access login-protected services

Username

Password

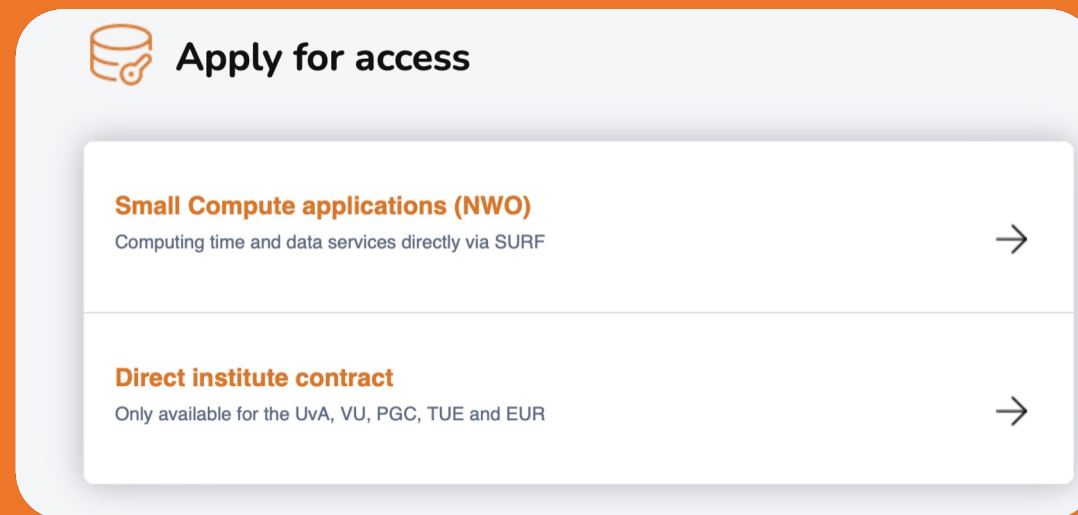
Login

A screenshot of the TU Delft login page. At the top is the TU Delft logo. Below it is the text "Please login to access login-protected services". There are two input fields: "Username" with the placeholder text "your NetID" and "Password" with the placeholder text "your password". A blue "Login" button is located at the bottom right of the form.

Apply for access

**Small Compute applications (NWO)**  
Computing time and data services directly via SURF →

**Direct institute contract**  
Only available for the UvA, VU, PGC, TUE and EUR →

A screenshot of the "Apply for access" page. It features a database icon and the title "Apply for access". There are two main sections, each with a right-pointing arrow. The first section is titled "Small Compute applications (NWO)" and has the subtext "Computing time and data services directly via SURF". The second section is titled "Direct institute contract" and has the subtext "Only available for the UvA, VU, PGC, TUE and EUR".

# SANE Project Request

Title of the project

[SANE] Socio-economic agent-based modeling

Description

Aa ▾ | B I ... | ☰ ▾ | 🔗 <> + ▾

Dear Servicedesk,

In the context of SANE on SURF Research Cloud I wish to request the following collaboration:

**Name:** Socio-economic ABM

**Description**  
Using the following datasets I wish to calibrate an agent-based model to discover new socio-economic phenomena....

<rest of description>

**Data provider contact person:** Martin Brandt, steward@data-provider.com

Kind regards,  
Ahmad Hesam

Technical project requirements

I expect to need 100GB of storage for the datasets, and about 10GB for my results, so 200GB in total should be enough.

I will need at least 20 (iterations) x 500 (core-hours per iteration) = 10k core hours to perform my analysis.

Resources

Snellius

Research Cloud - HPC Cloud

Data processing - Grid

Data processing - Spider

Signing Authority e-mail

john.doe@tudelft.nl

Recognizable institutional e-mail required

Signing Authority Staff Position within the organization

Professor

# | SURF processes request



Dear data provider,

**You have been invited** to the SANE

CO: "Socio-economic ABM"

Please accept this invite to start the  
SANE project





# Accept invitation to CO

Dear data provider,

**You have been invited** to the SANE  
CO: "Socio-economic ABM"

Please accept this invite to start the  
SANE project

Accept



Welcome to Socio-economic ABM

 You are invited to become **admin of** this collaboration

Please be aware that on becoming and admin of this collaboration, your personal data may be shared with the following services. Take a moment to review any policy documents provided by the service.

## Purpose of the collaboration

A collaboration to process sensitive data in the context of Socio-economic ABM

## We collaborate in 1 services



SURF Research Cloud

[Acceptable use policy](#)

| [Privacy policy](#)

|  [Contact](#)

A service used by this collaboration requires that you agree to their acceptable use policy.

I agree to the service's acceptable use policy

Proceed to My SANE CO Name

# Setup SANE environment

Welcome to SURF research cloud

 LOGIN

SURF HPC Cloud Network

SELECTED ✓

SURF HPC Cloud volume

SELECTED ✓



**SANE data server**

SANE data server



CHOOSE



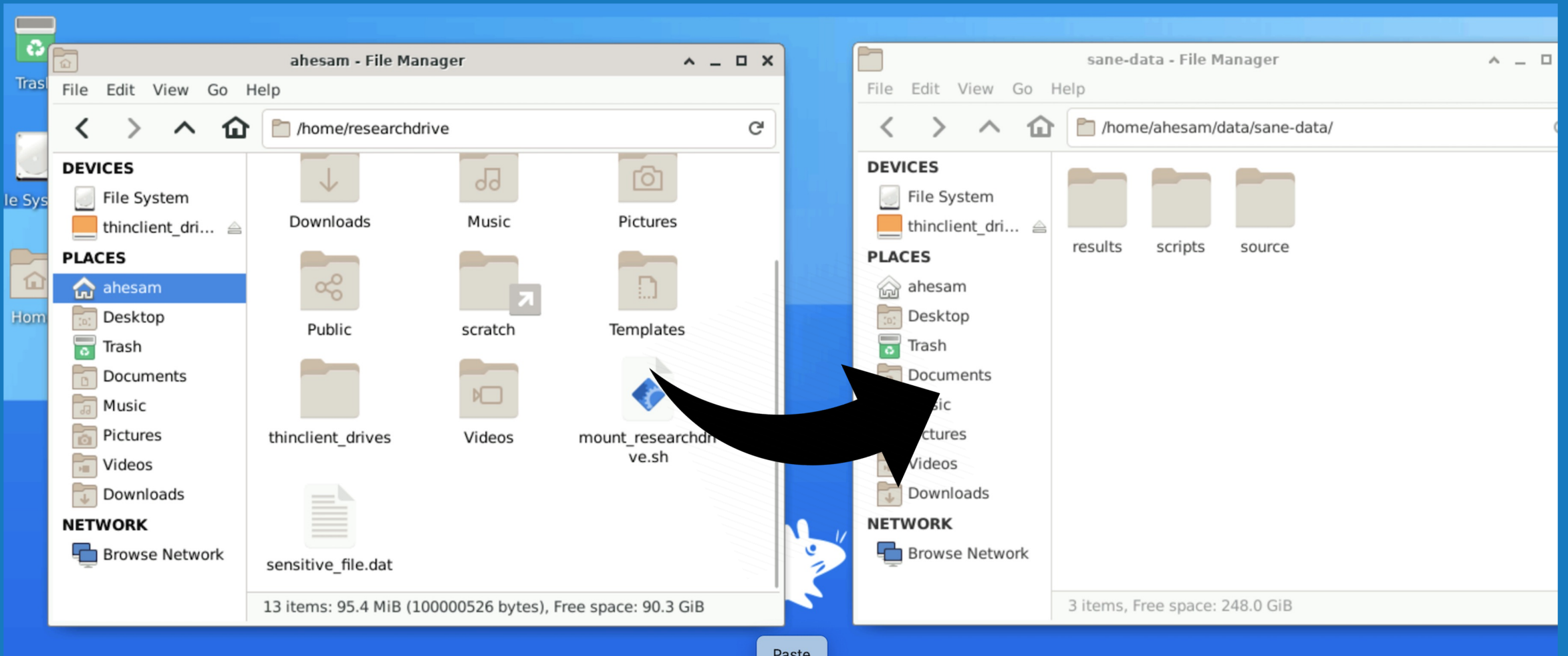
**SANE linux data owner portal**

Only access for data owner



CHOOSE

# Upload sensitive data



# Invite Researcher

Send invitations

---

Membership

Never expires

## Invitation details

Invitees ⓘ

Role ⓘ

Group membership ⓘ

Membership valid until ⓘ



Message ⓘ

# Accept invitation to CO

Dear researcher,

**You have been invited** to the SANE


CO: "Socio-economic ABM"

Please accept this invite to start the

SANE project

**Accept**

Welcome to Socio-economic ABM

 You are invited to become **member** of this collaboration

Please be aware that on becoming and admin of this collaboration, your personal data may be shared with the following services. Take a moment to review any policy documents provided by the service.

## Purpose of the collaboration

A collaboration to process sensitive data in the context of Socio-economic ABM

## We collaborate in 1 services



SURF Research Cloud

[Acceptable use policy](#)

| [Privacy policy](#)

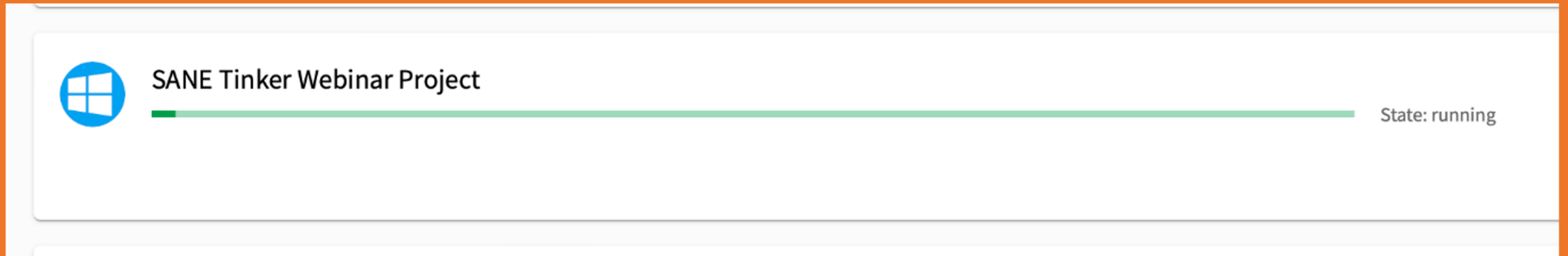
| [Contact](#)

A service used by this collaboration requires that you agree to their acceptable use policy.

I agree to the service's acceptable use policy

**Proceed to My SANE CO Name**

# Start SANE workspace

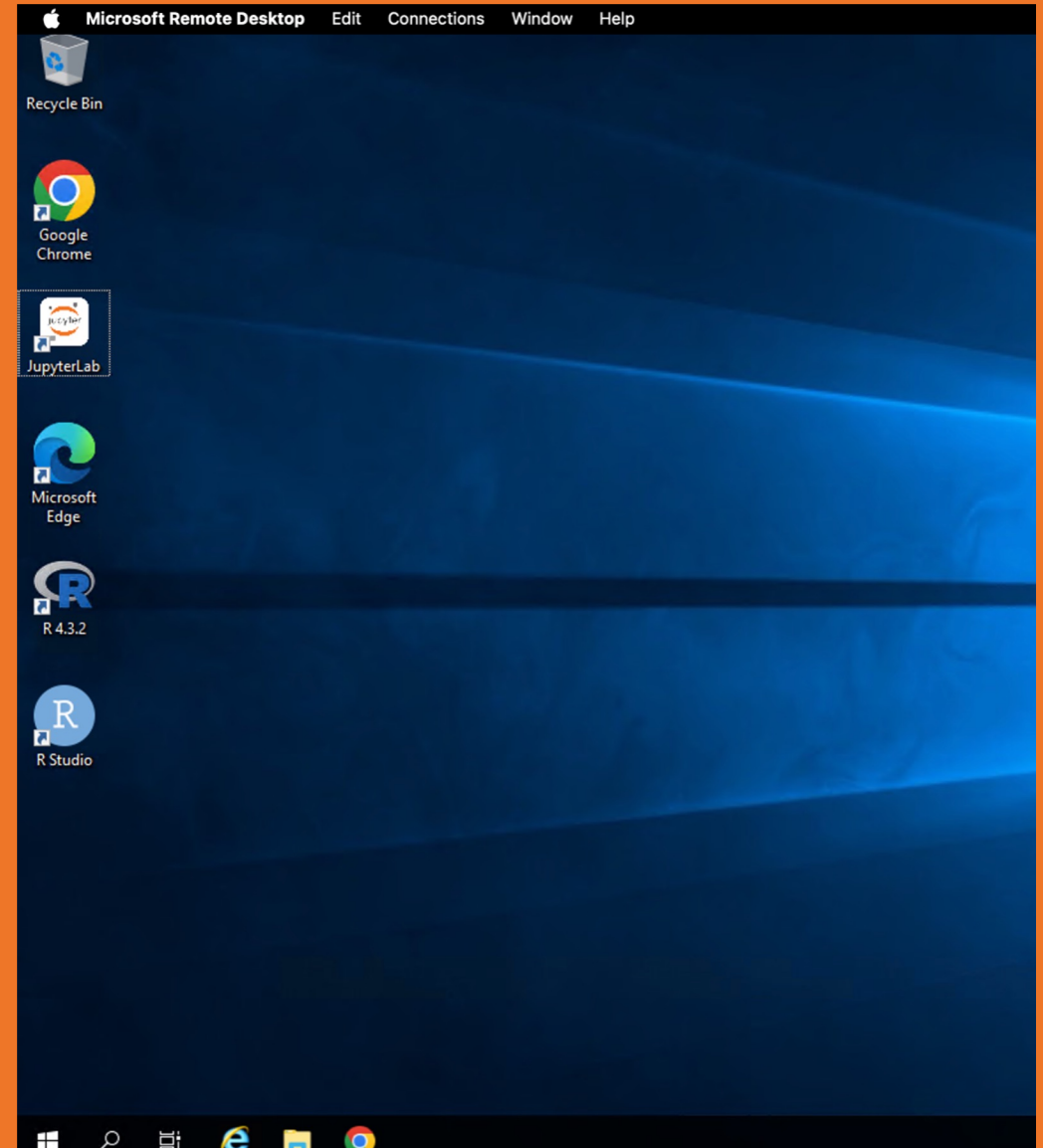


Use your favorite Remote Desktop Client to access your Tinker SANE workspace



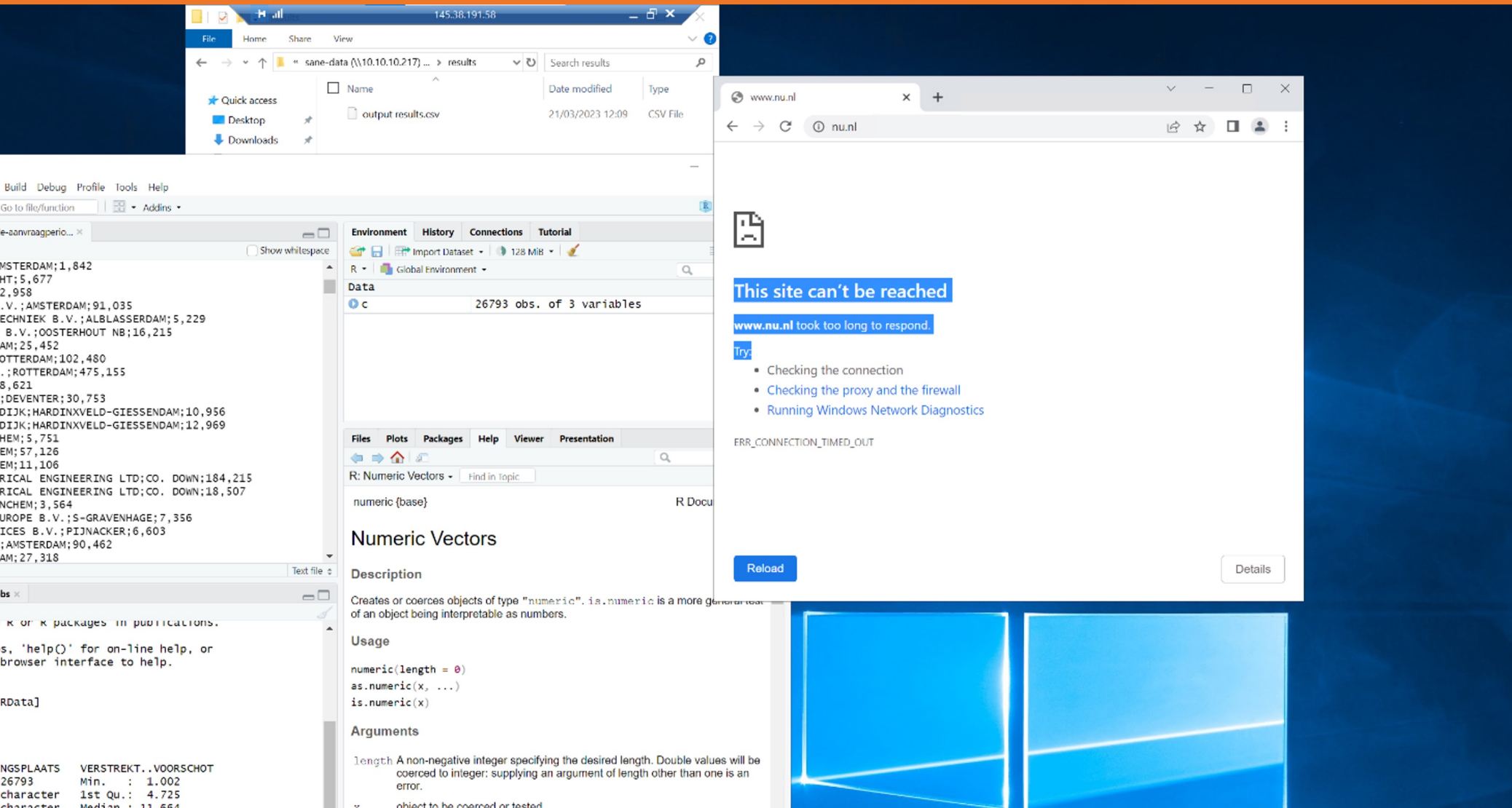


# Tinker SANE workspace





# Closed Down Environment



# Start Blind SANE workspace

## 1. Choose the expiration date of the machine

 10-02-2024 (22:16)

## 2. Workspace name, domain name and description

Name

SANE Blind Webinar Project

Hostname

saneblindwebin

Description

Description

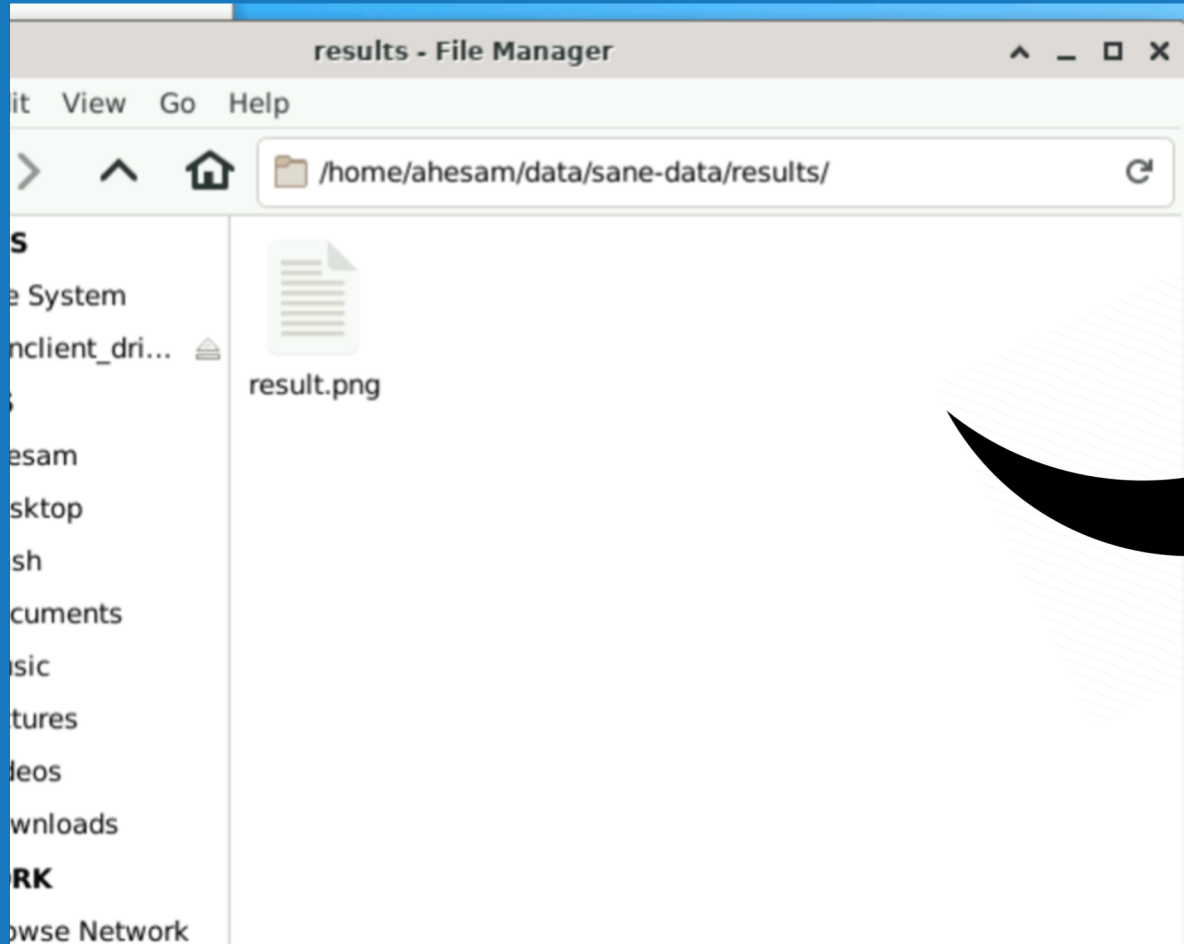
## 3. Workspace parameters

Script location: ( script subfolder or GIT repository url)

Script location



# | Inspect and release output results





# Relation to other projects

# Use Cases:

## FirmBackbone

Open-science principles and  
restricted data

Peter Gerbrands, Wolter Hassink,  
Daniel Oberski, Rutger Schilpzand,  
Arjen van Witteloostuijn



# Use Cases

## FirmBackbone



Universiteit Utrecht



VRIJE  
UNIVERSITEIT  
AMSTERDAM



Platform Digitale  
Infrastructuur  
Social Sciences  
& Humanities

## Lisa

Het werkgelegenheidsregister  
van Nederland



ODISSEI

## KVK

HOUVAST VOOR  
ONDERNEMERS

KB } nationale  
bibliotheek



## Clariah



Universiteit Utrecht

SURF

## Data Governance & Support

Contract Management

Data Protection

User Guides  
(Best Practices)

User Support  
(IT & Usage)

## Access Management

User Access Controls (SRAM)

## Data Ingestion (Bronze)

Data Retrieval  
(File, Web, ...)

Data Storage  
(Permanent)

Data Staging  
(Temporal)

## Data Processing (Silver)

Data Conversion  
(File Format)

Data Linking  
(Identities)

Data Extraction  
(Text Mining)

## Data Provisioning (Gold)

Data Pseudonymization

Data Cleaning

Data Aggregation

## Data Analysis (SANE)

Data Access

Data Tinkering

Blind Analysis

Metadata Extraction

Metadata Enrichment

Metadata Querying

Data Lineage

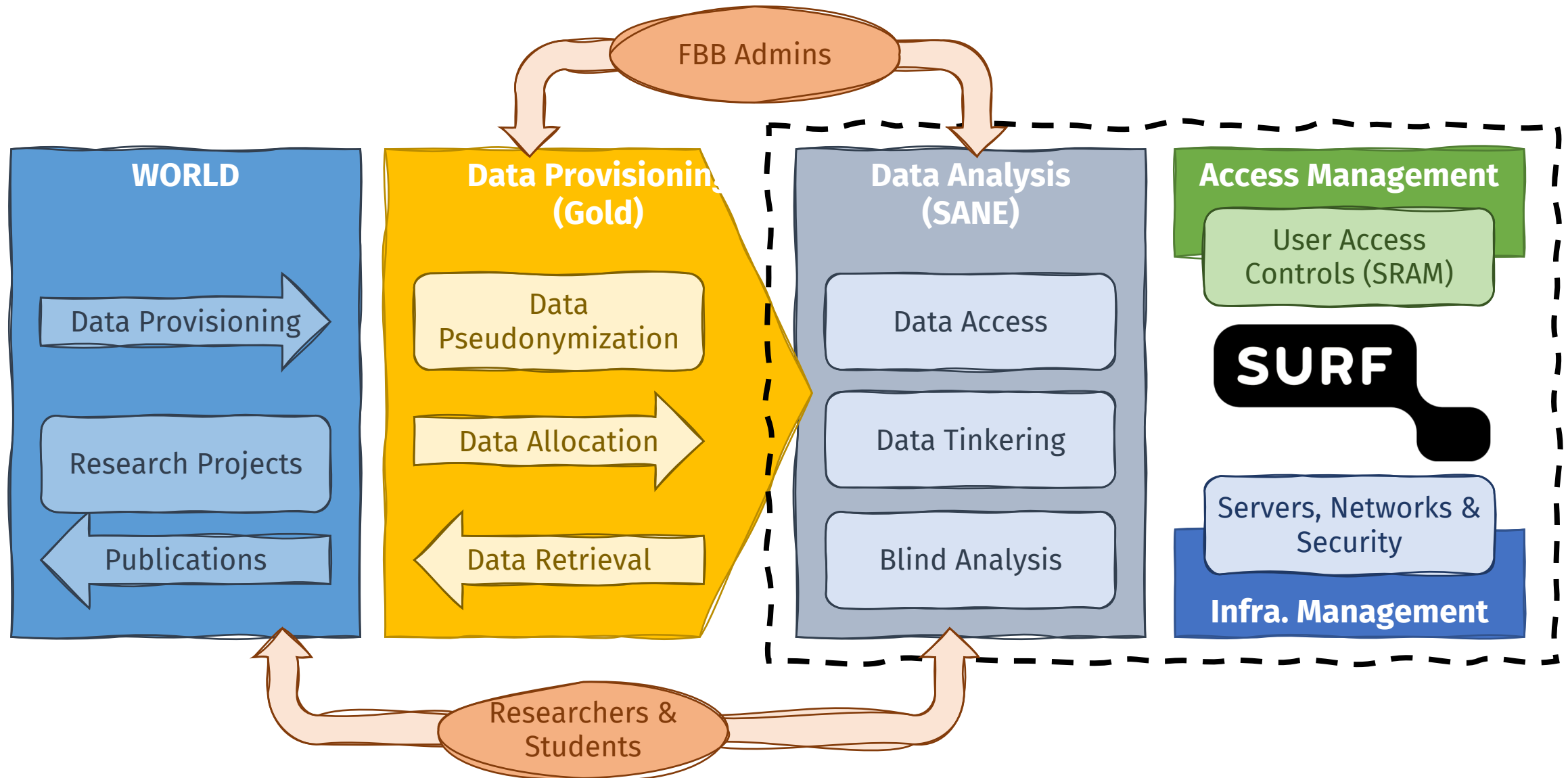
## Metadata Management

Servers, Networks &  
Security

## Infra. Management



# FirmBackbone – Architecture highlight



| Q&A

**Thank you for  
your attention!**

✉ [annette.langedijk@surf.nl](mailto:annette.langedijk@surf.nl)

✉ [lucas@odissei-data.nl](mailto:lucas@odissei-data.nl)

✉ [ahmad.hesam@surf.nl](mailto:ahmad.hesam@surf.nl)

✉ [freek.dijkstra@surf.nl](mailto:freek.dijkstra@surf.nl)

