gesis Leibniz Institute for the Social Sciences

Providing Safe Access to Sensitive Data

Deborah Wiltshire, Secure Data Center, Cologne

FORS Workshop
25 November 2022

Leibniz Association

gesis Leibniz Institute for the Social Sciences

Providing Safe Access to Sensitive Data

- Defining our data
- Types of secure data access solutions
- Structuring secure data services
- What‘s next for secure access
- Final thoughts

gesis Leibniz Institute for the Social Sciences

Thinking about secure data: Some key terms

IDENTIFIABLE DATA

Includes all the data; can directly identify individuals

PSEUDONYMISED DATA

Includes most of the data; direct identifiers removed but could be potentially indirectly identify individuals through jigsaw identification

ANONYMISED DATA

Data anonymised to protect confidentiality; risk of identifying individuals should be negligible

gesis Leibniz Institute for the Social Sciences

Types of data access solution

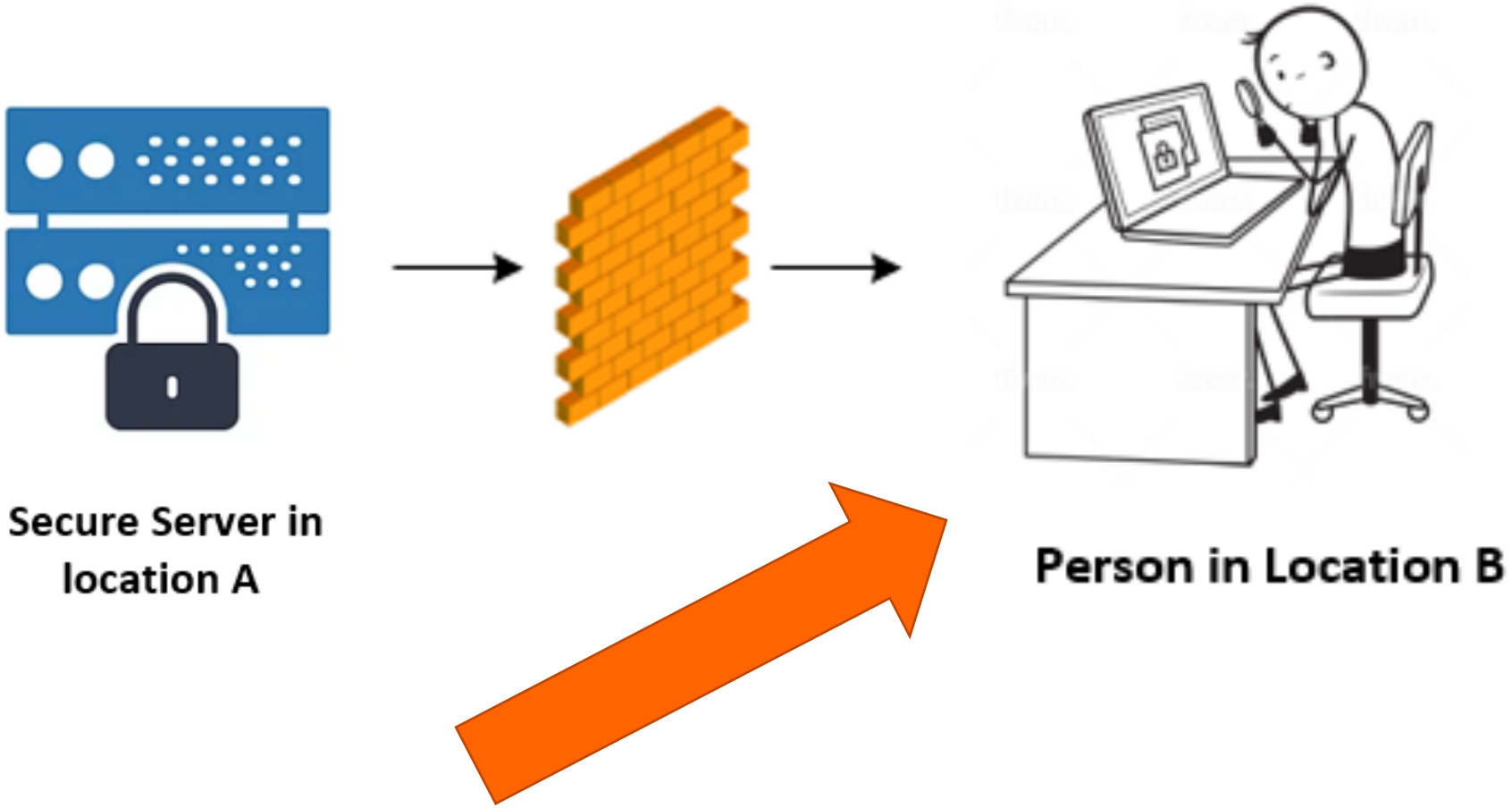
gesis Leibniz Institute for the Social Sciences

A quick history

- Initially access only via physical safe rooms/safe havens
- Move towards remote access
- 3 main models of remote access:
  - Remote Access
  - Remote Desktop
  - Remote Execution

gesis Leibniz Institute for the Social Sciences

The basic model of access



Pic credits: https://www.canstockphoto.com/illustration/secure-server.html & https://www.fiverr.com/fatfatma/a-high-quality-and-unique-stickman-figure

gesis Leibniz Institute for the Social Sciences

1. Safe Room Access

- A secure room in the premises of the data provider
- Number of physical controls are possible
  - Access controlled
  - Thin clients
  - Virtual environment sealed
  - No personal belongings

Pic credits: https://www.fiverr.com/fatfatma/a-high-quality-and-unique-stickman-figure

gesis Leibniz Institute for the Social Sciences

2. Remote Access

- Still based on Safe Room access
- Safe Room is at a partner organisation
- Access via bilateral agreements & secure technical connections
- Retains the physical controls of Safe Room access, but offers more flexibility

Location B
=
on premises of partner organisation

Pic credits: https://www.fiverr.com/fatfatma/a-high-quality-and-unique-stickman-figure

gesis Leibniz Institute for the Social Sciences

3. Remote Desktop Access

- Access is via secure encrypted internet connection from their own office
- E.g. The UK Data Service SecureLab
- Much more accessible for researchers
- But lose many of the physical controls of Safe Room access

Location B
=
Researchers own office

Pic credits: https://www.fiverr.com/fatfatma/a-high-quality-and-unique-stickman-figure

gesis Leibniz Institute for the Social Sciences

4. Remote Execution

- Location is still researchers office
- The main difference is that researchers don‘t work with the data
- Submit code & the results are returned to them
- E.g Statistics Canada *Real Time Remote Access*
- DataSHIELD

Location B

=

Researchers own office

gesis Leibniz Institute for the Social Sciences

Building a secure data service

gesis Leibniz Institute for the Social Sciences

The Human Model of Data Security

Based on the premise of the ´Accidental intruder´

Design system to address human factors

Well designed system AND well trained researchers, can allow greater access

Allows researchers to demonstrate that they can use data safely

Ritchie, 2017

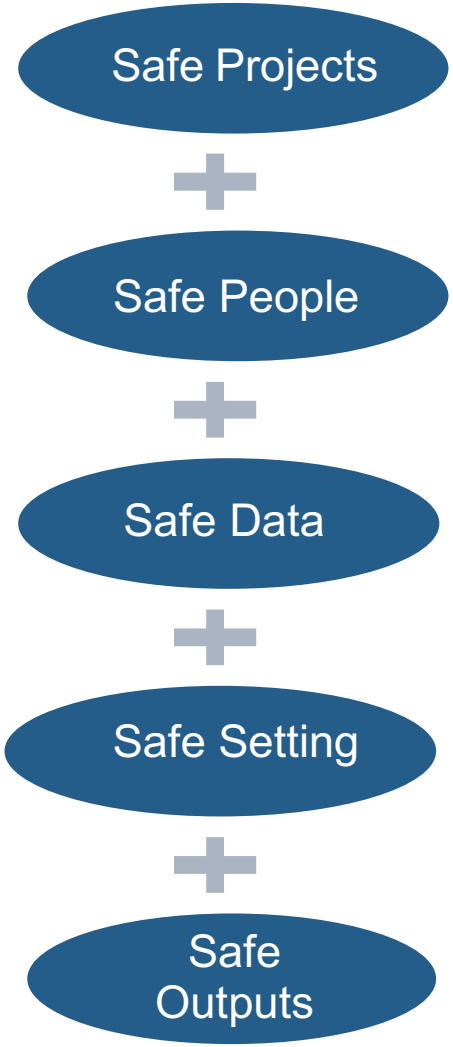
gesis Leibniz Institute for the Social Sciences

Introducing the Five Safes Framework

- Framework for managing access to secure microdata
- Provides a decision making process to ensure safe data use
- 5 simple to follow principles
- First developed at the Office for National Statistics in the UK
- Now used worldwide

gesis Leibniz Institute for the Social Sciences

The 5 Safes Framework



Clearly specified project with valid statistical purpose

The right knowledge & experience to work with the data

The level of detail in the data is appropriate for the setting

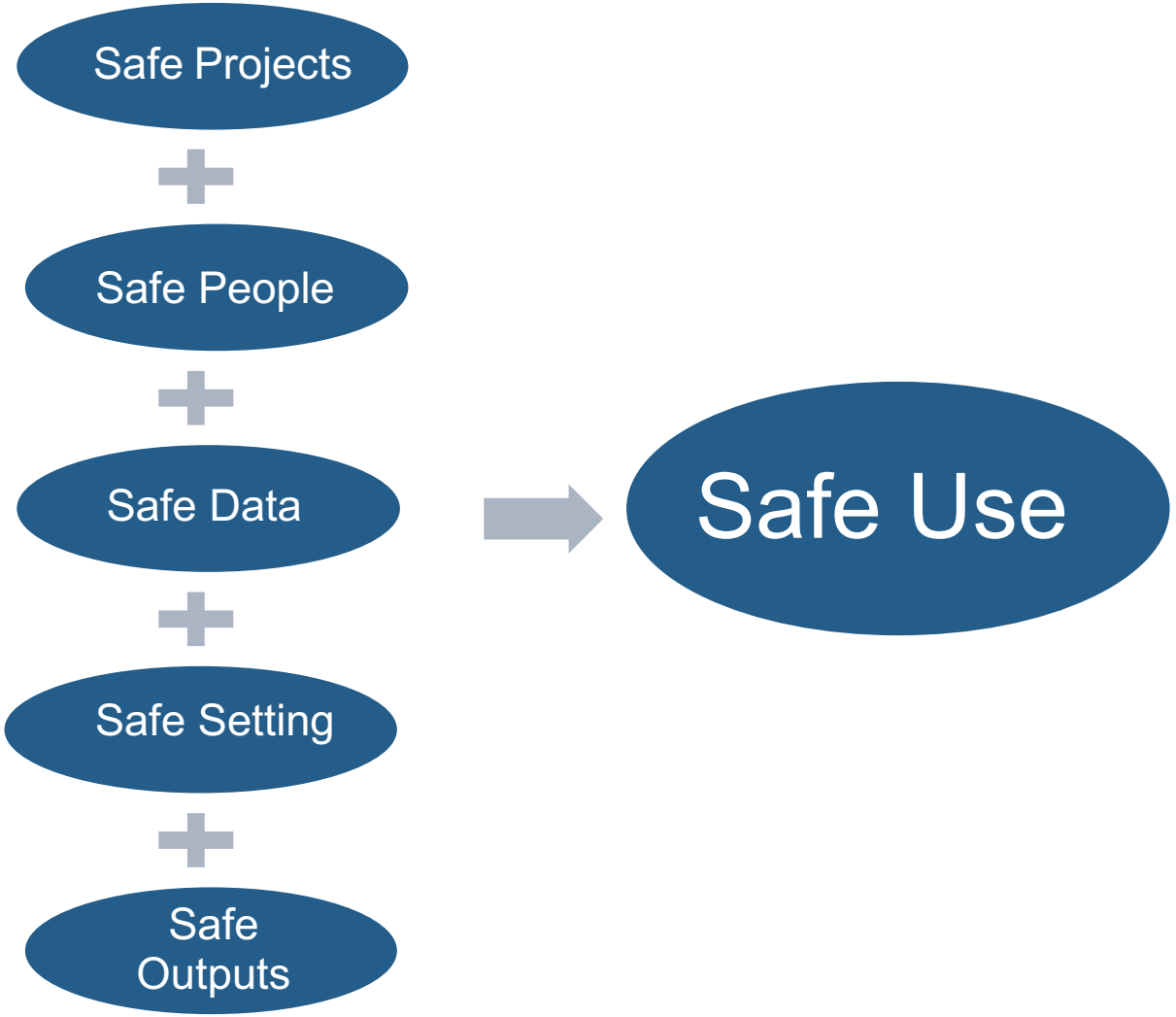
An environment with appropriate safeguards in place

The results of analysis intended for publication or presentation

From VML, 2004; Desai, Ritchie & Welpton 2016

gesis Leibniz Institute for the Social Sciences

The 5 Safes Framework



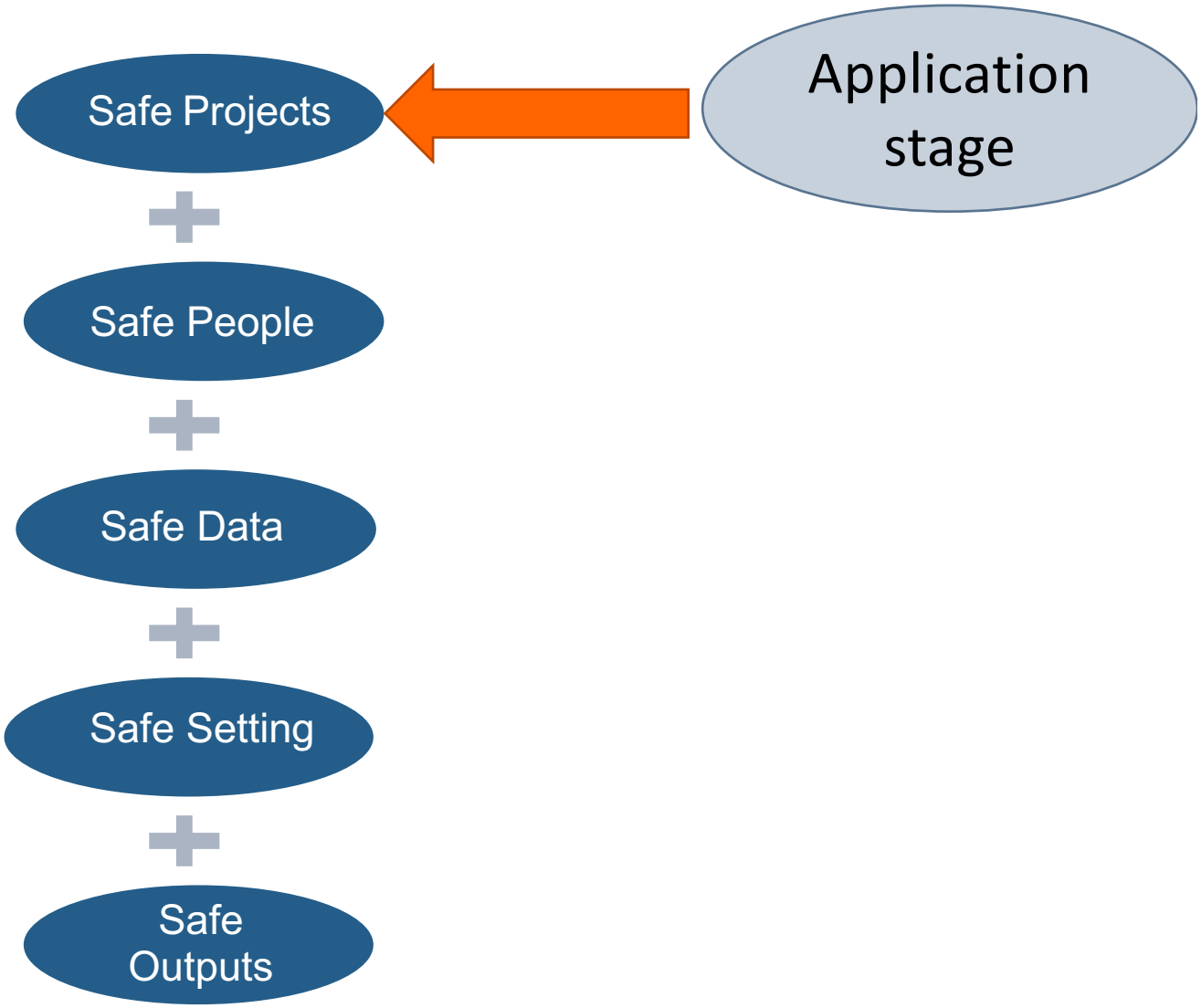
From VML, 2004; Desai, Ritchie & Welpton 2016

gesis Leibniz Institute for the Social Sciences

Structuring a secure data service within the Five Safes Framework

gesis Leibniz Institute for the Social Sciences

The 5 Safes Framework



From VML, 2004; Desai, Ritchie & Welpton 2016

Legal Gateways for Research

The legal gateways allow for:

- Specific researchers to carry out
  - specific research projects
  - during a specified period of time
  - using specific datasets

gesis Leibniz Institute for the Social Sciences

Safe Projects – Data Use Agreements

- Agreements between data service & researcher
- Set out –
  - who can access what data,
  - for what purpose
  - & for how long
- Terms & conditions
- Institutional agreements

gesis Leibniz Institute for the Social Sciences

Data Use Agreement

Regarding on-site access to the GESIS Secure Data Center

Contract number: _____
(provided by GESIS)

between

GESIS – GESIS – Leibniz Institut für Sozialwissenschaften
Quadrat B2,1
68159 Mannheim

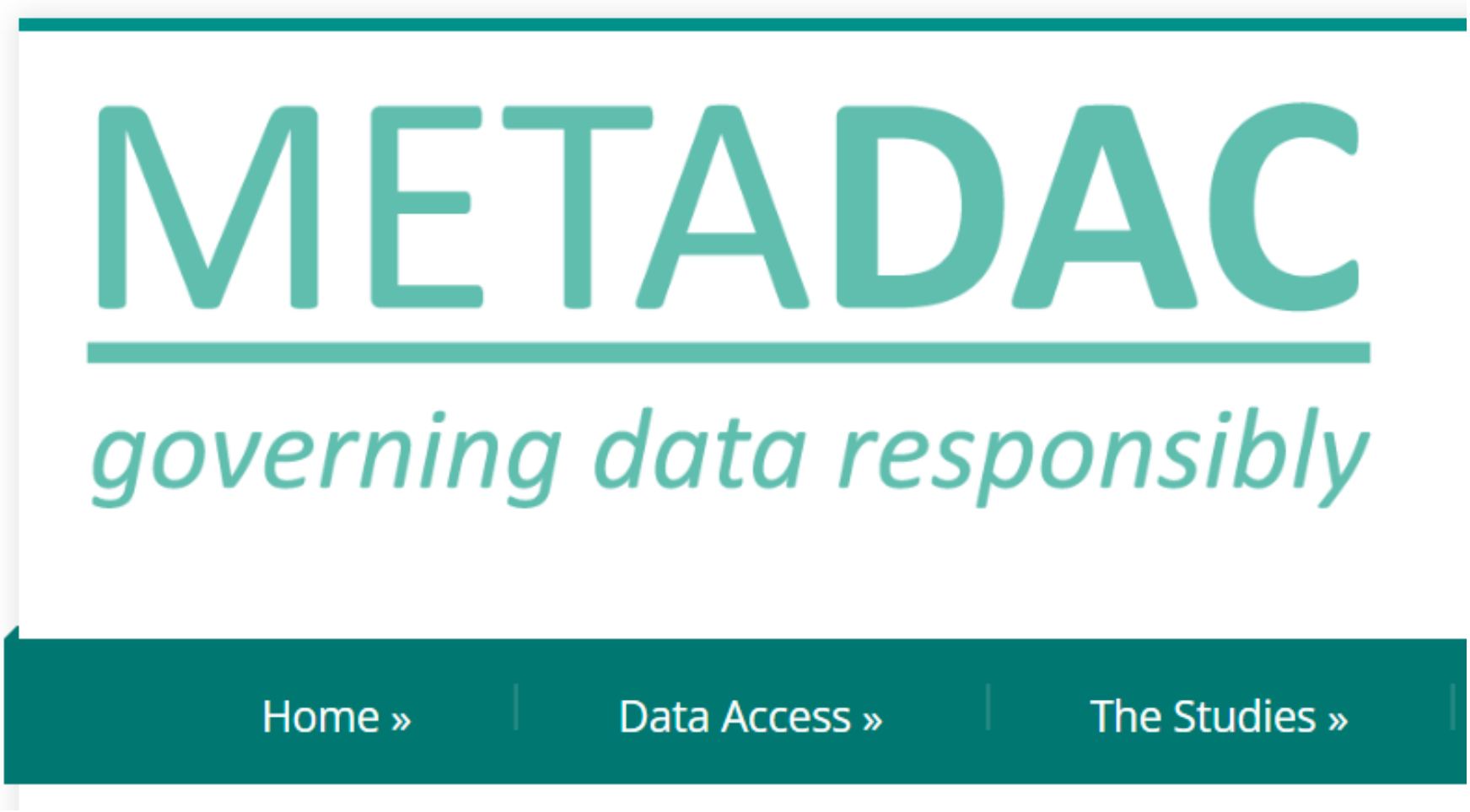
– hereafter referred to as GESIS –

and

Last name	
First name	
E-mail	
Telephone number	
Institution	
Business address	
Position of data recipient[1]	

gesis Leibniz Institute for the Social Sciences

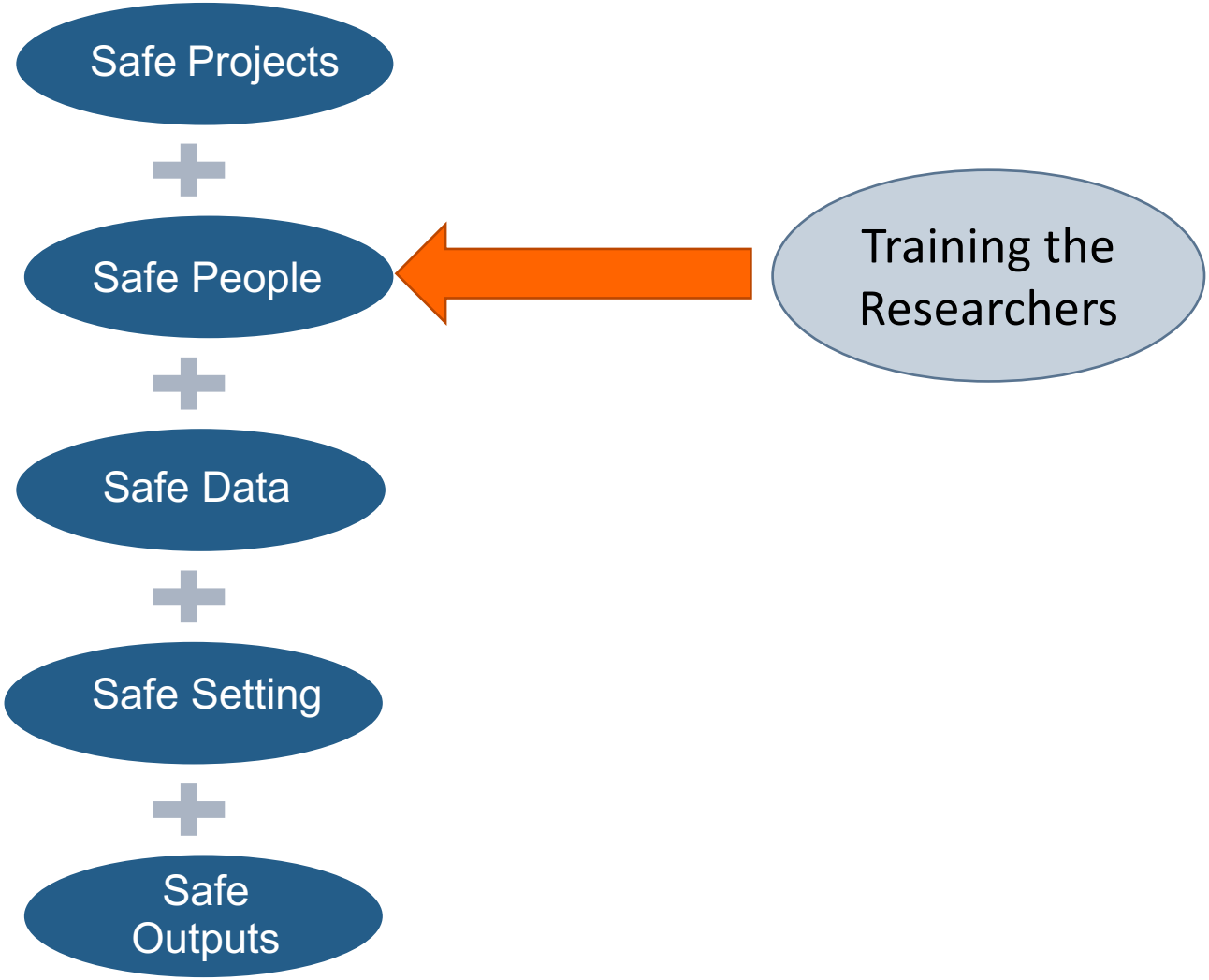
Accessing project applications



https://uksa.statisticsauthority.gov.uk/digitaleconomyact-research-statistics/research-accreditation-panel/ https://www.metadac.ac.uk/

gesis Leibniz Institute for the Social Sciences

The 5 Safes Framework



From VML, 2004; Desai, Ritchie & Welpton 2016

gesis Leibniz Institute for the Social Sciences

Safe People – the Researchers

- Researchers don't always have the knowledge necessary for working with secure data
- Researchers don't always read the instructions
- When researchers are trained - less likely to make mistakes that might prove harmful to data subjects
- The process of analysing sensitive data and publishing results from projects will be more efficient

gesis Leibniz Institute for the Social Sciences

Training your researchers

- Mandatory training?
- Safe Researcher Training
  - What influences data access
  - Legislation
  - Statistical disclosure
- Utilise existing resources
- Consortium training schemes
  - Safe Researcher Training, UK
  - MOReT –Germany

Office for National Statistics

ADR UK

The HMRC Datalab

UK Data Service

UNIVERSITÄT MANNHEIM

gesis Leibniz Institute for the Social Sciences

DEUTSCHE BUNDESBANK EUROSYSTEM

DESTATIS Statistisches Bundesamt

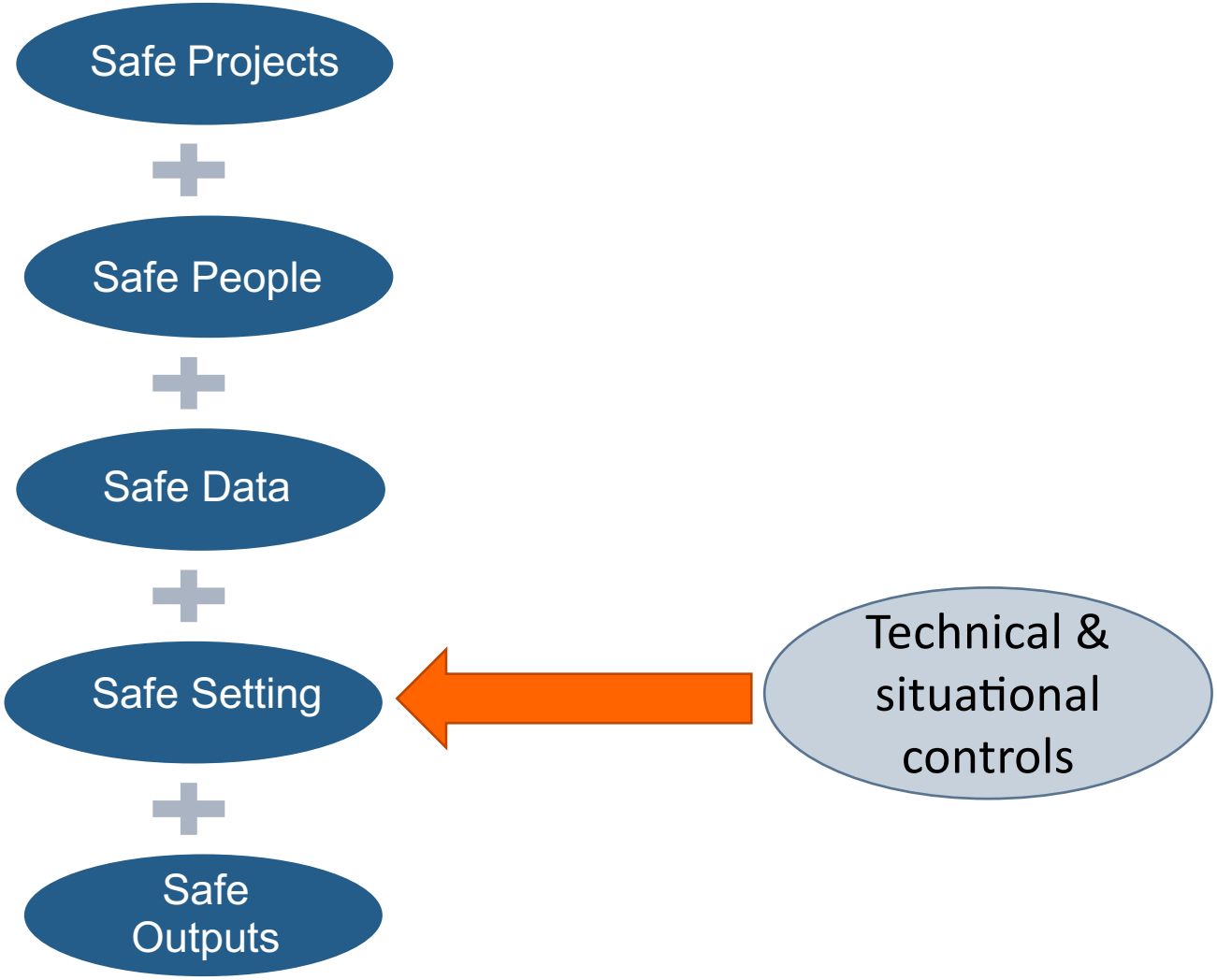
GERMAN HUMAN GENOME PHENOME ARCHIVE

SSHOC Training materials of workshop for secure data facility professionals. https://doi.org/10.5281/zenodo.5638596

SDAP https://securedatagroup.org/training/

gesis Leibniz Institute for the Social Sciences

The 5 Safes Framework



From VML, 2004; Desai, Ritchie & Welpton 2016

gesis Leibniz Institute for the Social Sciences

Safe Settings

The working environment:

- Thin clients
- Each researcher has unique login credentials
- Researchers log into a virtual desktop
- Virtual desktop is a completely sealed environment
- Screen lock automatically activated with inactivity
- Monitoring software

Physical Safe Room:

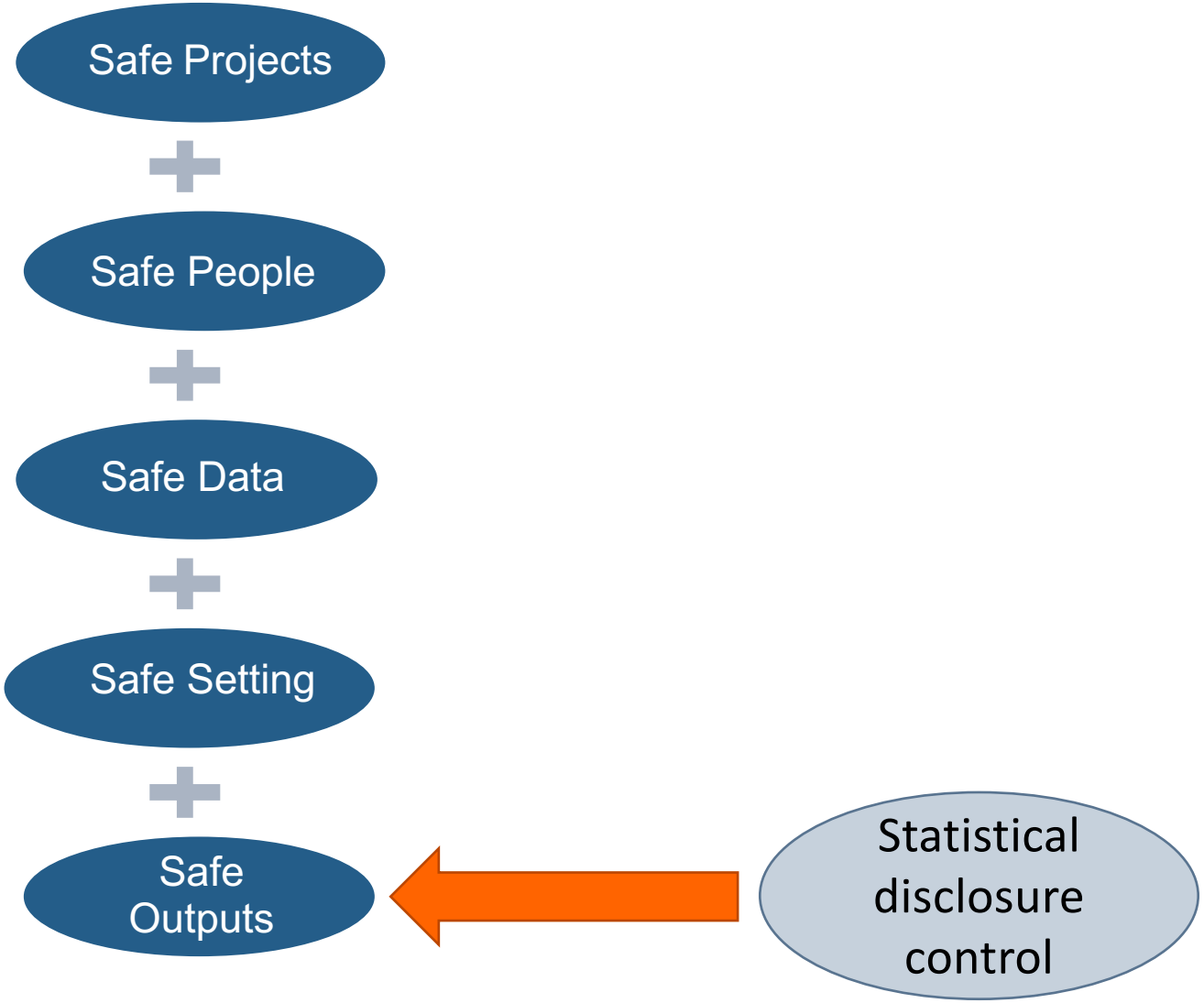
- Researcher id checks
- Privacy screen
- Locked room, restricted access
- Personal items not permitted (i.e. electronic devices)
- Taking notes - regulated

Remote Desktop:

- 2-factor authentication
- Requirements for work station
  - Private office
  - Fixed IP address
- Non-technical controls like training, legal agreements

gesis Leibniz Institute for the Social Sciences

The 5 Safes Framework



From VML, 2004; Desai, Ritchie & Welpton 2016

gesis Leibniz Institute for the Social Sciences

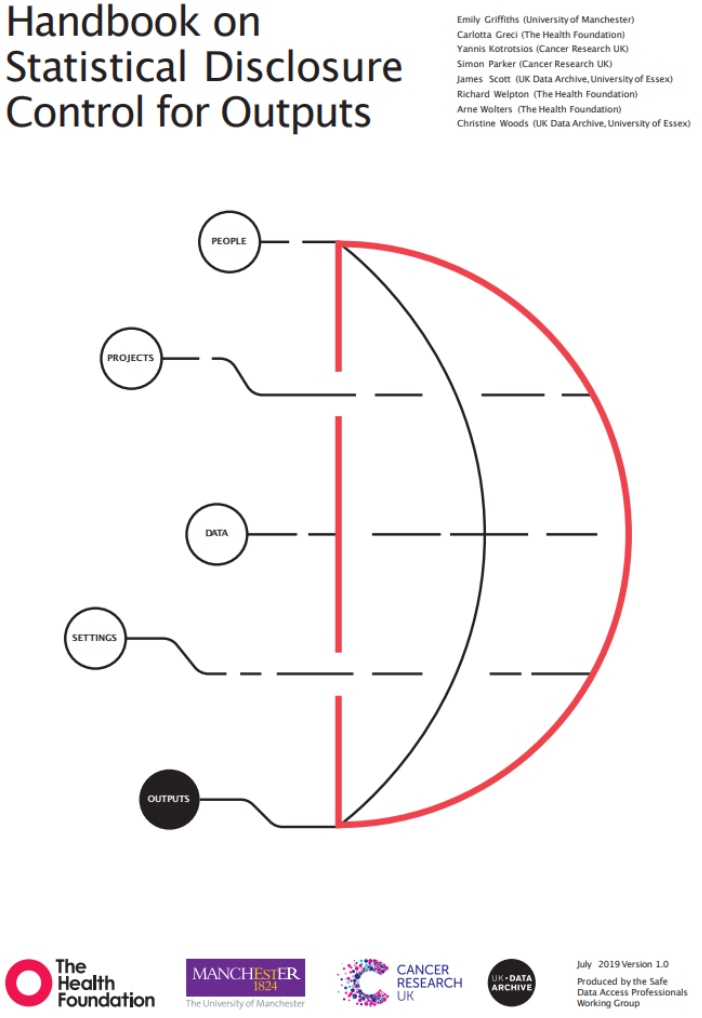
Introduction to Safe Outputs

- The aim is to minimise the risk of an individual being identified, or assigning an attribute to someone, from a piece of analysis
- Residual risk in published results
- Statistical Disclosure Control (SDC) is a key method of doing so
- “The unprovability of safety”
- The aim is to demonstrate that we’ve taken all reasonable measures to ensure the risk is minimal

gesis Leibniz Institute for the Social Sciences

Statistical Disclosure Control & statistical quality

- SDC is a set of rules that are applied to outputs before release
- Generally a ‘4 eyes’ approach is best practice
- SDC is applied to research outputs before release or publication
- SDC rules closely match principles of good research practices and statistical validity



[the SDAP SDC Handbook]

gesis Leibniz Institute for the Social Sciences

Extending the Five Safes

Safe Person

The Secure Data Access Professional or Data Steward!

- The success of your system relies heavily on your team!
- Specialist area – we're a relatively small group
- Often recruit people with little or no prior experience
- No formal training
- For smaller teams this can be a problem!

gesis Leibniz Institute for the Social Sciences

Safe Secure Data Stewards

Specific Training

- Data stewardship training
- UK output checkers training
- MOReT project

Professional & Support Networks

- SDAP
- International Network

gesis Leibniz Institute for the Social Sciences

What‘s next for secure access?

- Expansion of secure data access
  - More data available
  - More data linkage
  - New data forms - DVD data
- Dissolving of boundaries
  - International boundaries
  - Disciplinary boundaries

gesis Leibniz Institute for the Social Sciences

Some final thoughts

gesis Leibniz Institute for the Social Sciences

Leibniz Association

Deborah.wiltshire@gesis.org